

From Early Melanesian Pidgin to
Solomon Islands Pijin, Bislama and Tok Pisin.
A diachronic, comparative analysis of the emergence of
individual varieties of Melanesian Pidgin English

Inaugural-Dissertation
zur
Erlangung des Doktorgrades
der Philosophie des Fachbereiches 05
der Justus-Liebig-Universität Gießen

vorgelegt von
Damaris Schäfer

aus Asslar-Oberlemp

2021

Dekan: Prof. Dr. Thomas Möbius

1.Berichterstatter: Prof. Dr. Magnus Huber

2.Berichterstatter: Prof. Dr. Joybrato Mukherjee

Tag der Disputation: 03. Juni 2022

Table of contents

EIDESSTATTLICHE ERKLÄRUNG	VII
ACKNOWLEDGMENTS	IX
FIGURES	XI
TABLES	XV
LIST OF ABBREVIATIONS	XVII
ABSTRACT	XIX
ABSTRACT GERMAN	XXI
1 INTRODUCTION.....	1
1.1 GENERAL	1
1.2 PREVIOUS RESEARCH.....	3
1.3 MOTIVATION, AIM AND SCOPE OF THE STUDY	4
1.4 STRUCTURE OF THE STUDY	9
1.5 GENERAL REMARKS	10
1.5.1 <i>A note on terminology.....</i>	<i>10</i>
1.5.1.1 Dialects or languages	10
1.5.1.2 Individualisation and stabilisation	11
1.5.2 <i>Remarks on the names used to refer to the varieties.....</i>	<i>12</i>
1.5.3 <i>Remarks on the examples.....</i>	<i>13</i>
1.5.4 <i>Remarks on non-English extralinguistic quotes</i>	<i>13</i>
1.5.5 <i>Remarks on languages cited in this work.....</i>	<i>13</i>
2 STUDYING THE EMERGENCE OF PIDGINS & CREOLES FROM A DIACHRONIC PERSPECTIVE	15
2.1 HISTORICAL LINGUISTICS AND THE <i>BAD DATA PROBLEM</i>	16
2.2 CHALLENGES IN STUDYING PIDGINS AND CREOLES FROM A DIACHRONIC PERSPECTIVE	18
2.2.1 <i>Availability and accessibility of historical sources.....</i>	<i>18</i>
2.2.2 <i>Reliability of language examples extracted from historical sources</i>	<i>19</i>
2.2.2.1 Authenticity	20
2.2.2.2 Timespan between documentation and original language situation	21
2.2.2.3 Language attitudes	23
2.2.2.4 Duration of sojourn and travel activities to other Pacific Islands.....	24
2.2.2.5 Linguistic abilities of the writers	24
2.2.2.6 Editor revisions.....	25
2.2.2.7 Anglicisation	26
2.3 SUMMARY AND OUTLOOK	27
3 THE ORIGIN AND DEVELOPMENT OF MPE AND ITS GEOGRAPHICAL VARIATION.....	29
3.1 A SOCIO-HISTORICAL ACCOUNT OF LANGUAGE CONTACT IN MELANESIA	29
3.1.1 <i>The first explorations.....</i>	<i>33</i>
3.1.1.1 The exploration of the Solomon Islands	33
3.1.1.2 The exploration of Vanuatu	34
3.1.1.3 The exploration of Papua New Guinea.....	34
3.1.1.4 Summary	34
3.1.2 <i>The period of trading contacts.....</i>	<i>35</i>
3.1.2.1 Trading contacts in the Solomon Islands	36
3.1.2.2 Trading contacts in Vanuatu	37
3.1.2.3 Trading contacts in New Guinea	38
3.1.2.4 Summary	39
3.1.3 <i>Contact on overseas plantations</i>	<i>40</i>
3.1.3.1 Recruitment in adjacent islands.....	40
3.1.3.2 Linguistic diversity on the plantations.....	41
3.1.3.3 The development of plantation pidgins	42
3.1.3.4 Returning labourers and the diffusion of plantation pidgins.....	44
3.1.3.5 Labour recruitment statistics	45

3.1.3.6	Meta-linguistic evidence	48
3.1.3.6.1	Impact of overseas plantations on the development of Solomon Islands Pijin	48
3.1.3.6.2	Impact of overseas plantations on the development of Bislama	49
3.1.3.6.3	Impact of overseas plantations on the development of Tok Pisin	51
3.1.3.6.4	Summary	53
3.1.3.7	The end of the labour trade	53
3.1.4	<i>Developments in the home areas</i>	54
3.1.4.1	Developments in the Solomon Islands	54
3.1.4.2	Developments in Vanuatu	56
3.1.4.3	Developments in Papua New Guinea	57
3.1.4.3.1	German New Guinea	57
3.1.4.3.2	British New Guinea	59
3.1.4.4	Summary	60
3.1.5	<i>Missionary activities</i>	60
3.1.5.1	Missionary activities in the Solomon Islands	61
3.1.5.2	Missionary activities in Vanuatu	63
3.1.5.3	Missionary activities in New Guinea	65
3.1.5.4	Summary	66
3.1.6	<i>World War II</i>	66
3.1.6.1	World War II and New Guinea	67
3.1.6.2	World War II and the Solomon Islands	68
3.1.6.3	World War II and Vanuatu	69
3.1.6.4	Summary	69
3.1.7	<i>Post-World War events</i>	70
3.1.7.1	Post-World War events in the Solomon Islands	70
3.1.7.2	Post-World War events in Vanuatu	71
3.1.7.3	Post-World War events in Papua New Guinea	72
3.1.7.4	Summary	73
3.2	THEORIES ON THE GENESIS OF MPE AND THE ORIGIN OF THE INDIVIDUAL VARIETIES	74
3.2.1	<i>Origin of Tok Pisin on Samoan plantations</i>	74
3.2.2	<i>A 1970s early Melanesian Pidgin</i>	75
3.2.3	<i>Pacific-wide nautical jargon</i>	76
3.2.4	<i>Origin of MPE in Sydney and plantations in Queensland</i>	78
3.2.5	<i>South Seas Jargon as an element of Maritime Polynesian Pidgin</i>	79
3.2.6	<i>The end of the labour trade</i>	79
3.3	SUMMARY AND OUTLOOK	80
4	EMPIRICAL FOUNDATIONS: DATA AND METHODOLOGY	81
4.1	DATABASE COMPILATION	81
4.1.1	<i>Identification of archives and possible data sources</i>	81
4.1.1.1	<i>Deutsche Kolonialbibliothek, Frankfurt</i>	82
4.1.1.2	Online Archives	82
4.1.1.3	Bislama data collection by Philip Baker	83
4.1.1.4	Research Centre of the Australian War Memorial	83
4.1.1.5	Pacific Research Archives, Canberra	84
4.1.1.6	National Library of Australia, Canberra	84
4.1.1.7	Western Pacific Archives, Auckland	84
4.1.1.8	Pacific Manuscript Bureau, Canberra	85
4.1.1.9	Summary	85
4.1.2	<i>Focus during data extraction for reconstructing past stages of pidgins and creoles</i>	86
4.1.2.1	Language data	86
4.1.2.2	Testimonials about the similarities of the varieties	90
4.1.2.3	Information on the language situation, labour recruitment, and language attitudes	91
4.2	DATABASE DESIGN: THE THREE DATA COLLECTIONS	91
4.3	METHODOLOGICAL CONCERNS AND ANALYTICAL PROCEDURE	93
4.3.1	<i>Identification of potential differences and preparation of the spreadsheets</i>	95
4.3.2	<i>Choice and definition of linguistic variables for the analysis</i>	98
4.3.3	<i>The resulting data spreadsheets</i>	99
4.3.4	<i>Analysis of spreadsheet in terms of diachronic variation and change</i>	99
4.3.4.1	Dates of attestation in analysing the chronology of language change	100
4.3.4.2	Timeline and boxplot approach	103
4.3.4.3	Conditional inference trees	105

4.3.5	<i>Strategies in dealing with the reliability of sources</i>	107
4.4	SUMMARY AND OUTLOOK	111
5	META-LINGUISTIC STATEMENTS ON THE DIVERGENCE OF MPE VARIETIES	113
6	CASE STUDY: DEMONSTRATIVES	117
6.1	THEORETICAL BACKGROUND.....	117
6.2	DEMONSTRATIVES IN MPE TODAY	121
6.3	PREVIOUS DIACHRONIC STUDIES ON DEMONSTRATIVES IN MPE	123
6.4	METHODOLOGICAL CONSIDERATIONS.....	124
6.5	FINDINGS AND DISCUSSION.....	130
6.5.1	<i>General data distribution</i>	130
6.5.2	<i>Diachronic analysis of demonstratives in Solomon Islands Pijin</i>	132
6.5.2.1	Attested forms in SIP	132
6.5.2.2	Timeline and boxplot approach	134
6.5.2.3	Testing for the impact of the years of attestations on the choice of form	136
6.5.3	<i>Diachronic analysis of demonstratives in Bislama</i>	140
6.5.3.1	Attested forms in BIS	140
6.5.3.2	Timeline and boxplot approach	142
6.5.3.3	Testing for the impact of the years of attestations on the choice of form	144
6.5.4	<i>Diachronic analysis of demonstratives in Tok Pisin</i>	148
6.5.4.1	Attested forms in TP	148
6.5.4.2	Timeline and boxplot approach	149
6.5.4.3	Testing for the impact of the years of attestations on the choice of form	150
6.5.5	<i>Diachronic comparative analysis of demonstratives in MPE</i>	153
6.6	SUMMARY	154
7	CASE STUDY: RELATIVE CLAUSES	157
7.1	THEORETICAL BACKGROUND.....	157
7.2	RELATIVE CLAUSES IN MPE TODAY	161
7.3	PREVIOUS DIACHRONIC STUDIES ON RELATIVE CLAUSES IN MPE	163
7.4	METHODOLOGICAL CONSIDERATIONS.....	166
7.5	FINDINGS AND DISCUSSION.....	171
7.5.1	<i>General data distribution</i>	172
7.5.2	<i>Diachronic analysis of relative clauses in Solomon Islands Pijin</i>	174
7.5.2.1	Attested strategies and forms in SIP	174
7.5.2.2	Relativisation in SIP with a focus on strategies	176
7.5.2.2.1	Timeline and boxplot approach	176
7.5.2.2.2	Testing for the impact of the years of attestations on the relativisation strategy	177
7.5.2.3	Relativisation in SIP with a focus on concrete forms used to mark RCs	183
7.5.2.3.1	Timeline and boxplot approach	183
7.5.2.3.2	Testing for the impact of the years of attestations on the choice of form	184
7.5.3	<i>Diachronic analysis of relative clauses in Bislama</i>	187
7.5.3.1	Attested strategies and forms in BIS.....	187
7.5.3.2	Relativisation in BIS with a focus on strategies	188
7.5.3.2.1	Timeline and boxplot approach	188
7.5.3.2.2	Testing for the impact of the years of attestations on the relativisation strategy	189
7.5.3.3	Relativisation in BIS with a focus on concrete forms used to mark RCs	190
7.5.3.3.1	Timeline and boxplot approach	190
7.5.3.3.2	Testing for the impact of the years of attestations on the choice of form	191
7.5.4	<i>Diachronic analysis of relative clauses in Tok Pisin</i>	192
7.5.4.1	Attested strategies and forms in TP	192
7.5.4.2	Relativisation in TP with a focus on strategies	193
7.5.4.2.1	Timeline and boxplot approach	193
7.5.4.2.2	Testing for the impact of the years of attestations on the relativisation strategy	194
7.5.4.3	Relativisation in TP with a focus on concrete forms used to mark RCs.....	196
7.5.4.3.1	Timeline and boxplot approach	196
7.5.4.3.2	Testing for the impact of the years of attestations on the choice of form	196
7.5.5	<i>Diachronic comparative analysis of relative clauses in MPE</i>	196
7.6	SUMMARY	199

8 CASE STUDY: MODALITY	203
8.1 THEORETICAL BACKGROUND.....	203
8.2 MODALITY IN MPE TODAY.....	207
8.3 PREVIOUS DIACHRONIC STUDIES ON MODALITY IN MPE	210
8.4 VOLITION.....	211
8.4.1 <i>Methodological considerations</i>	212
8.4.2 <i>Findings and discussion</i>	214
8.4.2.1 General data distribution.....	214
8.4.2.2 Diachronic analysis of volition in SIP	216
8.4.2.2.1 Attested forms.....	216
8.4.2.2.2 Timeline and boxplot approach	218
8.4.2.2.3 Testing for the impact of the years of attestations on the choice of form	219
8.4.2.3 Diachronic analysis of volition in BIS.....	221
8.4.2.3.1 Attested forms.....	221
8.4.2.3.2 Timeline and boxplot approach	223
8.4.2.3.3 Testing for the impact of the years of attestations on the choice of form	224
8.4.2.4 Diachronic analysis of volition in TP	225
8.4.2.4.1 Attested forms.....	225
8.4.2.4.2 Timeline and boxplot approach	226
8.4.2.4.3 Testing for the impact of the years of attestations on the choice of form	227
8.4.2.5 Diachronic comparative analysis of volition in MPE.....	228
8.5 ABILITATIVE	230
8.5.1 <i>Methodological considerations</i>	230
8.5.2 <i>Findings and discussion</i>	234
8.5.2.1 General data distribution.....	234
8.5.2.2 Diachronic analysis of abilitative in SIP.....	236
8.5.2.2.1 Attested forms.....	236
8.5.2.2.2 Timeline and boxplot approach	239
8.5.2.2.3 Testing for the impact of the years of attestations on the choice of form	240
8.5.2.3 Diachronic analysis of abilitative in BIS	242
8.5.2.3.1 Attested forms.....	242
8.5.2.3.2 Timeline and boxplot approach	243
8.5.2.3.3 Testing for the impact of the years of attestations on the choice of form	244
8.5.2.4 Diachronic analysis of abilitative in TP.....	244
8.5.2.4.1 Attested forms.....	245
8.5.2.4.2 Timeline and boxplot approach	246
8.5.2.4.3 Testing for the impact of the years of attestations on the choice of form	248
8.5.2.5 Diachronic comparative analysis of abilitative in MPE	251
8.6 PERMISSIVE.....	252
8.6.1 <i>Methodological considerations</i>	253
8.6.2 <i>Findings and discussion</i>	255
8.6.2.1 General data distribution.....	255
8.6.2.2 Diachronic analysis of permissive in SIP	256
8.6.2.3 Diachronic analysis of permissive in BIS.....	259
8.6.2.4 Diachronic analysis of permissive in TP.....	260
8.6.2.5 Diachronic comparative analysis of permissive in MPE	260
8.7 SPECULATIVE	262
8.7.1 <i>Methodological considerations</i>	262
8.7.2 <i>Findings and discussion</i>	265
8.7.2.1 General data distribution.....	265
8.7.2.2 Diachronic analysis of speculative in SIP.....	266
8.7.2.2.1 Attested forms.....	267
8.7.2.2.2 Timeline and boxplot approach	267
8.7.2.2.3 Testing for the impact of the years of attestations on the choice of form	268
8.7.2.3 Diachronic analysis of speculative in BIS	271
8.7.2.3.1 Attested forms.....	271
8.7.2.3.2 Timeline and boxplot approach	271
8.7.2.3.3 Testing for the impact of the years of attestations on the choice of form	272
8.7.2.4 Diachronic analysis of speculative in TP.....	272
8.7.2.4.1 Attested forms.....	272
8.7.2.4.2 Timeline and boxplot approach	273
8.7.2.4.3 Testing for the impact of the years of attestations on the choice of form	273

8.7.2.5	Diachronic comparative analysis of speculative in MPE	273
8.8	SUMMARY	276
9	CASE STUDY: SELECTED PREPOSITIONS	279
9.1	THEORETICAL BACKGROUND.....	280
9.2	PREVIOUS DIACHRONIC STUDIES ON PREPOSITIONS IN MPE	284
9.3	PRIMARY METHODOLOGICAL STEPS	287
9.4	SELECTED PREPOSITIONS.....	288
9.4.1	<i>Comitatives and instrumentals</i>	288
9.4.1.1	Theoretical background	288
9.4.1.2	Comitatives and instrumentals in contemporary MPE varieties	290
9.4.1.3	Methodological considerations.....	291
9.4.1.4	Findings and discussion	296
9.4.1.4.1	General data distribution	296
9.4.1.4.2	Comitatives in SIP.....	298
9.4.1.4.2.1	Attested forms.....	298
9.4.1.4.2.2	Timeline and boxplot approach.....	299
9.4.1.4.2.3	Testing for the impact of the years of attestations on the choice of form.....	300
9.4.1.4.3	Instrumentals in SIP	302
9.4.1.4.3.1	Attested forms.....	302
9.4.1.4.3.2	Timeline and boxplot approach.....	303
9.4.1.4.3.3	Testing for the impact of the years of attestations on the choice of form.....	304
9.4.1.4.4	Comitatives and instrumentals in SIP.....	304
9.4.1.4.5	Comitative prepositions in BIS.....	306
9.4.1.4.5.1	Attested forms.....	306
9.4.1.4.5.2	Timeline and boxplot approach.....	307
9.4.1.4.5.3	Testing for the impact of the years of attestations on the choice of form.....	307
9.4.1.4.6	Instrumental prepositions in BIS	308
9.4.1.4.6.1	Attested forms.....	308
9.4.1.4.6.2	Timeline and boxplot approach.....	309
9.4.1.4.6.3	Testing for the impact of the years of attestations on the choice of form.....	310
9.4.1.4.7	Comitatives and instrumentals in BIS	311
9.4.1.4.8	Comitative prepositions in TP	312
9.4.1.4.8.1	Attested forms.....	312
9.4.1.4.8.2	Timeline and boxplot approach.....	314
9.4.1.4.8.3	Testing for the impact of the years of attestations on the choice of form.....	315
9.4.1.4.9	Instrumental prepositions in TP.....	320
9.4.1.4.9.1	Attested forms.....	320
9.4.1.4.9.2	Timeline and boxplot approach.....	320
9.4.1.4.9.3	Testing for the impact of the years of attestations on the choice of form.....	321
9.4.1.4.10	Comitatives and instrumentals in TP	323
9.4.1.4.11	Comparative analysis of comitatives and instrumentals in MPE.....	323
9.4.1.4.11.1	Comparative analysis of comitatives in MPE	324
9.4.1.4.11.2	Comparative analysis of instrumentals in MPE	325
9.4.1.5	Summary	326
9.4.2	<i>Terminatives</i>	328
9.4.2.1	Terminatives in contemporary MPE varieties.....	328
9.4.2.2	Methodological considerations.....	330
9.4.2.3	Findings and discussion	331
9.4.2.3.1	General data distribution	331
9.4.2.3.2	Terminatives in SIP.....	331
9.4.2.3.2.1	Attested forms	331
9.4.2.3.2.2	Timeline approach	333
9.4.2.3.2.3	Testing for the impact of the years of attestations on the choice of form.....	334
9.4.2.3.3	Terminatives in BIS	335
9.4.2.3.3.1	Attested forms	335
9.4.2.3.3.2	Timeline approach.....	336
9.4.2.3.4	Terminatives in TP.....	336
9.4.2.3.4.1	Attested forms	336
9.4.2.3.4.2	Timeline approach.....	337
9.4.2.3.4.3	Testing for the impact of the years of attestations on the choice of form.....	337
9.4.2.3.5	Diachronic comparative analysis of terminative prepositions in MPE	338
9.4.2.4	Summary	340

9.4.3	<i>Adessives</i>	341
9.4.3.1	Adessives in contemporary MPE varieties	342
9.4.3.2	Methodological considerations.....	343
9.4.3.3	Findings and discussion	345
9.4.3.3.1	General data distribution	345
9.4.3.3.2	Diachronic analysis of adessive prepositions in SIP.....	346
9.4.3.3.2.1	Attested forms	346
9.4.3.3.2.2	Timeline and boxplot approach.....	347
9.4.3.3.2.3	Testing for the impact of the years of attestations on the choice of form.....	348
9.4.3.3.3	Diachronic analysis of adessive prepositions in BIS	350
9.4.3.3.4	Diachronic analysis of adessive prepositions in TP	350
9.4.3.3.4.1	Attested forms	350
9.4.3.3.4.2	Timeline and boxplot approach.....	350
9.4.3.3.4.3	Testing for the impact of the years of attestations on the choice of form.....	351
9.4.3.3.5	Diachronic comparative analysis of adessive prepositions in MPE	351
9.4.3.4	Summary	354
10	CONCLUSION	355
10.1	SUMMARY OF THE RESULTS	355
10.2	EVALUATION OF THEORIES ON THE ORIGIN AND DEVELOPMENT OF MPE VARIETIES.....	362
10.3	THEORETICAL IMPLICATIONS FOR THE STABILISATION OF CONTACT VARIETIES	366
10.4	CONCLUDING REMARKS AND OUTLOOK	376
	REFERENCES	381
	APPENDICES	401
	APPENDIX I: ADDITIONAL BOXPLOTS	401
	APPENDIX II: ADDITIONAL CTREES	402
	APPENDIX III: GERMAN SUMMARY	409

Eidesstattliche Erklärung

Ich erkläre: Ich habe die vorgelegte Dissertation selbständig und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder nicht veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht.

Damaris Schäfer

Gießen, den 05.10.2021

Acknowledgments

First, I would like to thank my supervisor Magnus Huber for his support, feedback, encouragement, and guidance throughout the writing process of this thesis. With his invitation to join him to a visit to the colonial archives in Frankfurt, my interest in doing historical research on contact languages arose. He opened the doors for me to network and to exchange thoughts with other creolists by giving me the chance to speak at several conferences. I am also grateful to Joybrato Mukherjee, in whose team I work since 2021 and who agreed to be the second reviewer for my thesis.

I wish to express my deepest gratitude to Viveka Velupillai who gave me constant encouragement and advice. My work has greatly benefited from her theoretical and practical expertise in the field. In addition, she supported me through the ups and downs during the writing process and she always had a sympathetic ear for me.

I also want to express my gratitude to those people that helped me in collecting the early pidgin and creole language data. I am most grateful for the assistance of the librarians and archivists in doing my research. Special thanks go to Katherine Pawley of the Special Collections at the University of Auckland, Rachel Armstrong of the Pacific Research Archives in Canberra, and the executive officer for the Pacific Manuscripts Bureau, Kari James. Moreover, I would like to thank Magnus Huber and Viveka Velupillai for sharing the data they had collected in the Deutsche Kolonialbibliothek until 2015. I am also very thankful for Philip Baker who supported me with data and advice since my MA thesis. I would like to thank Doris Stollberg and Stefan Engelberg for sharing their early Tok Pisin and Bislama dictionaries and Peter Bakker for sharing his early discoveries with me. I am very thankful for the support of the DAAD which enabled me to collect historical data in New Zealand and Australia and, thus, to build the database for this study. Max, many thanks for your formatting help.

My thanks for assistance with various aspects of this project go to a number of people. Thanks go to Christine Jourdan who helped me to sort my research ideas in the early stages of this project. My thanks go out to Sara Roberts who offered her help and shared her knowledge on working with and dating historical data. I would also like to thank Stefan Th. Gries for introducing me to R and for taking his time for a one-on-one.

I am also truly thankful for my colleagues at the Department of English Linguistics at the Justus-Liebig-University in Giessen and for the other PhD students. I would like to particularly point out Barbara Gldenring and Kathrin Kircili who throughout the last two years did not only do a tremendous job as motivators but also turned into wonderful friends. In addition, they offered thoughtful commentary on the form of this work.

Without the support of my family and friends, this work would not have been possible. I am forever grateful for your patience and understanding and for indulging every whim. Mama und Papa, danke, dass ihr es mir ermglicht habt, diesen Weg zu gehen und dass ihr immer ein offenes Ohr in dieser Zeit fr mich hattet. Jennifer, Melissa, und Vanessa, danke, dass ihr mir immer Mut zugesprochen habt.

Finally, my greatest thanks go to my husband Samuel without whose support and patience this work would not have been possible: Danke, dass du, obwohl ich kaum Zeit fr dich hatte, immer sofort zur Stelle warst, wenn ich dich brauchte. Danke fr deine Liebe und unglaubliche Geduld in dieser herausfordernden Zeit.

Figures

Figure 1: Map of Papua New Guinea, Solomon Islands and Vanuatu.....	1
Figure 2: Categorisation of text types according to their proximity to speech (Schneider 2013: 61)	18
Figure 3: Continuum of reliability and influencing factors	20
Figure 4: Solomon Islander letter written by European (Nongosila; Oroar, P. 19.05.1906)	25
Figure 5: Map of the Pacific Islands, showing the place names mentioned in Section 3.1	31
Figure 6: Distribution of German New Guineans, Vanuatians, Solomon Islanders per plantation area	47
Figure 7: Visualisation of influence based on metalinguistic evidence	53
Figure 8: Early Melanesian Pidgin Development in Clark (1979: 48)	76
Figure 9: Extract from Court Record of the trial of Noeli Dili (Tulagi; Utuhia, L. 15.07.1920)	90
Figure 10: Extract from Double murder in the Solomon Islands by Harry Okea (Tulagi; Okea, H. 29.12.1911).....	90
Figure 11: Data retrieval & analytical steps.....	94
Figure 12: Extract from spreadsheet No. 1.3	97
Figure 13: Time lag influencing factors	101
Figure 14: Timeline of RC ST_FORM variants in BIS	103
Figure 15: Boxplot of RC ST_FORM variants in BIS	104
Figure 16: Different uses of demonstratives (Levinson 2018: 10)	120
Figure 17: Distribution of demonstrative datapoints across time per variety (based on 1,803 tokens).....	131
Figure 18: Relative frequencies of demonstrative variants per variety	131
Figure 19: Timeline of demonstrative variants in SIP.....	134
Figure 20: Boxplot of demonstrative variants in SIP	135
Figure 21: Boxplot of demonstrative variants in SIP depending on their syntactic context	136
Figure 22: Conditional inference tree for demonstratives in SIP ST_FORM~YEAR_DET	137
Figure 23: Conditional inference tree for demonstratives in SIP ST_FORM~YEAR_DET+FEATURE	138
Figure 24: Timeline of demonstrative variants in Bislama	142
Figure 25: Boxplot of demonstrative variants in Bislama	143
Figure 26: Boxplot of demonstrative variants in BIS depending on their syntactic context.....	144
Figure 27: Conditional inference tree for demonstratives in BIS YEAR_DET~ST_FORM	144
Figure 28: Conditional inference tree for demonstratives in BIS YEAR_DET~ST_FORM+FEATURE.....	145
Figure 29: Conditional inference tree for demonstratives in BIS ST_FORM~YEAR_DET+FEATURE+TEXT_	
TYPE_3.....	147
Figure 30: Timeline of demonstrative variants in TP	149
Figure 31: Boxplot of demonstrative variants in TP	150
Figure 32: Boxplot of demonstrative variants in TP depending on their syntactic context	150
Figure 33: Conditional inference tree for demonstratives in TP ST_FORM~YEAR_DET	152
Figure 34: Map of Melanesian Islands, showing attestations of demonstrative ST_FORM variants	156
Figure 35: Distribution of relative clause datapoints across time per variety	172
Figure 36: Relative frequencies of RC STRATEGY variants per variety	173
Figure 37: Relative frequencies of RC ST_FORM variants per variety	174
Figure 38: Timeline of RC STRATEGIES in SIP	177
Figure 39: Boxplot of RC STRATEGIES in SIP.....	177
Figure 40: Conditional inference tree for relative clauses in SIP STRATEGY~YEAR_DET	180
Figure 41: Conditional inference tree for relative clauses in SIP STRATEGY~YEAR_DET+FEATURE	181
Figure 42: Conditional inference tree for relative clauses in SIP STRATEGY~YEAR_DET+FEATURE+TXT_	
TYPE_3.....	182
Figure 43: Timeline of RC ST_FORM variants in SIP	183
Figure 44: Boxplot of RC ST_FORM variants in SIP	183
Figure 45: Conditional inference tree for relative clauses in SIP ST_FORM~YEAR_DET.....	184
Figure 46: Conditional inference tree for relative clauses in SIP ST_FORM~YEAR_DET+FEATURE.....	185
Figure 47: Conditional inference tree for relative clauses in SIP ST_FORM~YEAR_DET+FEATURE+TXT_	
TYPE_3.....	186
Figure 48: Timeline of RC STRATEGIES in BIS.....	188
Figure 49: Boxplot of RC STRATEGIES in BIS	189
Figure 50: Conditional inference tree for relative clauses in BIS STRATEGY~YEAR_DET	189
Figure 51: Conditional inference tree for relative clauses in BIS STRATEGY~YEAR_DET+FEATURE.....	190
Figure 52: Timeline of RC ST_FORM variants in BIS	191
Figure 53: Boxplot of RC ST_FORM variants in BIS	191
Figure 54: Conditional inference tree for relative clauses in BIS ST_FORM~YEAR_DET	192
Figure 55: Timeline of RC STRATEGIES in TP	193
Figure 56: Boxplot of RC STRATEGIES in TP.....	193
Figure 57: Conditional inference tree for relative clauses in TP STRATEGY~YEAR_DET	194

Figure 58: Conditional inference tree for relative clauses in TP STRATEGY~YEAR_DET+FEATURE	195
Figure 59: Conditional inference tree for relative clauses in TP STRATEGY~YEAR_DET+FEATURE+TXT_	
TYPE_3	195
Figure 60: Timeline of RC ST_FORM variants in TP	196
Figure 61: Boxplot of RC ST_FORMs in TP	196
Figure 62: Conditional inference tree for relative clauses in TP ST_FORM~VARIETY+YEAR_DET+FEATURE	
.....	199
Figure 63: Map of Melanesian Islands, showing attestations of relative clause ST_FORM variants	201
Figure 64: Classification of modality	204
Figure 65: Distribution of volition datapoints across time per variety	215
Figure 66: Relative frequencies of volition M_MORPH variants per variety	215
Figure 67: Timeline of volition variants in SIP	218
Figure 68: Boxplot of volition variants in SIP	219
Figure 69: Conditional inference tree for volition in SIP ST_FORM~YEAR_DET	219
Figure 70: Conditional inference tree for volition in SIP ST_FORM~YEAR_DET+FEATURE	220
Figure 71: Conditional inference tree for volition in SIP ST_FORM~YEAR_DET+FEATURE+TEXT_TYPE_3	
.....	221
Figure 72: Timeline of volition variants in BIS	223
Figure 73: Boxplot of volition variants in BIS	224
Figure 74: Conditional inference tree for volition in BIS ST_FORM~YEAR_DET+FEATURE+TXT_TYPE_3	225
Figure 75: Timeline of volition variants in TP	226
Figure 76: Boxplot of volition variants in TP	226
Figure 77: Conditional inference tree for volition in TP ST_FORM~YEAR_DET+FEATURE+TXT_TYPE_3	227
Figure 78: Map of Melanesian Islands, showing attestations of volition M_MORPH variants	229
Figure 79: Distribution of ability datapoints across time per variety	234
Figure 80: Relative frequencies of abilitative M_MORPH variants per variety	235
Figure 81: Relative frequencies of abilitative variants in affirmative contexts per variety	236
Figure 82: Relative frequencies of abilitative variants in negated contexts per variety	236
Figure 83: Timeline of ability variants in SIP	239
Figure 84: Boxplot of ability variants in SIP based on M_MORPH	240
Figure 85: Conditional inference tree for ability in SIP ST_FORM~YEAR_DET+FEATURE	241
Figure 86: Conditional inference tree for ability in SIP ST_FORM~YEAR_DET+FEATURE+TXT_TYPE_3	241
Figure 87: Timeline of ability variants in BIS	243
Figure 88: Boxplot of ability variants in BIS based on M_MORPH	244
Figure 89: Conditional inference tree for ability in BIS ST_FORM~YEAR_DET	244
Figure 90: Timeline of ability variants in TP	247
Figure 91: Boxplot of ability variants in TP based on M_MORPH	248
Figure 92: Conditional inference tree for ability in TP ST_FORM~YEAR_DET	248
Figure 93: Conditional inference tree for ability in TP ST_FORM~YEAR_DET+FEATURE+TXT_TYPE_3	
+AUTH_NAME	250
Figure 94: Map of Melanesian Islands, showing attestations of ability M_MORPH variants	252
Figure 95: Distribution of permissive datapoints across time per variety	255
Figure 96: Relative frequencies of permissive M_MORPH variants per variety	256
Figure 97: Timeline of permissive/prohibitive variants in SIP	257
Figure 98: Boxplot of permissive/prohibitive variants in SIP based on M_MORPH	258
Figure 99: Conditional inference tree for permissive/prohibitive in SIP ST_FORM~YEAR_DET	258
Figure 100: Conditional inference tree for permissive/prohibitive in SIP	
ST_FORM~YEAR_DET+TXT_TYPE_3+AUTH_NAME	259
Figure 101: Timeline of permissive/prohibitive variants in BIS	259
Figure 102: Timeline of permissive/prohibitive variants in TP	260
Figure 103: Conditional inference tree for permissive/prohibitive in the three MPE varieties ST_FORM~YEAR_	
DET +VARIETY+ TXT_TYPE+AUTH_NAME	261
Figure 104: Map of Melanesian Islands, showing attestations of permission M_MORPH variants	262
Figure 105: Distribution of speculative datapoints across time per variety	266
Figure 106: Relative frequencies of speculative variants per variety	266
Figure 107: Timeline of speculative modality variants in SIP	267
Figure 108: Boxplot of speculative modality variants in SIP	268
Figure 109: Conditional inference tree for speculative in SIP ST_FORM~YEAR_DET	268
Figure 110: Conditional inference tree for speculative in SIP ST_FORM~YEAR_DET+STRUCTURE	269
Figure 111: Conditional inference tree for speculative in SIP ST_FORM~YEAR_DET+STRUCTURE+TXT_	
TYPE_3+AUTH_NAME	270
Figure 112: Timeline speculative modality variants in BIS	271

Figure 113: Boxplot of speculative modality variants in BIS.....	272
Figure 114: Timeline of speculation modality variants in TP	273
Figure 115: Boxplot of speculation modality variants in TP.....	273
Figure 116: Conditional inference tree for speculative in the three MPE varieties ST_FORM~YEAR_DET +VARIETY+ STRUCTURE+TXT_TYPE_3.....	274
Figure 117: Map of Melanesian Islands, showing attestations of speculative M_MORPH variants.....	276
Figure 118: Prepositions in contemporary MPE varieties (Tryon & Charpentier 2004: 397).....	287
Figure 119: Types of instruments in early MPE	293
Figure 120: Distribution of comitative & instrumental datapoints across time per variety	297
Figure 121: Relative frequencies of instrumental prepositional variants per variety	297
Figure 122: Relative frequencies of comitative prepositional variants per variety	298
Figure 123: Timeline of comitative variants in SIP	300
Figure 124: Boxplot of comitative variants in SIP.....	300
Figure 125: Conditional inference tree for comitative prepositions in SIP ST_FORM~YEAR_DET	300
Figure 126: Conditional inference tree for comitative prepositions in SIP ST_FORM~YEAR_DET+TYPE	301
Figure 127: Timeline of instrumental variants in SIP	303
Figure 128: Boxplot of instrumental variants in SIP.....	303
Figure 129: Ctree analysis of instrumentals in SIP resulting in a stacked boxplot.....	304
Figure 130: Conditional inference tree for comitative & instrumental prepositions in SIP_ST_FORM~YEAR_DET +FEATURE.....	304
Figure 131: Conditional inference tree for comitative & instrumental prepositions in SIP ST_FORM~YEAR_DET+ FEATURE+TXT_TYPE_3+AUTH_NAME	305
Figure 132: Timeline of comitative variants in BIS.....	307
Figure 133: Boxplot comitative variants in BIS.....	307
Figure 134: Conditional inference tree for comitative prepositions in BIS ST_FORM~YEAR_DET.....	308
Figure 135: Timeline of instrumental variants in BIS.....	309
Figure 136: Boxplot of instrumental variants in BIS	309
Figure 137: Conditional inference tree for instrumental prepositions in BIS ST_FORM~YEAR_DET.....	310
Figure 138: Conditional inference tree for instrumental prepositions in BIS ST_FORM~YEAR_DET+TXT_ TYPE_3+AUTH_NAME	310
Figure 139: Conditional inference tree for instrumental & comitative prepositions in BIS ST_FORM~YEAR_DET +FEATURE.....	311
Figure 140: Timeline of comitative variants in TP	314
Figure 141: Boxplot of comitative variants in TP.....	315
Figure 142: Conditional inference tree for comitative prepositions in TP ST_FORM~YEAR_DET	315
Figure 143: Conditional inference tree for comitative prepositions in TP ST_FORM~YEAR_DET+TXT_TYPE_3	316
Figure 144: Conditional inference tree for comitative prepositions in TP ST_FORM~YEAR_DET+TYPE	318
Figure 145: Conditional inference tree for comitative prepositions in TP ST_FORM ~ YEAR_DET + TYPE + TXT_TYPE_3	319
Figure 146: Timeline instrumental variants in TP	321
Figure 147: Boxplot of instrumental variants in TP.....	321
Figure 148: Conditional inference tree for instrumental prepositions in TP ST_FORM~YEAR_DET	321
Figure 149: Conditional inference tree for instrumental prepositions in TP ST_FORM~YEAR_DET+TXT_ TYPE_3.....	322
Figure 150: Conditional inference tree for instrumental prepositions in the three MPE varieties ST_FORM~ YEAR_DET + VARIETY	326
Figure 151: Map of Melanesian Islands, showing attestations of instrumental prepositional variants	327
Figure 152: Map of Melanesian Islands, showing attestations of comitative prepositional variants	327
Figure 153: Distribution of terminative prepositional datapoints across time per variety	331
Figure 154: Timeline of terminative variants in SIP	334
Figure 155: Conditional inference tree for terminatives in SIP ST_FORM~YEAR_DET	334
Figure 156: Timeline of terminative variants in BIS.....	336
Figure 157: Timeline of terminative variants in TP	337
Figure 158: Conditional inference tree for terminatives in the three MPE varieties ST_FORM~YEAR_DET +VARIETY	338
Figure 159: Conditional inference tree for terminatives in the three MPE varieties ST_FORM~YEAR_DET +VARIETY+TXT_TYPE_3+AUTH_NAME	339
Figure 160: Map of Melanesian Islands, showing attestations of terminative prepositional variants	341
Figure 161: Distribution of adessive prepositional datapoints across time per variety	345
Figure 162: Relative frequencies of adessive prepositional forms in MPE	346
Figure 163: Timeline of adessive prepositional variants in SIP	347

Figure 164: Boxplot of adessive prepositional variants in SIP	348
Figure 165: Conditional inference tree for adessive prepositions in SIP ST_FORM~YEAR_DET	349
Figure 166: Conditional inference tree for adessive prepositions in SIP ST_FORM~YEAR_DET+ TXT_TYPE_3	349
Figure 167: Timeline of adessive prepositional variants in TP	351
Figure 168: Boxplot of adessive prepositional variants in TP	351
Figure 169: Conditional inference tree for adessive prepositions in the three MPE varieties ST_FORM~YEAR_DET+VARIETY	352
Figure 170: Conditional inference tree for adessive prepositions in the three MPE varieties ST_FORM~YEAR_DET+VARIETY +TXT_TYPE_3	353
Figure 171: Map of Melanesian Islands, showing attestations of adessive prepositional variants	354
Figure 172: Significant years for the choice of form based on ST_FORM~YEAR_DET	356
Figure 173: Significant years for the choice of form based on ST_FORM~YEAR_DET+ TYPE/FEATURE/STRUCTURE	356
Figure 174: Significant years for the choice of form based on ST_FORM~YEAR_DET+ TYPE/FEATURE/STRUCTURE+TXT_TYPE_3	358
Figure 175: Significant years for the choice of form based on ST_FORM ~ YEAR_DET+ TYPE/FEATURE/STRUCTURE+TXT_TYPE_3+AUTH_NAME	358
Figure 176: Complexity of contact language feature development (based on evidence of MPE varieties)	373
Figure 177: Boxplot of ability variants in SIP based on ST_FORM_F	401
Figure 178: Boxplot of ability variants in BIS based on ST_FORM_F	401
Figure 179: Boxplot of ability variants in TP based on ST_FORM_F	401
Figure 180: Conditional inference tree for relative clauses in BIS STRATEGY~YEAR_DET+FEATURE+ TXT_TYPE_3+AUTH_NAME	402
Figure 181: Conditional inference tree for relative clauses in BIS ST_FORM~YEAR_DET+FEATURE+ TXT_TYPE_3+AUTH_NAME	403
Figure 182: Conditional inference tree for speculative in SIP ST_FORM~YEAR_DET+STRUCTURE+ TXT_TYPE_3	404
Figure 183: Conditional inference tree for comitative & instrumental prepositions in BIS ST_FORM ~YEAR_DET+FEATURE+TXT_TYPE_3+ AUTH_NAME	405
Figure 184: Conditional inference tree for instrumental prepositions in TP ST_FORM~YEAR_DET+TYPE	406
Figure 185: Conditional inference tree for adessive prepositions in the three MPE varieties ST_FORM ~ YEAR_DET+VARIETY+TXT_TYPE_3+AUTH_NAME	407

Tables

Table 1: Morphosyntactic features investigated in the present study	8
Table 2: Number of Melanesian labourers on plantations in Queensland, Samoa, Fiji and New Caledonia	47
Table 3: Number of sources per archive per variety	86
Table 4: Orthographic realisations of <i>this fellow</i> and <i>one time</i>	92
Table 5: Linguistic variables	98
Table 6: Variables in spreadsheets	99
Table 7: Contrasting juxtaposition of timeline vs. boxplot approach	105
Table 8: Variable TXT_TYPE	109
Table 9: Linguistic coding for demonstratives.....	124
Table 10: Linguistic coding for relative clauses	171
Table 11: Tabular comparison of selected modality markers.....	210
Table 12: Linguistic coding for volition	212
Table 13: Frequencies of attested volition forms in SIP.....	216
Table 14: Frequencies of attested volition forms in BIS	222
Table 15: Frequencies of attested volition forms in TP.....	225
Table 16: Linguistic coding for ability	232
Table 17: Linguistic coding for permission	253
Table 18: Linguistic coding for speculation	262
Table 19: The semantic system of adpositions (Hagège 2010: 261-262)	283
Table 20: Development of Bislama prepositions (adapted from Crowley 1990a: 266)	285
Table 21: Linguistic coding for comitatives and instrumentals.....	292
Table 22: Fine-grained classification categories	294
Table 23: Linguistic coding for terminatives	330
Table 24: Linguistic coding for adessive prepositions	345
Table 25: Pre-1950 variants versus contemporary forms.....	357

List of abbreviations

ABES	abessive	LL	low language
ABIL	abilitative	LOC	locative
ADES	adessive	MASC	masculine
adn	adnominal	MODIF	macro-category modifier
Adp	adposition	MOT	motivative
ADV	adverbial	MPE	Melanesian Pidgin English
APiCS	Atlas of Pidgin and Creole Language Structures	MPP	Maritime Polynesian Pidgin
ART	article	N	noun
ASP	aspect marker	NCOMPL	noun complementiser
ATT	attributive	NCPF	New Caledonia Pidgin French
BIS	Bislama	NEG	negation
BSIP	British Solomon Islands Protectorate	negvolit	negated volition
COM	comitative	NGP	New Guinea Pidgin
COMP	comparative	NLA	National Library Australia
COMPL	completive	NP	noun phrase
CONJ	conjunction	NPC	noun phrase complement
CONT	continuity	NPm	noun phrase modifying
COP	copula	NSWPE	New South Wales Pidgin English
DEF	definite article	OBLIG	obligatory
DEM	demonstrative	PAP	Papua
DKB	Deutsche Kolonialbibliothek	PART	participle
DNG	Deutsch Neuguinea	PC	purpose clause
DU	dual	PE	Pidgin English
DUB	dubitative	PERM	permissive
DUR	durative	PERT	pertentive
EXCL	exclusive	PFV	perfective
FELO	Far Eastern Liaison Office	PL	plural
FLEx	Fieldworks Language Explorer	PM	predicate marker
FOC	focaliser	PMB	Pacific Manuscript Bureau
FP	Fiji Pidgin	POSS	possessive
FUT	future	PREP	preposition
GEN	genitive	PROG	progressive
HAB	habitual	PROH	prohibitive
HL	high language	prohib	prohibition
INABIL	inabilitative	pron	pronominal
INCL	inclusive	PST	past
INDEF	indefinite	PURP	purposive
INF	infinitive-like particle	QKM	Queensland Kanaka Mission
INST	instrumental	QLD	Queensland

List of abbreviations

QPPE	Queensland Plantation Pidgin English	temp	temporal
RC	relative clause	TERM	terminative
red	reduplicated	TMA	tense-mood-aspect
REFL	reflexive	TP	Tok Pisin
REL	relativiser	TR	transitive marker
rel.part	relative particle	TRI	trial
RES	resumptive pronoun	V	verb
SBJ	subject	VNC	variability-based neighbour clustering
SG	singular	VOL	volition
SIP	Solomon Islands Pijin	volit	volition
spat	spatial	WALS	World Atlas of Language Structures
SPEC	speculative	3	third person
SPPE	Samoan Plantation Pidgin English	2	second person
SRP	subject-referencing-pronoun	1	first person
SSEM	South Sea Evangelical Mission	Ø	zero
StE	Standard English	-	gap

Abstract

The present study analyses *when* and *how* the contact languages Tok Pisin (TP), Bislama (BIS) and Solomon Islands Pijin (SIP) developed out of Melanesian Pidgin English (MPE). Even though it has been previously assumed that the end of the labour trade represents the most decisive factor for the individual development of the three varieties (cf. Mühlhäusler 1978; Clark 1979; Jourdan 1985; Keesing 1988; Baker 1993; Jourdan & Keesing 1997; Sankoff 2021+), the current scientific knowledge still raises questions about the concrete time period in which the varieties localised, as well as about the linguistic features involved. To date assumptions have been predominantly grounded on extralinguistic events. A comparative study that is based on historical linguistic data and that combines qualitative and quantitative-statistical methods has not been conducted so far but is necessary to understand the linguistic development of the varieties.

To fill this academic void, the present study is based on historical linguistic data which was collected inter alia in the *German Colonial Archives*, the *Pacific Research Archives*, the *Western Pacific Archives*, and the *Pacific Manuscript Bureau*, and it uses qualitative and quantitative-statistical methods to learn *when* differences developed in the three varieties and *whether*, and if so, *to what extent* these differences can be traced back to the end of the labour trade. The collected linguistic data covers the time from the earliest detectable attestations until 1950.

In order to identify morphosyntactic differences that existed between the three early varieties, a morpheme-by-morpheme analysis was conducted and four linguistic categories – demonstratives, relative clauses, modality markers and prepositions – were selected for further analysis. To visualise how the features developed across time, boxplots were generated in RStudio (RStudio 2019) and timelines were created using Excel. To investigate whether the year of attestation had an impact on the realisation of a feature in the varieties, *conditional inference trees* (*ctrees*) were used. It was also tested whether other factors, such as the text type and author, had an impact on the choice of form as well.

While initially 41 significant time-based splits were identified with the help of *ctrees*, this number decreased when further predictor variables were considered so that when including all possible predictor variables, only nine dates remained significant. In addition, the results of the analysis showed that we cannot generalise per se when a divergence and/or stabilisation of the three MPE varieties took place. Even if the *ctree* analyses revealed that most of the statistically significant changes in the three varieties dated to the first half of the 20th century, which may be an indicator that the changes in the choice of feature forms might have been propelled by the end of the labour trade, the significant dates, in the first instance, show significant *changes* in the data.

They do not specify per se whether the respective changes involve a reduction of variants and/or the introduction of new, diverging forms. Furthermore, the study showed that a differentiation between the notions of *divergence* and *stabilisation* is required. Though most attested forms only began to stabilise after the end of the labour trade with overseas plantations, the attestation of forms such as *might* (dubitative modality), *with him* (comitative), *catch him* (terminative), *close up long* (adessive), and *where* (relative clause particle), which were exclusively attested in SIP and BIS and not in TP, made clear that the labour recruitment years (and not the end of the labour trade) seem to have had a major impact on the emergence of diverging forms. In addition, the study revealed that, as in non-pidgins and non-creoles, the forms used to encode the individual morphosyntactic features developed and stabilised at different points in time. This shows that the divergence and stabilisation was a gradual rather than abrupt process.

In sum, the qualitative and quantitative-statistical analysis of historical linguistic data revealed that the labour recruitment histories and the concomitant impact of overseas plantation pidgins, the influence of mission varieties, substrate reinforcement, and the degree of exposure to the lexifier can be regarded as the major reasons for *diverging forms* in the three varieties. In contrast, the end of the labour trade, the spread and use of the varieties to further domains in the home areas, and substrate influence represent major reasons identified for the *stabilisation* of each of the individual varieties and thus, for the reduction of – or at least a preference of – particular variants.

The findings pose an important contribution and advance for the field of creolistics since they provide information about the processes involved in the origin and development of contact languages, and about grammaticalisation processes and language universals. The results suggest that it is likely that several of the mechanism which have been put forward in individual theories were involved and interacted during the emergence and stabilisation of contact varieties and that none of the existing theories alone is sufficient to explain the complex development of the MPE varieties.

Abstract German

Die vorliegende Studie untersucht die Fragestellung, *wann* und *wie* sich die Kontaktsprachen Tok Pisin (TP), Bislama (BIS) und Solomon Islands Pijin (SIP) aus dem melanesischen Pidgin-Englisch in ihren individuellen Formen herausentwickelt haben und welche sprachstrukturellen Unterscheidungen im diachronen Verlauf vorhanden sind. Auch wenn die Beendigung des Arbeiterhandels im Pazifik meist als ausschlaggebend für die individuelle Entwicklung der drei Varietäten angenommen wird (vgl. Mühlhäusler 1978; Clark 1979; Jourdan 1985; Keesing 1988; Baker 1993; Jourdan & Keesing 1997; Sankoff 2021+), weist der Wissensstand über die Entwicklung und Entstehung der drei Kontaktsprachen bislang Fragen hinsichtlich des genauen Zeitraumes und der linguistischen Veränderungsparameter auf. Bislang wurden Annahmen mit extralinguistischen Ereignissen begründet – eine auf frühem Sprachmaterial vergleichende qualitative und quantitativ-statistische Studie, die zur Klärung und Offenlegung des linguistischen Sachverhalts unabdingbar ist, blieb jedoch aus.

Um diese Forschungslücke zu schließen, wurde in der vorliegenden Arbeit historisches Sprachmaterial, welches unter anderem in der *Deutschen Kolonialbibliothek*, den *Pacific Research Archives*, den *Western Pacific Archives*, und dem *Pacific Manuscript Bureau* gesammelt wurde, qualitativ und quantitativ-statistisch untersucht, um herauszufinden, wann sich Unterschiede in den Varietäten entwickelt haben und inwiefern diese auf das Ende des Arbeiterhandels zurückgeführt werden können. Das gesammelte Sprachmaterial deckt den Zeitraum von den am frühesten auffindbaren Belegen bis 1950 ab.

Um sprachstrukturelle Unterschiede der Varietäten zu identifizieren, wurde eine interlineare Morphemanalyse durchgeführt und vier linguistische Kategorien – Demonstrativpronomen, Relativsätze, Modalitätsmarker und Präpositionen – zur weiteren Analyse ausgewählt. Zur Veranschaulichung der Sprachmerkmale in ihrem Zeitverlauf und um Veränderungen in den Varietäten festzustellen, wurden sowohl Boxplots in RStudio (RStudio 2019) generiert, als auch Zeitstrahlen mit Hilfe von Excel erstellt. Um herauszufinden, inwiefern beobachtete Veränderungen in den Kodierungsformen auf das Jahr der Attestierung zurückgeführt werden können beziehungsweise ob das Jahr der Attestierung einen statistisch signifikanten Einfluss auf die Realisierung hat, wurden *Conditional Inference Trees* (*Ctrees*) verwendet. Zudem wurde auch der Einfluss anderer Parameter, wie z.B. des Texttyps und des Autors untersucht, welche sich, gemäß der Analyse, als weitere Einflussfaktoren erwiesen haben.

So konnten mit Hilfe der *Ctrees* zunächst 41 Zeitpunkte identifiziert werden, an denen die Daten signifikante Änderungen aufzeigten. Die Anzahl verringerte sich jedoch, sobald zusätzliche Einflussvariablen, wie z.B. der Texttyp und der Autor mit berücksichtigt wurden,

sodass bei Berücksichtigung aller potentiellen Einflüsse nur noch neun signifikante Zeitpunkte verblieben. Die Ergebnisse zeigten zudem, dass nicht per se generalisiert werden kann, wann die individuelle Entwicklung der drei Varietäten stattgefunden hat. Auch wenn die Ctree-Analyse ergab, dass die meisten signifikanten Veränderungen auf die erste Hälfte des 20. Jahrhunderts datiert werden können, was darauf hinweisen könnte, dass das Ende des Arbeiterhandels tatsächlich der Auslöser für die Individualisierung der Varietäten war, so muss bedacht werden, dass die in Ctrees identifizierten Splits in erster Linie auf signifikante Veränderungen in den Daten hinweisen. Sie geben somit noch keinen Aufschluss darüber, inwiefern diese Veränderungen mit einer Verringerung von Variation und/oder neuen Formen einhergehen. Die Studie macht zudem deutlich, dass zwischen den Prozessen *Individualisierung* und *Stabilisierung* differenziert werden muss. Auch wenn sich die meisten Formen erst nach Ende des Arbeiterhandels stabilisierten, haben Formen wie *might* (dubitativ Modalität), *with him* (Komitativ), *catch him* (Terminativ), *close up long* (Adessiv), und *where* (Relativsatzpartikel), welche ausschließlich in SIP und BIS, nicht jedoch in TP vorkamen, gezeigt, dass die Rekrutierung auf überseeische Plantagen einen großen Einfluss auf divergierende Formentwicklungen hatte. Zudem wurde deutlich, dass für die untersuchten Sprachmerkmale unterschiedliche Zeitpunkte signifikante Änderungen hervorriefen. Dies spricht dafür, dass wie auch bei natürlichen Sprachen (d.h. Nicht-Pidgin und Nicht-Kreolsprachen) die Entwicklung und Stabilisierung von einzelnen grammatikalischen Merkmalen zu unterschiedlichen Zeitperioden und somit graduell stattfand.

Zusammenfassend hat die qualitative und quantitativ-statistische Analyse des historischen Sprachmaterials gezeigt, dass sowohl die unterschiedlichen Arbeiter-/Rekrutierungshistorien der Gebiete und der damit einhergehende Einfluss von Plantagenkontaktsprachen, als auch der Einfluss von Missionen, Substrateinfluss und die Menge des Kontaktes mit dem Lexifier zur *Individualisierung* der Varietäten beigetragen haben. Als Gründe für die *Stabilisierung* der Varietäten wurden das Ende des Arbeiterhandels, die Verbreitung und Verwendung der Varietäten in den Heimatgebieten, und Substrateinfluss identifiziert.

Die Ergebnisse der Dissertation bedeuten einen wichtigen Fortschritt für den Erkenntnisstand der Varietätenforschung und Kreolistik, da sie Aufschluss über den Ursprung und die Entwicklung von Kontaktsprachen, Grammatikalisierungsprozesse und Sprachuniversalien geben. Sie zeigen deutlich, dass die Prozesse, die zur Entstehung von Pidgin- und Kreolsprachen führten, weitaus komplexer waren, als einzelne Theorien bislang vermuten ließen und es wahrscheinlich ist, dass ein Zusammenspiel der verschiedenen Theorien und Mechanismen zur Entwicklung und Stabilisierung der Varietäten führte.

1 Introduction

Die Babylonier, die an Sprachverwirrung zugrunde gingen, kannten nicht das Pidgin-Englisch; sonst wäre es ihnen nicht passiert!

(Schellong 1934: 98)

1.1 General

Melanesia is characterised by a high degree of language diversity. Dutton (2006: 207) describes it as “linguistically one of the most diverse areas of the world, if not the most diverse”. In Papua New Guinea alone, more than 840 languages are spoken; in Vanuatu there are around 110 living languages and in the Solomon Islands there are around 73 living languages (Ethnologue 2019, <https://www.ethnologue.com>, last access 30 November 2020). To communicate in such a linguistically diverse area, the inhabitants of the island groups make use of the contact language *Melanesian Pidgin English*.

Melanesian Pidgin English is the cover term used to refer to the Pidgin English varieties spoken in the Melanesian states Papua New Guinea, Vanuatu, and the Solomon Islands. It has three national varieties which are *Tok Pisin* spoken in Papua New Guinea, *Bislama* spoken in Vanuatu and *Solomon Islands Pijin* spoken in the Solomon Islands (cf. Figure 1).



Figure 1: Map of Papua New Guinea, Solomon Islands and Vanuatu¹

Melanesian Pidgin English can be regarded as the offspring of the trade jargons and plantation pidgins that served as the main medium of intercommunication between islanders and foreigners

¹ If not otherwise indicated, the maps in the present study were made in Microsoft Excel with the Bing Map add-in. Bougainville and Buka are highlighted in purple as they differ from the other Solomon Islands. Geographically they belong to the Solomon Islands, politically they represent an autonomous region of Papua New Guinea. Therefore, the Pidgin English spoken in Bougainville and Bouka is considered to be closer to Tok Pisin than to Solomon Islands Pijin.

in the Pacific in the 19th century.² Although the first contact between Europeans and Pacific Islanders dates much earlier, the 19th century can be regarded as the decisive century for increasing intercommunication between the demographic groups. While the initial encounters are assumed to have led to a range of trading jargons, more intensive contact arose with the establishment of plantations in Queensland, Samoa, New Caledonia and Fiji from 1860 onwards. To satisfy the high labour demand, labour was recruited from adjacent island groups – inter alia from the Solomon Islands, Papua New Guinea and Vanuatu. Due to the high language diversity in the Pacific area, islanders with very diverse linguistic backgrounds came into contact on the plantations and had to communicate not only with each other but also with their European (“white”) supervisors.³ Out of these contact situations, several plantation pidgins developed which were then brought back to the labourers’ places of origin when their contracts expired.

Based on the socio-historical circumstances, it can be assumed that the pidgin varieties which were imported by returning labourers to Papua New Guinea, the Solomon Islands and Vanuatu were quite similar. At the same time it is possible that pidgins spoken on different plantations in different areas may have been distinct. If a high number of labourers from one region went to Samoan plantations and a high proportion of labourers from another region went to Queensland plantations, distinct features in the varieties may have developed. There is further the chance that varieties spoken on the plantations differed but that it came to a process of *levelling* and, along with it, homogenisation with the return of the labourers.

At the beginning of the 20th century, the indentured labour system in the Pacific came to an end, and with it, the intense contact between Pacific Islanders from various areas. The end of the labour trade, and, as a consequence thereof, the isolation of the three areas under investigation, is assumed to have furthered the individual development of the three varieties.

Several scholars agree with the assumption that the end of the labour trade was *one* of the *decisive* factors for the development of Melanesian Pidgin English into its three sub-varieties Tok Pisin, Bislama and Solomon Islands Pijin (cf., for instance, Mühlhäusler 1978; Clark 1979; Jourdan 1985; Keesing 1988; Baker 1993; Jourdan & Keesing 1997; Sankoff 2021+). However, a diachronic, comparative analysis including language data of all three varieties based on qualitative as well as statistical methods has so far been lacking.

² Contact jargons were already in use before the first Europeans arrived. The Pacific is one of the linguistically most diverse areas in the world and even inhabitants of neighbouring villages did not necessarily speak the same language. Thus, local contact varieties had already developed and were in use before the Europeans arrived (cf. Dutton 1996a; Mühlhäusler et al. 1996).

³ At the time the term “white” was used to refer to Europeans and people of European descent. As this term is ideologically loaded, the term “European” will be used instead with which I refer to people of European ancestry.

1.2 Previous research

Most historical studies discussing the development of Solomon Islands Pijin, Tok Pisin and Bislama have focused on a single variety only. Peter Mühlhäusler (1978, 1979, 1985a, 1985b) conducted impressive historical research studying the origin and development of Tok Pisin. In terms of Solomon Islands Pijin, historical studies have been conducted inter alia by Christine Jourdan (1989, 1995, 1996) and Judith Bennett (1987). Jourdan analyses Solomon Islands Pijin from a predominantly language sociological perspective, while Bennett's approach is historical and non-linguistic. Terry Crowley (1990a) has done research on the history and development of Bislama and treats the grammatical development of the variety in detail from the 1840s to the present.

Variety-specific as well as comparative historical studies have mostly been restricted to the question of to what extent Queensland and Samoan plantation pidgins influenced the individualisation of the varieties. They further focus on just one or two of the varieties (cf. Mühlhäusler 1978), mix historical data with post-1970 language data of surviving speakers (cf. Mühlhäusler 1978; Keesing 1988), and/or focus only on features that are different from standard varieties of English (cf. Baker 1993). Even though the studies claim Samoan and/or Queensland plantation pidgins to be precursors of Solomon Islands Pijin, Bislama and Tok Pisin, they also assume that the end of the labour trade was the starting point for the localisation and the individual feature development of the respective varieties (cf. Mühlhäusler 1978; Clark 1979; Jourdan 1985; Keesing 1988; Baker 1993; Jourdan & Keesing 1997). Yet, these assumptions are based on extralinguistic events only. In how far the varieties in fact further standardised and localised with the end of the labour trade has not been studied so far.

There are also several studies that use extralinguistic evidence as the indicator for when the individual development of the varieties took place, and that connect these external events to features in which the varieties differ nowadays (cf., for instance, Wurm 2012). Simons (1983) compares only Solomon Islands Pijin with Tok Pisin and the analysis is based on present day material. The comparative study by John Lynch (2010), which focuses on transitive markers in the three varieties but with special emphasis on Bislama, also analyses the features based on present day language material. Siegel (2008: 180-181) agrees that the dialect differentiation took place with the return of labourers, but not solely in that the islanders from the various areas did not have contact any longer, but also, and predominantly because, from then on substrate influence could have a greater impact on the stabilisation of specific features. As the focus of Siegel's study is on substrate influence and not on when the varieties diverged, he only provides a small number of historical examples. By selecting features in which the varieties differ

nowadays and attributing these differences to the substrate languages spoken in Papua New Guinea, Vanuatu, and the Solomon Islands, Siegel succeeds in showing that substrate influence played a role in the individual development of the varieties. However, his study is not informative in terms of *when* the substrate influence took place and therefore does not answer the question whether a localisation took place after the end of the labour trade.

There are also several studies available that more generally research the historical interrelationships of Pacific (and Atlantic) pidgins. These studies frequently do not go beyond mere lists of features and their earliest attestations. An exception is provided by Baker & Huber (2001) who applied a statistical, feature-based approach and thus were the first to consider quantitative measurements. However, Baker & Huber (2001) analysed earliest attestations only and did not differentiate between the individual Melanesian Pidgin English varieties.

The most extensive research on the origin and development of pidgin varieties in the Pacific is Tryon & Charpentier's (2004) *Pacific pidgins and creoles: Origins, growth and developments*. These scholars focus especially on the relationship between the Pidgin English spoken on the Australian plantations and the contact varieties that developed in the Pacific area. They look at different time periods and the history of contact, and they present language examples for each of the time periods. A linguistic comparative analysis, however, does not take place. It is only in chapter nine that the three Melanesian Pidgin English varieties are compared, focussing on differences which exist between the varieties today.⁴

To date, there are still questions regarding the timeframe in which the linguistic divergence of the varieties took place and the features it affected. Assuming that the end of the labour trade was one of the decisive factors, individual features (or at least a preference thereof) should be observable by the beginning decade of the 20th century.

1.3 Motivation, aim and scope of the study

Although the historical events support the general hypothesis that a development into individual varieties was caused by the end of the labour trade, it rests on speculative assumptions. No attempt has been made so far to compare the three varieties and to explore the starting point of their individual development by focussing on an analysis of early language data that goes beyond the comparison of earliest attestations and that includes statistical methods to investigate *when* selected features of the varieties began to individualise.

Therefore, the present study will examine morphosyntactic variation and change in Solomon Islands Pijin, Bislama and Tok Pisin from a diachronic perspective by combining

⁴ With *today*, Tryon & Charpentier refer to the year 2004.

qualitative and quantitative-statistical methods. It adds to the previous studies by providing a new methodological approach in order to add to the knowledge we have so far. Most of the studies based on the varieties were conducted more than 25 years ago. Working with historical sources implies that, independent of how much effort a researcher puts into the archival research, information on early contact languages might still be hidden in old manuscripts. Bringing together newly collected data with material compiled by earlier researchers and comparing the results of the analysis with earlier findings will contribute to a better understanding of the emergence and the individual development of the varieties.

The study will also help evaluate theories that have been put forth in order to discuss the origin of pidgin and creole languages in general and their relationship to one another. One of the major debates in the field of creolistics has evolved around the question of whether pidgin and creole languages are typologically different from non-pidgin and non-creole varieties. In this debate, John McWhorter (2002) has put forth the idea of a *Creole Prototype*. At the core of this theory, which has also been referred to as *Creole exceptionalism*,⁵ it is argued that creoles are “a synchronically definable typological class” and that “this class is demonstrably the result of the pidginization of lexifier sources” (McWhorter 1998: 790; cf. also McWhorter 2011: 5). McWhorter’s theory postulates that if languages use little or no tone, little or no inflectional morphology and little or no “noncompositional combinations of derivational markers and roots”, they could be classified as creoles “even without recourse to information about their histories” (McWhorter 2011: 5-6). The foundations for the theory, however, were already laid by Derek Bickerton’s *Language Bioprogram Hypothesis* (LBH) which is based on Chomsky’s model of universal grammar. By comparing Hawai’i Creole English with other creoles around the world, Bickerton identified what he called prototypical features of creole languages. Bickerton (1981, 1984, 1988) argues that children who are born in situations of intense language contact are assumed to have a pidgin as their primary language input, and thus are assumed to resort to an innate program to build an adequate language. In this framework, similarities between different creoles can be explained by the fact that this innate language bioprogram is the same for all humans. Thus, both the LBH and Creole Prototype Theory suppose that creoles form a special class of languages. While the idea of the *Creole exceptionalism* has been inter alia supported by Bakker et al. (2011), Bakker & Daval-Markussen (2017) and Daval-Markussen (2018), there has also been criticism against this theory.

⁵ The term “Creole exceptionalism” was coined by Michel DeGraff in 2003 as an umbrella term for all those theories that postulate that a creole can be identified through the structure of the language alone without knowing its socio-history (cf. DeGraff 2003).

One of the counter-theories of the LBH and the Creole Prototype Theory is the *Founder Principle* which assumes that creoles started as varieties of the lexifier language but diverged from it with increasing numbers of labourers of different linguistic backgrounds arriving in the contact situation. While the first labourers would have been in direct contact with the lexifier, new arriving labourers would have had to learn the lexifier language from other labourers who themselves spoke varieties of the lexifier. According to this theory, creoles thus represent restructured forms of the lexifier language. This theory, which was introduced by Robert Chaudenson (1992, 2001), was further continued by Salikoko Mufwene (1996, 2001). Mufwene placed the focus on the role which the founder population plays in a contact situation. He argued that non-standard varieties of the founder population were in use and could establish themselves as they were removed from the metropolitan/standard variety. Therefore, the theory assumes that creoles represent restructured forms of the lexifier, whereby the lexifier itself already represented non-standard varieties.

Based on the framework of the Founder Principle, the theory of an *evolutionary account of creole formation* was developed which argues that the linguistic system of any contact language is dependent on the *linguistic ecology* of a specific contact situation and the languages involved in that situation (cf., for instance, Aboh 2009; Ansaldo 2009a, 2009b; DeGraff 2014; Yakpo 2021+). In contrast to the LBH and the Creole Prototype Theory, the evolutionary approach of creole formation suggests that socio-historical information about the contact situation in which a variety developed is required since linguistic features alone are not sufficient to categorise a language as a creole.

The importance of the linguistic ecology for the development of contact languages is strongly connected to the idea of *feature pools* and *feature selection*. Mufwene (2001, 2006) suggested in his *Feature Pool Hypothesis* that the linguistic ecology of a contact situation will lead to a unique *feature pool*, which contains the linguistic systems of all individuals involved in the contact situation. As speakers interact in the contact situation, they exchange utterances (cf. Ansaldo 2009a: 275) and can resort to the feature pool. When producing utterances, speakers may replicate a feature identically or they may replicate it with some form of alteration (cf. Ansaldo 2009a: 275; cf. also Croft 2000). According to Aboh & Ansaldo (2007), identical replications are more likely in monolingual environments, whereas altered replications could be expected in multilingual environments, as is the case in most of the situations in which pidgins and creoles developed. For a feature to be selected it needs to be typologically common and frequently used in the contact situation (Aboh & Ansaldo 2007). Thus, supporters of this theory claim that contact

languages are the product of a “recombination of linguistic features [...] from typologically different languages” (Aboh 2009: 340).

Other creolists (cf. Muysken 2008; Ansaldo 2009b, 2010; Yakpo & Muysken 2017) have pointed towards the importance of *multilingualism* in the development of contact languages (cf. also the chapters in Aboh & Vigouroux 2021). Ansaldo (2009b, 2010) argues that contact languages are the result of “identity alignment in a multilingual context” (Ansaldo 2010: 616). In other words, he assumes that populations in multilingual ecologies try to negotiate their position in the society by “shifting and mixing codes, i.e. selection of features from different grammatical systems and recombination in a new grammar” (Ansaldo 2010: 619). Thus, it is not assumed that the target was the colonisers’ language but instead, the new developing contact variety is considered to be the creative product of “a rich multilingual competence” (Ansaldo 2010: 619). Given that contact languages arise in *multilingual ecologies* (Yakpo & Muysken 2017: 3), it is claimed that phenomena such as code-switching and code-mixing, as well as borrowing may have contributed to the formation of the contact varieties (cf. Muysken 2008, 2021+, and the chapters in Aboh & Vigouroux 2021+).

Yet another theory assumes that creole languages are the result of *relexification*, with relexification being defined as “a mental process that builds new lexical entries of an already established lexicon and replacing their phonological representations with representations derived from another language” (Lefebvre 1998: 16). The idea that a relexification process contributes to the formation of contact languages was already postulated by Musyken (1981, 1988) who focussed on the emergence of the mixed language Media Lengua. Lefebvre (1998, 2004) has expanded this idea to creole languages focussing on the contact situation in Haiti. The theory assumes that people speaking different languages coming together in contact situations had to find a way to communicate. Since they were exposed to the language of the colonisers, they would have relexified the lexicon of their mother tongues with the lexicon of the coloniser’s language, all the while keeping the structures of their mother tongues (Lefebvre 1998: 35).

All of the above-mentioned theories are dependent on historical data, since it is, in the end, the postulated *developments* of pidgins and creoles that are being discussed. The present study adds to the types of studies listed above in that it provides in-depth historical data to assess these theories and to shed light on the processes involved in the formation of contact languages. In the narrower sense, the study aims at broadening our understanding of when the three Melanesian Pidgin English varieties started to develop their unique features by using first-hand data drawn from historical sources. Instead of covering all attested differences in the varieties, in-depth case studies of four morphosyntactic domains, namely demonstratives, relative clauses, modal markers

and selected prepositions, were conducted. The linguistic variables were selected based on observed differences recorded in the early data collected for the three varieties. The differences observed in each of the domains led to the focus on the features displayed in Table 1.

Linguistic domain	Feature
demonstratives	adnominal demonstrative pronouns
	pronominal demonstrative pronouns
relative clauses	subject relative clauses
	object relative clauses
modality	abilitative
	volition
	permission
	speculation
prepositions	comitative
	instrumental
	terminative
	adessive

Table 1: Morphosyntactic features investigated in the present study

Possible reasons for the divergence of specific features in the varieties will be examined in a second step and will be discussed in light of the data results. The main research question can be subdivided into the following questions:

Research Question #1: How were the four linguistic variables realised in Solomon Islands Pijin, Bislama and Tok Pisin over time?

The present study will investigate, how the three varieties Tok Pisin, Solomon Islands Pijin and Bislama differ in realising the linguistic variables in the timeframe from their earliest attestations until 1950. This means that not only different variables and their variants will be analysed but also their dates of attestation will be of importance.

Research Question #2: Was variation receding, stabilising or increasing over time within the individual varieties and across the varieties?

As the varieties were not fully developed yet, it would be fair to assume that they displayed a high level of variation. This study will analyse to what extent variation *within* each variety, as well as *across* the varieties, changed over time. I will investigate whether the three varieties show similarities in *when* they came up with a choice of a variant or a choice of co-existing variants.

Research Question #3: Which external influences may have had an impact on the individual developments of the varieties regarding the features under investigation?

Despite the fact that many scholars argue that the end of the labour trade can be regarded as the decisive factor for the individual development of the three varieties, there are other possible factors that may have had an impact on the development of the varieties (such as Queensland and

Samoan plantation pidgins, World War II, missions, language planning and policies, substrate influence). Based on the results of the qualitative and quantitative analysis, this study will investigate whether it makes sense to assume that the end of the labour trade was the decisive factor for the varieties' individualisation.

Research Question #4: What do the results tell us about the factors at work in feature selection as well as grammaticalisation processes in the stabilisation of contact languages in general?

The functional purpose of initial contact jargons is to allow communication between speakers that do not share a common language. Therefore, they arise out of a pragmatic need for speakers to make themselves understood by some means. While these early contact varieties are characterised by a high degree of variability, features will grammaticalise when contact languages begin to stabilise and turn into the main languages for a community. As shown above, the processes at work during the development and stabilisation of contact varieties are still highly debated. The present study will thus contribute to discussions about the processes at work during the stabilisation of contact languages. By analysing specific grammatical features and their development and stabilisation in Solomon Islands Pijin, Bislama and Tok Pisin, the study will shed light on processes at work in grammaticalisation and feature selection. Since the study focusses not only on linguistic data, but also on socio-historical sources, it will also give insights to the role that external factors (wars, the missions, etc.) and internal factors (substrates, superstrates, etc.) played in the formation and stabilisation of contact varieties. Moreover, the study will provide information on the role that language sociology plays in language development. Thus, it will show how factors such as language attitudes, language planning, language policy, using a language in the spoken and/or written medium, codification and diffusion may influence the development of contact varieties.

1.4 Structure of the study

Chapter 2 begins with some general remarks about challenges as well as opportunities researchers are confronted with when studying languages, and more specifically pidgins and creoles, from a diachronic perspective. The discussion is necessary not only to understand the potential of historical data for the reconstruction of a language's past, but also to highlight the necessity of a proper and careful handling of early language data material.

Chapter 3 will move the focus towards the varieties under investigation. In Section 3.1 the sociolinguistic history of the three Melanesian Pidgin English varieties will be reconstructed by providing evidence from pre-colonial, colonial and post-colonial sources. This background

information will not only serve to understand how languages came into contact in the Pacific but it will also provide the basis for understanding the various theories that have been put forth to explain the development of these varieties, which are introduced in Section 3.2.

Chapter 4 will then outline the empirical foundations of the present study. It starts with a detailed description of how and where the language material was collected before the methodological steps are summarised. This chapter will also give information on coding processes, feature selection and strategies in dealing with the reliability of the sources.

Because metalinguistic information may also contribute to understanding when SIP, BIS and TP developed their individual characteristics, Chapter 5 will summarise the main statements that could be extracted from early sources regarding the similarity and dissimilarity of the three MPE varieties at different points in time. Chapters 6-9 will then present the analysis of demonstratives (Chapter 6), relative clauses (Chapter 7), modality (Chapter 8) and selected prepositions (Chapter 9) across time to learn more about the divergence of the MPE varieties. The individual case study chapters are structured similarly. They all begin with some theoretical background information on the variable, give information about the variable's realisation in the contemporary MPE varieties, introduce previous diachronic research, outline methodological considerations before, finally, the findings are presented and discussed.

The final chapter (Chapter 10) will summarise the key results of the individual case studies and will attempt to answer the research questions introduced above. Moreover, areas that deserve further research in the future will be outlined.

1.5 General remarks

1.5.1 A note on terminology

1.5.1.1 Dialects or languages

Focussing on the origin and development of the individual varieties raises the question of whether Tok Pisin, Solomon Islands Pijin and Bislama should be considered dialects of one and the same language or whether they should be considered languages in their own right. Most researchers refer to the varieties as (national) dialects (see, for instance, Crowley 1991: 51; Jourdan & Keesing 1997: 402; Siegel 2008: 175). Even if Lynch does not use the word *dialect*, he argues that the three varieties belong “recognizably to the same language” (Lynch 1998: 223) and also Tryon & Charpentier claim that “[s]trictly, they should be considered three dialects of a single language, for there is considerable mutual intelligibility between them” (2004: 393).

It is a well-known approach to use the intelligibility criterion to classify varieties as either *languages* or *dialects* (cf. Gooskens 2018: 206). According to the criterion, varieties can be

considered dialects of a single language if speakers of different varieties are able to communicate successfully and without comprehension difficulties despite differences in vocabulary, grammar and pronunciation (cf., for instance, Edwards 2009: 63). However, some linguists such as Hammarström (2008) contradict this view and argue that the degree of intelligibility alone does not serve as a reliable criterion to distinguish between languages and dialects. One criticism that is voiced involves the question of the extent of mutual intelligibility which needs to exist in order to classify a variety as either being a dialect or a language.⁶ In addition, intelligibility is a subjective criterion which varies across speakers. That mutual intelligibility alone cannot be taken as the critical factor can also be seen when taking a look at the Scandinavian languages Swedish, Danish and Norwegian. These are considered to be individual languages despite the speakers' ability to communicate with each other. Examples like these demonstrate that politics, power relations and national identity can have an impact on the classification of varieties as languages or dialects as well (cf. Edwards 2009: 64).

While Solomon Islands Pijin, Bislama and Tok Pisin are frequently referred to as dialects, online databases such as *Glottolog* (www.glottolog.org, last access 01 May 2021) and *Ethnologue* (www.ethnologue.com, last access 30 November 2020) list them as individual languages. The present study is not aimed at determining whether Tok Pisin, Bislama and Solomon Islands Pijin are languages, or whether they should be considered dialects of Melanesian Pidgin English. Rather, this study is concerned with when the varieties started to diverge from each other. Therefore, the neutral term *variety* will be used throughout the study.

1.5.1.2 Individualisation and stabilisation

It is necessary to define the two notions *individualisation* and *stabilisation* as used in the present work. The aim of the present study is to investigate *when* the three MPE varieties Solomon Islands Pijin, Bislama and Tok Pisin developed their individual features. Starting from the assumption that MPE was used throughout the Pacific as a lingua franca, it is assumed that the early MPE was similar in the areas where it was spoken but that individual varieties began to emerge, when the islanders of the three areas were no longer in continuous contact. This process is understood as *individualisation* in this study.

Sometimes the term *localisation* may be used instead to indicate that different features became established in the three Melanesian areas. In addition, I will also use the term *divergence* to describe the gradual individualisation of the MPE varieties.

⁶ For further criticisms see Edwards (2009: 63-64) or Gooskens (2018: 206-208).

It has been assumed that the *individualisation* of the three MPE varieties came along with their *stabilisation*. The notion of stabilisation describes the vague idea about having a grammatical norm and reduced variation in a language. However, it is important to note that it does not entail elimination of variation since variation can be a stable component in languages as well. Thus, it would be misleading to assume that a variety can be considered ‘stable’ if only a single form is attested to encode a feature, with all other variants disappearing. Creolists such as Mühlhäusler (1997: 138) describe stabilisation as the “the gradual replacement of free variation and inconsistencies by more regular syntactic lexical structures”, whereby “a pragmatic mode of speaking begins to give way to a grammatical one”. However, this definition is not without its problems. It remains unclear how many variants are still allowed to coexist and whether changes in the frequencies of feature forms might be enough to specify it as a beginning of a stabilisation of forms. Moreover, it would also be wrong to look at contemporary forms of a feature and consider a feature to have stabilised if only the contemporary forms are attested and all other variants are ruled out. The problem with such an approach is that forms can stabilise in a specific time period, i.e. can be in use for a period of time before another form stabilises and turns into the common form. Drechsel (2014: 30) argues that “as a result of stabilization, a jargon’s extended linguistic variation and pragmatic way of speaking yield to increased grammaticalization and an expanding lexicon, by which speakers develop strategies independent of their first languages and thus a stable grammar of its own”. Nonetheless, it should be noted that there is no fixed rule which can be consulted to clearly decide whether a variety is stable or not.

1.5.2 Remarks on the names used to refer to the varieties

All early language examples which were found for the Eastern Papua New Guinea area are referred to as *Tok Pisin*, those found for the Solomon Islands area are called *Solomon Islands Pijin* and those for the Vanuatu region are referred to as *Bislama*. Though the terms Solomon Islands Pijin (SIP), Bislama (BIS) and Tok Pisin (TP) will be used, this does neither mean that the varieties were fully developed nor does it mean that there was just a single Pidgin English variety spoken on each island. In contrast, it can be assumed that regional variation was the norm and that several varieties with different levels of stability coexisted. The terms are nonetheless used to disambiguate between the three varieties and the locations where the examples were attested.

1.5.3 Remarks on the examples

The examples given throughout this work will be printed with the orthography in which they were found in the early sources. This means that most examples will show an English spelling even though morphemes may have deviated from the English forms in speech. The spelling and the fact that most of the vocabulary of the three varieties derives from English may lead to the false assumption that the meaning of the examples can easily be detected. As Velupillai (2015: 7-8) argues:

[t]he wide-spread and unfortunate practice of generally refraining to give morpheme-by-morpheme glossing for pidgins and creoles, especially the English-lexified ones, thus not only risks perpetuate the impression that these are simply somewhat odd versions of the lexifier language (rather than languages in their own right), but also risks obfuscate what the examples actually show.

English, as the lexifier-language of the three varieties, contributed a great deal to the varieties but several morphemes and words have gained new and additional meanings or have grammaticalised. This is the reason why the examples will not only be supplemented by their translations but also by morpheme-by-morpheme glosses, using the international standard of the Leipzig Glossing Rules (<https://www.eva.mpg.de/lingua/resources/glossing-rules.php>, last access 29 September 2021).

1.5.4 Remarks on non-English extralinguistic quotes

As the Solomon Islands, Vanuatu and Papua New Guinea were possessed not only by English powers, but also by the Germans and the French, some of the underlying source texts were written in French or German. Examples that are directly quoted from these sources are translated into English and the original of the translation is provided in the footnote. All examples were translated by me.

1.5.5 Remarks on languages cited in this work

Since it is very important in typology to know the genealogical affiliation of a language as well as the area where it is spoken (cf. Velupillai 2012), I will provide this information the first time a language is referred to, according to the format ‘language name (language family (language genus): location)’. The language locations are based on the information obtained from the *World Atlas of Language Structures* (WALS), *Glottolog* (<https://glottolog.org/>, last access 22 May 2021) and *Ethnologue* (<https://www.ethnologue.com/>, last access 22 May 2021).

2 Studying the emergence of pidgins & creoles from a diachronic perspective

The present study is situated in the field of *historical creolistics* (Arends 1995: ix), which is a sub-discipline within the study of pidgin, creole and mixed languages that focuses on the external and internal history of contact varieties. Historical creolists assume that “the combination of early language data and detailed extralinguistic information [...] provides [...] a unique opportunity to study creole genesis from a historically realistic perspective” (Arends 1995: x) and, thus, to learn about their formation processes.

Although interest in the history and structure of pidgins and creoles emerged with the beginnings of creolistics in the 19th century, the use of historical data was disregarded for a long time. Most of the scholars were interested in questions concerning whether contact languages around the world show structural similarities and if so, what these similarities can tell about the languages’ emergence. They thus approached the subject matter theoretically and focussed on synchronic data only. Consequently, several theories have been put forth which try to explain the underlying formation processes of pidgins and creoles without considering diachronic language data (cf., for instance, the Language Bioprogram Hypothesis by Bickerton (1981, 1984) and the notion of a Creole Prototype by McWhorter (2005)).

A major reason for the “rooted ahistoricism” (Arends 1995: ix) is that in their earliest stages, pidgins and creoles were considered by their contemporaries to be non-standard varieties that were unworthy of a detailed examination. Thus, there are almost no early comprehensive descriptions of the varieties available and the documents that exist attest to earlier language stages only in a fragmented manner (cf. Huber & Velupillai 2018: 131). Therefore, many researchers (cf., for instance, McWhorter 1997) have avoided the use of early documents in the study of pidgin and creole languages, since they assume that the contemporaries will not have documented the by then highly stigmatised varieties adequately enough. Another reason, which Huber & Velupillai (2018: 132) consider the *de facto* reason for the ignorance of historical data, is that the collection and analysis of early language data demands considerable time and effort as the material is hidden in archives around the world.

Fortunately, there are several language specific and cross-linguistic studies available which are based on diachronic data and that have shown the added value which historical data creates in learning about the emergence and development of pidgin and creole languages (cf., for instance, Baker 1987, 1993; Keesing 1988; Bruyn 1995a; Huber 1999; Baker & Huber 2001).

This does not imply that the use of historical data is not without its hazards. The present chapter illuminates the challenges which are associated with the collection and use of early linguistic data. As the study lies at the intersection of creolistics and historical linguistics, and

since historical creolists face similar challenges as historical linguists, Section 2.1 will outline the general difficulties that historical studies are confronted with when studying past stages of language development. Section 2.2 will refer to challenges which especially creolists encounter in accessing the language of the past.

2.1 Historical linguistics and the *Bad Data Problem*

“[W]e will have to make do with what is left of the written record, with all the implications this has for the representativeness of the data.”

(Knoolhuizen 2006: 3)

Studies focussing on the history of language and language change are dependent on diachronic data. The further back in time the focus of a study goes, the more restricted the possibilities available to the researcher and the more likely the encounter with the *Bad Data Problem*. The designation, introduced by William Labov, refers to the circumstance of historical data surviving “by chance, not by design” (Labov 1999: 11).

Historical researchers are unable to gather data aligned to their research question but must work with “fragments of the literary record” which are “beyond the control of the investigator” (Labov 1982: 20). Thus, although language variation and change starts in speech rather than in writing (cf. Elspass 2012: 157; Schneider 2013: 57), historical studies usually must rely on the written record. It is nearly impossible for historical researchers to obtain spoken language data, at least, when studies focus on a time prior to the introduction of audio-recording devices. Even if the latter were introduced, the amount of data that survived until today is often relatively small (Labov 1999: 11). Most historical studies are thus restricted regarding what can be studied. The accessibility of spoken versus written data will determine whether studies on phonology are possible or whether they must be restricted to grammar and lexis (cf. Raumolin-Brunberg 1996: 18).

Even if most historical studies depend on fragments of the written record, it does not imply that the history of *speech* cannot be studied from the latter, as it is generally agreed that the “traditional distinction between ‘spoken language’ and ‘written language’ is simplistic and even misleading” (Elspass 2012: 157). Following models such as Koch & Oesterreicher’s proximity-distance model (1985, 1994), it is assumed that even written texts can be speech-like, though different text types might vary in their closeness to speech. Scholars do not necessarily agree on the text type they consider to be “maximally speech-like” (Nevalainen & Raumolin-Brunberg 2012: 23). Thus, while Elspass (2012: 165) suggests diaries to be proximate to speech, Nevalainen & Raumolin-Brunberg (2012: 32) promote the use of letters (= personal correspondence) and Baker (2010: 77-78) suggests trial proceedings or drama texts as the most adequate text type.

Independent of the chosen text type, scholars need to be aware that “written language tends to be more conservative, normative, and formal than oral language” (Hernández Campoy & Schilling 2012: 68) since the original speech act becomes filtered during the process of textualisation:

[T]he written record functions as a filter, as it were: it provides us with a representation of a speech act that we would have liked to have listened to and recorded acoustically and that without the written record would have been lost altogether; but at the same time the rendering of the speech event is only indirect and imperfect, affected by the nature of the recording context in certain ways. (Schneider 2013: 58)

The linguist working with historical data not designed for linguistic purposes needs to understand that the data may contain several filters which may have had an impact on the depiction of the language. While it is a necessity to be aware of a possible alienation, it does not purport that the past data is useless. The early sources represent the only possibility to reconstruct a language’s past and should therefore not be rejected. Instead, the historical researcher needs to be aware of the challenges the data contain and will have to “remove the filter” as much as possible to eliminate wrong interpretations, or, as Schneider summarises (2013: 58), to:

assess the nature of the recording process in all possible and relevant ways and to evaluate and take into account its likely impact on the relationship between the speech event and the record, to reconstruct the speech event itself, as accurately as possible.

The number of filter-layers a text contains and the concomitant question of reliability is, according to Schneider (2013: 60), dependent on three factors, namely the “reality of a speech event”, “the relationship between the speaker and the person” documenting the utterance and the time gap between the speech event and its documentation. Based on an evaluation of these criteria, texts could be categorised into five text groups, which are 1) recorded, 2) recalled, 3) imagined, 4) observed and 5) invented, which he sorts along a continuum depending on their proximity to speech (cf. Figure 2). Schneider claims that the closer a text is to speech, the more reliable the data would be. However, as pointed out above, other scholars might make use of different classification schemes.

Next to discussing the question of *proximity to speech*, historical linguists are restricted in *whose* language they can study (cf. Raumolin-Brunberg 1996: 18). Being dependent on the written record, factors such as literacy as well as printing opportunities need to be considered. Thus, though contextual information might be difficult to gather, it is a prerequisite for the reconstruction of past societies and for the evaluation of language samples regarding their closeness to reality and their suitability for historical linguistic studies.

<i>Category</i>	<i>Reality of speech event</i>	<i>Speaker–writer identity</i>	<i>Temporal distance speech – record</i>	<i>Characteristic text types</i>
Recorded	real, unique	different	immediate	interview transcripts, trial records
Recalled	real, unique	different	later	ex-slave narratives
Imagined	hypothetic, unique	identical	immediate	letters, diaries
Observed	usu. real, unique	different	later	commentaries
Invented	hypothetic, unspecified	n. a.	unspecified	literary dialect

Figure 2: Categorisation of text types according to their proximity to speech (Schneider 2013: 61)

2.2 Challenges in studying pidgins and creoles from a diachronic perspective

Creolists focussing on the origin and development of early contact varieties are acquainted with the above-mentioned challenges. When studying the emergence of pidgin and creole varieties that have their origin in the past, available data material is mostly restricted to written documents. Early recordings, such as Rudolf Poech’s (1904) early phonogram record of TP, represent an exception. The creolist needs to work with “what is left of the written record” (Knoolhuizen 2006: 3) so that several filters will have to be removed. However, as the following section will demonstrate, the historical analysis of contact varieties bears hazards which go beyond the challenges which historical linguists encounter who focus on Standard English varieties.

2.2.1 Availability and accessibility of historical sources

While historical linguists focussing on English (Indo-European (Germanic): UK) struggle predominantly with the oral versus written text accessibility, historical creolists may struggle to find sources of any kind which document early contact varieties. The historical documentation of early pidgins and creoles is limited and dependent on the variety under investigation. For instance, Huber & Velupillai (2018: 133) mention that while Sranan (Creole (English-lexified): Suriname) is well documented over time, for Eskimo Pidgin (Pidgin (Eskimo-lexified): Canada) there is only a limited amount of short texts available.

For contact languages that have their origin in colonial or pre-colonial eras and that developed in societies in which literacy was not widespread, early evidence needs to be searched for in documents such as travel accounts, government reports, sojourn reports, missionary journals and diaries, as these sources represent the only possibility to obtain access to the *earliest* forms and use of contact varieties. The amount of Pidgin English the sources contain varies from single words, phrases, and sentences to larger texts, and it is impossible for the researcher to know beforehand whether there will be any samples of early pidgin/creole in these sources at all. Most

of the early sources were written by the colonisers and consequently, in the coloniser's language. The variety they encountered was only, if at all, occasionally quoted or referred to.

A major problem for the reconstruction of the earliest stages of most pidgin and creole languages is that predominantly European or superstrate-influenced documents are available. As McWhorter (2000: 235) points out in a review of Jacques Arends' *The Early Stages of Creolization* "we must ask ourselves how confidently we would chart the history of English based solely on materials written by Chinese missionaries whose competence in English was 'good'". Also, Roberts argues that databases should contain "texts from all [...] groups involved in the creole formation process" (2005: 41). However, studies on the diachronic development of pidgins and creoles are dependent on the surviving records and texts from *all groups* may not have survived for the early stages in a pidgin development.

Written records such as letters and trial proceedings that encompassed longer passages completely written in the contact varieties are scarce as pidgins are usually oral languages. They served as a *spoken* medium and only started, if at all, to be used in a written form in later developmental stages. On the one hand, this has the advantage that in the earliest documents, and especially in travel reports, a differentiation between *spoken* vs. *written* is not a prevailing problem. As a written form did not exist, it can be assumed that the authors introduced the readers to the variety as it was *spoken* in the colonies. It can thus be assumed that the written documentation contains several hints on the oral form of the languages. On the other hand, the conditions under which the texts were produced generate other filters that need to be considered and removed in order to make judgments about the *reliability* of the early sources.

2.2.2 Reliability of language examples extracted from historical sources

Studies that analyse the development of pidgin and creole languages based on historical sources need to critically evaluate how reliable the extracted data is. Baker & Winer (1999) as well as Huber & Velupillai (2016) critically engaged with the question of reliability and highlighted several factors that may have an impact on the portrayal of the early language data.

Based on Baker & Winer (1999) and Huber & Velupillai (2016), Figure 3 illustrates the different factors that can have an impact on the reliability of early sources. I propose a continuum of reliability, as there are no clear-cut borders in terms of when a source can be considered trustworthy or not and to demonstrate the varying degrees that early sources show in terms of their trustworthiness. While most of the reliability-influencing factors can be portrayed as continua as well, the factor of *authenticity* represents an exception (Section 2.2.2.1). Furthermore, we must differentiate between author-based factors and more general factors. The author-based

factors include the timespan between a documentation and the original language situation, language attitudes, duration of sojourn, travel activities and the linguistic abilities of the writer (Section 2.2.2.2-2.2.2.5) while the more general factors include editor revisions and anglicisation (Section 2.2.2.6-2.2.2.7). All factors will be explained and supplemented by examples in the following. It should be noted that all mentioned reliability factors can interact and influence each other.

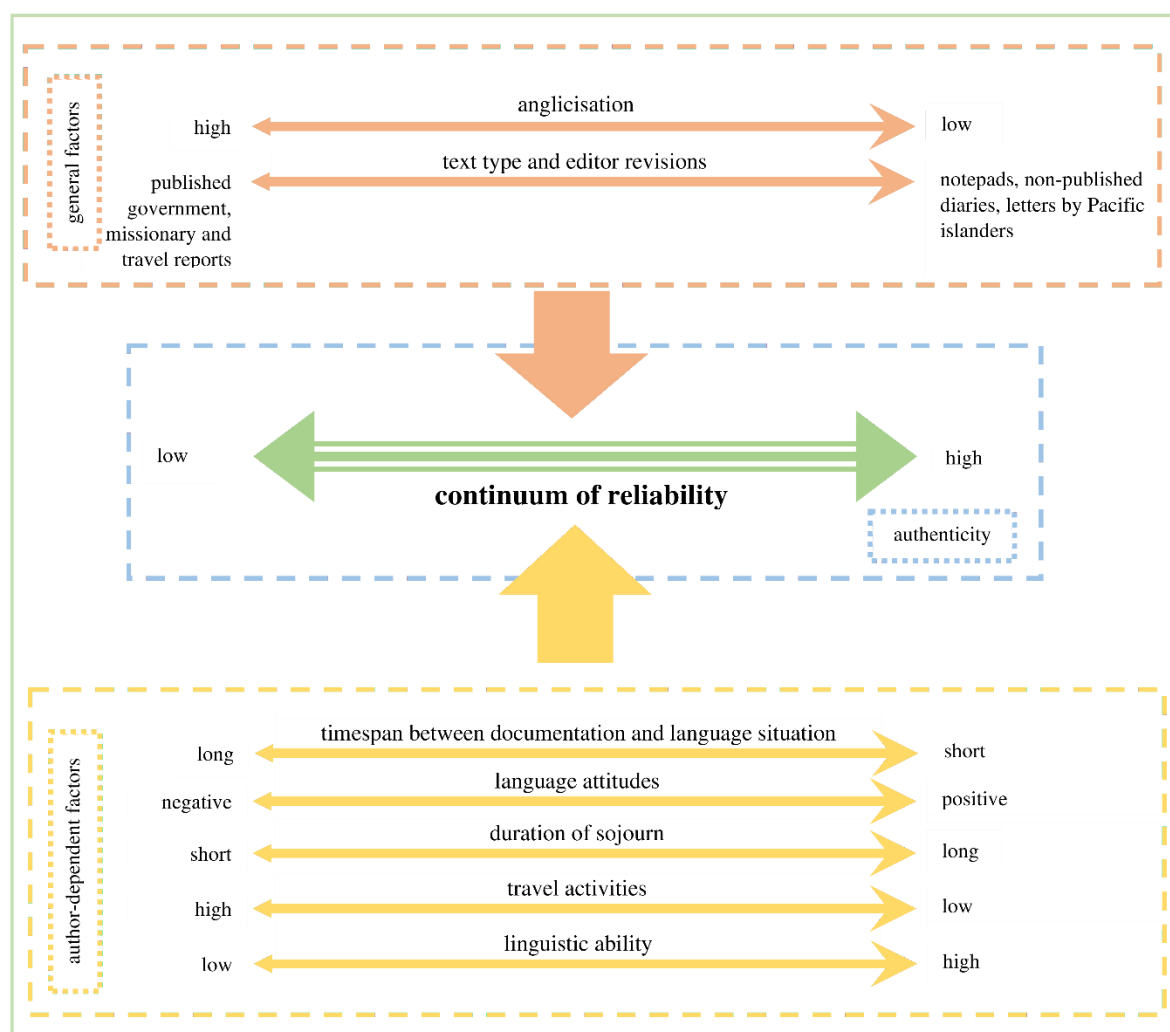


Figure 3: Continuum of reliability and influencing factors

2.2.2.1 Authenticity

Working with early attestations of pidgins and creoles reveals that language examples were frequently doubled by reporting the same situation in more than one source. There are cases in which one and the same author reproduces identical language examples because of reporting about a situation in more than one source. For instance, Collinson reports in two sources about his experiences with a Solomon Islander, resulting in the doubling of language examples (cf.

Collinson 1926: 229 vs. Collinson 1929: 24). There is also the possibility that a language example is documented in the unprinted diary or the notes of an author and was later published.

Additionally, the same language example may be found in sources of different authors. Here, two types can be distinguished. In some cases, reference to the original author is made. Many early newspaper articles quoted out of travel reports (cf. Kalgoorlie Miner, Anonymous 19.07.1897: 2 vs. Woodford 1897: 30) and missions reprinted letters in their mission reports.⁷ Another example represents the proclamation of September 1914 about the annexation of German possessions, which was reprinted more than four times (cf., for instance, Government Gazette 01.11.1914; Reeves 1915: 77; Cameron 1923: 292-293; Idriess 1941: 31-32). As no fixed orthography existed, the reduplications sometimes differed in their spellings and thus a comparison of the sources may provide important information about early pronunciation.⁸

The second type is when language examples are printed in more than one source and no reference to the original author is made, which represents an early form of plagiarism (cf. Baker & Winer 1999:103; Huber & Velupillai 2016: 133-134). In addition, outsiders that report “from anecdotes told at the dinner table” need to be differentiated from insiders with “first-hand experience” (Baker & Winer 1999: 104). If authors had never been in the area in which a contact variety was spoken, their language examples will either be made up or copied from other writers of that time and will thus not be considered authentic.

2.2.2.2 Timespan between documentation and original language situation

One of the factors that could be classified as author-based is the timespan between the documentation or publication of an utterance and the original language situation (cf. Huber & Velupillai 2016: 133), a factor which also Schneider (2013: 60) considers to be of importance.

The process from the speech event to the actual documentation sometimes consisted of a short period only, for instance, if authors made immediate notes on the circumstances of a speech act. The archival documents of Charles Morris Woodford, for instance, contain several small notepads in which he recorded speech situations. Due to the shortness of time between the documentation and the actual language situation, the citations can be assumed to be reasonably accurate.

Unpublished diaries represent a further quite reliable source, as speech events were usually documented on the same day, albeit a timeframe between the actual language situation and the

⁷ Cf., for instance, *Letter by Crystal Ufaria to Gwaeta* which can be found in *Not in Vain* Issue 83 September 1941 (Ufaria 1941) and in the *Solomon's Soldiers News* Number 7, 1944 (Deck 1944: 3; AU PMB DOC 442).

⁸ Cf., for instance, *belonga* (Reeves 1915: 77; Idriess 1941: 31-32) vs. *belongina* (Government Gazette 1.11.1914; Cameron 1923: 292-293).

documentation needs to be assumed. In addition, they can be considered reliable data sources as unpublished diaries were not written with a specific audience in mind but, normally, for the authors themselves.

The degree of reliability decreases when turning to published sources, for which the timespan from the speech event to the actual documentation was usually longer. A comparison of the publication date with the date to which the speech event refers will shed light on the number of years it took an author to write about his journey. The longer the timespan, the higher the danger of an indistinct recollection of events, and thus, the higher the chance of inaccurate language examples (cf. Huber & Velupillai 2016: 133). However, since authors may have based their publications on diaries and notes they made during their sojourn, a long time lapse is not necessarily paralleled by misremembrance. Huber & Velupillai (2016: 133) cite Otto Schellong who, publishing his memoirs 46 years after his stay in German New Guinea, used his diary as the basis of his writings. A risk remains, as Schellong himself admits, that the documentation of events that lie in the past will be imperfect and influenced by blurred memories (cf. Schellong 1934: 4).

The ability of the authors to recite the original speech act is not solely dependent on the timespan between documentation/publication and the actual speech event. Even if the timespan is rather short, introductory sentences and the context into which a quote is embedded may yield information about the reliability. Phrases such as “In substance, Simon said” (Deck July 1919: 4; PMB 1253) or “and his words were something like this” (Deck July 1928: 5; PMB 1253) should catch the attention of the observer as they indicate that the following oration was potentially altered or an approximation. Phrases which speak for a more truthful rendering of the original speech act are, for instance, “I give it in ‘pigeon’ English, as it was told me” (Deck March 1910: np; AU PMB DOC 440).

Thus, it is important to consider the publication date of the source and to compare it to the date of when the speech event itself had taken place.⁹ Moreover, it is important to observe whether the context provides information regarding the reliability and accuracy of the language sample. Factors such as whether the publication process was based on extensive notes, diaries or mere memory need to be considered to evaluate the accuracy of the written documentation. The longer the time interval between the two dates, the higher the risk of misremembrance.

⁹ It needs to be noted, however, that the timing of early sources represents a challenge as well (see Chapter 4).

2.2.2.3 Language attitudes

Positive as well as negative attitudes by the author towards the contact variety may have had an impact on the choice of language examples which were printed (cf. Baker & Winer 1999: 106; Huber & Velupillai 2016: 131-132). Authors with negative attitudes may have focussed on what they considered *peculiarities* which may have resulted in inauthentic biased language documentations. In contrast, authors with positive attitudes may have referred to and described the morphosyntactic elements of the varieties from a neutral point of view.

Attitudes may be politically or ideologically fixated (cf. Baker & Winer 1999: 106). The author's role in the colony had an impact on the motivation for their writings and will thus have contributed to how accurately they portrayed the contact varieties. Since, for instance, the Pacific Islanders were frequently described as primitive people, the Pidgin English was sometimes described as reflecting this simplicity (cf. Jacques 1922: 96). Authors supporting such views are more likely to rely on stereotypical language examples to show that "primitive vocabulary means primitive expression" (London 1911: 297). If the aim of a mission or the government was to promote a language other than Pidgin English, it is fair to assume that the portrayal of the pidgin variety was not a very positive one, which may have been reflected in the language examples provided.

Huber & Velupillai (2016: 132) mention circumlocutions as an example for stereotypical language documentation by writers with negative attitudes. Circumlocutions cannot be considered inauthentic per se, as a case study conducted in 2017 proved that they represented an important strategy to fill lexical gaps and to guarantee a successful communication between Pacific Islanders and European colonisers (Schäfer 2017a). They may in fact contain important information on grammatical structures. Furthermore, in addition to the stereotypical examples found in several travel reports, more sophisticated examples are attested in early dictionaries.¹⁰ However, if writers made use of circumlocutions to show that the pidgin language would be "a language based on the apprehension abilities of the coloured race" (Daiber 1902: 229),¹¹ researchers need to be careful. It needs to be decided on a case by case basis which of the circumlocutions reflect authentic examples and which cannot be considered real attestations.

¹⁰ Although some of the circumlocutions found in early accounts were most probably invented by European writers in their attempt to illustrate the 'inadequacy' of the pidgins, it seems reasonable to assume that many in fact are genuine examples, resulting from an important strategy in contact situations which allows speakers to communicate when a lexical item for a referent is lacking. That it was a common strategy which was not only mentioned by opponents of Pidgin English can be attested for all three areas. While circumlocutions are common in early contact situations, they are said to be replaced in stable pidgins, for instance, by "phrase-like formulas for the description of new concepts" (Romaine 2005: 1094). In addition, they are said to "give way to compounding" (McMahon 1994: 263) while developing from a pidgin to a creole.

¹¹ "Die Verkehrssprache des Europäers mit dem Südsee-Insulaner ist das sogenannte Pidgin- oder Business-Englisch, eine für das Begriffsvermögen der dunkeln Rasse hergestutzte Sprache [...]" (Daiber 1902: 229).

2.2.2.4 Duration of sojourn and travel activities to other Pacific Islands

The amount of time that the early writers spent in a specific region varied to a great extent. While some were permanent residents, others were short-term visitors. The longer an author was exposed to or in contact with a pidgin, the better the writer's ability to report about the variety. The duration of stay in a specific colony may thus have influenced the pidgin's portrayal (cf. Huber & Velupillai 2016: 134).

Next to the duration of sojourn, travel activities of the authors might influence their documentation of the language (cf. 2016: 134). Many of the early writers (depending on their role in the colony) travelled to several islands in the Pacific region and were not bound to a specific island group. Especially those working on recruiting vessels, as well as missionaries, will have encountered the different contact varieties that were spoken in the areas. For instance, the recruiter and later trader of the British Solomon Islands Protectorate John Cromar (1935) reports about events that took place in Vanuatu and the Solomon Islands. The missionary George Brown (1908) visited, among others, New Guinea, the Solomon Islands, Samoa and Fiji, because it was common that missionaries preached the gospel in various regions. Since the majority were neither linguists nor language experts, it is debatable whether they were able to recognise differences in the regional contact varieties. Even if they were able to recognise or sufficiently reflect on differences, due to the timespan between the documentation and the speech event itself, there is a risk that the authors were unable to connect the language examples with the correct area or that they quoted speakers in the contact variety which was most common to them, producing biased descriptions of the varieties spoken. Hence, the duration of sojourn and the travel activities represent further author-based factors which can have a bearing on the reliability of the sources and need to be considered during the analysis.

2.2.2.5 Linguistic abilities of the writers

As pointed out by Huber & Velupillai (2016: 132), the reliability of early sources can further be influenced by the linguistic abilities of the writers. Referring to an example of Buchner's description of Cameroonian Pidgin English (Creole (English-lexified): Cameroon), they demonstrate how authors sometimes translated Pidgin English falsely (cf. 2016: 132). Thus, translations as well as the pidgin language examples as such need to be critically judged, as they may vary in reliability depending on the authors linguistic abilities.

2.2.2.6 Editor revisions

Editor practices may further transform the original speech act (cf. Huber & Velupillai 2016: 133) and represent a non-author-based factor. Printer and editor interference may cause words of the original speech act to be adjusted and an orthography closer or further away from Standard English (cf. Section 2.2.2.7). Depending on the text type, a speech act may be removed to varying degrees from the original speech form. While direct notes and unpublished diaries represent genres which have no intervening editorial steps in between the speech act and their documentation, texts that were published or intended for public consumption put an editor filter on the original speech.

Even letters which are written by Pacific Islanders may have been transformed. For instance, in the Pacific it was not before the missions started to implement schools in which they taught Pacific Islanders that literacy started to spread. Prior to that, if Pacific Islanders wanted to write letters, it was a common practice that European missionaries or other European officials were asked to write the letter for them. The letters were dictated by the islanders to the European and just signed with X by the Pacific Islander. For instance, Clarker reports that the “Atchin natives asked [him] to write down” (Clarker 02.03.1914) and another letter by four Atchin inhabitants starts with “We man-Lalip ask Mr. Parker, one missionary stop along Atchen, him write one letter along you” (Meltiknumba et al. 30.03.1914). Figure 4 shows that sometimes it only becomes visible at the end that a letter was written by a European with the bracketed “(signed)” and “per”.

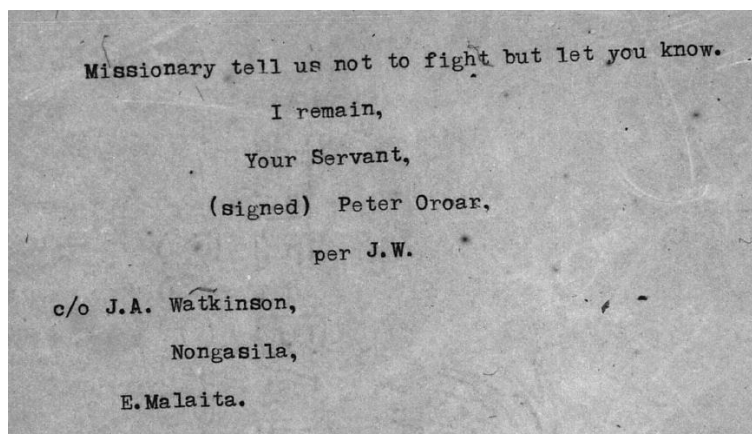


Figure 4: Solomon Islander letter written by European (Nongosila; Oroar, P. 19.05.1906)

There is also the chance that letters were altered if reprinted in a journal or an edited report. The contextual information may provide information about whether words were transformed or whether “no words have been altered” (Deck October-November 1937; AU PMB DOC 440).

2.2.2.7 Anglicisation

The reliability of the earliest attested contact language samples is further influenced by the fact that they were found in sources which were written by Europeans. As mentioned above, most of the Pacific Islanders, at least in the early 19th century, were illiterate. Printing was foreign to the Pacific region so that travel, sojourn and government reports were printed in European cities. This will have increased the timespan between the publication of an utterance and the original language situation since it was very time consuming to bring the material from the Pacific to the other side of the world. At the same time the printing in European cities implied that the earliest samples of the contact varieties were printed from a European perspective. The early writers quoting Pidgin English could not resort to a writing system. This, together with the fact that English was the major lexifier of the contact varieties in the Pacific, led most of the writers to use an English orthography instead of a more phonetic spelling. Especially authors with negative attitudes deployed an English spelling to highlight the “crippled” and “corrupt” English” of the Pacific (Kunze 1897: 70;¹² Cormack 1944: 130) and to avoid that “mutilated English words were considered to be of native origin” (Stephan & Graebner 1907: 20).¹³ Depending on the text type and its audience, a writer may have adapted the language portrayal of the pidgin language to meet the perceptions the audience or readers could recognise.

That the use of an English orthography may shift a variety towards the lexifier and may obscure what the language was really like, can be learned from Shelton-Smith (1929: 10) who, referring to New Guinea, claims that “the English phonetic does not even nearly approximate the native pronunciation, but it is necessary to use it for the sake of clarity”. For the Solomon Islands, evidence is found in Hogbin (1939: 163) who argues that “although the vocabulary is very largely English the constructions are Melanesian pronunciation” and Jacomb, referring to the New Hebrides claims that “of course in pronunciation and tonality there are expressive differences” (1914: 91). Some of the early sources point towards the differences between the English and Pidgin English phonology of words. English-derived lexicon is alienated, for instance, by consonant cluster reduction, by absence or replacement of the interdental fricatives [θ] and [ð], and by replacement of palato-alveolar sibilants, to mention a few (cf. Army Education Branch 1944: 2-3).

The anglicisation of Pidgin English examples further covers the fact that though the bulk of the lexicon may have its origin in English, words have undergone semantic changes and

¹² “in seinem krüppelhaften Englisch” (Kunze 1897: 70)

¹³ “Eine weitere Gefahr [...] ist die, daß man verstümmelte englische Worte für eingeborene hält” (Stephan & Graebner 1907: 20).

expansion so that the “language is as unintelligible as though English words were not used” (Norden 1926: 42). An early example of semantic expansion represents the word *grass* which is attested with the meaning ‘hair’ as early as 1888.

The greatest issue in terms of anglicisation with regard to the present study is that the anglicisation might obscure variation and grammatical developments in the analysed varieties. Grammaticalisation processes are frequently accompanied by changes in the form. For instance, *this fellow* is a common noun phrase in some Standard varieties of English. In the Pidgin English of the Pacific *this fellow* grammaticalised into the demonstrative and can thus be followed by a noun (e.g. *this fellow kaikai* ‘this food’). In this grammaticalisation process the form *this fellow* develops into *dispela*. Thus, on the one hand, the English orthography shows the etymological origin of a Pidgin lexeme. On the other hand, it might be that the form *this fellow* turned into *dispela* much earlier, but that the development is obscured by the English orthography being used.

2.3 Summary and outlook

The present chapter outlined the theoretical background of studying pidgins and creoles from a diachronic perspective. While Section 2.1 discussed the challenges which any historical study must deal with, Section 2.2 referred to specific challenges which historical creolists encounter. Apart from the problem of the availability of early language samples, it was demonstrated that historical language data needs to be carefully evaluated in terms of its reliability. Possible reliability-influencing factors were introduced which need to be revisited in the methodological framework (Chapter 4).

As pointed out in the quote by Arends (1995: ix) at the beginning of this chapter, in order to make the best out of the fragmented data, language-internal data needs to be analysed in combination with detailed extralinguistic information. Thus, the following chapter will give a socio-historical account of language contact in Melanesia and will provide the basis for understanding language internal developments in Solomon Islands Pijin (Pidgincreole (English-lexified): Solomon Islands), Bislama (Pidgincreole (English-lexified): Vanuatu) and Tok Pisin (Pidgincreole (English-lexified): Papua).¹⁴

¹⁴ At present, the three varieties can best be described as *pidgincreoles*. The term, which was coined by Philip Baker, expresses that the varieties represent former pidgins that are nativising and thus become the mother tongues for some of their speakers. They cannot be considered creoles yet since they are not the mother tongues for the whole community (cf. Bakker 2008: 131).

3 The origin and development of MPE and its geographical variation

This chapter focusses on the origin and development of Melanesian Pidgin English (MPE). It starts by exploring the sociolinguistic history of language contact that fostered the development of contact varieties in Melanesia and thereabouts (Section 3.1). The historical developments will form the basis in understanding the main theories that have been put forth regarding the origin and development of Melanesian Pidgin English(es), which will be introduced in Section 3.2.

3.1 A socio-historical account of language contact in Melanesia

If a diachronic analysis of language is to be satisfactorily comprehensive, it must also take into account the on-the-ground, contemporary social realities in which the relevant linguistic production originated.

(Soukup 2017: 673)

Contact languages arise if members of a society need to communicate but do not have a shared language available. While people involved in occasional contacts might get along and communicate with ad hoc solutions (cf., for instance, Velupillai 2015: 19), more permanent contact situations between people without a shared language necessitate what Bakker refers to as, “normative systems of communication” (2008: 131). This can be explained by the fact that languages need to meet “the different and multifarious social needs of the communities that use them” (Lyons 1968: 43). As soon as the social needs of a community change, their language will “tend to change to meet the new conditions” (1968: 43). Thus, if the amount of communicative functions for which a contact variety is used increase, the variety will elaborate and might change on all linguistic levels.

Therefore, language development cannot be analysed detached from its social settings (cf. Soukup 2017: 673). To understand the origin of Melanesian Pidgin English and its further development into Solomon Islands Pijin, Bislama and Tok Pisin, a look at its socio-history is a prerequisite. Various sociolinguistic factors, such as the interaction and ratio of contact between Europeans vs. non-Europeans and between different Pacific Islanders need to be taken into consideration to understand the varieties’ development (cf. Velupillai 2015).

Thus, the present chapter will give an account of the social and historical circumstances of language contact which led to the development of contact varieties in the Pacific. It will only present an *overview* of the sociolinguistic developments in the three island groups – which themselves consist of thousands of smaller islands, provinces, regions and villages – so that a detailed account of the history and developments cannot be provided in the scope of this work. As the present study concentrates on English-lexified contact varieties, the focus is placed on

contact situations in which Papua New Guineans, Solomon Islanders, Vanuatians¹⁵ and Europeans were involved. It should be noted, however, that non-European contact varieties existed as well (see, for instance, Drechsel 2014). The descriptions of the social and historical conditions of language contact are to a high degree based on my examination of pre-colonial and colonial sources but were supplemented by recent research on the history of Melanesia. They will help to understand and interpret the results of the diachronic analysis of the varieties in Chapters 6-9. The map in Figure 5 shows the place names mentioned throughout the chapter.

¹⁵ Since 1980 inhabitants of Vanuatu refer to themselves as *Ni-Vanuatu*. Throughout this thesis, I will nonetheless refer to them as *Vanuatians* as the study focusses on developments prior to 1980.

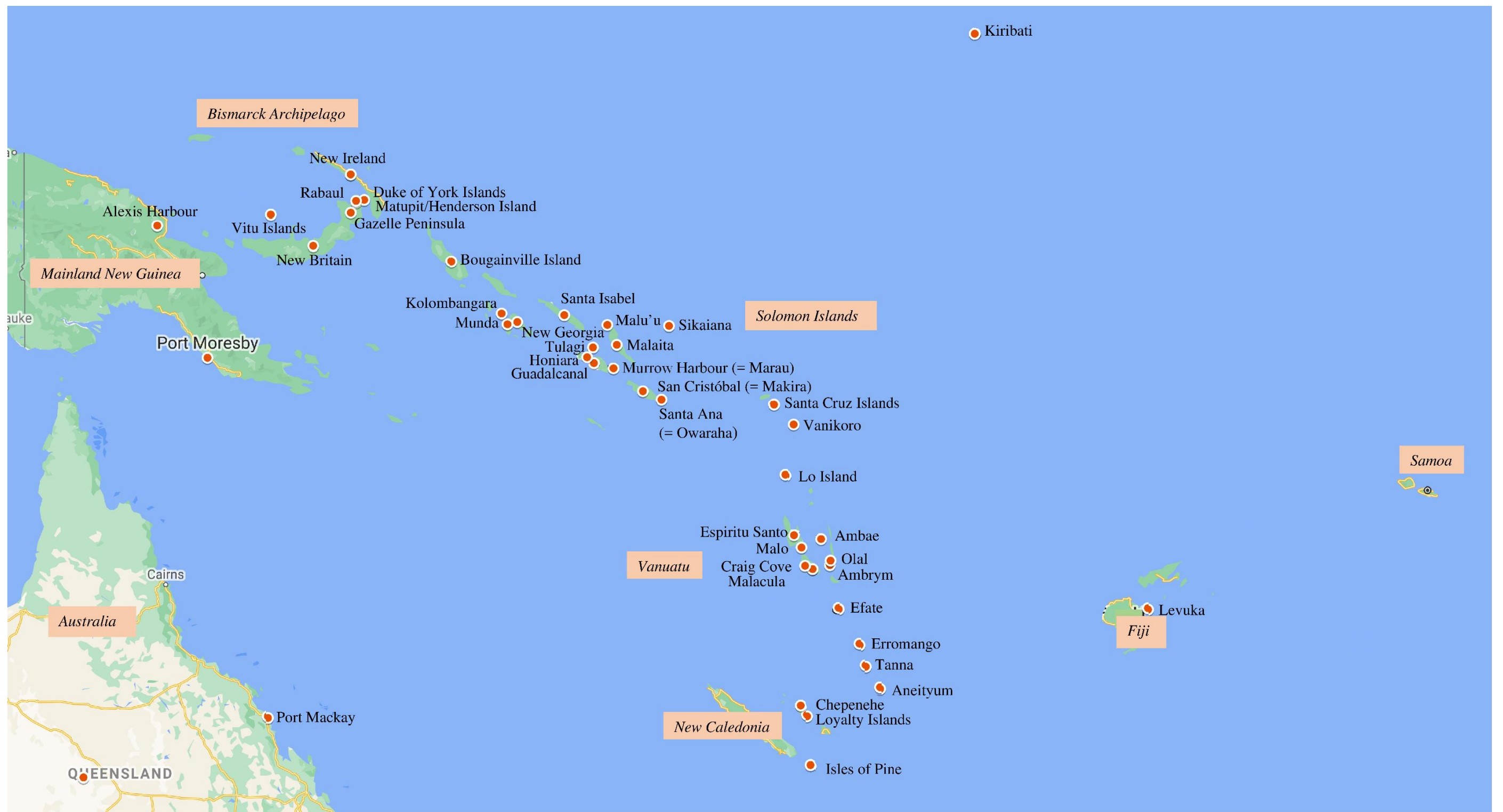


Figure 5: Map of the Pacific Islands, showing the place names mentioned in Section 3.1 (created with Google Maps)

3.1.1 The first explorations

The earliest contacts between Europeans and Pacific Islanders can be dated to the 16th and 17th centuries when Europeans passed by the Solomon Islands, Vanuatu and Papua New Guinea on their discovery and exploratory voyages in search for treasures, goods and unexplored land. The Europeans had to communicate with the Pacific Islanders in order to barter goods for food. This begs the question of how the parties communicated with each other. This chapter gives an account of the first (documented) encounters between Europeans and Pacific Islanders in the Solomon Islands (3.1.1.1), Vanuatu (3.1.1.2) and Papua New Guinea (3.1.1.3) with a special focus on how communication occurred in the initial situations of contact.

3.1.1.1 The exploration of the Solomon Islands

The first European explorer that arrived in the Solomon Islands was the Spaniard Álvaro de Mendaña de Neyra who was sent by the governor Lope García de Castro in an expedition of two ships to search for the Ophir Islands, which were said to be the islands to which the Biblical King Solomon had sent his treasures. The crew left Callao, near Lima in Peru in 1567 and arrived on what they named Estrella Bay on Santa Isabel in November 1568 (cf. Braumann 2018). During their first exploration, they gave Spanish names to the Solomon Islands which are still used today (cf., for instance, San Cristóbal, Guadalcanal, Santa Ana) (cf. Hackney & Thompson 1901). Since the primary encounters with the Solomon Islanders were frequently hostile, they returned to Peru. One year later Mendaña started a second expedition with the aim to establish settlements. In his reports about the second journey, there is evidence that communication with Solomon Islanders took place. Mendaña mentions that they communicated “by signs” (1901: 120) as no shared language existed.¹⁶ However, there is also evidence that words were interchanged between the groups. The following account describes a situation on Santa Isabel on 13 November 1568:

I made him sit down, and began to ask him what they called the sun, the moon, the sky, and other things ; and he named them all in his tongue, which is such that it may easily be learned by us [...] They seemed very eager to learn our words, and asked us to teach them, at which we were greatly rejoiced [...]. (1901: 113)

Moreover, communication took place by taking local inhabitants as interpreters (cf., for instance, Sarmiento or Medaña in Hackney & Thompson 1901: 63, 88, 90, 123).

¹⁶ Cf. “All this conversation was carried on in a few words of his language, and by signs, which they are very quick at understanding” (Medaña edited and translated by Hackney & Thompson 1901: 130).

3.1.1.2 The exploration of Vanuatu

The first European to arrive on Santo, an island of Vanuatu, was Pedro Fernando de Quirós on 3 May 1606 (cf. Markham 1904: xxv). Only a century later, the French Louis Antoine de Bougainville arrived in Ambae and Malo, sailing also to Malekula and Santo. Antoine de Bougainville was followed by James Cook who travelled to the island group as the commander of *HMS Resolution* and arrived in 1774.

Reports such as “I then made signs (for we understood not a word of their language) that we wanted wood ; and they made signs to us to cut down the trees” (Cook 1842: 498), show that the use of signs and pointing were the usual tools of communication in the initial encounters between Europeans and non-Europeans in Vanuatu. Clark (1979) reports that it was also a common practice of European ship crews to learn words of the local languages but that there was no attempt to teach the Pacific Islanders the European languages.

3.1.1.3 The exploration of Papua New Guinea

The earliest contacts between Europeans and Pacific Islanders in Papua New Guinea date back to 1526 when the Portuguese Jorge de Meneses arrived on the west coast of what is today Papua New Guinea. Meneses named one of the islands *Ilhas dos Papuas* to describe it as the land of frizzy-haired people (cf. de Barros 1777: 179ff.). Nineteen years later, the Spanish explorer Inigo Ortiz de Retes passed the north coast and named it *Nueva Guinea* because of the “similarity of its inhabitants with those of Guinea in Africa” (Sociedad Geográfica 1885: 239;¹⁷ Waiko 1993: 17). No information about communication in these initial contact situations could be obtained. Around the turn of the century Dutch ships are said to have arrived in Melanesia; however, no attempts were made to settle permanently on the island group before 1793.

3.1.1.4 Summary

In sum we see that despite the early exploratory voyages made by Europeans to the Melanesian Islands, permanent contact did not exist. Instead, interactions were sporadic and short-term, and signs and body language were used to guarantee that the parties involved understood each other. Moreover, to facilitate communication, individual words were occasionally taught to the other party and interpreters were employed where possible.

¹⁷ “[...] reemplazando con el nombre de Nueva Guinea, por la semejanza de sus habitantes con los de la Guinea africana” (Sociedad Geográfica 1885: 239)

3.1.2 The period of trading contacts

Even if the first contacts between Europeans and Pacific Islanders started with the exploratory voyages described above, the 19th century can be regarded as the *real* beginning of contact and communication. This was the time when whaling, sandalwood and trepang trade saw a boost. Even though the contact between Europeans and Pacific Islanders might still have been sporadic, the involved parties needed to communicate during trading interactions on the coastal areas. Moreover, trading vessels needed ship crews and collected their crew members from diverse Pacific Islands so that onboard, Europeans and Pacific Islanders with different linguistic backgrounds came together and had to find a tool of communication.

When the trading vessels arrived in the three island groups, they were confronted with what is frequently referred to in the early sources as a *Babel of tongues* in reference to the Biblical story of the *Tower of Babel* (cf. Forbes 1875: 198; Steel 1880: 312; Inglis 1887: 7; Adams 1890: 66; Watt 1896: 118; Grimshaw 1907: 229 to mention just a few). The multilingual complexity fostered the development of contact varieties even before the first European traders arrived in the Pacific, leading to Oceanic based contact languages (cf. Mühlhäusler et al. 1996). However, contact jargons between Europeans and Pacific Islanders developed for the first time during the trading period. As Schuchardt (1883: 1/151; translation: Gilbert 1980: 15) states:

[t]he whalers were the first to have contacts (loose, to be sure) with the islanders; but the fact that they touched on such different and remote coasts and stayed away so long from civilized countries furthered to a large extent the formation of a jargon, usually known simply as ‘whalers’ jargon’.

Thus, a Maritime jargon developed in the Pacific, although its usage was restricted to coastal areas and ship crews.

Both contact between the various Melanesian islands and between the ports in China and Australia (Sydney) developed, establishing a triangular trade in which “Pacific island products were acquired for Western manufactured goods, and then exchanged for Chinese silk and tea” (Marks 2012: 225). Thus, contact varieties that had developed by that time in China (= Chinese Pidgin English)¹⁸ and New South Wales (= New South Wales Pidgin English)¹⁹ may have had an impact on the jargon(s) spoken in Melanesia as well.

It is well attested that trading took place in the Pacific, but it needs to be clarified to what extent trading was relevant in the Solomon Islands (3.1.2.1), Vanuatu (3.1.2.2) and Papua New Guinea (3.1.2.3). If the areas were involved in trading, it is essential to determine whether the usage of a contact jargon came along with the trading.

¹⁸ (Pidgin (English-lexified): China)

¹⁹ (Pidgin (English-lexified): New South Wales)

The book *Log of Logs* by Ian Nicholson (1990) lists the ships that travelled into Australian and New Zealand waters. The list also contains information regarding the various places the ships visited on their route and the purpose of their voyages (e.g. recruiting, trading, procuring sandalwood, whaling, missionary voyage). Nicholson's work was consulted to investigate whether ships from Australia and New Zealand went to the Solomon Islands, Vanuatu and Papua New Guinea for whaling, sandalwood or trepan trade. Further evidence was gained from the collected diachronic data at hand.

3.1.2.1 Trading contacts in the Solomon Islands

The Solomon Islands were actively involved in the early whaling activities in the Pacific. Evidence is inter alia found in Cheyne (1852: 67) who mentions that Malaita and the Bougainville Islands have been involved in whaling contacts. From surviving logbooks, it can be learned that whaling cruises were made to other Solomon Islands as well. For instance, Nicholson (1990: 15) lists the *Alfred* that went to the Santa Cruz Islands in 1827/1828 and Richards (2010: 15) refers to the *Wallaby* which went on a sperm whaling voyage ending up at Murrow Harbour, where it "lay trading for three weeks" in 1840.

The whaling contacts were paralleled by the development of contact jargons of which evidence can be found, inter alia, in a book by Andrew Cheyne, a trader who went to the Solomon Islands in the 1840s. He reports about Sikaiana, an atoll of the Solomon Islands that the inhabitants "can nearly all speak more or less broken English, which they have picked up through their intercourse with whale ships, who often visit them to get supplies of cocoa-nuts and pigs" (1852: 53). This attests to the fact that the Solomon Islands were involved in the whaling activities and proves concurrently that a contact variety was used in interactions between whaling crews and the island inhabitants. Elsewhere Cheyne reports to have come to the Solomon Islands to "form[...] establishments for collecting and curing biche de mer [= sea cucumbers] for the China market" (1852: 54), substantiating that the island group was also involved in the beach-la-mar trade. Webster (1863: 51-52), referring to Sikaiana in the year 1851, provides further evidence that the contact variety was used in beach-la-mar trading activities:

We were surprised to hear several of them address us in very good broken English. They informed us that a party of Europeans had been some time on their Island, collecting beche-le-mer, which abounds on the reefs. From this party they had gained considerable proficiency in our language.

Webster (1863) also provides evidence for an early trade-based contact jargon on the island of Makira referring to the year 1851: "He gave his name as Jerobo, and could speak a few words of English, having been on board a whaler for a short time off this coast. [...] By the assistance of Jerobo's broken English, we had a long conversation with the party" (1863: 102).

Besides the usage of contact jargons, interpreters were employed in order to trade with Solomon Islanders (cf., for instance, Cheyne 1852: 67). These interpreters were usually Solomon Islanders who were involved in the early trading contacts, had picked up a knowledge of the trade jargon(s) and were later able to apply their knowledge in further contacts with foreigners.

3.1.2.2 Trading contacts in Vanuatu

In Vanuatu trading played an important role even if whaling activities were of less importance than the sandalwood trade. This is reflected in the *Log of logs* book which lists almost no whaling ships that called at the New Hebrides, as they were known by then. In 1872 the reverend John Kay reports, however, that “[f]or a number of years past, there has been one whaling establishment on Aneityum” (1872: 34), and *The Nautical Magazine* informs us that there were “voyages, in the whaling service, between the years of 1828 and 1834” (1839: 603) to the island group. To what extent contact jargons were used in these contact situations cannot be proven with the data at hand.

Nevertheless, the use of early jargons can be traced back to the trepan and sandalwood trade. Early sources referring to the 1930s inform that “[i]n the northern districts of Tanna, and also on Erromanga, is abundance of sandalwood” (Heath 1941: 96; cf. also Gordon 1862: 3) and that on the New Hebrides, New Caledonia, and the Loyalty Islands there are “from ten to twenty vessels being constantly engaged in the lucrative trade in sandalwood, and beche-le-mer, with China” (cf. *The Nautical Magazine* 1850: 425).

The lucrative sandalwood trade resulted in the establishment of the first permanent station on Aneityum by James Paddon in 1844 (cf. Hilliard 1970: 122).²⁰ In addition, it is reported that Tannese Islanders were recruited to work at sandalwood stations on the Isle of Pines (New Caledonia) for the duration of three months between 1848 and 1861 (cf. Shineberg 1967: 84; Tryon & Charpentier 2004: 108). On the sandalwood stations labourers with different linguistic backgrounds had to work and communicate with each other, which evoked a need for contact jargons. That frequent contact between New Caledonia and Vanuatu existed is evident from the following quote, which shows the answer given by a Vanuatuan about whether he knew about trepan: “Yes [...] [N]ew Caledonia big canoe come, stop for hims [...] Cap’n use native, and him give guns” (Munro 1867: 201-202).

Further evidence for the use of contact jargons can be found in Erskine, who reports on a Tannese in the 1840s who “spoke some words of English with a very distinct pronunciation” (1853: 307). In another passage the reader is informed about a boy who had worked on a ship and

²⁰ Some sources mention 1843 as the year of the establishment (cf. Tryon & Charpentier 2004: 108).

“spoke English tolerably” (1853: 393) so that by the time bishop Selwyn came to Vanuatu in 1851, “‘sandalwood English’ was widely understood” (Hilliard 1970: 122).

The earliest language example found for the region of Vanuatu dates back to the year 1830. From 1829 until 1831 George Bennett travelled in the South Pacific and reports that he met Elau, a child from Erromanga, of whom he quotes several language examples, such as “Ungka no like play now” (1883: 3).

3.1.2.3 Trading contacts in New Guinea

New Guinea seems to have been only scantily involved in early trading contacts. The previously mentioned *Log of logs* does not list whaling, sandalwood or other trading vessels going to Papua or New Guinea between 1800 and 1850.²¹ There is only one exception, namely the whaling ship *Hydrus*, which got lost on the coast of New Guinea. As the “officers and the rest of the crew were murdered by the natives” (*Shipping Gazette and Sydney General Trade List* 01.03.1945: 59) there was no oral interaction involved in the short contact situation. This corroborates Tryon & Charpentier (2004: 79-80) who, based on Cumpston (1963) and Nicholson (1977), did not identify South Pacific voyages to New Guinea.

The early data collected contains information about the exploring brig *Margaret Oakley*, in which the earliest attestation of an English-lexified contact variety on the Witu islands can be found. The speaker uses, inter alia, the words “Me no speak lie! me real Darco [...]” (Jacobs 1844: 80), demonstrating that the area probably was visited earlier by trading vessels.

Despite the scarce attestations of trading contacts in New Guinea, it can be assumed that New Ireland was involved as it “was situated on the shipping route from Australia to China” (Tryon et al. 1996: 483).²² No testimonies were found for the years 1800 to the 1850s. Evidence for New Ireland only dates to the year 1876 and was found in a book by the German H. Strauch

²¹ The territory that is known as *Papua New Guinea* today has a long history of renaming that is strongly connected to its complex administrative history. When the British established a protectorate over south-east New Guinea in 1884, they named the territory *British New Guinea*, while the northern part of New Guinea, which was annexed by Germany, became to be known as *German New Guinea*. However, when the British had to transfer the control of British New Guinea to the Commonwealth of Australia in 1906, the territory was renamed into *Territory of Papua*. Fifteen years later, in 1921, Australia was granted a mandate to run German New Guinea as well and the territory was henceforth called the *Mandated Territory of New Guinea*. Even if both territories were governed by Australia by then, the areas were still governed separately. Only after World War II Australia established a joint administration over both regions and the territory was renamed into *Territory of Papua and New Guinea*. It was not before 1971 that the territory was named *Papua New Guinea*.

²² Similar as with the name for Papua New Guinea, the individual islands and areas of Papua New Guinea became renamed several times. For example, the island that is referred to as *New Ireland* today was formerly known as *Neumecklenburg* during the German administration. What is today’s *New Britain* was called *Neupommern* during the German administration. Even if it is more common to use the names as they were at the time of the events, since, for example, location boundaries may have differed in the past, I decided to use the contemporary designations to refer to the areas. This will make it easier to locate the islands/areas/towns/villages for readers that are not familiar with the historical names.

who states that due to the contact between New Irelanders and trading ships, some of them had a small knowledge of English: “The inhabitants of the small village were in contact with ships that were regularly abiding there; some of them even had a pleasant knowledge of the English language” (1876: 406).²³

Moreover, when the missionary George Brown reports about his encounter with a Matupit Islander in 1875, who served as an interpreter speaking “the best kind of English that was then spoken there by the few who knew it” (Brown 1908: 93), this might be an indicator that earlier contact situations had existed.²⁴ Nonetheless, the area that became German New Guinea was only marginally involved in trading contacts.

3.1.2.4 Summary

For the trading period in the beginning of the 19th century we thus see that trade-based interaction between Europeans and Pacific Islanders took place in all three areas under investigation – albeit to varying degrees. Trading contacts were more common in Vanuatu and the Solomon Islands than in New Guinea. They promoted the development of contact jargons as people with diverse linguistic backgrounds had to communicate in the coastal areas and, especially, on boards of the trading vessels.

The trading vessels landed not only on the shores of various Pacific Islands, but also on the shores of China, Australia and New Zealand. As such, linguistic input from the different areas may have influenced the jargon(s). This involves not only European, Pacific and Asiatic languages, but also other contact varieties that had developed by that time (cf., for instance, Chinese Pidgin English or South Wales Pidgin English, as well as non-English-based contact languages). As Schellong points out: “The crews might come onboard of a different vessel after some time, where they will meet new islanders and will adopt new words” (1934: 98).²⁵ When being relocated, ship crews may have contributed to the development of early jargons and possibly spread the jargon(s) to new places so that *diffusion* represents an important influencing factor.

In the case of Vanuatu, the establishment of shore stations on Aneityum and the presence of Vanuatuan labourers at stations in New Caledonia promoted the development of the early

²³ “Die Bewohner des kleinen hier befindlichen Dorfes waren jedenfalls vielfach in Berührung mit dort anscheinend regelmässig verweilenden Schiffen gewesen, einige besaßen sogar eine erfreuliche Kenntniss der englischen Sprache” (Strauch 1876: 406).

²⁴ When Brown quotes the Matupit Islander, it becomes clear that “the best kind of English” refers to an early form of a contact variety: “Missionary no come Matupit, ah! Topulu he no come. Missionary come, oh! Topulu he come. He go house belong Matupit” (Brown 1908: 93).

²⁵ “Die Mannschaften kommen nach einiger Zeit vielleicht auf ein anderes Schiff, treffen hier wieder mit anderen Insulanern zusammen und nehmen wiederum andere Worte an” (Schellong 1934: 98).

contact variety as linguistically diverse labourers were forced to work and communicate with each other. Due to the higher intensity of language contact in Vanuatu and the Solomon Islands, a more advanced development of contact jargons can be expected in comparison to New Guinea. However, the jargon(s) can be assumed to have been “narrowly restricted in essential conditions to one or at most two European men, and the few communities of islanders with which they were in intimate contact” and to have been characterised by variability (Churchill 1911: 8). The primary aim in the initial trading encounters was to trade efficiently; the trade jargon(s) were thus only used by those people who were involved in the actual trade interactions. As ship crews and the participants of the trading interactions changed frequently, it may be assumed that the jargons were unstable and “ha[d] to be reinvented for each situation and by every user” (Velupillai 2015: 19).

3.1.3 Contact on overseas plantations

Contact between Europeans and Melanesians as well as between Melanesians from different geographical areas intensified with the establishment of the first plantations in Queensland, Samoa, New Caledonia and Fiji. In 1863 plantation systems for growing cotton developed in Australia. Due to the American Civil War, the British mills had to cope with a shortage of cotton which resulted in the establishment of cotton plantations in Australia (cf. Holm 1988: 527). One of the first major entrepreneurs active in Australia was Robert Towns, who had already been involved in the Sandalwood trade on the New Hebrides. In need of labourers, Towns introduced the idea of recruiting his labour force in the New Hebrides and “arranged with resident Tanna planter Ross Lewin to recruit on his behalf” (Tryon & Charpentier 2004: 174), resulting in 66 Vanuatuans working for one year on Queensland plantations. Plantation owners in other countries copied Towns’ idea and started to recruit on surrounding islands (cf. Schnee 1904: 58). The demand for manpower from the Pacific Islands was increasing, especially with the end of the American Civil War, as the cultivation of cotton plantations declined in Queensland, and sugar was increasingly cultivated (cf. Tryon & Charpentier 2004: 175). Small plantations grew into large-scale agricultural units, and “the greater the expansion of the industry, the greater the demand for ‘kanakas’” (Corris 1968: 85), causing the labour trade to prosper. This led to a new and intensified form of contact between Pacific Islanders and Europeans, which is outlined in the following section.

3.1.3.1 Recruitment in adjacent islands

Queensland, Samoan, New Caledonian and Fijian plantation owners sent out recruiting schooners to the New Hebrides, the Loyalties, the Solomon Islands, Kiribati, New Guinea and other smaller

island groups in search for labourers who *make paper* (i.e. ‘sign a contract’; cf., for instance, Tropical 1921: 7). The islanders were bound per indentured labour contracts to work on a plantation for a limited amount of time. The early sources show that the most common contract duration was for three years (cf. Jacques 1922: 72); however, two- and five-year contracts existed as well (cf. Idriess 1941: 88; Tryon & Charpentier 2004: 180). As recruiting vessels did not always succeed in getting the islanders aboard the plantation schooners, they started to force them to come along and kidnapped them – a practice called “blackbirding” (cf. Schnee 1904: 58). According to Jacques (1922: 73), the violent procedure is called “pullen” which probably derives from the English verb ‘to pull’. It thus symbolises the practice of recruiters, who, when failing to recruit willing workers, lured people to the waterfront and pulled them into the boats. Even if some workers were recruited by their own choice, Parkinson makes clear that in most cases they did not understand the incurred liabilities of their contracts (cf. 1887: 28).

Early evidence shows that jargons served as the medium of communication for recruiters on board of the recruiting vessels. In Vanuatu, they are reported to have employed islanders using a contact variety saying “Yes; suppose you let him some boy go along a Queensland, we buy him altogether [...]” (Giles 1968: 41) and in Forbes (1875: 251) recruiters address the Vanuatuan with “You likee come work Fiji?” Parkinson reports about a Samoan recruiter who asked New Guineans in, what Parkinson defines as typical South-Sea English: “You like go Samoa?” (Parkinson 1887: 29). Evidence that the early contact varieties were used in recruitment interactions between Solomon Islanders and the recruiters can be found in Cromar (1935: 138): “We were here to obtain recruits for Queensland. [...] Then he asked: ‘You buy’em boy along shoota”.

3.1.3.2 Linguistic diversity on the plantations

Once arrived on the plantations, Pacific Islanders with linguistically diverse backgrounds had to work together under European supervision. The areas were characterised by a high level of language diversity (cf. Lynch 1923: 26; Frommund 1926: 51-52; Collinson 1929: 20). The situation thus required a medium of communication through which the labourers could communicate “horizontally, among the labourers” but also “vertically, between the workers and the white owners and supervisors” (Mühlhäusler 1979: 60).

Compared to trading encounters, the contact situation on the plantations was no longer short-term but individual workers were exposed to the contact situation for at least two years. As shown in Section 3.1.2, contact jargons with their origin in trading activities were in use before the establishment of the first plantations. It is reasonable to assume that labourers with a

knowledge of such a jargon, due to previous contact with Europeans, will have made use of it. As jargons were also used on recruiting vessels, they may have provided a significant input for the resulting plantation pidgins that developed in the second half of the 19th century.

3.1.3.3 The development of plantation pidgins

The plantations provided a fertile soil for the development of pidgins since trading jargons came into renewed contact with both superstrate and substrate languages. Demographic factors, such as the origin of the population groups, will have determined the degree to which the workers' mother tongues influenced the resulting pidgin. As Velupillai highlights, it can be assumed that "the substratal input language(s) with many speakers had more influence on the outcome of the contact language than the substratal input language(s) with few speakers" (Velupillai 2015: 109). A similar claim was already made by Schellong as early as 1934. He argues that the "tribe which was predominant during the encounter of different islanders will also receive linguistic predominance" (Schellong 1934: 98).²⁶ Since labourers left the plantations when their contracts expired and *new chums* ('newcomers') were consistently arriving on the plantations, the pidgin varieties might have changed depending on the ecology of the contact group. The early sources provide evidence that it was also common that after the contract expired, workers were reemployed to work on plantations in another overseas area. For instance, a worker who had previously worked on Samoa or in Queensland could be reemployed to a Fiji plantation (cf. Colonial Office 1883-84: 218; Norden 1926: 96). Thus, the plantation pidgins may have influenced each other as well.

On the plantations in Queensland, the Pacific jargons and English that were used as the medium of communication developed into *Queensland Plantation Pidgin English* (QPPE; Pidgin (English-lexified): Queensland)).²⁷ On Samoan plantations, the jargons developed into *Samoa Plantation Pidgin English* (SPPE; Pidgin (English-lexified): Samoa) (cf. Mühlhäusler 1978: 70). It is also possible that a form of German (Indo-European (Germanic): Germany) served as an additional means of communication on plantations in Samoa. For instance, an author of the *Samoaanische Zeitung* advocates for a control of Pidgin English and the "prevalent German vernacular with more or less corrupted English expressions" (Anonymous 1913: 1).²⁸ In addition,

²⁶ "Der Volksstamm, der bei dem Zusammentreffen verschiedener Inselbewohner gerade das numerische Übergewicht hat, wird wohl auch in gewissem Grade das sprachliche Übergewicht bekommen" (Schellong 1934: 98).

²⁷ There has been considerable research on the variety. For further information, see Dutton (1980), Dutton & Mühlhäusler (1984) and Mühlhäusler (1996).

²⁸ "Sie richten sich in zweiter Linie natürlich auch gegen die sehr überhand nehmende Durchsetzung der deutschen Umgangssprache mit mehr oder weniger verballhornisierten englischen Ausdrücken, deren Gebrauch durch das starke Vorhandensein des Pidgin-Englisch sehr gefördert wird" (Anonymous 1913: 1).

Zieschank reports that the German on Samoa was “strongly mixed with English chunks” (Zieschank 1918: 57).²⁹ Unfortunately, there is no information available on who in fact spoke this “colloquial” German and where exactly it was used. Although there is no language data available, the quotes make clear that the German language was not completely absent. At the same time, the small amount of evidence we have reflects that SPPE represented the dominant contact variety used on Samoan plantations.

On the plantations in Fiji, the predominant language was Standard Fijian (Austronesian (Oceanic): Fiji) (cf. Siegel 1987: 73). In the initial years, plantation owners sought manpower on the Fiji Islands only. It was not before 1865 that Vanuatians, Solomon Islanders, New Guineans and Kiribati Islanders were recruited as well, while the recruitment of labourers on Fiji Islands was sustained. This may have been one reason why the use of Fijian, instead of the development of a Pidgin English, was possible. Another reason may be that recruits from the same kin-groups were accommodated together. Forbes reports that “[e]ach tribe of imported labourers on every plantation has its separate hut, or collection of huts. A Tasmanian will not fraternise with a Malacoba man, nor an Erromango man with an Ambhoym man” (1875: 61). Nonetheless, it seems likely that situations occurred in which Fijians needed to communicate with other Pacific Islanders. For instance, Brewster (1937: 101) reports that the various groups communicated “in the beche-de-mer or pidgin English which with Fijian forms the lingua franca” (in Tryon & Charpentier 2004: 190). Moreover, there is evidence that earlier trade jargons were used to recruit workers from the various islands (cf. Forbes 1875: 251). Nonetheless, evidence for Pidgin English being used on the plantations in Fiji is scarce. Siegel concludes that:

in the first half of the plantation era, MPE was used by Pacific Islanders in Fiji only for communication with Europeans (when either or both did not know Fijian). Therefore, it was used only in a dual contact situation. As it was not used among the islanders themselves in Fiji (as it was, by contrast, in Queensland and Samoa), MPE did not stabilize further into a distinct variety, and a “Fiji Plantation Pidgin English” never developed. (Siegel 1987: 81)

On the plantations in New Caledonia that were owned by English speakers, an early form of New Caledonia Pidgin English (Pidgin (English-lexified): New Caledonia) was used (cf. Tryon et al. 1996: 476; Mühlhäusler & Baker 1996: 511). However, when the French increasingly gained control “labourers spoke French among themselves and pidgin French developed” (Hollyman 1976: 44 in Tryon & Charpentier 2004: 205) so that English- and French-lexified pidgins coexisted in New Caledonia.

²⁹ “Daß aber die englische Sprache vorherrschend blieb und auch unser Deutsch hier stark mit englischen Brocken vermengt wird, ist fast die ausschließliche Schuld der alten Ansiedler selbst, die sich über den Mangel an deutscher Art beklagen” (Zieschank 1918: 57).

3.1.3.4 Returning labourers and the diffusion of plantation pidgins

The origin of the labourers was not only of importance for the development of pidgins on the plantations, but also had an impact on their diffusion. Since the indentured labourers, when their contracts expired, were repatriated either to their places of origin (cf. Parkinson 1887: 27; Cromar 1935: 117) or to places other than their home areas, they will have taken with them a knowledge of the plantation variety. The latter occurred if no exact place of origin could be determined, and labourers were brought to places which were thought to be nearest to their home areas (cf. Laycock 1970: x). Other islanders were scared to return to their home villages, as can be learned from Ivens (1918: 227), who reports about men being landed in places other than their homes “owing to a fear of reprisals for some act of wrong-doing which they had committed and which had led to their recruiting”.

Even if repatriates presumably did not speak the contact language with their kin-group members, there is evidence that plantation pidgin varieties were used in later contacts with Europeans. For instance, the German Krämer-Bannow reports that in their interactions with natives they made use of the Pidgin English of the Bismarck Archipelago, which would have been brought to the area by returned workers (cf. 1916: 20). Wendeland provides an example in which the recruit forgets his home language. When he meets a compatriot, who tries to talk with him in his mother tongue, Wendeland claims “[e]ven though he understood him, he could not answer in his native language, because it had disappeared from his memory” (1939: 76-77).³⁰ Even if it is likely that the returning Pacific Islanders regained the knowledge of their mother tongue quickly, the returnees might have made use of the pidgin language until the full knowledge of their native tongue came back. Moreover, even if the episode might have been exaggerated, Pacific Islanders may have used the contact variety for trading purposes with neighbouring villages. In pre-colonial times it was a common practice that a boy from one village was raised in a family of a neighbouring village to acquire their language and to serve as an interpreter in trading situations (cf. Wendeland 1939: 18). The learned pidgin variety will have provided an alternative medium for trading communication between neighbouring villages.

Reports about returned labourers having taught the pidgin variety to other men in the islands is a further indicator for that the plantation pidgins were taken along and used by the returnees in interaction in their home areas:

³⁰ “Auf meine teilnehmende Frage, was ihn so traurig stimmte, schluchzte er: „*Master! Me lose him talk b'long place b'long me; suppose me come back belong place b'long me, me no more save talk.*” (Herr! Ich habe meine Heimatsprache vergessen; wenn ich nach meinem Platz zurückkomme, kann ich nicht mehr die Sprache reden.) [...] Heute war er einem neuangeworbenen, eben mit dem Schiff eingetroffenen Landsmann begegnet, der ihn freudig in seiner Muttersprache begrüßt hatte. Obwohl er ihn verstand, konnte er ihm nicht in seiner Heimatsprache antworten, weil sie ihm aus dem Gedächtnis geschwunden war” (Wendeland 1939: 76-77).

Children pick up South Sea English very quickly ; and I have known boys who came on board my vessel converse fluently, having acquired the language from returned labourers and by visiting trading and labour vessels. (Wawn 1893: 41)

The plantation varieties of Queensland, Samoa, Fiji and New Caledonia may thus have provided an important foundation for the formation of the MPE varieties. The amount of Solomon Islanders, Vanuatians and New Guineans that went to work on the particular overseas plantations is thus important to determine the impact of the overseas plantation pidgins on the development of the individual MPE varieties.

3.1.3.5 Labour recruitment statistics

Several researchers have collected information regarding the origin of recruits that have served on the plantations in Queensland, Samoa, Fiji and New Caledonia. There is a compilation available by Moses for Samoa (1973) and another one for Queensland by Price & Baker (1976). A compilation for Fiji and New Caledonia is available in Tryon & Charpentier (2004).

Queensland plantations recruited labourers from the Loyalties, Vanuatu, the Solomons, Papua New Guinea, Kiribati and other smaller islands, whereby the majority of workers was provided by Vanuatu in the period 1863-1906. Tryon & Charpentier (2004: 178) report that 65% of workers came from Vanuatu, 30% from the Solomons and fewer than 5% from the Bismarck Archipelago in New Guinea.

A similar picture is given for the Fiji plantations. For the period 1876-1911 it is reported that around 52% of workers were recruited in Vanuatu, 30% in the Solomon Islands and 6% in German New Guinea. In the beginning years of the Fiji plantations, the labour-recruiting in Vanuatu played a major role. However, from 1887 onwards the number of labourers recruited per year in the Solomon Islands surpassed the number of workers recruited per year in Vanuatu (cf. Tryon & Charpentier 2004: 182-187). It needs to be mentioned that the number of workers from other Fiji Islands is not included even though Fijian Islanders provided a great number of labourers as well.

Workers for New Caledonia plantations were recruited predominantly in Vanuatu and only about 1,000 labourers were recruited in the Solomon Islands and around 100 in Kiribati. The recruitment started in the year 1865 and lasted until 1929 (cf. Tryon & Charpentier 2004: 199-205).

The recruitment procedures of Samoa differ the most from the other areas. In the years 1867-1877 almost all recruits came from Kiribati (cf. Tryon & Charpentier 2004: 196) and Samoa did not start to recruit labour on the New Hebrides and Solomon Islands before 1878 (cf. Moses 1973: 102). By then, the areas had already served for 15 years as the major recruiting grounds for

the Queensland plantations. Thus, there is a chance that Vanuatians and/or Solomon Islanders, who were recruited starting in 1878 to work on Samoan plantations, had worked on plantations in Queensland and been in contact with the Queensland plantation variety before. Until 1885, Kiribati provided around 46% of workers on Samoan plantations, Vanuatu provided 25%, followed by New Guinea with 14%, the Solomon Islands with 13% and the Carolinas with roughly 2%.

The year 1885 represents the turning point in the labour recruiting history of Samoa. When north-eastern New Guinea and the Bismarck Archipelago were annexed by the Germans in 1884, the German commissioner Gustav von Oertzen prohibited the exportation of labour out of the Archipelago, with the exception that German plantations could still recruit their labourers in German New Guinea (cf. Schnee 1904: 60). Thus, from 1886 onwards, the labour recruitment to Queensland and Fiji decreased and finally stopped, while the exportation of labourers to the Samoan plantations and with it, mutual linguistic influence, could continue.

Table 2 gives a summary of these statistics, showing the number of Melanesian labourers from German New Guinea, Vanuatu and the Solomon Islands and the plantations to which they went. A note needs to be made regarding the number of Vanuatian labourers on plantations in Queensland, Fiji and New Caledonia. The figures suggest that the Vanuatu labourers represented the most dominant group of the founding population on the plantations during the period 1863-1885 in Queensland and Fiji and during the period of 1863-1929 in New Caledonia. It can thus be assumed that Vanuatu substrate languages may have played a major role during the formation of Pidgin English varieties on the plantations, which then spread to other areas.

In addition, Figure 6 visualises that the year 1885 represented a major turning point in the recruiting history. As the majority of workers recruited in German New Guinea had worked on Samoan plantations and the majority of workers recruited in the Solomon Islands and Vanuatu had been to Queensland plantations, Figure 6 and Table 2 beg the question whether QPPE may have had a major impact on the development of Bislama and Solomon Islands Pijin and only influenced the development of Tok Pisin in the early years.³¹ From 1885, SPPE seems to have turned into the major influencing variety for the development of Tok Pisin.

³¹ The influence might have either directly come from New Guineans working on Queensland plantations or indirectly due to the recruitment of Vanuatian and Solomon Islanders that had previously worked on plantations in Queensland and brought QPPE to Samoan plantations.

plantation labourer's origin	Queensland				Samoa				Fiji				New Caledonia	
	1863-1885		1886-1906		1863-1885		1886-1914		1863-1885		1886-1911		1863-1929	
		%		%		%		%		%		%		%
German New Guinea	2,809	7.76	0	0	693	14.27	5,307	100	1,618	7.09	0	0	0	0
Vanuatu	27,028	74.67	12,947	49.66	1,201	24.73	0	0	13,471	59.04	727	15.10	ca. 14,000	92.72
Solomon Islands	5,118	14.14	13,099	50.24	618	12.72	0	0	5,030	22.05	3,198	75.93	ca. 1,000	6.62
Others	1,241	3.42	27	0.10	2,345	48.28	0	0	2,696	11.82	287	6.81	ca. 100	0.66
Total	36,196		26,073		4,857		5,307		22,815		4,212		15,100	

Table 2: Number of Melanesian labourers on plantations in Queensland, Samoa, Fiji and New Caledonia³²

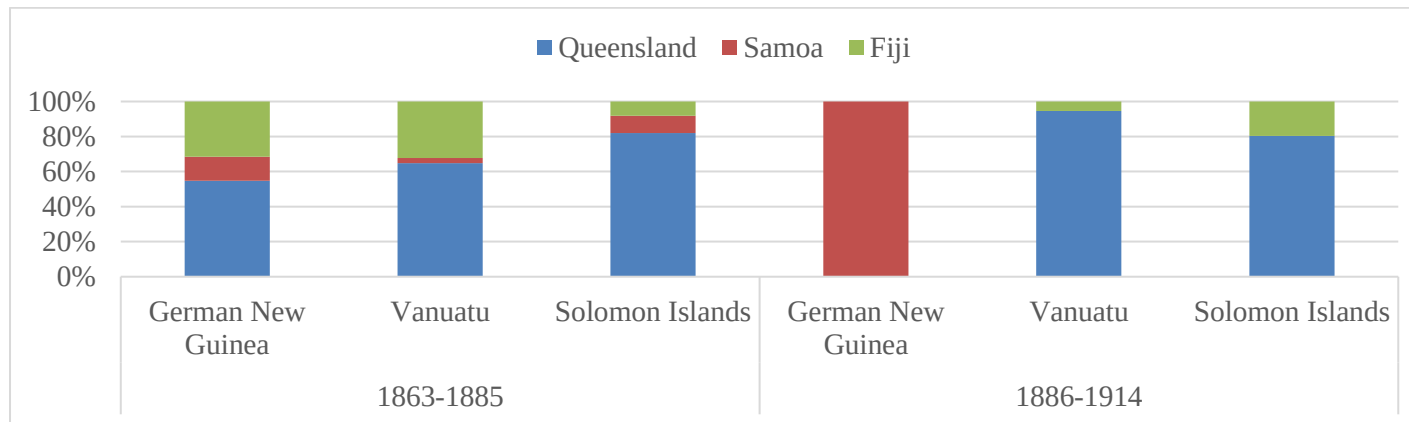


Figure 6: Distribution of German New Guineans, Vanuatuans, Solomon Islanders per plantation area

³² The differing end dates for the second period are based on the differing years in which the labour trade came to an end in the individual areas and the availability of data, respectively. The percentages are rough approximations: the exact number of labourers that worked on the different plantations as well as their places of origin cannot be determined, and it is sometimes not clear when exactly workers from a specific area went to the plantations. For Fiji, for instance, the number of labourers from the Fiji Islands was not available and could therefore not be included into the total population. For QLD it is known that 191 labourers were recruited from Kiribati, but it is not known when they arrived on the QLD plantations. Therefore, they could not be included into the statistics.

3.1.3.6 Meta-linguistic evidence

The figures indicate a possible influence of the plantation varieties on MPE. Meta-linguistic data found in the early colonial sources provides further information on the impact of the four plantation areas on the development of Solomon Islands Pijin, Bislama and Tok Pisin and will be presented in the following.

3.1.3.6.1 Impact of overseas plantations on the development of Solomon Islands Pijin

There exist multiple reports about Solomon Islanders who had served on Queensland plantations. Several writers give accounts of meeting islanders “who had spent several years in Queensland” (Woodford 1890: 26) and reports of vessels that “land[ed] Boys from Queensland” (Young 1925: 187) are prevalent. Sometimes, the Solomon Islanders report in pidgin to have worked “alonga plantation longa Queenslan’ one time” (Abbott 1908: 59), indicating that the Pidgin English was learned on the Queensland plantations. Several of those statements by the Solomon Islanders are present in Dickinson’s report: “Me been work along Queensland for make um sugar” (1927: 64) and “Me been work sugar-cane along Queensland” (1927: 108), just to mention two. Philip referring to being on board of the Makira in 1912 states that “[t]he returned Queensland ‘boy’ is much in evidence here” and quotes one of them saying “*me* been alonga Queensland” (Philip 1978: 87-88). The strongest evidence that Solomon Islanders brought QPPE with them when they returned is provided by Cormack who states that “a class of boys of a school leave their village and go off in a body to Queensland, to return someday in black clothes, hats, and boots, with pidgin English” (1944: 142). The quote makes clear that the ability to speak Pidgin English was acquired, at least by some, on the Queensland plantations. A similar report can be found in Hogbin (1939: 167) who claims that “[r]eturned plantation labourers in particular are often swollen-headed, loud-mouthed, and bumptious, and parade any information they may have in pidgin English, to the intense annoyance of those who have remained at home”.

Although a similarity between the contact jargons and the reported plantation pidgin can be recognised, the earlier contact variety differed from QPPE. Cromar (1935: 137) referring to Malaita argues that “[s]ome of the men could speak a little *bêche-de-mer*, but one was very fluent, and said that he had been to Queensland”. This shows that the contact variety was used more permanently in Queensland and thus developed and was perceived as more fluent.

The number of sources which contain attestations of Solomon Islanders to have worked on plantations in Samoa are less common. In one of his journals, Woodford (1884: 101-102; ANUA 481-267) reports about a man who:

went to Fiji where he worked six months for a white man in Sevuka, after which he returned home. He was then recruited for Samoa where he worked four years for a German planter and was then re-engaged by his employer as overseer and remained with him a further period of six years.

The quote demonstrates that the Solomon Islander had been to Samoa and shows once again that it was a common practice for Pacific Islanders to be reemployed and work on plantations in various areas. The only other account of a Solomon Islander who had worked on Samoan plantations was found in Woodford's diary for the year 1886 (cf. Woodford 1886; AU ANUA 481-10).

Solomon Islanders were also recruited for Fiji plantations. Evidence can be found, for instance, in Norden (1926: 96) who was in contact with a Solomon Islander who "was recruited for Queensland, and later for Fiji". However, most of the early sources claim that Fijian was the language used on Fijian plantations: "[T]he Fijian tongue, [is] a language understood by the men who had served their term on the Fiji plantations" (Guppy 1887: 53). Only two sources could be identified in which an English-based pidgin was learned on Fiji. Rannie reports coming into contact with "only one English-speaking native at Vanikoro, and he had picked up a very indifferent smattering of the language during a stay in Fiji" (Rannie 1912: 172) and the bishop of Melanesia reports that he "met a returned labourer from Fiji, who was most voluble in very bad English" (Melanesian Mission 05.08.1877: 23; AU PMB DOC 216). Regarding the former quote, it is not sure where exactly the person had picked up the language, i.e. whether the knowledge was really obtained on a plantation.

So far, no attestations reporting Solomon Islanders' stay in New Caledonia were found, which reflects the fact that only a comparatively small number of labourers were recruited for the area.

3.1.3.6.2 Impact of overseas plantations on the development of Bislama

The number of reports about Vanuatians having served on plantations in Queensland is high. The European writers tell about having "boys" or coming into contact with men who have "been to Queensland" (Thomas 1886: 314). Vanuatians report about village members serving on Australian plantations: "Oh, he stop along Queensland, by-and-by he come back" (Colonial Office 1883-84: 223). Moreover, in Lamb (1905: 159) an exchange between a European and a New Hebridean is found, in which the French asks "You been along Queensland?" and the Vanuatian responds with "Yes".

There are many reports about returned Queensland labourers who are able to speak a pidgin variety, as, for instance, "Me speakee English, my name belong Black John, me been Porter Mackai" (Coote 1883: 80). In Cromar's *Jock of the Islands* it is reported about returnees

which “came by steamer from Northern Queensland” to “be landed at various places in the New Hebrides” (1935: 117). When Cromar asks them about what they have done with their money, the response by the returnees is in Pidgin English: “Me fellow keep him [...] By and by me fellow buy’em gun along man-we-we and German man” (1935: 117-118).

That the Pidgin English knowledge was gained during their stay on Queensland plantations can be learned from Thomas (1886: 291) who recalls meeting a Tannese who had served in Maryborough and spoke with him in Pidgin English. Thomas claims that “without the knowledge of English this man had acquired, owing to the labour trade” a conversation would not have been successful. Moreover, Coote (1883: 80) draws a direct connection between the Pidgin English knowledge and the Queensland plantations claiming that:

There have evidently been many labour vessels here [Lo Island] from time to time, for we found that several men could speak a little “sandal-wood English,” as it is called ; none of them, however, appeared at all pleased with their experience of civilization. The place they had been to was Port Mackay in Queensland, the centre of the sugar district.

The early accounts also reflect that around 14,000 Vanuatuans had served on Fiji plantations, which make 20.18% of the total number of 69,374 Vanuatuans that went to plantations in Queensland, Samoa, Fiji, and New Caledonia. As argued by Siegel (1987), the predominant language used on Fiji plantations was Fijian along with Pidgin Fijan which is also echoed in historical data found in Wawn reporting about “a returned [Vanuatuan] labourer from Fiji, who could speak a mixture of English and Fijian” (1993: 75) and in Schuchardt it is reported that workers in Fiji would have picked up “a barbarian Fijian, but not mixed with English” (1889: 162).³³ Though this supports the idea that a Pidgin English was not used on the plantations, evidence can be found that it served as the primary tool in recruiting workers. In cases in which it is reported that Fijian recruiting vessels came to Vanuatu the islanders are said to have responded in Beach-la-Mar with:³⁴ “Fiji no good man ; too muchey work, Fiji” (Romilly 1886: 179), “Too muchy work Fiji ; no good” (Forbes 1875: 251). Moreover, an English-based contact variety must have existed in Fiji for at least a certain amount of time. There are reports about Tanna men being in Fiji stating “No more Tanna men come Fiji - we no like him eat we” (Great Britain 1869: 1024). On Levuka, Wawn observes a conversation between a European storekeeper and plantation labourers. Only one of the plantation labourers is said to be able to speak English (cf. 1893: 122). While the conversation is first carried on in an English-based contact variety, the

³³ “Die fremden insulaner, die nach Fidschi kämen, lernten Fidschiisch, nicht Englisch; sie nähmen ein barbarisches, aber nicht mit Englisch vermischtes Fidschiisch nach ihren inseln zurück [...]” (Schuchardt 1889: 162).

³⁴ Since the English-lexified contact variety was used in the trade of beach-la-mar, the pidgin language was named after the object of the trade.

European switches into Fijian when he became aware of Wawn observing them (1893: 123). The fact that only one of the plantation workers was able to speak pidgin (and only because he had worked previously as a house servant) shows that pidgin was not used as the language for intertribal communication on the plantations, but it might have been used in other domains.

As Vanuatu provided most workers for the plantations in New Caledonia, it is no surprise that attestations of Vanuatuans having served in New Caledonia can be found. In terms of the language used and acquired, one European reports about meeting a Vila inhabitant who had served a term in New Caledonia. “He presented, consequently, a burlesque imitation of his former employers. In particular, he had learned how to jabber and gesticulate as well as any Frenchman” (Wawn 1893: 143). What follows is a conversation in Pidgin English. Thomas (1886: 257) informs about a Tannese, who had worked in Queensland and New Caledonia stating “No good man-a-wee-wee!” Evidence is also found in Le Chartier (1885) that the contact situation in New Caledonia “allowed any indigene to learn French and English sufficiently in order to provide travellers with basic information” (Le Chartier 1885: 119).³⁵ This shows that the Vanuatuan Islanders did not only have contact with an English pidgin variety but also with a kind of Pidgin French as the following sentence shows: “Toi grand chef Ambrym, beaucoup popinées jolies, pas besoin femmes blanches” (Le Chartier 1885: 257).

There is only a single source identified so far, in which it is referred to a Vanuatuan who had worked on a plantation in Samoa and who afterwards was reemployed to work on a plantation in Efate (Colonial Office 1883-84: 218). From the meta-linguistic sources, it is thus reasonable to assume that SPPE had little to no influence on the development of Bislama.

3.1.3.6.3 Impact of overseas plantations on the development of Tok Pisin

In the early colonial sources for New Guinea evidence can be found that New Guinea islanders had served on plantations in Fiji, Samoa and Queensland. Schellong reports that when recruiting vessels from mainland New Guinea arrived at the Bismarck Archipelago, the inhabitants from the islands did not know how to classify New Guinea but that they knew about “Fiji, Samoa or Makay (a sugar plantation in Queensland) [...] of these places they had heard or they had been there before” (1934: 171).³⁶ This shows that returning workers brought back a pidgin variety. For

³⁵ “[...] permis à quelques-uns des indigènes d’apprendre le français et l’anglais suffisamment pour fournir quelques renseignements élémentaires au voyageur désireux de se les procurer” (Le Chartier 1885: 119).

³⁶ “[M]it “Neuguinea” wissen sie nichts Rechtes anzufangen; hätten wir ihnen Fidschi, Samoa oder Makay (eine Zuckerpflanzung in Queensland) genannt, dann hätten sie sogleich gewußt, woran sie waren; von diesen Orten hatten sie gehört oder waren selbst einmal dort gewesen” (Schellong 1934: 171).

instance, Behrmann states that “some people had already worked on the plantations of the whites, one could communicate with them in pidgin-English” (1922: 309).³⁷

In contrast to the early data of the Solomon Islands and New Hebrides, several reports about ex-Samoan workers were found for New Guinea. Jung, referring to New Britain plantations, reports that labourers who had previously worked on Samoan plantations were reemployed (cf. 1885: 285). Stephan & Graebner claim to have worked with a Pidgin speaking interpreter during their stay in German New Guinea who had come to Samoa as a young boy and had served for a long time as a sailor and plantation worker (cf. 1907: 21). Krämer-Bannow reports about a man named Anis von Tano, who would have worked in Samoa together with other people and, consequently, the workers would be able to express themselves in Pidgin (cf. 1916: 20). Additionally, a report by Krämer (1913: 406) in the *Deutsche Kolonialzeitung* informs about Pidgin English-speaking New Irelanders who would have acquired their knowledge of the language during their stay in Samoa and in the Bismarck Archipelago.³⁸

Reports about New Guineans that had worked on Queensland plantations exist as well. Schellong informs about a voyage of a New Guinea Company ship, aiming to recruit workers for mainland Kaiser-Wilhelmsland, stating “in another area of the Archipelago the captain achieved to recruit six blacks [...]; they had worked once in the sugar plantations of Queensland and spoke a good Pidgin English” (1934: 90).³⁹ This quotation illustrates both that the pidgin language learned in Australia was brought back and spread by the returned labourers, and that the returned labourers were reemployed on the plantations in German New Guinea. Stephan & Graebner mention not only a Samoan speaking interpreter but also another interpreter who “was a sailor in Queensland before the German occupancy and acquired a considerable amount of pidgin vocabulary” (1907: 22-23),⁴⁰ which shows that the pidgins formed or learned on the ships travelling to Queensland were diffused into the home areas.

In terms of the plantations in Fiji, evidence exists that New Guineans served on the plantations in Fiji (cf., for instance, Parkinson 1887: 35). Schellong reports about New Guineans in New Ireland who had worked on the plantations in Fiji and who are able to use a form of early

³⁷ “Einige Leute hatten bereits in den Plantagen der Weißen gearbeitet, man konnte sich mit ihnen auf pidgin-englisch verständigen” (Behrmann 1922: 309).

³⁸ “Trotz der Jungfräulichkeit des Landes hinsichtlich weißer Besucher waren mehrere etwas Pidgin-Englisch radebrechende Eingeborene vorhanden, die vordem als Arbeiter auf Planzungen im Archipel und auf Samoa ihre Kenntnisse erworben hatten” (Krämer 1913: 406).

³⁹ “[...] denn an einer anderen Stelle des Archipels gelang es dem Kapitän, sechs Schwarze auf redliche Weise anzuwerben; sie haben bereits einmal in den Zuckerplantagen Queenslands gearbeitet und sprechen ein gutes Pidgin-Englisch” (Schellong 1934: 90).

⁴⁰ “Tompuan aus Lamassa mochte ungefähr 40 Jahre zählen. Er war noch von der deutschen Begriffsergreifung als Matrose in Queensland gewesen und hatte sich einen bedeutenden Pidgeon- Wortschatz erworben” (Stephan & Graebner 1907: 22-23).

pidgin, stating that: “master he speak two fellow yam; me stop here; by and by me go back Fidji” (1934: 171). Whether they had acquired the knowledge of the like on the plantations in Fiji, remains open to question.

3.1.3.6.4 Summary

Based on meta-linguistic data, it has become evident that Queensland Plantation Pidgin English was brought to the Solomon Islands, Vanuatu and German New Guinea by repatriated labourers and is thus likely to have influenced the varieties spoken in those areas (see Figure 7).⁴¹ Moreover, Samoan Plantation Pidgin English was brought to New Guinea by returned labourers, but only little evidence is found for the variety being brought to the Solomon Islands and Vanuatu. For all three areas there are reports of islanders having worked in Fiji. While Pidgin English was used to recruit labour for the Fijian plantations, Fijian or Fijian Pidgin was used on the plantations instead. The English-based pidgin variety used on Fiji is thus not likely to have had a great influence on the development of Melanesian Pidgin English. Still, it might be that Pidgin Fijian (Pidgin (Fijian-lexified: Fiji) or Fijian has left lexical or structural traces. Similarly, the New Caledonian Pidgin French (Pidgin (French-lexified: New Caledonia) is likely to have had only a minor impact, if at all, on the development of Bislama.

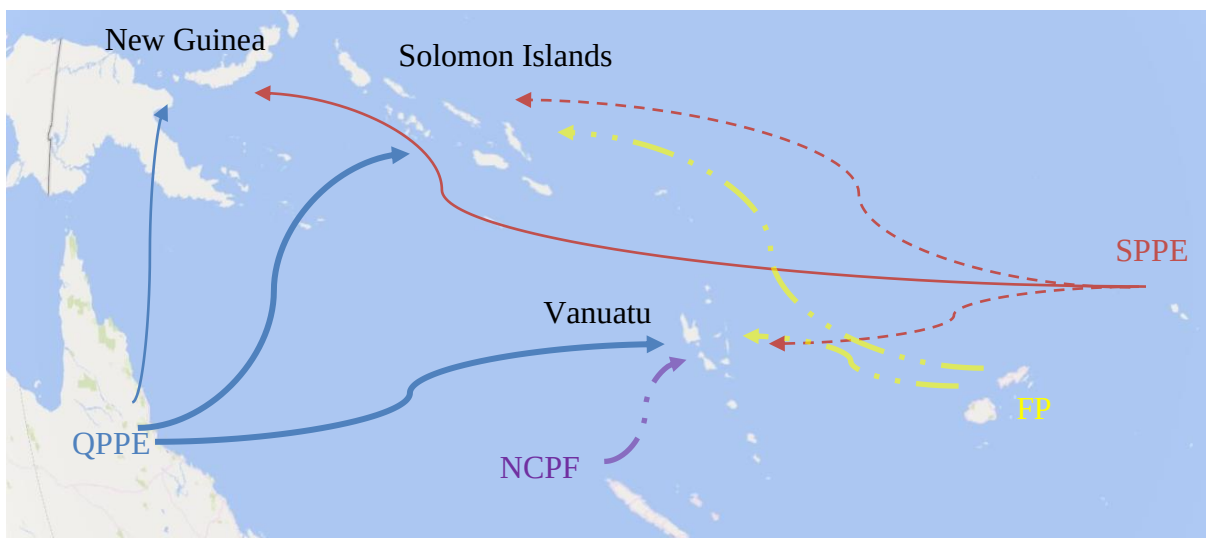


Figure 7: Visualisation of influence based on metalinguistic evidence

3.1.3.7 The end of the labour trade

In January 1901 the *Commonwealth of Australia* was established. By that time Australia’s “non-white” population had grown to a great extent, which was negatively evaluated by the “white”

⁴¹ Due to the numerical dominance of Vanuatuan labourers on QLD plantations (cf. Table 2), it is likely that Vanuatu substrate languages had a major impact on this early form of QPPE which diffused to the other areas.

Australian population. Racist attitudes prevailed and voices were being raised for an immigration regulation. Therefore, one of the first acts which the newly formed government passed was the *Immigration Restriction Act*, which was accompanied by the *Pacific Island Labourers Act*. The laws represent the establishment of the *White Australia Policy* and included inter alia the reduction of labour recruitment until 1904 and the deportation of all Pacific Island labourers by 1906. As Munro reports, the planters and islanders tried to resist the new law, but nonetheless “over 4,000 Kanakas were compulsorily repatriated” (Munro 1995: 609). Thus, the labour recruitment of the plantations in Queensland came to an end in 1906. Samoa stopped the labour recruitment in 1914, when German New Guinea became the Australian Trust Territory. Fiji differed from the other colonies in recruiting not only on the Pacific Islands but also in India, from where more than 60,000 Indian indentured labourers are said to have arrived (cf. Tryon & Charpentier 2004: 192). In 1916, however, Fiji also stopped its indentured labour recruitment.

Before 1906, labourers from German New Guinea, the Solomon Islands and Vanuatu were leaving to work on overseas plantations and could bring with them the overseas pidgin varieties when returning home. As soon as the labour trade stopped, the interaction between Solomon Islanders and Vanuatians on plantations and recruitment vessels came to an end, along with their linguistic interaction. Though German New Guinean contact with Solomon Islanders and Vanuatians had stopped earlier, the end of the recruitment for Samoan plantations in 1914 may have had an impact on the development of early Tok Pisin. New Guineans no longer interacted with Samoans on Samoan plantations and no longer came into contact with the shipboard varieties. It can thus be assumed that the end of the labour trade and, as a consequence thereof, the isolation of the areas under investigation favoured the separate development of the three varieties.

3.1.4 Developments in the home areas

While New Hebrideans, Solomon Islanders and New Guineans were recruited to work on overseas plantations, the development in the home areas was influenced by the European arrival as well. The events in the home areas and their impact on the development of MPE will be outlined in the following sections.

3.1.4.1 Developments in the Solomon Islands

Even though traders had settled in the Solomon Islands by 1868, no major plantations were established before 1896 (cf. Bennett 1993). From 1877-1893 the Solomon Islands were under a loose jurisdiction by the British High Commissioner in Fiji and no “central regulating authority”

was present (Bennett 1987: 63). Thus, contact between Solomon Islanders and Europeans existed but the number of permanent European residents was growing only very slowly as the trading was based on “[t]emporary residents” (1987: 59) and “traders acted as individuals rather than as an organized group” (1987: 63). In general, it can thus be said that while the Solomon Islands provided, together with Vanuatu, the majority of labourers for overseas plantations, they did not cultivate plantations on their home island group before the 1890s. As Jourdan & Selbach point out, as there was no “social *raison d’être*, it [= the Pidgin English brought back by returning labourers] remained largely unused, except for affect” (2008: 164).

When the island group became the *British Solomon Islands Protectorate* in 1893, the situation changed. The previously learned plantation pidgin varieties were used in intercommunication with British Colonial officers (cf. Jourdan 1995: 139; Jourdan & Selbach 2008: 164). Charles Morris Woodford became the first acting resident commissioner of the Solomon Islands in 1896 and Pidgin English represented an important tool in the legal administration of the state.

With the rapidly increasing establishment of further plantations in the Solomon Islands from 1905 onwards, Pidgin English gained in importance. Following the common Pacific practice of recruiting labourers, inhabitants of other Solomon Islands (especially Malaita) were recruited. Solomon Islanders who had previously worked on overseas plantations were among the first being recruited to work on the home plantations. The previously learned QPPE and SPPE varieties were reutilised on the plantations and “others learned it very quickly, *sure le tas*, at work or in daily casual contacts with Pijin speakers” (Jourdan 1990: 168). Moreover, intermarriages became more common (cf. Bennett 1987: 179). As Bennett (1987: 190) points out:

Through traders and the recruitment of some ‘old hands’ on the new plantations, Pidgin English spread, reinforced by English-speaking planters, while Pidgin Fijian atrophied. Gradually, as thousands of men came in and out of plantations, the language Pijin both stabilized and spread, [...] By the 1930s it had achieved a fairly standard form and was so widely known that recruiting vessels were no longer required by law to carry interpreters. ‘New chums’ had to learn the language quickly as both the boss-boy and the master used it. As a result, men from the bush of Fataleka (Malaita) could soon talk to men from the south coast of Guadalcanal. Villagers near plantations picked up the language as they visited plantations to sell produce.

Although both Pidgin Fijian and Pidgin English were brought home by returning labourers, it was Pidgin English that survived and spread with the establishment of plantations in the Solomon Islands. Even though the plantations contributed to the stabilisation of early Pijin, the variety was, as described by Jourdan (1995: 139), “everyone’s language, but no one’s language”. It served as a lingua franca for interethnic communication, but it was only used if the situation required its usage. Daily social activities were preferably performed in the vernacular language. Moreover, Pijin was predominantly used by males since females were not recruited to work on plantations

and were thus not exposed to the pidgin variety (cf. Bennett 1987: 118). Women with a knowledge of Pijin can usually be shown to have been in contact with mission schools.

It needs to be pointed out that from 1913 onwards Chinese traders settled in the Solomon Islands (cf. Bennett 1987: 206) and may have influenced the development of the variety as well – even if to a lesser degree.

3.1.4.2 Developments in Vanuatu

The establishment of cotton plantations in the New Hebrides started already in 1867, when the British founded the first plantation on Tanna. Thus, the cultivation of plantations began simultaneously with the exportation of labourers to overseas plantations. Tryon & Charpentier state that “[r]ight from the beginning, the custom of employing ‘off-island’ labour was invoked, with obvious consequences for the need for and the development of a *lingua franca*” (2004: 215). By 1874, the British had already established 16 plantations on Aneityum, Tanna, Erromango and Efate (cf. 2004: 216) and, as it was common to recruit labourers from surrounding islands, the development of early Bislama was propelled.

French economic interest on Vanuatu began to rise in 1872, when the first French plantation was established on Efate. Ten years later, the *Compagnie Calédonienne des Nouvelles-Hébrides* was founded by John Higginson with the intention to purchase land in Vanuatu and to lobby for a French take-over (cf. Aldrich 1993: 196). Due to this practice, several French families moved to the island group (cf. Tryon & Charpentier 2004: 217). At the same time, Australian interest in the island groups grew and also French Catholic missionaries and Anglo-Australian Presbyterians began to compete in the Vanuatu area. To protect French and British citizens in the New Hebrides, a loose jurisdiction, namely the *Anglo-French Joint Naval Commission* was established by a convention in 1887. Four years later Vanuatu counted 51 British and 70 French settlers (cf. Tryon & Charpentier 2004: 218). In 1906 the *Anglo-French Condominium* was established, meaning that Vanuatu henceforth was administrated by both the French and British powers. Inter alia, it was decided that a Joint Court should be established. Tryon & Charpentier (2004: 301) report that by that time Vanuatu counted 401 French citizens and 228 British citizens which means a ratio of 2/3. It is important to note that despite the numerical dominance of French citizens, the influence of French on Bislama seems to have been short-lasting and/or limited. One possible reason might be that the varieties that were brought to Vanuatu by the first returning labourers have shaped the resulting contact variety to a greater degree than later contact with French. This may explain why Bislama is English-lexified despite the fact that it was also French governed.

French as well as the British were actively promoting the agricultural land use which involved a need for further labourers. Between 1912 and 1939 alone, around 32,000 Vanuatuan labourers were engaged to work on the plantations in Vanuatu, three quarters worked on French plantations and one quarter on British plantations (cf. 2004: 303). It is frequently argued that due to the great amount of Vanuatuan internal labour recruitment, the contact variety “could well have developed fully even if the recruitment of labour for work on the overseas plantations in Queensland, Fiji, and Samoa had never taken place” (Tryon & Charpentier 2004: 303).

The recruitment of Vanuatuan labourers saw a decline from 1920 onwards, as New Hebrideans’ interest arose in producing and selling copra on their own. Thus, labourers were recruited in Vietnam so that around 21,915 Vietnamese came to Vanuatu starting in 1921 until 1940 (cf. Tryon & Charpentier 2004: 306).

3.1.4.3 Developments in Papua New Guinea

3.1.4.3.1 German New Guinea

The first European companies on the Bismarck Archipelago were established in the second half of the 19th century. For instance, the German commercial enterprise of J. C. Godeffroy & Sohn, a company which has its origin in Hamburg, was established on Mioko in 1876 (cf. Tryon & Charpentier 2004: 241) and a year later, the company Robertson and Hensheim established on the Duke of York Islands (cf. Mosel 1979: 164). From 1878 onwards, other trading centres were set up on the neighbouring islands. The establishment of European companies in the Bismarck Archipelago intensified the contact between Europeans and New Guineans and led to an increased knowledge of the early contact variety. Schuchardt (1883: 6/154; translation: Gilbert 1980: 18), referring to Hensheim reports that in “New Britain, where upon his arrival seven years ago [1876] no native had been able to understand a European language, nowadays nearly everybody, above all the children, speak this variety of English”. Thus, when the north eastern part of the mainland of New Guinea (Kaiser-Wilhelmsland) as well as the offshore islands of the Bismarck Archipelago (New Britain, New Ireland, and other surrounding islands) were officially annexed by the Germans in 1884, an early form of Pidgin English was already prevalent (cf., for instance, Schafroth 1916: 19).⁴²

⁴² The negotiations regarding the status of the Solomon Islands continued until 1899. While in 1886 the German and British colonial powers divided the Solomon Islands so that Bougainville, Buka, the Shortland Islands, Choiseul, Santa Isabel, Ontong Java and parts of the Florida Islands came under German protection, the border was moved again in a further declaration in 1899-1900. Only Bougainville and Buka remained under the German protectorate, whereas the remaining Solomon Island areas became part of the British Solomon Islands Protectorate (cf. Griffin 2005: 74-75).

Under German administration, the country developed into a “plantation-based economy” (Tryon & Charpentier 2004: 325) and plantations spread from the Bismarck Archipelago to mainland Kaiser Wilhelmsland. In order to handle the need for workers, the plantations recruited their labourers from other German New Guinea areas. Frequently, the recruited labourers had previously served on the plantations in Queensland or Samoa and thus had brought a knowledge of the there-spoken plantation varieties with them. Furthermore, Chinese, Javanese, people from Makassar, other Malays and Indians were recruited (cf. Hagen 1899: 43; von Beck 1903: 556; Frommund 1926: 63). Therefore, the varieties may have influenced the development of Tok Pisin to a certain extent as well.

While in the Solomon Islands the use of the early pidgin variety was rather restricted to the plantation grounds, the pidgin variety quickly spread and developed in German New Guinea. New Guineans worked not only on plantations but also as house boys and in “carpentry, medical assistance, police work and village administration” (Goulden 1990: 18). Von Hesse-Wartegg (1902: 99-100) asserts that Pidgin English was no longer only used on the plantations but in Herbertshöhe – the capital of East New Britain – in general. The pidgin which formerly was “the general language between Europeans and coloured men” (Jacques 1922: 96) turned into the communicative tool for New Guineans as well: “Also among themselves the natives made use of it if they came from different areas” (Jacques 1922: 96).⁴³ Schnee (1904: 305) further mentions that the contact variety served as the main medium of communication in interethnic marriages.⁴⁴

The development of the variety in German New Guinea was further influenced by being withdrawn from its major lexifier under German administration. Pidgin English was considered to be a language as any other language of the world (cf. Friederici 1911: 95).⁴⁵ Therefore, no attempts were made to assimilate the pidgin to English (cf. Mühlhäusler 1979: 82-83). Due to the already existing English-based contact variety, there was no need for a further lingua franca to enable communication. Pidgin English already represented the “most widespread lingua franca” (von Hesse-Wartegg 1902: 52)⁴⁶ in German New Guinea. Even though efforts were made by the

⁴³ “Dies Pidgin war die allgemeine Sprache zwischen Europäern und Farbigen. Auch unter sich benutzten es die Eingebornen, wenn sie aus verschiedensprachigen Gegenden waren” (Jacques 1922: 96).

⁴⁴ “Auch für den Verkehr der aus den verschiedenen Sprachgebieten stammenden Eingeborenen scheint es sich durchweg als genügend zu erweisen, selbst unter Ehepaaren [...] Doch ist mir nicht ein einziger Fall bekannt geworden, in dem der Mann die Sprache der Frau oder die Frau die des Mannes erlernt hätte. Zur Verständigung zwischen beiden diente lediglich das Pidginenglisch” (Schnee 1904: 305).

⁴⁵ “[A]uch das Pidgin-Englisch ist eine lebende Sprache, die sich entwickelt, die ihre Dialekte hat; und Niemand wird sich brauchbar in ihr verständigen können, der sich etwa einbilden wollte, sie lediglich von einem anderen Europäer lernen zu können” (Friederici 1911: 95).

⁴⁶ “[...] das Pidgen-Englisch war bereits die verbreitetste Verkehrssprache, als die Deutschen hierherkamen, sie ist es auch auf den anderen Inseln der Südsee, und man konnte sie begreiflicherweise nicht einfach wegdekretieren und durch die deutsche ersetzen” (von Hesse-Wartegg 1902: 52).

colonial government to implement German in order to meet the administrative and educational needs of the colony, nothing was *actively* done to promote the establishment of German. In fact, the use of Pidgin English was further promoted. Friederici reports that the tribal chiefs were supported by a *tultul*, an interpreter, who, due to German policies, needed to speak Pidgin English fluently (cf. 1911: 93). Employees in other administrative positions, such as a *luluai* or *kukurai*, needed to have a knowledge of Pidgin English as well (cf. 1911: 93).⁴⁷

In 1914, German New Guinea was occupied by an Australian military administration for seven years, before it became the *Australian Trust Territory of New Guinea* in 1920. Under Australian administration the planting and trading industry was further promoted so that the number of New Guineans leaving their home areas was rising. Tryon & Charpentier (2004: 327), referring to the *Report of League of Nations* of the years 1920-1921, claim that New Guineans left their home areas with the effect that “[t]his labour mobility had obvious consequences favouring the spread of New Guinea Pidgin English”.⁴⁸ The language was further made use of in the gold industry which began in the 1930s.

3.1.4.3.2 British New Guinea

The south-eastern parts of Papua New Guinea have a different history, as they were occupied by the British in 1884 and, four years later, turned into a British colony.

When William McGregor became the first administrator of the area, he needed to form a “suitable police force with which to help extend government influence over an increasing area and to enforce law and order” (Dutton 1996b: 226). Therefore, he established the *Armed Native Constabulary*, which initially consisted of “two Fijians and twelve Solomon Islanders” (1996b: 226) who were responsible for the recruitment of further Papuans. In the police force, Papuans with diverse linguistic backgrounds came together so that the situation was in many ways comparable to the multilingual situation on the plantations. Though an early form of English contact variety existed as well, it was a form of Motu which represented the contact variety prevalent in the Port Moresby area. A foreigner-talk version of Motu had already developed when the first Europeans were arriving in the villages around Port Moresby in pre-McGregor times (cf. Dutton 1996b: 225). Being reemployed in the police forces Pidgin Motu developed and stabilised into *Police Motu* which turned into “the unofficial language of administration” (1996b: 227).

⁴⁷ The comparatively small influence of German on Tok Pisin may also be based in that English-based contact jargons were already in use when the Germans arrived. As German did not form part of the initial linguistic founder ecology, this may support the claim made by Chaudenson (1992, 2001) and Mufwene (1996, 2001) that the founding population had a greater impact on the emerging contact variety.

⁴⁸ Further figures about the areas to which New Guineans went can be found in Tryon & Charpentier (2004: 327-330).

Though interaction between expatriates and Papuans were in place in the trading industry “Papua never served as a major reservoir of labour for plantations elsewhere in the Pacific, and did also not “attract[...] labourers from outside” (cf. Mühlhäusler & Baker 1996: 501). Thus, the English-lexified pidgin, usually referred to as *Papuan Pidgin English*, was present but never stabilised, “remain[ing] a local solution to local problems” (1996: 501).

British New Guinea was transferred to Australia in 1902 and was renamed into *Territory of Papua*. Even though some plantations had been established in the Territory of Papua by then, it was not before 1906-1914 that its “commercial development” was promoted (Tryon & Charpentier 2004: 332).

3.1.4.4 Summary

As the present section has shown, the Melanesian Islands not only provided workers for overseas plantations but saw the establishment of plantations in the home areas. The labourers were recruited from different areas within the own island groups, which resulted in multilingual workforces. Repatriated labourers that had worked on overseas plantations and had a knowledge of a plantation pidgin were frequently reemployed on the home plantations.

However, differences between the island groups can be observed. Vanuatu plantations were established almost simultaneously with the establishment of overseas plantations, and in German New Guinea the plantation economy started rather early as well. While in these regions Pidgin English was used on the plantations and also spread to other domains, the early form of Solomon Islands Pijin only gained in importance and was more commonly used by the beginning of the 20th century when more plantations were established.

3.1.5 Missionary activities

The island groups were visited not only by trading and recruiting vessels but also by – what in the early Pidgin English was called – “ship belong Jesus” (cf., for example, Norden 1926: 123). Missionaries arriving on the islands from 1839 onwards were immediately confronted with the language diversity of the South Sea and had to come up with a solution to cope with it. The mission societies had the aim to preach the bible, for which a tool of communication was a prerequisite. Moreover, with the establishment of mission schools they had to decide upon the language of instruction. The language choice question was addressed in different ways by the different missions varying from using English, teaching German, using Pidgin English, to learning the local languages. Those missions who decided on the use of Pidgin English will have contributed to the varieties’ further development. The following sections will give overviews of

the mission societies active in the Solomon Islands (3.1.5.1), Vanuatu (3.1.5.2) and New Guinea (3.1.5.3) based on the socio-historical information found in the early sources.

3.1.5.1 Missionary activities in the Solomon Islands

The first missionary voyage to San Christoval was conducted by Epalle a bishop of the *French Mission of the Society of Mary* in 1845 (cf. Chaurain 1846: 399). As Epalle was murdered when the mission vessel arrived, the “mission in that quarter ha[d] [...] been abandoned” (*The London Quarterly Review* July 1854: 97). It was not before 1898 that the *Catholic Mission* returned to the Solomon Islands, which meant that the Solomon Islanders by then already had a knowledge of Samoan (Austronesian (Oceanic, Polynesian): Samoa), Fijian or Pidgin English (cf. Tryon 1996: 619), which is why in the early years the Catholic Mission used Fijian and Pidgin Fijian as the medium of communication in their missionary activities.⁴⁹

In 1849 the first voyage of the *Melanesian Mission* by bishop Selwyn to the Solomon Islands took place. The mission was from then onwards active not only in the Solomon Islands, but also in the New Hebrides. While the initial language used by the mission was English and an attempt of learning the local languages was made, with the establishment of the main mission school in Mota, bishop Patteson decided that the “Christian mother tongue of the mission” (Wilson 1911: 16) should be Mota (Austronesian (Southern Oceanic): Banks Islands).⁵⁰ Pidgin English was not used by the mission and negative attitudes towards the variety were overtly expressed. Thus, in *The New Review* 8 the writer reports “[t]he first man I met addressed me in ‘pidgin English’ (language abhorred by the Mission and sternly tabooed)” (Montgomery 1893: 551).

From the beginning of the 20th century onwards, the *Methodist Mission* also established mission stations in the Solomons. The mission decided for Roviana (Austronesian (Oceanic): New Georgia) as their lingua franca and missionaries had to have an additional knowledge of English “if they are to take a part in the larger life of the Islands” (Metcalf 1947: 68; PMB 80). However, the official decision for a mission language other than Pidgin English did not mean that the missionaries were not making use of the variety. The missionary John Metcalf states in a letter (24.05.1937; PMB 68):

⁴⁹ A similar observation can be made in other areas as well, such as, for example, in Hawaii (cf. Roberts 2005). Missionaries were pragmatic and wanted to spread *The Word* as fast and well as possible. It was only later that language attitudes and moralising came into play.

⁵⁰ Cf., in addition, MacQuarrie (1948: 109).

English only a few really understand well enough to follow a sermon however simple the language used, Roviana I do not know well enough to speak, and the Choiseul talk is not known by the great majority, and certainly not by the Bougainvillians. I therefore decided to desecrate the sacred precincts of the Methodist Cathedral by using ‘Pidgin’, to which Mr. Leadley agreed.

This shows that Pidgin English was applied in situations in which a knowledge of local languages or English was not shared by the people involved.

The *Seventh Day Adventist Mission* became active in the Solomon Islands in 1914 when the missionaries Mr. and Ms. Jones opened a mission station in New Georgia (Cormack 1944: 147). The available early documents did not contain a great amount of information about the language(s) used by the mission. Tryon (1996: 621) states that the mission used Marovo (Austronesian (Oceanic): New Georgia) as a medium of instruction but taught English in their establishments.

The *South Sea Evangelical Mission* has contributed most to the development and spread of Pidgin English (PE) in the Solomon Islands. The mission was founded by Florence Young as the *Queensland Kanaka Mission* (QKM) in Bundaberg in 1886. Young started to do missionary work already in 1882 among Vanuatuans and Solomon Islanders who worked as indentured labourers on the sugar plantation of her brother (cf. Burt 1994: 105). The first mission converts returned to their homes in the Solomon Islands and New Hebrides by the end of the century and were assigned to spread the word of God in their home countries. The best-known example is that of Peter Abu’ofa who established a missionary school on Malu’u (cf. Moore 2017: 232). With the abolishment of the labour trade in Australia and the subsequent announcement that all labourers should be returned to their places of origin, the QKM formed a Solomon Islands Branch in 1904 and in 1907 moved completely to the Solomon Islands and therefore changed its name to *South Sea Evangelical Mission* (SSEM). As was argued in a newspaper article in 1933, the SSEM did “not trouble to learn the numerous dialects of the islanders, but content themselves with ‘pidgin English,’ which is common to all tribes” (*The Brisbane Courier* 02.06.1933: 4). Although some missionaries, such as Norman Deck, also learned local languages and tried to translate Bible stories into the vernacular, PE was the major medium of instruction. Thus, the mission seems to have contributed a great deal to the establishment and spread of PE in the Solomon Islands.

However, according to Mühlhäusler & Mühlhäusler (2005: 4; 11), the mission had to revisit its language policies by the early 1920s due to pressure from the government to implement English as the medium of instruction. In fact, the archival SSEM documents inform that “missionaries [were] trying to teach the boys and girls to speak and understand purer English, so that they may enter into the meaning of [the] English Bible” (Anonymous 1945; PMB 442). As PE had served as a helpful tool of communication in the area characterised by language diversity,

no abrupt change followed. Instead, PE was used “as a basis for creating a language more similar to acrolectal English” (2005: 12). Mühlhäusler & Mühlhäusler (2005: 9) claim that the mission developed a “simplified form of English, a solution that seemed obvious to those who regarded Pidgin English as a means of promoting a gradual transition to more acrolectal forms of English”. This ‘simplified’ English would therefore have consisted of several varieties “ranging from anglicising Pidgin English to removing perceived difficult passages from Standard English texts” and each missionary would have “relied on their own intuitions” concerning how much PE the simple English contained (2005: 12).

To summarise, by the time the Solomon Islands became a British Protectorate, five mission societies actively spread the gospels along the Solomon Islands (cf. Department of Geological Survey 1931: 11). Even if not all the missions decided for PE as the main medium of instruction, the usefulness of the variety could not be denied and they relied on it in situations in which they could not successfully communicate in the local or in the mission language.

3.1.5.2 Missionary activities in Vanuatu

The *Melanesian Mission* was active not only in the Solomon Islands but also in Vanuatu. As mentioned above, the mission regarded Pidgin English as an inadequate tool for missionary work. Instead of learning all local varieties, they tried to teach in and to establish Mota as the lingua franca of the mission (cf. Hogbin 1939: 259).

Moreover, the *London Missionary Society*, which later became the *Presbyterian Church of the New Hebrides*, established mission stations in Vanuatu. The mission activities started in 1839 when the reverend John Williams arrived in Erromanga and the first permanent mission station was established on Aneityum by John Geddie in 1848. The *London Missionary Society* did not make use of Pidgin English but used Nakanamanga (Austronesian (Oceanic): Vanuatu) and Aulua (Austronesian (Oceanic): Malakula) as their mission languages and aimed at translating the word of God into the local varieties (cf. Inglis 1887: 346; Tryon 1996: 622) since these were considered to be the best “vehicle capable of conveying to the native mind a clear conception of the truths of the Bible” (Inglis 1887: 100). Thus, evidence can be found that Michelsen and MacDonald were actively involved in the translation of hymns and the bible into local languages (cf. Michelsen 1893: 34, 169, 179).

In 1868 the Presbyterian Church of the New Hebrides also became active on Tanna through the reverend William Watt and his wife Agnes. The report given by Agnes Watt about their work shows that the missionaries aimed to learn the local varieties as well (cf. Watt 1896: 43, 64, 71-72). Even if “[a] missionary has many months of hard study before he can declare to

the people on their own tongue the wonderful works of God” (1896: 118), it was considered the best way to communicate the scripture.⁵¹ However, especially in their initial encounters with Pacific Islanders and when not yet being proficient in the local varieties, the missionaries made use of Pidgin English. Thus, when visiting Tanna for the first time, Watt reports that due to the language diversity their “principal means of communication is in broken English; or as it is better known here as sandal-wood English (being the broken English used by the sandal-wood traders in past years), which is very unsatisfactory” (Watt 1896: 127). Reverend Paton’s journal entry from 30 August 1893 provides further evidence that even if PE was not the language of instruction, it was present in the mission schools and used in daily interactions (PMB MS 32).⁵² Moreover, for the year 1894 he reports about a service being conducted “in what is here termed Sandalwood English – a sort of peculiar broken English, which traders use with the Natives all over the Islands” (Paton 1894: 6-7). Interpreters were in general to be avoided (cf. Inglis 1887: 127) but were considered a better tool for communication than using Pidgin English (cf. Watt 1896: 127).

When Samuel MacFarlane reports about his first voyage to Tanna he claims that “[t]he conversation was carried on in broken English, many of the natives being able thus to express themselves from frequent intercourse with foreigners” (1873: 106). In another passage of his report (even if referring to Chepenehe, Loyalty Islands), he explains that he “felt greatly embarrassed at not being able to speak to the people in their own language, although the difficulty was considerably lessened by a number of the young men being able to speak broken English” (1873: 73). This shows that even if the mission’s language of instruction was not PE, it was used when they had no other shared language.

A further mission active in Vanuatu was the *Roman Catholic Mission*. As Tryon states, the mission used local languages but none as a lingua franca (1996: 622). The early documents show that when the father J. B. Jamond, who previously had been active in the Roman Catholic Mission at Olal, went to Craig Cove and attended a message, he “improvised a small catechism in bichelamar for them” (Jamond 1906-1912, AU PMB MS 53-11).⁵³ This demonstrates that he used the contact pidgin to deliver the script. From his report it can be learned that it was also

⁵¹ “Would people only [...] consider whether they would thus eagerly learn a Foreign language from longing to be instructed about a new God! Or, even, how would we like to go back only to pre-Reformation days, and have our Bible practically a sealed book, by seeing it exclusively in the Latin tongue! If a Chinaman wanted to come and instruct us in the doctrines of Confucius, he would certainly begin by learning our language, and adapting himself to our modes of thought, not by trying to teach us Chinese and throwing open to us the treasures of Chinese Literature” (Paton 1894: 115).

⁵² Paton reports about a situation which had happened in the mission school. When quoting the boy Tommy, Paton uses Pidgin English.

⁵³ “je leure improvisai un petit catéchisme en ‘bichelamar’” (Jamond 1906-1912, AU PMB MS 53-11).

common to teach in the local languages. Moreover, the Marist missionary Pionnier published useful sentences for a baptism in Pidgin English in 1913, showing that Pidgin English was used by the Catholic Mission (Pionnier 1913: 9-17, 184-198).

3.1.5.3 Missionary activities in New Guinea

In New Guinea, as in the other island groups under investigation, the different missions came up with individual solutions in terms of the language question. Most of the missions aimed to learn the “native dialects” (cf., for instance, Collinson 06.10.1923: 21).

The *Methodist Mission* arrived in New Guinea as early as 1875. As an early form of Tok Pisin was not widely spread by that time (cf. Section 3.1.2), the mission made use of the Oceanic languages Ramoaaina (Austronesian (Oceanic)) in the Duke of Yorks and Tolai (Austronesian (Oceanic)) in New Britain in its initial years and decided to use Tolai in all mission schools in 1896 (cf. Ross 1996: 598). Tolai was the language that was spoken on the Gazelle peninsula of *Neupommern*, which is today’s New Britain. Due to the work of the missionaries, dialects of Tolai were also in use on Neulauenburg (today’s Duke of York Islands) and Südneumecklenburg (today’s South New Ireland) (cf. Schnee 1904: 312-316). Though the use of PE was officially denied, there is evidence of services being conducted in PE. For instance, Brown reports about having “conducted the first part of the service in Fijian, and then addressed them in pigeon English”, although the sermon was then further translated for the listeners (Brown 1908: 135).

The *Catholic Mission* established its first mission station on the Bismarck Archipelago in 1889 using Tolai as the language of instruction. However, Ross (1996: 599) reports that the Catholic missionaries did not use Tolai a lot but learned and produced scriptures in the local languages. The Catholic Mission on mainland New Guinea, which were “staffed by German Missionaries of the Divine Word order” (1996: 602), used German as the medium of instruction in their initial years (cf. Friederici 1911: 96; Krämer-Bannow 1916: 263). For instance, in the Catholic missionary schools in Alexishafen, or on the Gazelle peninsula, German was the language taught in the mission schools (cf. Werner 1911: 239; Behrmann 1922: 309).⁵⁴ Behrmann reports an incident in which he is greeted by a Pacific Islander in High German explaining to his readers that “this was a student of the Catholic mission who teach their alumni the German

⁵⁴ As a consequence, the missionary school students in some mission schools started to use the German language outside the classroom setting for intertribal communication, which led to the emergence of a colonial variety of German. The variety came to be known as Rabaul Creole German, also called Unserdeutsch (Maitz et al. forthcoming).

language” (1922: 309).⁵⁵ When Franz Wolf realised in 1923 that there were many Pacific Islanders able to speak Boiken (Sepik (Ndu): Papua New Guinea), the mission intended to use the language as a lingua franca (cf. Ross 1996: 602). As the attempts failed, the language was replaced by Tok Pisin only seven years later. From then on, the Roman Catholic Mission published several documents in Pidgin English such as, to mention a few, the *Bigfelo Katolik Baibel* (1939a), *Buk raring na singsing* (van Klaarwater 1934), *Skul Bilong Evangelio* (1939b) and reports can be found which state that prayers were taught “in pidgin-English, [...] to the natives of the missions of New Guinea by the Missionary Fathers of the Divine Word” (Anonymous 31.03.1938: 32).

The *Lutherans* were present in New Guinea with two missionary societies, namely the Neuendettelsau Mission and the Rhenish Mission. As Ross reports, the missionaries “learned the local vernacular, taught in it, then sent out some of their converts as local evangelist, to neighbouring areas” (1996: 599). As they frequently came into contact with Yabem (Austronesian (Oceanic)), Kâte (Trans-New Guinea (Huon)) and Gedaged (Austronesian (Oceanic)) speakers, these languages turned into the lingua francas of the mission.

3.1.5.4 Summary

The present section has shown that while the majority of missions promoted the use of a vernacular language as the mission’s lingua franca, only some decided to use PE. Despite the ambitions by some missions to use a language other than PE, its usefulness could not be denied which was why, “[a]lthough the missionary usually works in the local vernacular, pidgin English allows him to converse with the natives of other language groups” (Cormack 1944: 130). From the early collected socio-historical information it is apparent that many of the missionary societies which officially denied the usage of PE for their activities used the variety, nonetheless. Those missions that actively encouraged the use of PE were of importance as they promoted the production of written PE material and, consequently, contributed to the varieties’ stabilisation.

3.1.6 World War II

The outbreak of World War II led to renewed contact between Pacific Islanders and Australian, New Zealand and American soldiers. Allied troops were established in operational zones and surviving reports show that Melanesian Pidgin English became “one of 40-odd languages and

⁵⁵ “Als ich zu einem Nachbardorfe ging, begegnete mir sogar ein Schwarzer mit einem schönen, roten Lendentuch, der mich plötzlich auf gut hochdeutsch anredete: “Guten Morgen, Master!” [...] Es war ein Schüler der katholischen Mission, die ihren Zöglingen die deutsche Sprache beibrachte” (Behrmann 1922: 309).

dialects [...] being taught by the Army's Special Service Division" as "a matter of practical necessity for soldiers in New Guinea, the Solomons and thereabouts to know" (Wickware 1943: 67). Also the *Sydney Morning Herald* reports on 8 January 1943 that in order "[t]o familiarise men likely to be fighting in the Pacific islands with as many aspects as possible of the new conditions they will meet, special instructors are being appointed to teach soldiers 'pidgin English'" (8 January 1943: 9). Language guides for soldiers were written and dictionaries and grammars were produced.⁵⁶

The present section outlines to what extent the Trusted Territory of New Guinea (3.1.6.1), the Solomon Islands (3.1.6.2) and Vanuatu (3.1.6.3) were involved in World War II and what consequences it had on the pidgin varieties spoken in the areas.

3.1.6.1 World War II and New Guinea

The Trusted Territory of New Guinea was getting involved in World War II when on 4 January 1942 the Japanese bombed and, 19 days later, captured Rabaul. Rabaul was of strategic importance for the Japanese, as New Britain was on the way of Allied shipping routes to Australia. In May, the Japanese forces occupied the complete Bismarck Archipelago (cf. Hinz 1995: 37).

The events of the Second World War had a marked impact on Tok Pisin in terms of its functions and domains of use. When New Guinea became drawn into in the Second World War, this led to a break with the colonial plantation economy. Australian and American forces launched an offensive against the Japanese in 1943 and the area became a place of heavy battles. Many of the plantations were destroyed, and with them the *raison d'être* of the contact variety.

While Tok Pisin had been used in a master-servant relationship and for intertribal communication on plantations and in the mine industry before the war, New Guineans were now recruited as carriers or were enrolled in the forces and elsewhere (e.g. missionary work). Contact arose between Tok Pisin speakers and those that spoke the "more anglicized Papuan Pidgin English" (Mühlhäusler et al. 2003: 7). Australian authorities, as well as American, New Zealand and Japanese troops, recognised the usefulness of Tok Pisin in communicating with the indigenous population in an area characterised by a high level of language diversity. Allied troops were taught the pidgin (cf. Wickware 1943) and several language guides, dictionaries and grammars were produced for soldiers. For instance, the *Booklet on Pidgin English as used in the Mandated Territory of New Guinea* by Helton (1943) was written to help the troops to "speak the language of the country invaded" (1943: 6). Next to a vocabulary list, it provides the user with

⁵⁶ Cf., for instance, the booklet *A Pocket Vocabulary of Malay, Pidgin English, and Japanese Phrases* (Johns 1942), which was produced for U.S. troops in the Pacific Areas.

commands that may be of importance during war.⁵⁷ In addition, Sayer (1943) writes in the preface of the second edition of his text and vocabulary book called *Pidgin English*:

As a result of the war in the Pacific areas, the interest in the U.S. during 1941 and 1942 greatly increased, and in this year 1943, it is noted the U.S. government has even supplied certain members of the armed forces with data upon the subject, so that confronted with Pidgin talk in certain areas the Army, Navy, or Air Force Personnel will be able to know something about it.

There is also evidence that Japanese soldiers learnt and made use of the variety. Evidence can be found in an interview which was conducted by Dr. Iwamoto Hiromitsu in connection with a project called *Remembering the war in New Guinea* in which New Guineans were asked whether the Japanese spoke Tok Pisin. The answers given by the Islanders show that those Japanese that spent a long time in New Guinea were able to speak Tok Pisin, but that others were not able to use the language.⁵⁸ Furthermore, there is evidence that schools were established in which Pidgin English was taught to the Japanese:

On Karkar Island, they established a school, where native children were taught Japanese and Japanese servicemen were taught Pidgin English [...] They taught the soldiers Pidgin English, which was essential if there were to be any sort of administration (Lawrence 1964: 106-107).

Pidgin English was not exclusively used as a spoken medium between the military and indigenous population. All sides used the language for written war propaganda. Millions of leaflets were produced in pidgin and dropped over New Guinea and the northern parts of the Solomon Islands (cf. Kerr 1985; *National Australian Library* NLA MS 9002 and *Australian War Memorial*). The Australians even set up a special unit called the *Far Eastern Liaison Office* responsible for producing propaganda leaflets.

3.1.6.2 World War II and the Solomon Islands

After having captured Rabaul, the Japanese also occupied Tulagi and North Guadalcanal in March 1942. In November 1942 the Americans began the fight to recapture the areas. Japanese established air bases at Munda and Kolombangara. Even though the Allied forces recaptured most of the Solomon Islands area by June 1943, the Japanese were present in the island group until 1944. In contrast to the Australian Trust Territory, dictionaries and booklets of Solomon Islands Pijin were not found during my search for early data. This may be explained by the fact that the varieties were not considered to differ a great deal and were considered by the Americans to be

⁵⁷ For example, “Bring im me long other pella place”, meaning ‘Bring me to another village’ (Helton 1943: 16).

⁵⁸ “Sampela. Ol i save toktok pisin. Ol i bin stap olsem planti mun nau long mipela na ol i save toktok long tok pisin na i kamap klia gut liklik. Sampela ino kisim yet tok pisin” (Hiromitsu & Australian National University 2000: online).

characterised by mutual intelligibility, despite “certain variations in in vocabulary and pronunciation from village to village and from island to island” (cf. Wickware 1943: 67).

The World War, however, led to renewed contact between Pacific Islanders and foreigners. The *Solomon Island Labour Corps* was established and consisted of European officials as well as returned plantation managers. There are many reports in which Solomon Islanders state that they worked in the Labour Corps for about two years: “Time war come me work for two year long Labour Corps in Guadalcanal for American” (*The Stories of the Crew* 1947: 11; AU PMB DOC 439). War songs in Solomon Islands Pijin appeared and the district officer Lennox Barrow (1942-1947) provides evidence that the troops consisting of Americans and Solomon Islanders communicated in Pidgin English (AU PMB MS 517).

3.1.6.3 World War II and Vanuatu

Even though the area of Vanuatu was not attacked and invaded by the Japanese forces, renewed contact between Vanuatians and foreigners arose in Vanuatu as an estimated number of 5,000 American troops “were quartered or passed through Vanuatu in their struggle with the Japanese in the Solomon Islands” (Tryon & Charpentier 2004: 319). Vanuatu hosted airfields and naval anchorages and thus became an important military base for American troops as well as New Zealanders and Australians. In addition, the *New Hebrides Defense Force* was established by the Australian military on the island of Malekula. Vanuatians were recruited on three-month contracts to work for the Americans (cf. Crowley 1990a: 107; Tryon & Charpentier 2004: 320) and it can be expected that early Bislama was influenced through these interactions. If this proves to be true, however, it raises the question why French did not have a stronger impact on the development of Bislama. If three-month interactions can have an impact on the language, one would expect French to have had a stronger impact as well. The fact that soldiers used written Pidgin English for mass communication may provide one explanation, since transcriptions will have contributed to the stabilisation of the PE varieties. Power relations and imbalances in the degree of power may represent another reason. Thus, sociological factors might play a role in terms of whether and to what degree languages could have an impact on each other.

3.1.6.4 Summary

The war time interaction between Pacific Islanders, Australians, New Zealanders, Americans and Japanese represents an important phase in the development of MPE. Prior to the Second World War, *written* forms of the varieties were used for religious purposes only. By using Tok Pisin in a written form for mass communication during wartime, the language was exposed to

standardising processes and the attitude towards the language changed as soldiers learned Melanesian Pidgin English as a language in its own right.

Moreover, American English is likely to have left its traces in the pidgin varieties. *The Daily Telegraph* in March 1944 reports that “PIDGIN English is getting so mixed up with American slang and jive talk that missionaries, planters, and traders will need a new vocabulary to get by if the war lasts much longer” (*The Daily Telegraph*, 9 March 1944: 8). Furthermore, in *Life* it is stated that:

[u]ndoubtedly U. S. slang will be broadened and enriched after the war when our soldiers come home with Melanesian pidgin on their tongues. On the other hand, it's equally likely that the pidgin language even now is undergoing changes as the natives are exposed to potent Americanisms (Wickware 1943: 70).

The renewed contact was accompanied by the introduction of new war items for which no vocabulary existed so far. It seems reasonable to assume that the renewed contact situation left traces in the pidgins' lexicons and grammars. That this assumption is not farfetched is supported by my study focussing on *The impact of Second World War propaganda leaflets on Tok Pisin grammar and lexicon* (Schäfer 2017b), in which I was able to show that the newly arising contact led to the introduction of new lexical items and to the semantic expansion of already existing lexicon.⁵⁹

3.1.7 Post-World War events

Since the scope of the present study focusses on the individual feature development until the end of the 1940s, this section will only briefly refer to a limited amount of post-war historical events.⁶⁰

3.1.7.1 Post-World War events in the Solomon Islands

The post war era started in the Solomon Islands with the Maasina Rule movement. During World War II the Solomon Islanders felt forsaken by the British and came to appreciate the friendly cooperation with the Americans. Thus, the movement emerged out of the aim to create a “unified front to negotiate with the British on local matters” (Bennett 1987: 293). Solomon Islands Pijin provided the main means of communication and “it is through it only that intertribal meetings could be held [...] [and] that the political ideology of the movement was disseminated” (Jourdan 1985: 51; cf. Jourdan & Selbach 2008: 166). Although the movement came to an end when its

⁵⁹ For instance, the pre-war meaning of *umbrella* was ‘umbrella’. Due to World War II the meaning was expanded and the word became used to refer to parachutes as well.

⁶⁰ For a detailed description of the historical developments, see Bennett (1987) for the Solomon Islands, MacClancy (1981) for Vanuatu, Waiko (1993) for Papua New Guinea and Tryon & Charpentier (2004) for the Pacific Islands in general.

nine leaders were imprisoned by the British administration in 1952, this represents the first time that Pijin was used for political purposes in the Solomon Islands.

Another decisive turning point was the movement of the capital from Tulagi to Honiara in 1953. In the capital, people from various ethnic backgrounds came together which made the knowledge of Solomon Islands Pijin a prerequisite. As Jourdan (1995: 140) summarises, Pijin “became the ‘natural’ language of the urban world, superseding the vernaculars”. Thus, the use of Pijin was expanded to further domains and was no longer reserved for the male population, but became used by women as well, and turned into a necessity for all those who “want[ed] to have a social life which goes beyond the limits of the ethnic group” (Jourdan 1995: 140). The variety’s use was further promoted by the establishment of multi-lingual schools and inter-island trade, which both required a unifying tool of communication.

Despite its spread, however, the variety suffered from a “systematic denigration [...] by the local administrative elite” (Jourdan 2018: 83), which was perpetuated when the Solomon Islands gained independence in 1978. English, the official language of the country, was and is regarded as the high language (HL) which is used in education and for written communication. Solomon Islands Pijin, even if by now accepted as a helpful political tool for wider communication and being the “de facto national language” (Jourdan 1990: 168) is regarded as the low language (LL) variety.

From its beginnings to today, acrolects, mesolects and basilects of Solomon Islands Pijin can be differentiated and further variation is observed depending on the “geography, social class, gender and age” of a speaker (Jourdan 2008: 468). The highest degree of variation is attested between urban and rural varieties.

3.1.7.2 Post-World War events in Vanuatu

In Vanuatu, the labour recruitment saw a new revival in the post war era as Vanuatuans’ interest in their own copra production rose, resulting in recruiting labourers from other Asian areas (e.g. Tahitians, Italians, Wallisians) (cf. Tryon & Charpentier 2004: 321). Multi-ethnic interactions occurred not only on the plantations but also in the fish industry and in manganese mining (cf. 2004: 321).

Similar to the Solomon Islands, urbanisation played an important role in the further stabilisation of Bislama. Crowley reports that it increased between the years 1955 until 1967 so that in the cities “urban-based inter-island families” were no exception, the children of which acquiring Bislama as their mother tongue (Crowley 1990a: 107).

Bislama turned also into an important tool for the distribution of news and reports. From 1961 onwards, the *Bulletin d'information de la Residence de France*⁶¹ was distributed, a large part of it printed in Bislama. Charpentier reports that only ten years later British newspapers started to include a section written in Bislama as well (cf. 1979: 163). Even today, the pidgin is used in some newspapers, while television and radio programs are also sometimes broadcasted in the variety (cf. Meyerhoff 2013a: 224).

Vanuatu gained independence in 1980. In contrast to Solomon Islands Pijin, Bislama represents one of the official languages along with English and French. It is not used as the medium of instruction in schools, albeit initial attempts have been made to implement Bislama for basic education in urban centres (cf. Meyerhoff 2013a: 224).

From its beginnings to today, a standardised form of Bislama has yet not developed (cf. Meyerhoff 2013a: 224). Similar to the Solomon Islands, differences can be observed between rural vs. urban speakers of the language. Since there are, however, no Vanuatians that spend their lifetime only in either a rural or an urban area (due to changing job opportunities, lack of land or housing, access to education and health care services, etc.), “there is constant feeding between the two, and it is possible to recognize only the ends of a continuum” (Crowley 1990a: 20). Next to the rural vs. urban distinction, a differentiation between more-educated and less-educated forms (= edulects) of Bislama can be made. Differences can also be observed based on whether Vanuatians experienced a French- or English-based education (cf. Crowley 1990a: 20).

3.1.7.3 Post-World War events in Papua New Guinea

The World War events had a remaining impact on how Australians regarded Papua New Guinea. As they started to recognise the country's importance, they aimed at helping the New Guineans to “run their own country” (Mühlhäusler 1979: 98) which led to a reduction of the strong pre-war master-servant distinction. While the use of English was, prior to the war, restricted to the masters, and pidgin was regarded as the language of the servants,⁶² English was promoted among New Guineans after 1945. This had the effect that “certain groups had free access to English and could change the character of NGP [= New Guinea Pidgin] in its direction” (Mühlhäusler 1979: 99).

Urbanisation is also of importance for the further development of Tok Pisin. The establishment of urban centres had the effect that many New Guineans migrated to the cities in search for better jobs. The gathering of New Guineans with multilingual backgrounds led to the necessity of Tok Pisin for successful communication.

⁶¹ *New Hebrides Service de l'Information* and *New Hebrides Résidence de France*

⁶² As the servants were called *boys* the variety was also referred to as *tok-boi* (cf. Mühlhäusler 1979: 98).

Furthermore, prior to the war, Tok Pisin was predominantly used as a spoken language and the missions represented the only institutions that published in Tok Pisin. After the World War, the government had to “choos[e] between effective communication with the largest possible number of people or promotion of the territory-wide use of English” (Mühlhäusler 1979: 101). The government decided in favour of the former so that several newspapers were produced in Tok Pisin and radio stations broadcasted in the variety. The usage of Tok Pisin in the media was considered a tool “of bringing NGP [= Tok Pisin] closer to English” (Mühlhäusler 1979: 101).

The Trust Territory of New Guinea and the colony of Papua were united in 1975 and became what is nowadays known as Papua New Guinea with a single administration. Besides Hiri Motu and English, Tok Pisin is one of the three official languages of the country. Even though English represents the primary language of education, the communities can, since 1990, decide about the language of instruction to be used in elementary schools. This has led to Tok Pisin being implemented in the early education years in some areas (cf. Smith & Siegel 2013a: 215).

As is the case with Solomon Islands Pijin and Bislama, a standardised form of Tok Pisin does not exist (cf. Wurm 2012: 424) and urbanisation has led to the differentiation between a rural and an urban variety. Moreover, Mühlhäusler (1979: 147) distinguishes between Tok Masta and bush variants, the former describing the form of Tok Pisin used by expatriates, and the latter describing the variants which are spoken “in remote areas of Papua New Guinea (Mühlhäusler 1979: 149).⁶³

3.1.7.4 Summary

This section provided an overview of post-war socio-historical and sociolinguistic developments in the Solomon Islands, Vanuatu and Papua New Guinea. We saw that the use of MPE was expanded to further domains in all the varieties. At the same time, we saw that while Bislama in Vanuatu and Tok Pisin in Papua New Guinea are officially recognised and have blossomed in the past years, Pijin in the Solomon Islands does not receive the same recognition. It became clear that there are substantial arguments to assume that the varieties will have changed in the years after 1950 as well.

⁶³ For a closer elaboration of the four different dialects, see Mühlhäusler (1979: 140-154).

3.2 Theories on the genesis of MPE and the origin of the individual varieties

Several hypotheses have been proposed which try to explain how Melanesian Pidgin English and its individual varieties came into being. While some of these theories are based on the socio-historical events described above, other theories represent mere speculations and will therefore not be further referred to.⁶⁴

An ongoing and heated scholarly discussion has risen out of the question whether and to what extent the plantations had a bearing on the diverging feature development in the three varieties Tok Pisin, Bislama and Solomon Islands Pijin. The role of the Queensland and Samoan plantations plays a prominent role in this discussion. In the following, an overview of the different theories regarding the origin of Melanesian Pidgin English will be outlined.

3.2.1 Origin of Tok Pisin on Samoan plantations

Tok Pisin is the most commonly analysed variety of MPE and so theories about its origin have been put forth since as early as 1966. Wurm (1966: 51) was one of the first linguists who saw a direct link between labour recruitment and the development of early Tok Pisin, claiming Tok Pisin to be the “direct result of indentured native labour on the sugarcane plantations of North Queensland”. While his viewpoint was *inter alia* supported by Laycock (1970: x), the theory caused immediate opposition. To regard QPPE as the major input variety of Tok Pisin was opposed by Salisbury (1967) who pointed towards the historical fact that New Guinea was only involved in the Queensland labour trade until 1885 and provided only a small number of labourers.

Peter Mühlhäusler (1978) supported Salisbury’s claim that QPPE might have been the major source for what became Solomon Islands Pijin and Bislama but was of almost no importance for the development of Tok Pisin. Mühlhäusler, whose study focussed on Tok Pisin and did not include the other MPE varieties, claims that the origin of Tok Pisin goes back to Samoan plantations and the variety spoken there. He attributes a decisive role in the individual development of Tok Pisin to the recruitment for Samoan plantations from 1884 onwards and argues that Tok Pisin had developed exogenously on Samoan plantations and was brought to New Guinea by returning labourers.

⁶⁴ In Helton (1943: 5), for instance, it is argued that Tok Pisin developed because the Germans did not succeed “to teach the natives to speak German” and “taught them the various English names of the articles they were using” instead. That this theory is based solely on speculations can be learned from the early records which show that by the time the Germans were aiming to teach German to the islanders, an early form of Tok Pisin had already been established.

To prove this, Mühlhäusler conducted fieldwork in Western Samoa in 1975 and used the interview material together with colonial sources of SPPE and early Tok Pisin to qualitatively compare the varieties with each other. Pointing out the similarities between the two, Mühlhäusler concluded that SPPE and early Tok Pisin show the “closest structural relationship” (1978: 109) in comparison to all other Pacific varieties. He even goes as far as to claim that “[u]p to 1914 the speakers of these two pidgins [=SPPE and Tok Pisin] constitute[d] a single speech community” (1978: 109). Thus, the end of the labour trade between Samoa und German New Guinea in the year 1914 is regarded as crucial for the individual development of Tok Pisin and its divergence from SPPE. According to Mühlhäusler, “[i]t was only when SPP[E] was taken to New Guinea and the Bismarck Archipelago that it could develop from a restricted plantation pidgin into an extended pidgin serving as a means of communication over a wide range of topics” (1978: 86).

Mühlhäusler provides evidence not only through linguistic attestations but also through sources which mention that labourers in German New Guinea had learned their pidgin on the plantations in Samoa. In their article *New Evidence of a Samoan Origin of New Guinea Tok Pisin* Mosel & Mühlhäusler (1982) further support their claim by providing an interview in which an informant confirms that he learned the pidgin in Samoa.

Even though Mühlhäusler does not exclude the possibility that other factors had an influence on the formation of TP, he concludes from the qualitative analysis between SPPE and early TP that these two varieties are closest to each other, without including and directly comparing the pidgin spoken on the plantations in Queensland to the analysis, claiming that a comparison with other varieties would be “not possible at present, since the available data are neither reliable enough nor complete enough” (1978: 87). Even if he does not deny an indirect influence of QPPE (1978: 69), he does not put it into concrete terms. Problematic might also be that the Samoan data used for the analysis consists of only a small amount of colonial data and the pidgin as obtained from interviews in 1975. All languages develop over time and this is especially true for pidgins. The 1975 attested SPPE is thus likely to differ from its pre 20th century form. The low availability of SPPE may be due to the lack of surviving records but it may also be an indicator that Samoan Plantation Pidgin English did not have wide currency.

3.2.2 A 1970s early Melanesian Pidgin

Four years after Mühlhäusler had published his study on the origin of Tok Pisin on Samoan plantations, Ross Clark consulted early sources focussing on 30 features common in pidgin and creole languages to investigate how these features spread. In his article *In search of Beach-lamar: Towards a history of Pacific Pidgin English* (Clark 1979), he claims that South Seas Jargons

were introduced in the Pacific in the 1840s. According to Clark, the initial jargons stabilised into Sandalwood English due to sandalwood activities in New Caledonia, the New Hebrides and Loyalty Islands. With the beginning of labour recruitment, Sandalwood English would have been brought to Queensland and Fiji plantations and developed quickly into “an early form of Melanesian pidgin” by the 1870s (1979: 49). As Figure 8 shows, he assumes early MPE to have further developed into individual varieties.

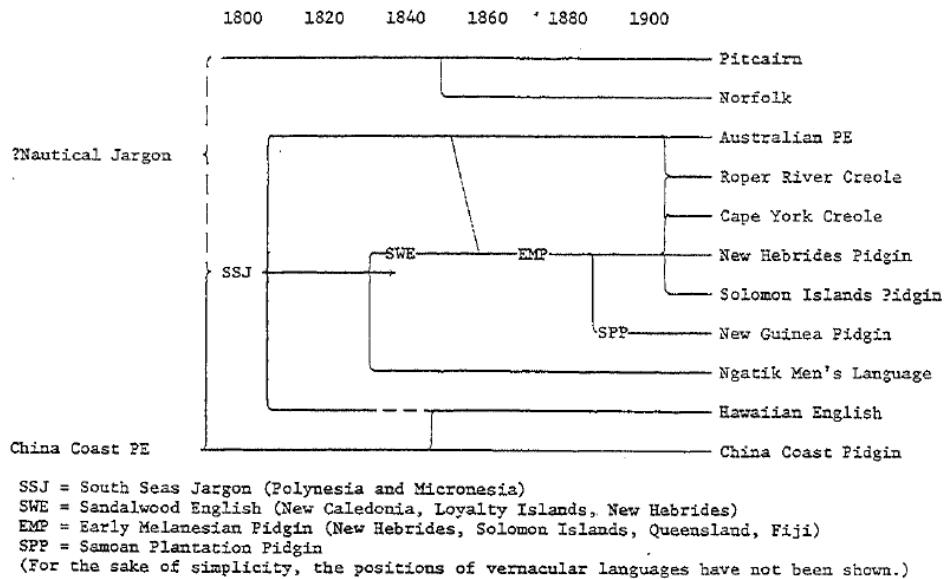


Figure 8: Early Melanesian Pidgin Development in Clark (1979: 48)

Like Mühlhäusler, he ties the individual character of Tok Pisin to SPPE. While the figure indicates that the differentiation of Tok Pisin from Solomon Islands Pijin and Bislama thus took place around 1880, he dates the differentiation of Bislama from Solomon Islands Pijin shortly after 1900, and thus, to the end of the labour trade. At the same time, he suggests that Australian Aboriginal Pidgin had an impact on Sandalwood English in the early labour trade as “there is evidence that at least four of the comparative features were present in Australia at least as early as or earlier than in Melanesia” (1979: 43).

3.2.3 Pacific-wide nautical jargon

While Mühlhäusler and Clark assume that English-lexified pidgins developed simultaneously in different areas of the Pacific, Roger Keesing (1988) argues for the existence of a single Pacific-wide nautical jargon. This “pidgin” as Keesing calls it, “although not yet fully stable, was relatively uniform” (1988: 4) and was used during trade interactions and labour recruitment on board of trading and recruitment vessels until the late 1880s. The pidgin spoken on the different

plantations, independent of whether in Queensland, Samoa, New Caledonia or Fiji, would have been the same and would have belonged to a widespread speech community:

[I]n the early 1860s there was a single dialect of Pacific Pidgin, largely shipboard-based, which provided the linguistic input into plantations in Queensland, Samoa, New Caledonia, Fiji, the Marshalls, and other areas [...] this shipboard-based dialect already incorporated many of the grammatical patterns later recorded in Samoa and Queensland. (1988: 53)

According to Keesing, local dialects might have developed in the 1870s as ships no longer represented the primary locus of contact but plantations took over that role, allowing local lexemes of the plantation areas to enter the vocabulary. Nonetheless, Keesing argues that the differences were only “minor elaborations and modifications” until the end of the 1880s (1988: 4). Thus, he contradicts Mühlhäusler and argues that it would not make sense from a historical nor a linguistic perspective to assume that Samoan recruitment ships used a specific PE that would have differed from the other areas and which would make “foremen, and recruiters, and plantation managers [...] learn a new dialect” (Keesing 1988: 57).⁶⁵ Keesing thus hypothesises that the stabilisation of MPE took place in the period between 1860 to 1890 and that it was already grammatically developed at that time (cf. Keesing 1988: 25). With the establishment of plantations in Queensland, Fiji, Samoa and New Caledonia, and later on in Papua New Guinea, Vanuatu and the Solomons, its usage would have been narrowed down to the southwestern Pacific. He argues that the modification and differentiation of Tok Pisin from the other two varieties was due to “superstrate influence of German and the substrate influences of the Oceanic Austronesian languages of the Bismarcks and the New Guinea coast, and of Papuan languages” in the late 1880s (1988: 61). Solomon Islands Pijin diverged from Bislama due to the “separation of the Solomons and New Hebrides pidgin speech communities” (1988: 171, 173) with the end of the labour trade. Keesing considers substratum influence as the decisive factor which evoked the individual feature development.

In a later article by Jourdan & Keesing (1997), the authors argue that Tok Pisin, Bislama, and Solomon Islands Pijin were connected closely until the late 1880s in a single speech community. Although Tok Pisin “became disconnected from the Oceanic substratum languages” (Jourdan & Keesing 1997: 402) at that point in time, several syntactic patterns were already well formed and could thus be found in all three varieties, “despite the separate developmental history proposed for the latter by Mühlhäusler” (Jourdan & Keesing 1997: 403). They argue that the further differentiation between the varieties was caused by English forms being replaced in Tok Pisin “with vernacular labels from Bismarck languages”, resulting in “a special and different

⁶⁵ Keesing (1988: 57) claims: “I see no evidence, linguistic or historical, for separating Samoan pidgin of the latter 1880s from the pidgin spoken elsewhere in the southwestern Pacific.”

developmental trajectory and sociolinguistic environment” from Bislama and Solomon Islands Pijin (Jourdan & Keesing 1997: 403). The latter two varieties remained in contact through their substrate languages (cf. Jourdan & Keesing 1997: 403). Their differentiation would have started when “[i]n the early 20th century, after the Solomon plantation workforce had been cut off from their New Hebridean counterparts through the closure of Queensland and Fiji recruiting”, which would have resulted in syntactic patterns more similar to the grammar of Southeast Solomonian languages (Jourdan & Keesing 1997: 404).

3.2.4 Origin of MPE in Sydney and plantations in Queensland

The idea of a Pacific-wide nautical jargon was contradicted and refuted by Philip Baker (1993) who, comparing the earliest dates of attestation of 107 features in Chinese Pidgin English, Queensland Plantation Pidgin English and New South Wales Pidgin English with the earliest dates of attestation in Melanesian Pidgin English, demonstrated that there is no evidence to suppose a universal pidgin. In his view, the Melanesian pidgins derive from QPPE that, in turn, was a continuation of NSWPE. Since Sydney served as the main trading area to which European and American ships came, Baker (1993: 61) argues that:

its pidgin [New South Wales Pidgin English] was not only the recipient of more features from pre-existing pidgins and creoles of the Atlantic region and Asia than other pidgins which were subsequently to develop in the southwestern Pacific, but it was also the most important donor of features to the latter.

Melanesian island workers serving on Queensland plantations would have learned the variety during their stay on the plantations and would have brought back their knowledge when repatriated. As the pidgins spoken in Melanesia up to that point would have been less developed, it “would have been rapidly absorbed into [Queensland Plantation Pidgin English] to create what may properly be termed MPE” (Baker 1993: 56).

Baker claims that diverging features in the MPE varieties would have started to develop due to substrate influence once the Pacific Islanders were repatriated to their places of origin. Thus, indigenous languages of the individual island groups will have influenced the varieties’ further development (1993: 62).

As Baker showed that many of the features attested in MPE varieties and in QPPE could be found in the NSWPE of the beginning of the 19th century, he assumed that NSWPE had a greater impact than previously assumed.

3.2.5 South Seas Jargon as an element of Maritime Polynesian Pidgin

A relatively recent study by Emanuel J. Drechsel (2014) focussing on *Language Contact in the Early Colonial Pacific* provides evidence that a Polynesian Pidgin may have had an impact on the development of Pacific Pidgin Englishes as well. Drechsler takes an opposing view to that of Ross Clark, who assumed South Seas Jargons to be English-lexified. Though Clark (1977: 35) does not deny the fact that traces of “broken Polynesian” can be found in an early form of the jargon, he claims that “broken Polynesian” could not be clearly separated from “broken English”, as the basic word order would have followed English. Therefore, he assumes the Polynesian items to reflect local vocabulary being relexified and turning into Sandalwood English at the beginning of the 19th century (1977: 45).

Drechsel contradicts this view and shows that a Maritime Polynesian Pidgin (MPP) was used in the first interactions between Pacific Islanders and Europeans and thus provided the basis for South Seas Jargons. His archival research indicates that a Maritime Polynesian Pidgin was in use much earlier than Pacific Pidgin English (2014: 300) and was inter alia used in maritime trade situations focussing on whaling, sandalwood and trepang trade (2014: 86). Even when English was introduced in the Pacific and English-lexified contact varieties developed, the latter were not used exclusively and did not represent the “primary interlingual medium”, but instead represented “only one among many other means by which native and foreign residents interacted in crosscultural contacts at the time” (2014: 300). Drechsel thus does not exclude the possibility that Maritime Polynesian Pidgin was used by indentured labourers on those plantations “with an indigenous pidgin as a sociolinguistically dominant medium” (2014: 86). MPP may have served “as the linguistic springboard for European and American settlers as well as immigrant laborers” in contact situations in the Pacific (Drechsel 2014: 282).

3.2.6 The end of the labour trade

Despite the various theories regarding the origin of Melanesian Pidgin English, the researchers basically agree that Pacific pidgins further diverged and developed their individual features with the end of the labour trade (cf. Mühlhäusler 1978; Clark 1979; Keesing 1988; Baker 1993; Jourdan & Keesing 1997; Sankoff 2021+). Such an assumption is reasonable when external, socio-historical events are considered (cf. Section 3.1.3).

While some authors differentiate between several pidgin varieties, Tryon & Charpentier (2004) assume that “a generally uniform Pacific Pidgin [was] spoken throughout Island Melanesia by the time that labour recruiting stopped, in 1906 in Queensland, and as late as 1911 in Fiji and 1914 in Samoa [...]” (Tryon & Charpentier 2004: 349), although they add that

competing forms existed in the idiolects. The end of the labour trade would have “provid[ed] an environment which resulted in different linguistic choices being made in the three areas” (Tryon & Charpentier 2004: 350) that, in turn, would have led to the development of “particular characteristics” (2004: 198) by the end of the labour trade period.

3.3 Summary and outlook

The first section of this chapter outlined the socio-historical circumstances that led to the need and development of pidgin varieties in New Guinea, the Solomon Islands and Vanuatu. It was shown that, from a socio-historical perspective, labour recruitment and the end of the labour trade may have had a dominant impact on the development of Melanesian Pidgin English into its three varieties, i.e. Solomon Islands Pijin, Tok Pisin and Bislama. At the same time, diverging features may have originated many years prior to or after the end of the labour trade since the socio-history suggests that other possible factors may have influenced the individual varieties’ developments as well. It is imaginable that it was not a unidirectional development from similar to non-similar but that the historical circumstances changed the degree of proximity in both directions over time.

Section 3.2 introduced various theories that have been put forth to explain the origin of Melanesian Pidgin English. The section has shown that our knowledge about past stages of MPE is still far from complete. Scholars placed their focus on the development of MPE and the role which jargons and overseas plantations pidgins had on its development. Although all theories make the indirect claim that the end of the labour trade led to the localisation of the three MPE varieties, and, although the historical events support the general hypothesis, none of the studies back up their claims with historical comparative linguistic data.

As outlined in the introduction, a diachronic, comparative analysis of early language material of the three varieties that goes beyond the documentation of earliest attestations has not been conducted so far. However, it is essential for the clarification and disclosure of the linguistic circumstances. A data-driven approach may add clarity regarding *which* linguistic features diverged *at what point in time* and will contribute to the identification of *which external events* might have been responsible for the individual development of the varieties. The following chapter will provide a detailed framework of the data and methodology used in the present study to answer the research question *when* a localisation of *demonstratives*, *relative clause markers*, *modality markers* and selected *prepositions* took place in Solomon Islands Pijin, Bislama and Tok Pisin.

4 Empirical foundations: Data and methodology

To reconstruct previous stages of languages and their development, early language data is required. The present chapter documents how and where data was collected for the present study (4.1) and introduces the data which served as the basis for the investigation (4.2). Moreover, the methodology employed is introduced, while the processes involved from the feature selection to the final qualitative and quantitative analysis are described (4.3). The chapter closes with a comment on the strategies which were implemented to prove the reliability of the collected data regarding the factors introduced in Section 2.2 (i.e. anglicisation, editor revisions, language attitudes, authenticity etc.).

4.1 Database compilation

Although researchers have devoted considerable time collecting early attestations and descriptions on contact varieties to reveal how they emerge, the collections are scattered among individual researchers around the world (cf. Huber & Velupillai 2018: 133). Only in recent years have researchers started to develop electronic databases which make the collected material available for a larger audience.⁶⁶ However, as databases for Tok Pisin, Bislama and Solomon Islands Pijin do not exist, the data for the present study had to be searched for and collected in various archives. The following sections provide detailed information on how and where data was searched for (Section 4.1.1) and what type of data was extracted (Section 4.1.2).

4.1.1 Identification of archives and possible data sources

The first step involved identifying potential archives and data sources which might contain travel accounts, court proceedings, memoirs or other types of documentation referring to the Solomon Islands, Vanuatu and Papua New Guinea.⁶⁷ The time period for which data was collected had to be restricted. The main aim of the study was to investigate when the MPE varieties diverged from each other and whether the assumption that the end of the labour trade was the start of the divergence could be supported. Thus, I focused on data covering the time period from their earliest attestations until the year 1950. This does not imply, however, that linguistic changes did not happen after 1950. All languages change over time, and the varieties will have continued to diverge or converge until today. Nevertheless, it is beyond the scope of this study to examine the

⁶⁶ Cf., for instance, the databases *Early Suriname Creole Archive* (SUCA, suca.ruhosting.nl, last access 13 May 2021) and the *Negerhollands Database* (NEHOL, <http://www.clarin.nl/node/162>, last access 13 May 2021).

⁶⁷ Initially, the aim was to have a fourth corpus for Torres Straits, which is frequently considered to belong to Melanesian Pidgin as well. This could not be realised as too few sources could be obtained.

post-1950 changes. In order to remain as theory-neutral as possible with respect to the point of departure for these varieties, I aimed at including data from as early as possible at the potential points of origin, in order to test the suggestions that the early 20th century was in fact the time of divergence.

The difficulty in searching for Pidgin English evidence in early sources is that most search functions will not provide information on whether an author quotes a Pidgin English sentence or not. The items must be individually examined to ascertain the presence or absence of language examples. Language material was sourced mainly from the following archives.⁶⁸

4.1.1.1 *Deutsche Kolonialbibliothek, Frankfurt*

The *Deutsche Kolonialbibliothek* (= German Colonial Library; henceforth *DKB*) housed at Goethe University Frankfurt served as a starting point for the data collection. The German Colonial Society, which was founded in 1887 to promote German colonial activities, was amalgamated with smaller colonial libraries into the German Colonial Library in 1936 and represents one of the most important libraries of German colonial sources.

The library comprises inter alia diaries, letters, travel accounts, biographies, missionary and government reports of early colonial activities in the Pacific region. In 2009 Huber & Velupillai started with student group excursions to the *DKB* to collect early colonial data in search for evidence and information on languages used in colonial contact situations. For German New Guinea 94 sources had already been excerpted and kindly provided by Huber & Velupillai. The data was supplemented by excerpting further German New Guinea sources not analysed so far, and excerpts of Pidgin English used in the Solomon Islands and Vanuatu were added. A greater number of sources could be obtained for New Guinea than for the Solomon Islands or Vanuatu due to the fact that the latter regions were only visited occasionally by Germans, while New Guinea was possessed by the Germans for 30 years.

4.1.1.2 Online Archives

The archive material collected in the *DKB* was supplemented by sources available in online archives. Books, travel reports and journals made available by the *Project Gutenberg* (<http://www.gutenberg.org/>, last access 3 July 2019) or the *Internet Archive* (<https://archive.org/index.php>, last access 13 May 2021) were searched through to find additional early attestations of PE. Further online resources made available by the *National Library of*

⁶⁸ It should be noted that from other German libraries, as, for instance, the research library Erfurt/Gotha, only scattered items were accessed which is why they will not be listed explicitly.

Australia via Trove (<https://trove.nla.gov.au/>, last access 29 September 2021) were searched through as well, including newspaper articles that reported on PE usage. Besides, cross references, which were found in the colonial sources of the DKB, were searched for online.

4.1.1.3 Bislama data collection by Philip Baker

In the 1980s Philip Baker assembled early attestations of MPE varieties for his research involving the impact of Australian Pidgins on the development of MPE varieties. While he had lost most of his material in the 2004 Tsunami in South and South East Asia, he saved a copy of his early data collection on Bislama, which he kindly provided for the purpose of the present study. The data by Philip Baker was annotated with information regarding the speaker, year and place of utterance. The example sentences were not embedded in their original contexts, which, however, is of importance for the interpretation and analysis of the language data. Thus, the sources were searched for in order to embed the language examples into their original contexts.

4.1.1.4 Research Centre of the Australian War Memorial

Research showed that the Australian War Memorial possesses a collection of pamphlets by the *Far Eastern Liaison Office* (FELO) and the *Enemy Leaflet Collection*. Around 159 pamphlets from the Far Eastern Liaison Office (FELO) and 11 from the Enemy Leaflet Collection were obtained from the Australian War Memorial.

Summing up the material of the DKB, the Online Archives, the Australian War Memorial and the material by Philip Baker, language data was extracted out of 649 sources.⁶⁹ However, the data collected could not sufficiently cover the time period from 1840 to 1950. While a great amount of material documenting early Tok Pisin could be collected (containing, for instance, dictionaries and grammars), the data for Solomon Islands Pijin and Bislama was scarce. Additionally, almost no language data stemming from the years 1930 to 1950 was collected so far. A six-week research stay in Canberra and Auckland supplemented the data basis with material from the *Pacific Research Archives* (Canberra), the *National Library of Australia* (Canberra), the *Western Pacific Archives* (Auckland) and the *Pacific Manuscript Bureau*.

⁶⁹ The number solely refers to the sources from which *language data* was extracted. It does not include sources with none-attestations or those providing meta-linguistic information. On average, only one out of ten sources contained data for the present study so that about ten times as many sources were originally consulted.

4.1.1.5 Pacific Research Archives, Canberra

The *Pacific Research Archives* in Canberra contains unpublished and published research material from and about the Pacific Islands. Material out of ten collections was looked at, but the most promising collection was the one compiled by Charles Morris Woodford, who was a naturalist and the first resident commissioner of the British Solomon Islands Protectorate in the years 1896-1915. The collection includes his diaries, correspondence, research notes, and also other documents related to the administration of the protectorate. Another promising source was the *Burns Philp and Company Limited* collections, which contain company records such as “minutes, correspondence, reports, legal and financial records, share registers, shipping records, staff records, photographs, maps and plans, printed material and newspaper collection” (cf. AU Deposit N115 & N145).⁷⁰ While a lot of Pidgin English was attested in the Woodford Collection, little Pidgin English data was found in the Burns Philp collection.⁷¹

4.1.1.6 National Library of Australia, Canberra

During the data collection in Germany, it became apparent that several of the early books and reports could not be accessed from Germany but were in the holdings of the National Library of Australia. Thus, a list of 52 sources was created which were searched for in the National Library of Australia. In total, 49 of these sources contained Pidgin English. The items looked at include hymn books and gospel translations, but also travel reports. Additionally, the National Library has a collection of propaganda war leaflets which were consulted and compared to the pamphlets obtained by the Australian War Memorial in 2016.

4.1.1.7 Western Pacific Archives, Auckland

The *Western Pacific Archives*, which is housed in the Special Collections of the University of Auckland, contains the records of the British colonial administration in the Western Pacific from 1877-1978 and was thus promising with respect to files in which early PE might be documented. The archives contain records of the Western Pacific High Commission (1875-1978), the New Hebrides British Service (1902-1975) and the British Commissioner and Consul, Tonga (1862-1968). The former two were of relevance for this study, especially as the databases for Solomon Islands Pijin and Bislama were still small. Due to the restricted amount of time which I was able to spend at the archives, only a limited amount of material could be consulted. The selection of

⁷⁰ <https://archivescollection.anu.edu.au/index.php/burns-philp-and-company-head-office-deposit> (last access 29 September 2021)

⁷¹ I am very grateful to Rachel Armstrong and her team for their support in the Western Pacific Archives.

archive material was driven by best guess which of the sources might contain files in which Solomon Islanders and Vanuatuans may have used Pidgin English. Thus, the focus was on letters and proceedings about native outrages, trials, native complaints and land dispute.

A focus was laid on records of the *Western Pacific High Commission*. The collection *Inwards Correspondence* consists of letters and comments *from* and *to* the High Commissioner and Resident Commissioner, including trial proceedings and court records dating to the years 1889-1942. While 1,023 files on 303 microfilms were investigated, language examples could only be extracted out of 101 files.

In addition, a part of the records of the *New Hebrides British Service*, which covers the years 1907-1929, was investigated. The *New Hebrides British Service* was established under the Anglo-French New Hebrides convention of 1906 and the collection contains letters and other official documents of colonial administration, focussing on land records, civil and criminal legal cases, native administration and district agent reports. Data was found in 32 files out of 314 file folders looked at. The amount of Pidgin English found in the Western Pacific Archives ranges from single words to several pages pure Pidgin English.⁷²

4.1.1.8 Pacific Manuscript Bureau, Canberra

The *Pacific Manuscript Bureau* makes archives, manuscripts and printed material relating to the Pacific available on microfilms. Their microfilm collection is “the most extensive collection of non-government primary documentation on the Pacific Islands” (<http://asiapacific.anu.edu.au/pambu/catalogue/>, last access 29 September 2021). The online available finding aids for Papua New Guinea, the Solomon Islands and Vanuatu were used, and 54 collections were consulted on microfilm.⁷³ Most PE data was found in documents of the South Sea Evangelical Mission.

4.1.1.9 Summary

The above-mentioned archives were selected because they were the most promising for providing sources that contain early documentations of Solomon Islands Pijin, Bislama and Tok Pisin. Table 3 gives an overview of the number of sources from which data was extracted for the present study per variety and archive.⁷⁴ In total, the data for the present study was extracted from 984 sources.

⁷² Special thanks to Katherine Pawley and Stephen Innes of the Special Collections for their help to find possible data sources.

⁷³ Special thanks to Kari James for her assistance.

⁷⁴ The table only lists those sources, in which demonstratives, relative clauses, modality markers and/or the selected prepositions were identified. Sources which contained PE examples that did not contain one of the features under analysis are not listed in the table.

Archives/Data Sources	SIP	BIS	TP
Deutsche Kolonialbibliothek	3	2	32
Online Archives	45	10	36
Philip Baker	0	38	0
Baker/Online Archives	0	47	0
National Library of Australia/Australian War Memorial	0	0	116
Pacific Research Archives, Canberra	23	1	11
National Library of Australia, Canberra	4	2	6
Western Pacific Archives, Auckland	102	70	0
Pacific Manuscript Bureau	324	46	12
Others (e.g. general libraries, Leibniz-Institut für Deutsche Sprache ⁷⁵)	20	8	26
Total	521	224	239

Table 3: Number of sources per archive per variety

4.1.2 Focus during data extraction for reconstructing past stages of pidgins and creoles

Data was extracted from various text types, including travel reports, travel journals, government reports, missionary reports, diaries, letters, trials and proceedings. All sources, independent of whether they were computer-searchable or not, were read page by page. I refrained from searching with the help of keywords as this bears the risk of failing to notice important data. Furthermore, given that early sources were printed in different fonts (e.g. Gothic typeset font), keywords might not lead to results because of OCR errors. Moreover, a search for Pidgin English keywords is problematic as no common orthography existed. The keyword search was avoided to guarantee the targeted accuracy. The sources were looked through in search for language data (4.1.2.1), testimonials about the varieties' similarity (4.1.2.2) and further meta-linguistic information about the contact situation (4.1.2.3).

4.1.2.1 Language data

Emphasis was laid on the extraction of *actual language data*. The language examples were reproduced in the way in which they appeared in the early sources. This implies that no 'corrections' were made, and typos were kept. Sentences which preceded or followed the Pidgin English example were extracted as well, which proved to be helpful for the analysis, interpretation and translation which was undertaken at a later stage.

Every effort was made to further supplement each piece of language data with socio-biographical speaker information, including annotations on the ethnicity, gender, and age of the

⁷⁵ Special thanks to Dr. Doris Stolberg and Prof. Dr. Stefan Engelberg who provided me with some of their early Tok Pisin and Bislama dictionaries which they collected as part of their project *Wortschatz deutschen Ursprungs im Tok Pisin* (<http://lwp.ids-mannheim.de/doc/tokpisin/start>, last access 13 May 2021).

speaker. Even though most early sources will not provide this information, it was added where obtainable.

Next to speaker information, I searched for contextual information about *where* and *when* the sentence was uttered or reported. This is inter alia important for the clear assignment of a Pidgin English example to one of the three areas Solomon Islands, Papua New Guinea or Vanuatu. Some sources could not be clearly identified with respect to the areas under observation, as the source only indicated it to be Melanesian Pidgin English without specifying the concrete area of recording. Those sources were extracted as well but not included in the database for further analysis. The year to which a Pidgin English example refers had to be noted down, since the year of publication does not necessarily reflect the year in which a communicative situation took place. In a small amount of cases the year the language example refers to is directly stated. In most cases, however, no direct reference is made. In these cases, a closer look into the preface and introduction was necessary, as these occasionally provided information about the time in which the author was in the region. However, if no information was found at all, further research about the author was conducted. In addition, I searched for contextual information in the sources, such as for references to historical events, in order to identify the period in which a communicative interaction took place. If no information was available at all, I used the year of publication/documentation as the year of attestation since the speech situation must have taken place at some point prior to the publication/documentation date. The dating of language samples as well as their localisation is quite problematic as the sources varied a great deal in their explicitness on socio-biographical, as well as contextual information. Cases in which all contextual and socio-biographical speaker information is obtainable are an exception. The individual extracted data files were supplemented by the full bibliographical reference, name of the library/archive and the signature of the source.

A problem with the language data excerpts was that language examples did not always represent ‘pure’ Pidgin English. Intermediate forms of English and Pidgin English seem to have existed in the areas, whereby forms closer to Standard English interspersed with Pidgin English forms, as well as Pidgin English forms with borrowings from English could be attested. However, these forms do not necessarily represent lectal differences. It is also imaginable that an originally basilectal sentence gives the impression of being acrolectal due to author and editor modifications. Therefore, we must be aware of the risk of several filters applied to any given PE example. However, as even a single pidgin feature attested in an otherwise acrolectal sentence may give information on the presence of that feature in the Pidgin English variety, seemingly acrolectal sentences with a small number of non-standard features were included as well. The

following two examples show sentences that have been subjected to author or editor modifications:

- (1) *I get this present **where** Miss Deck **bring** him.*

(Manna Kwoi 1909; *Letter by Overi*, SSEM 1909: 37; AU PMB DOC 439)

- (2) *I wonder how many Christians **savey** this lesson **goodfella**?*

(One Pusu 1930; Read 29.01.1930 *Letter to Friends*; PMB 1150)

Though both sentences sound quite acrolectal, they contain some PE features. In sentence (1) *where* as a relative particle and the transitive marker *him* can be observed, whereas sentence (2) contains the verb *savey* ‘know’ and the modifier *fella* attached to the adjective. Moreover, it cannot be excluded per se that, for instance, the demonstrative particle in the sentences is English, as *this* might function as the demonstrative in the Pidgin English variety as well. The mixture of Pidgin English and English-like features may just reflect one variant of the variety how it was spoken at a given time or it may have been anglicised by the author or the editor. Excluding it from the analysis would have risked manipulating data to show solely Pidgin English features.

A note needs to be made regarding the language data collected in personal letters, notes and circular letters of the SSEM for the SIP database. During the database compilation it became obvious that a mixing of Pidgin English and StE features was also present in many (though not all) of the contact language examples extracted from SSEM sources. In some sources StE features were identified only occasionally, whereas others almost seemed as if code switching occurred. As outlined in Section 3.1.5.1, the mission used Pidgin English as the main medium of instruction but, according to Mühlhäusler & Mühlhäusler (2005: 12), developed a ‘simplified’⁷⁶ English by the 1920s that could reduce the amount of PE features it contained. The language examples found in the SSEM sources seem to mirror this.

The question arising out of this was how to treat the language data collected in early SSEM sources. As the language data varied in the amount of PE and StE features it contained, I tried to classify the material among a continuum depending on the amount of PE and StE features. This decision was made for each sentence and not whole texts. Those texts and passages that showed a clear closeness to Pidgin English were included and those in which StE dominated were excluded. Moreover, I focussed on language data which appeared in quotation marks and was portrayed as direct quotes of spoken or written Pidgin English.

There are two major reasons why the SSEM sources were considered to be of importance for this study and why they should not be completely ignored. First, anglicisation represents a

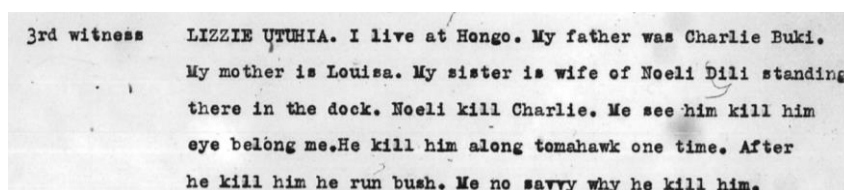
⁷⁶ It should be noted that what exactly is meant by ‘simplified’ is vague and depends on how the concept of linguistic complexity is defined.

challenge that is present in most of the early sources independent of whether the data derives from the SSEM or not. Thus, if all data containing anglicised sources had been ignored, little data would have been left for the analysis. Even if the SSEM sources may contain *anglicised* Pidgin English, the PE features therein may provide information about what PE was like. Another reason why the PE or PE similar language data found in SSEM sources was included was unintentionally provided by Norman Deck who argued that even if the mission “gradually teach[es ...] scholars at the training schools better English, [...] this will have little influence upon the dialect spoken by the average returned labourer” (Deck July 1919: 6; AU PMB MS 1253). Missionary life went beyond the mission stations and depended on the interaction between Christian converts and unconverted Pacific Islanders. Thus, the mission variety could only change provided it was still intelligible enough for intercommunication with the “average returned labourers” that were not exposed to “better English” but had acquired the PE variety. At the same time, it should not be forgotten that when schoolboys returned to their homes, they brought the mission variety back home. Thus, even if Mühlhäusler & Mühlhäusler (2005) argue that the ‘simplified’ English was artificial, the *varieties* were taught and used by the mission and should therefore not be ignored as they might have had an impact on the development of PE in general. Although it is true that the teaching of a ‘simplified’ English may have led to a situation in which some Pacific Islanders shifted from PE to English, as was the case with Sardius, who “almost dropped his pidgin English for English” (*That Bell Again* June 1947: 13; AU PMB MS 1253), a quote like this shows the possibility that the introduction of English may have led to several intermediate varieties of PE. Depending on the degree to which (‘simplified’) English was used and taught on the mission stations, various lects may have emerged, ranging from basilectal to acrolectal PE. Moreover, it is likely that intra-speaker variation existed. Though a careful treatment of the data is required, those sentences that are very close to PE should not be ignored.

As shown in Section 2.2.2.7, the English orthography might prevent the researcher from identifying Pidgin English features, which will only become visible by closer inspection. Sentence (3) indicates such a sentence in which the adverbial construction *I think* is attested which could be interpreted as 1SG and the verb *think*. Only when considering the context of the utterance it becomes clear that *I think* is used as an adverb. If a phonetic spelling would have been used instead (for instance, *ajtɪŋk*) the identification of the feature would have been easier.

(3)	<i>I</i>	<i>think</i>	<i>tomorrow</i>	<i>you</i>	<i>die</i>	
	*1SG	*think	tomorrow	2SG	die -->	‘I think you will die tomorrow.’
	probably		tomorrow	2SG	die -->	‘You might die tomorrow’
- (*The Stories of the Crew*; NIV June 1947: 9; AU PMB DOC 439)

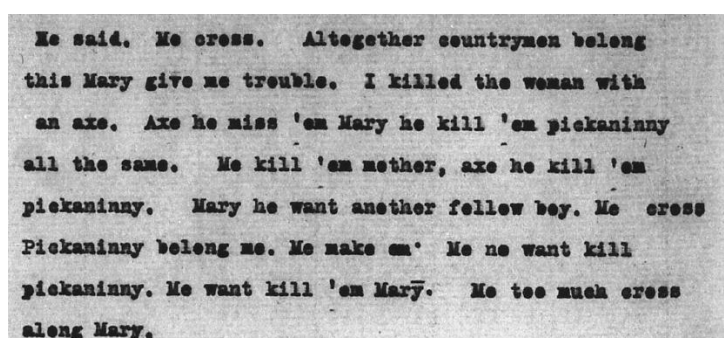
In addition, code switching was attested to be a common phenomenon. Knibbs' reports about a Solomon Islander that "[h]is English was a little weak, and when excited he would lapse altogether into pidgin" (Knibbs 1929: 234, referring to pre-1929). Code switching was also common when court proceedings were written down. Figure 9 shows an extract of the *Forwards copies of Court Record of the trial of Noeli Dili* from 15 July 1920. While the witness starts in English, from the sixth sentence onwards Pidgin English is used.



3rd witness LIZZIE UTUHIA. I live at Hongo. My father was Charlie Buki.
My mother is Louisa. My sister is wife of Noeli Dili standing
there in the dock. Noeli kill Charlie. Me see him kill him
eye belong me. He kill him along tomahawk one time. After
he kill him he run bush. Me no savvy why he kill him.

Figure 9: Extract from Court Record of the trial of Noeli Dili (Tulagi; Utuhia, L. 15.07.1920)

In clear cut cases such as this, I only extracted the sentences which were in Pidgin English. In other sources, it occasionally occurred that a single English sentence appeared in an otherwise Pidgin English witness (cf. Figure 10, sentence "I killed the woman with an axe"). If this was the case, I only extracted the Pidgin English sentences and omitted the English sentence.



He said. Me cress. Altogether countrymen belong
this Mary give me trouble. I killed the woman with
an axe. Axe he miss 'em Mary he kill 'em pickaninny
all the same. Me kill 'em mother, axe he kill 'em
pickaninny. Mary he want another fellow boy. Me cress
Pickaninny belong me. Me make em' Me no want kill
pickaninny. Me want kill 'em Mary. Me tee much cress
along Mary.

Figure 10: Extract from Double murder in the Solomon Islands by Harry Okea (Tulagi; Okea, H. 29.12.1911)

4.1.2.2 Testimonials about the similarities of the varieties

Apart from language data, testimonials about the similarities of the varieties were extracted. Initial descriptions of Pidgin English frequently claim that the variety was "alike in British, French, or other possessions" (Alexander 1927: 213). If the three varieties diverged over time, it is possible that early travellers, residents and missionaries might have noticed a change in the similarities of the varieties and may have reported about it. If such reports were found, they were extracted as well, since the testimonials *might* be an indicator of when the varieties diverged.⁷⁷ The results of the collected testimonials are presented in Chapter 5.

⁷⁷ 'Might' because it is probable that the authors' language capacities influenced their evaluation of the intelligibility between the MPE varieties.

4.1.2.3 Information on the language situation, labour recruitment, and language attitudes

Information on the general *language situation* and the *labour recruitment* was extracted for the documentation of the socio-historical circumstances under which the contact varieties emerged and developed. As pointed out, external circumstances may have led to the development and/or manifestation of individual linguistic features in the three Melanesian Pidgin English varieties. *Language attitudes* were extracted, since positive or negative attitudes towards languages may have advanced or hampered the development of Pidgin English. They may further contribute in evaluating how trustworthy a Pidgin English example is (see Section 2.2.2.3).

4.2 Database design: The three data collections

The present section refers to the extracted *language samples* only, as these provide the basis for the analysis in Chapters 6-9. As pointed out in Section 2.2, it is important to remember that historical linguistic data consists of written records. The design of a database consisting of historical archival data will therefore by necessity differ from a contemporary data corpus.

The data collection consists of data from 984 different text files and is subdivided into a Tok Pisin, a Bislama and a Solomon Islands Pijin subset. Data attested for British New Guinea and for Bougainville and Buka was collected in two separate files. Bougainville and Buka have a special role as these areas came under German protection in 1886 and remained politically a part of New Guinea ever since, even though they geographically belong to the Solomon Islands. The data collected for Bougainville and Buka, however, was too small to be used as an individual subset in the comparative analysis. The areas provided labourers for Fiji, Queensland as well as Samoan plantations until the islands were annexed by the Germans in 1886 (cf. Mühlhäusler 1985a: 46). The annexation of islands usually served as a means for the European powers to “exclude or limit participation of outsiders in local labour recruiting” and to monopolise their labour supplies (Munro & Firth 1990: 92). Consequently, labourers from Bougainville and Buka “may still be recruited there only for German plantations” (1990: 94). In addition, the Germans had a competitive advantage against other European powers in recruiting labourers from Bougainville and Buka, because the British government had prohibited supplying arms and ammunition to Pacific Islanders. This was, however, the preferred payment for the islanders (1990: 94). As Bougainville and Buka were treated as part of the Bismarck Archipelago and were under the German imperial government and as the islanders were recruited to work on German plantations in German New Guinea and German Samoa, the Pidgin English on Bougainville and Buka is more likely to resemble the PE of German New Guinea. Therefore, I decided to include

the few sources that could be obtained for these areas into the TP subset. Nonetheless, it is likely that regional differences existed.⁷⁸

British New Guinea, in contrast, was planned to be treated separately as the south-eastern parts of Papua New Guinea have a completely different history than the north-eastern parts and the Bismarck Archipelago (cf. Chapter 3.1.4.3.2). This is likely to have led to differences in the pidgin varieties used in the areas. The data collected for British New Guinea was rather small which may have to do with the fact that the English-lexified contact variety was not widely spread in the area (cf. 3.1.4.3.2). Thus, the data was not included or further considered in the present study.

The size of the complete data amounts to about 154,648 Pidgin English words (excluding the number of words of three dictionaries). The number of words is unequally distributed among the three varieties. While the TP data consists of 66,714 words (= 43%) collected from 239 sources (the word counts of three dictionaries not included), the SIP data consists of 68,220 words (= 44%) collected from 521 sources and the BIS data consists of only 19,749 words (= 13%) extracted from 224 sources. For the purpose of word counts, based on *Microsoft Word*, a word was defined as an uninterrupted string of characters delimited by punctuation or white space.

Word counts, however, cannot be considered an appropriate tool for comparing the amount of data included for each variety. As a standardised spelling system did not exist, a single word or even a single morpheme was spelled in many different ways. For instance, the demonstrative pronoun *this fellow* is found with 24 different spelling variants in the data collection, and *one time* has 12 spelling variants (cf. Table 4).

standard form	orthographic realisations
this fellow	<i>despela; dis fala; dis fela; dis fella; dis feller; dis fellow; dis pala; dis pela; dis pella; disfala; disfela; disfele; disfelo; disfēla; disfēla; dispala; dispela; this fella; this feller; this fellow; this pella; this-fella; thisfella; tis pella</i>
one time	<i>one time; vantaime; onetaime; ontaine; ont aim; vantaem; vantaum; wantaim; wantaem; wan taim; wontaim; wəntajm</i>

Table 4: Orthographic realisations of *this fellow* and *one time*

Depending on the spelling variant used, words were counted as one or two words following the word count conventions in *Microsoft Word*. The diversity of spelling was attested in all three varieties. Thus, spelling variants will have an impact on the total amount of words and do not provide useful information about the dataset's sizes and their comparability. Following this, the

⁷⁸ The *Melanesian Pidgin English language guide* for soldiers, for example, describes the PE as used in the eastern part of New Guinea, in New Britain and Bougainville and Buka with the hint for the learners that “differences that are important in the usage of the Solomons are noted at the end” (Army Education Branch 1944: 2).

word counts can only, if at all, be taken as a rough indicator of the amount of data present in this work.

Comparing the number of texts included per genre is inadequate given that the amount of language data found in the individual sources varied a great deal from single words to phrases to large texts so that the numbers of texts per genre do not provide a useful information. Instead, the number of tokens found per year per variety for the investigated features will be stated in the individual analysis chapters with the help of a dot plot.

Nonetheless, some general tendencies of the overall dataset have to be outlined. It became clear during the database compilation that the data is distributed unequally among the three varieties and among the different time periods. While some time periods are proportionally underrepresented, others are overrepresented. For Solomon Islands Pijin and Tok Pisin, for instance, the majority of data was found for the first half of the 20th century, while the dataset of Bislama contains almost no attestations from after 1940, but, in contrast, has more pre-1900 attestations available in comparison to the other two varieties.

Another issue in terms of representativity of the datasets is that there is a social imbalance between the material which derives from European writers compared to that of non-European Pacific Island writers. The language samples are also unequally distributed among the text types from which they were extracted. Not all text types can be found throughout the complete period under investigation and one or the other may be dominant. For instance, missionary writings dominate the Solomon Islands Pijin dataset in the first half of the 20th century which might lead to a one-sided portrayal of the variety. For the Tok Pisin data, for example, the war pamphlets make up a large amount of input of the database.

All this raises the question of the representativity of the data collection as the individual datasets differ in their size, the material available per year, and, as will be seen in Chapters 6-9, in the material available per investigated feature. These are factors which need to be considered during the analysis and especially during the interpretation of the results. The datasets are far from ideal, but they are the best that linguists can obtain in order to investigate and test the assumptions made about the origin and development from Melanesian Pidgin English to the individual varieties Solomon Islands Pijin, Bislama and Tok Pisin.

4.3 Methodological concerns and analytical procedure

As pointed out in the introductory chapters, previous studies focussing on the similarities and differences between the three MPE varieties principally look at differences that are prevalent in the pidgins today. For several reasons this study refrains from starting with the known differences

existing between the varieties. For one thing, differences attested today need not necessarily have developed in the 19th and early 20th century but may have originated at some remote period. For another thing, different variations of a given variable may be attested in the early data but might no longer be present in the three pidgins. As Tryon & Charpentier point out, “regionalisms have been largely neutralised today with the much more extensive use of the pidgins on radio, and latterly on television” (2004: 350). Thus, to work backwards and start with the known differences that exist between TP, BIS and SIP today, and to match early attestations to the like, might obscure differences that existed in prior developmental stages of the varieties. In addition, there is no guarantee that the early data will contain attestations of a feature that is selected based on the varieties’ present-day language structures.

This study therefore takes a data-driven approach in order to discern variables and their variants for the present study. This section describes the data retrieval and coding process as well as the analytical procedure. The following figure gives an overview of the individual steps that will be discussed in the following sections.

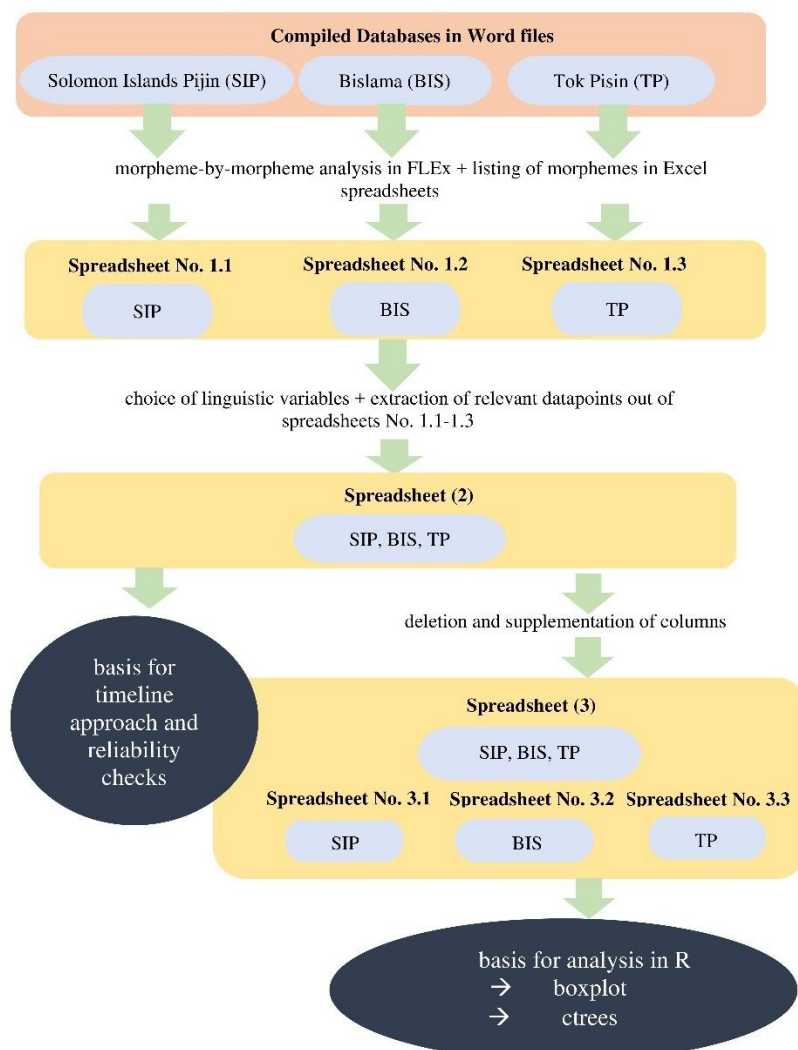


Figure 11: Data retrieval & analytical steps

4.3.1 Identification of potential differences and preparation of the spreadsheets

During the database compilation, the language data was collected in *Word* files in its original immediate literary context. In a first approach to identify potential differences and similarities between the varieties, a morpheme-by-morpheme analysis was conducted.⁷⁹ I followed the principles set up in the *Leipzig Glossing Rules* (<http://www.eva.mpg.de/lingua/resources/glossing-rules.php>, last access 29 September 2021). For the interlinear glossing, the language examples were integrated into *Fieldworks Language Explorer* (FLEX), a tool developed by SIL International for linguistic fieldwork (<https://software.sil.org/fieldworks/>, last access 29 September 2021). The tool provides software for the analysis of linguistic data and allows interlinearisation, glossing and morphological analysis. I supplemented the data entries with information on the date of attestation, speaker, place and source.

Simultaneously, three *Excel* spreadsheets, one for each pidgin, were prepared. The individual morphemes and their attestations were entered in the respective spreadsheet. All morphemes (independent of whether the same or not as in StE) were integrated. This differed from the more common approach to solely focus on features that deviate from StE,⁸⁰ since the procedure implicitly assumes that the language structures of the lexifier of today are the same as those being used at the time of contact. It is to be expected, however, that several varieties of the lexifier language were used in the early contact situations and “that they did not speak the present-day varieties of the input languages” (Velupillai 2015: 139). Moreover, even StE features may have been used as variants of a variable and should thus be integrated into the analysis.

The resulting spreadsheets, as indicated in Figure 12, were coded for the following information: The column ID lists a unique numeric id for each unique row in the data frame and the column VARIETY serves to clearly distinguish between the three varieties when the three datasets are compared.

Next to the column TOKEN, in which the morpheme was listed as attested in the early source, a column ST_FORM was created, in which the attested morpheme was listed with a standardised English spelling. The choice of using an English orthography for the rendering of the standardised form is solely for the purpose of convenience.

⁷⁹ A morpheme-by-morpheme analysis also proved helpful, since it was not possible to make use of search programs to search for features and forms in the collected datasets. Since the varieties are characterised by multifunctionality in that forms can fulfil different functions and due to various spelling variants, the morpheme-by-morpheme analysis turned into a prerequisite.

⁸⁰ Cf., for instance, Baker (1987, 1993); Baker & Huber (2001).

Further columns added were GLOSS in which the gloss abbreviation was written and TRANSLATION providing the translation into English. As the grammatical context in which a form is attested may also be of importance, the column STRUCTURE was added, and the column SENTENCE lists the complete sentence from which a morpheme was extracted.

The data was further coded for *source-based information*, such as the publication date (YEAR_PUBL), the exact reference (SOURCE), and the title of the source (TXT_TITLE) from which the data was extracted. Author based information such as the name (AUTH_NAME), origin (AUTH_ORIGIN), and the authors role in the colony (AUTH_ROLE) were supplemented. In addition, it was decided to add a column called TXT_TYPE which will be defined in more detail in Section 4.3.5.

Furthermore, *contextual information* was encoded in the spreadsheet. The column YEAR_ATT provides information about the year to which an utterance refers. Entries varied from concrete dates to specific years to time periods. Since statistical programs such as R Studio require a single date per data point, a concrete year was determined for each of the datapoints as well (= variable YEAR_DET). If periods were documented under YEAR_ATT, the year in the middle of the period was chosen as YEAR_DET.

The coding of geographical information, in order to provide *contextual information about where a sentence/word was uttered*, was challenging. Place names mentioned in the early sources do not necessarily exist any longer. Moreover, places in different island groups sound similar (cf., for instance, *Tongoa* in Vanuatu and *Tongona* in the Solomons) and various ways of spelling place names were attested. To identify whether a place belonged to the Solomon Islands, Vanuatu or Papua New Guinea, I searched for the place names via <https://cartographic.info/names/> (last access 29 September 2021). Each of the areas consists of several islands which form island groups, which then again can be clustered into provinces. Therefore, a fine-grained system of place variables was required. The variable LOCATION does not refer to the geographical place but rather the room or space a person is in. For instance, it includes information such as ‘school’, ‘plantation’, ‘court’, ‘harbour’, or ‘sea’. The variable PLACE indicates the village or city name in which the language example was used. Under REGION/ISLAND the name of the island/mainland is listed, and PROVINCE matches the islands to their provinces.⁸¹ COORDINATE_1 and COORDINATE_2 were added to be able to link datapoints in the database with geographical coordinates in a dynamic, interactive database.

⁸¹ There are 19 provinces distinguished in Papua New Guinea, nine provinces in the Solomon Islands, and six provinces in Vanuatu.

								SOURCE-BASED INFORMATION				AUTHOR-BASED INFORMATION		
ID	VARIETY	TOKEN	ST_FORM	GLOSS	TRANSLATION	STRUCTURE	SENTENCE	YEAR_PUBL	SOURCE	TXT_TITLE	TXT_TYPE	AUTH_NAME	AUTH_ORIGIN	AUTH_ROLE
85	tp	rauss	raus	throw.out	threw (her) out	rauss O/TR	me cross' long woman, me rauss him!	1907	Stephan & Graebner 1907: 29	Neu-Mecklenburg (Bismarck Archipel): Die Küste von Umuddu bis Kap St. Georg	expedition / travel report	Stephan, Emil & Graebner, Fritz	German	doctor ethnologist
86	tp	rauss	raus	throw.out	throw (her) out	rauss O/TR	he no work, me rauss him	1907	Stephan & Graebner 1907: 110	Neu-Mecklenburg (Bismarck Archipel): Die Küste von Umuddu bis Kap St. Georg	expedition / travel report	Stephan, Emil & Graebner, Fritz	German	doctor ethnologist
87	tp	run him	run	chase	chases	run 3SG/TR	Woman he look him, he run him, he kitch this fellow man, he speak him puss-puss; oh he puss-puss plenty!	1907	Stephan & Graebner 1907: 123	Neu-Mecklenburg (Bismarck Archipel): Die Küste von Umuddu bis Kap St. Georg	expedition / travel report	Stephan, Emil & Graebner, Fritz	German	doctor ethnologist
88	tp	sabe	save	know	know	sabe	Suppose me kitch him grass, he m die, he stink, suppose you kitch him, he allright: You sabe too much, kiab!	1907	Stephan & Graebner 1907: 25	Neu-Mecklenburg (Bismarck Archipel): Die Küste von Umuddu bis Kap St. Georg	expedition / travel report	Stephan, Emil & Graebner, Fritz	German	doctor ethnologist
89	tp	sabe	save	know	knows	sabe	Suppose this fellow man he sabe , he die, all finish	1907	Stephan & Graebner 1907: 25	Neu-Mecklenburg (Bismarck Archipel): Die Küste von Umuddu bis Kap St. Georg	expedition / travel report	Stephan, Emil & Graebner, Fritz	German	doctor ethnologist
90	tp	sileep	sleep	sleep	sleep	sileep	Woman he hear him, bäl belong him he mowe, he no sileep , he come, he puss-puss belong this fellow.	1907	Stephan & Graebner 1907: 123-124	Neu-Mecklenburg (Bismarck Archipel): Die Küste von Umuddu bis Kap St. Georg	expedition / travel report	Stephan, Emil & Graebner, Fritz	German	doctor ethnologist

CONTEXTUAL INFORMATION								SOCIO-BIOGRAPHICAL SPEAKER INFORMATION					
YEAR_ATT	YEAR_DET	LOCATION	PLACE	REGION/ISLAND	PROVINCE	COORDINATE_1	COORDINATE_2	SP_NAME	SP_GENDER	SP_ROLE	SP_AGE	SP_ORIGIN	
1904	1904	NA	Vunapo	New Britain	East New Britain Province	- 4.35	152.28	NA	NA	NA	NA	Lamassa	
1904	1904	NA	Lamassa	New Ireland	New Ireland Province	- 4.68	152.77	NA	m	NA	NA	Lamassa	
1904	1904	NA	Lamassa	New Ireland	New Ireland Province	- 4.68	152.77	NA	m	NA	NA	NA	
1904	1904	NA	Lamassa	New Ireland	New Ireland Province	- 4.68	152.77	NA	m	NA	NA	Lamassa	
1904	1904	NA	Lamassa	New Ireland	New Ireland Province	- 4.68	152.77	NA	m	NA	NA	Lamassa	
1904	1904	NA	Lamassa	New Ireland	New Ireland Province	- 4.68	152.77	NA	m	NA	NA	NA	

Figure 12: Extract from spreadsheet No. 1.3

Since the author of a source is not necessarily the speaker of a PE utterance, the datapoints were coded for socio-biographical speaker information, such as the speaker's name (SP_NAME), gender (SP_GENDER), role (SP_ROLE), age (SP_AGE) and origin (SP_ORIGIN). Unfortunately, the social variables were only available for a limited amount of language samples.

4.3.2 Choice and definition of linguistic variables for the analysis

Based on the morpheme-by-morpheme analysis and the data filtering in Excel, features from four linguistic domains, namely *demonstratives*, *relative clauses*, *modality* and *prepositions*, were selected for further analysis. The selection of features was based on the following criteria:

- The varieties had to show differences in the choice of forms used to express the feature **AND**
- The feature had to be attested *frequently* **OR**
- The feature had to be disregarded in previous studies.

With these criteria in mind, this study focusses only on those features of a linguistic domain in which differences were attested. For instance, since the varieties did not show any differences in their choice of forms used to express obligation, the feature will not be included into the analysis. Table 5 shows the selection of linguistic features that will be analysed in Chapters 6-9.

Due to the fact that the features investigated raise different questions in their analyses, the analysis chapters will contain a brief section explaining the methodological approach for each feature. Due to the number of different variants that can be used to express a certain feature, the variants will also be introduced in the analysis chapters.

Linguistic domain	Feature	Coded
demonstratives	adnominal demonstrative pronouns	dem_adn
	pronominal demonstrative pronouns	dem_pron
relative clauses	subject relative clauses	rel_sub
	object relative clauses	rel_obj
modality	abilitative	mod_abil mod_inabil
	volition	mod_volit mod_negvolit
	permission	mod_perm mod_prohib
	speculation	mod_spec
prepositions	comitative	prep_com
	instrumental	prep_inst
	terminative	prep_term
	adessive	prep_ades

Table 5: Linguistic variables

4.3.3 The resulting data spreadsheets

Two spreadsheets provided the basis for the final analysis of the features. Table 6 lists and summarises the variables they include. Spreadsheet (2) served as the basis for the timeline approach and for the data reliability checks. Spreadsheet (3), by contrast, only contains those columns which are necessary to perform a statistical analysis in R Studio.

The resulting data spreadsheet (2) contains 5,104 tokens, while the spreadsheet (3) contains 4,972 tokens due to ambiguous datapoints which could not be included into the analysis in R. The analysis chapters will provide an overview of the number of attestations for a given feature and will outline how it was dealt with ambiguous datapoints. They will also provide a closer definition of the variants of each feature (= ST_FORMs). If, according to the feature, an additional variable had to be added, this will be outlined in the chapters as well.

Spreadsheet (2)	Spreadsheet (3)
ID	ID
VARIETY	(VARIETY)
FEATURE	FEATURE
TOKEN	-----
ST_FORM	ST_FORM
STRUCTURE	STRUCTURE
TRANSLATION	-----
GLOSS	-----
SENTENCE	-----
SOURCE	-----
TXT_TYPE	TXT_TYPE
TXT_TITLE	TXT_TITLE
YEAR_PUBL	-----
AUTH_NAME	AUTH_NAME
AUTH_ORIGIN	-----
AUTH_ROLE	-----
LOCATION	-----
PLACE	-----
REGION	-----
AREA OF REGION	-----
COORDINATE_1	-----
COORDINATE_2	-----
SP_NAME	-----
SP_GENDER	-----
SP_ROLE	-----
SP_AGE	-----
SP_ORIGIN	-----
YEAR_ATT	YEAR_DET

Table 6: Variables in spreadsheets

4.3.4 Analysis of spreadsheet in terms of diachronic variation and change

The spreadsheets served as the basis for the analysis of language change. The process of language change is usually described as occurring in three stages. In the initial stage a new feature (= innovation) is introduced in a society. As the newly introduced feature spreads at different times

among different people, its usage will still be restricted to a small part of the society, leading to the coexistence and co-usage of “old and new” forms in the second stage. The third stage is assumed to be achieved when “the older feature falls out of usage” (Wright 2012: 62).

This linear development of three stages cannot be readily applied to the development of TP, BIS and SIP from MPE, nor to any other variety. Different scenarios are possible during the process of their *regionalisation*, or in other words, during the development of individual features in the three regions. It is likely that in early MPE various forms to express a feature co-existed and that the process of localisation was paralleled by the choice of one or a few of these forms through dialect levelling. It is also imaginable that completely new forms were introduced in one, two or even all the varieties, either through borrowings from the local languages or through grammaticalisation processes. Depending on whether completely new forms for a feature are introduced or whether a process of dialect levelling takes place, two or three stages may be observed.

Independent of whether new forms are introduced, or a process of dialect levelling takes place, it is a matter of debate when we can speak of a case of language change. As Nevalainen & Raumolin-Brunberg (2017: 56) claim:

Can we speak of language change, although these processes have not been completed? If we can, [...] what then is the limit, how much linguistic and nonlinguistic variation is allowed? If we look at the question from the angle of the speech community, how large a percentage of speakers should have adopted the innovation or used it as their main variant before we can say that language change has taken place?

The present study looks at the attestations of individual features to determine “when the variation leading to a shift begins” (2017: 56). It is assumed that the innovations introduced into a language, as well as the decision for one or several of the possible, previously co-existing forms, are important indicators of linguistic change and thus indicate at least the *beginning of linguistic change*.

The following sections outline the methodological framework used in the present study to identify the beginnings of language change in the three varieties by applying qualitative (Section 4.3.4.1 and Section 4.3.4.2) as well as quantitative analytical tools to the data (Section 4.3.4.3).

4.3.4.1 Dates of attestation in analysing the chronology of language change

Since the aim of the study is to investigate the chronology of language change in TP, SIP and BIS, dating the early language examples is essential. First attestations and changes in frequency of occurrence are assumed to be important indicators of linguistic change. Especially in creolistic studies, the attestation and non-attestation of features has been used to reconstruct past stages of contact varieties and to analyse their historical relationships (cf., for instance, Clark 1979; Baker

1993, 1995; Roberts 1998, 2005; Baker & Huber 2001). Nonetheless, the provisional nature of first attestations needs to be clearly stressed. Section 4.1.2.1 already pointed to the problem of identifying concrete dates of attestation for the early language examples.

Even if dates of attestation have been determined with greatest accuracy, their explanatory power needs to be addressed. As highlighted in Chapter 2, studies investigating past stages of language development are frequently restricted to written sources only. Hence, the first problem that arises is that dates of attestation can only demonstrate that “certain innovations happen to turn up in writing for the first time at a certain date” (Adams 2013: 25), while it remains unclear how long they were used in speech before being documented in the written record. As Alinei (2004: 5) states, the written evidence “cannot represent a term a quo (‘from which’), i.e. the real ‘beginning’. At most, it represents, [...] a term ante quem (‘before which’).” Thus, it is reasonable to assume a time lag between the innovation of a feature in speech and its first written attestation. It is not possible to determine the concrete length of a time lag. As Alinei states:

[i]t certainly cannot be argued that the earliest written evidence of a language, dated to moment X, proves that it existed only, for example, half a century, or 1, or 3, or 5, or 10 centuries before X, in compliance with some statistical law or the supposed molecular clock of organic change. It may be even supposed that it existed whole millennia before its first attestation.

Though a concrete length cannot be determined, there are certain factors that will have an impact on the duration of a possible time lag (see Figure 13).

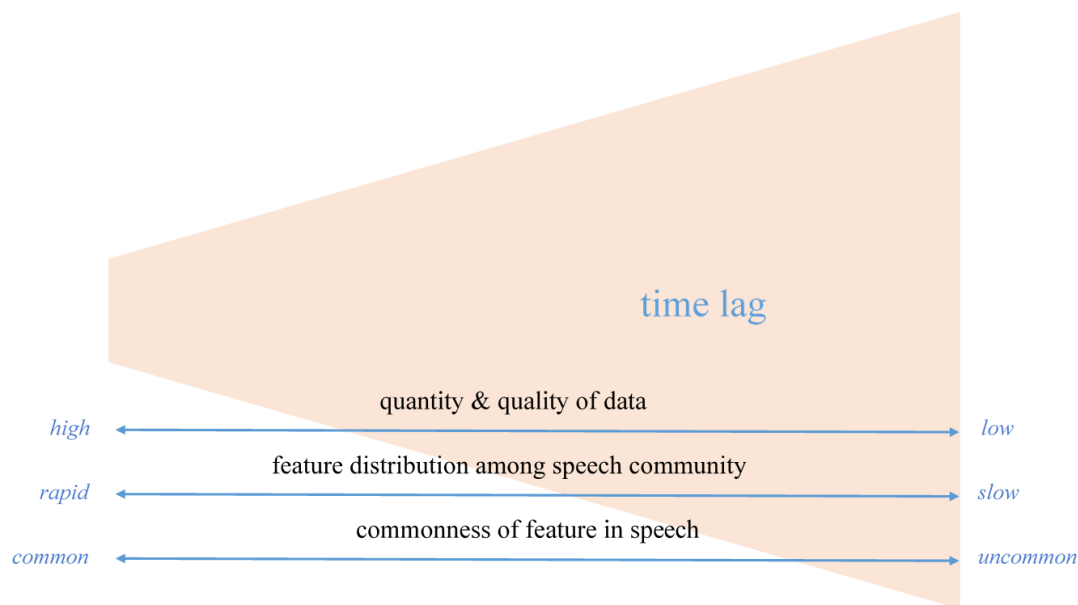


Figure 13: Time lag influencing factors

The most important factor is the quantity and quality of the available sources which will have an impact on the (non-) attestation of features. It is relatively common that a greater amount of texts is available the closer one moves to the present. The later texts are frequently more extensive and complex and therefore able to better represent languages structures. For instance, most of the data

collected for this study covering the second half of the 19th century consists of short sentences. Only the post-1900 data consists of longer texts which are therefore better able to reflect the linguistic complexities of the varieties. Thus, the availability and quality of the sources may distort the data. For instance, in Solomon Islands Pijin the period from its beginnings to 1900 is only scarcely attested in comparison to the years 1900-1950. Thus, I expect that the time lag is greatest in the first period, which then shortens as the database increases in its representation of the linguistic variety. For Bislama, data for the period 1940-1950 was almost not available. Therefore, the non-attestation of features in this period need to be treated carefully. For Tok Pisin, again another structure of the dataset can be observed with a good documentation of the years 1935-1945. A greater time lag can be expected in the pre-1900 years.

A time lag may be further influenced by how a feature spread and was distributed among the speech community. Innovations are frequently described as occurring in form of an S-curve, predicting a “slow spread in the initial stages of a change, followed by a period of (rapid) acceleration in which the change catches hold, and finally the trajectory of the diffusion slows down before its completed” (Deumert 2004: 216). Thus, a feature may remain unobserved for a time in natural speech, before it turns into a widely available option. If a feature spread only slowly and thus was not widely distributed among the speech community, it limits the chance that the feature was documented in the written record and thus a greater time lag can be expected.

The features that are investigated might have an impact on the time lag as well. Certain language structures will be more commonly used in speech than others. If the feature under investigation is a common feature, the time lag for the feature can be considered to be shorter than for a feature which is less commonly used in speech. Moreover, features might be used only by a small part of a society as they are restricted to certain domains. Therefore, dates of attestation will have a higher validity if the databases contain language examples from various domains.

It needs to be pointed out again that just because a feature is absent in a variety this does not necessarily mean that the feature did not exist. It is possible that the innovation existed but had simply not been recorded. This may be especially true for earlier time periods, for which only a limited amount of language data of the varieties is available. Moreover, it is likely that further historical records are still hidden in the archives which might alter the analysis results.

Taking the above factors into consideration, one might be tempted to avoid the use of historical attestations (and their *years* of attestation) to reconstruct a language’s origin and development. In fact, if the study aimed at showing when MPE really was developing into its individual varieties, the study would surely fail. However, even though the language change cannot be linked to *concrete* dates, it can be an indicator for how the varieties were *likely*

developing based on the surviving historical records. For this, it is important to have a thorough knowledge of the composition of the databases. The composition will shed light on the explanatory power of the sources and on possible biases. To conclude with the words of Adams, “the evidence is often sufficient to reveal, if not the dates of a change (a concept of little meaning anyway, because [...] it is usually impossible to pin down the start of a development), at least its stages [...] providing a relative chronology” (2013: 26). The use of dates of attestations is the only way to discover possible developmental stages which otherwise remain unknown.

4.3.4.2 Timeline and boxplot approach

The historical attestations of the selected features and their dates of attestations were visualised in two different manners. In a first step, non-frequency-based *timelines* were created in Excel as exemplified in Figure 14. Based on all attestations obtained for a feature (FEATURE), it was checked for each year (YEAR_ATT), whether a form (ST_FORM) was present or absent. Each dot in the timeline indicates that the feature under observation was attested, while the x-axis indicates when it was attested. The colour of a data point clearly assigns it to a specific form. The frequencies of a form per year were not taken into consideration. For example, in Figure 14 we can observe that the relative particle *that* was attested in 1929, however it is not indicated how frequently it was attested.

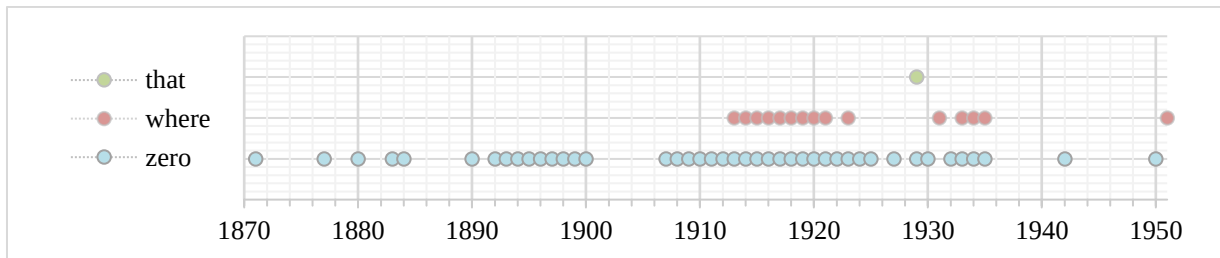


Figure 14: Timeline of RC ST_FORM variants in BIS

For the allocation, the finest temporal resolution (YEAR_ATT) was carried out. The entries for the variable YEAR_ATT vary from concrete dates to specific years to time periods. If a form was attested within a period, it was visualised as if the form was attested in each year of the period. This might be misleading in that a feature might seem more common than it was. However, the procedure avoids dedicating a specific year to the attestation, which might be misleading as well.

The timeline approach will give a first impression about the years in which the different variants of a feature were attested in the three varieties. As pointed out, the timeline does not consider frequencies but solely shows *if* and *in which years which variants* were attested. Frequencies may be important indicators in terms of when a feature became the most dominant form. However, due to the quantity and quality of the data, and due to the fact that linguistic

innovations only slowly spread in the initial stages, the presence or absence of a feature may be more conclusive than the frequencies. A clear disadvantage of the described approach is that it obscures the unequal data distribution, a knowledge of which is needed for the interpretation of the results.

Thus, in a second step, boxplots were created for each feature in each variety. The boxplots were made with the statistical language R (R Core Team 2020) in RStudio (RStudio 2019). A boxplot visualises the distribution of the different forms used to realise a feature (according to the variable `ST_FORM`) across time (`YEAR_DET`). As the example in Figure 15 indicates, it consists of boxes, vertical lines, horizontal lines, and dots.

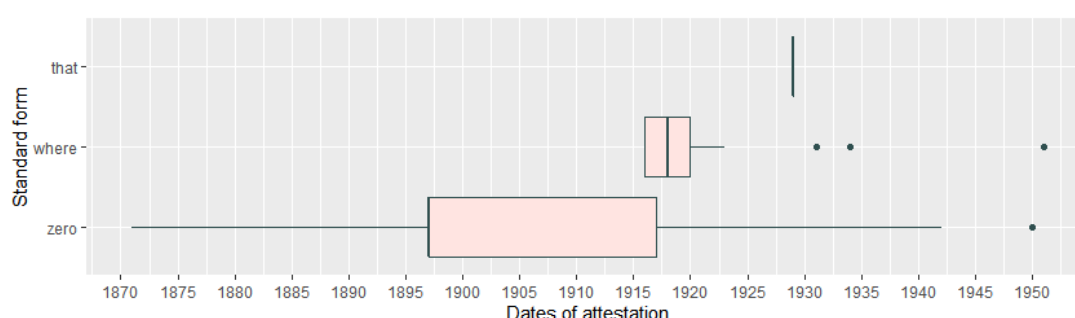


Figure 15: Boxplot of RC `ST_FORM` variants in BIS

The box portion of the boxplot indicates the area that contains the data from the 25th percentile to the 75th percentile and is also referred to as the inter-quartile range. The vertical lines indicate the median of the data (= the 50th percentile). The horizontal lines that go out from the boxes are called whiskers and represent scores outside of the middle 50% and finally, the individual dots represent outlier points, which are datapoints that behave completely different from the remaining data (cf. Keen 2010: 102). The boxplot is also called box-and-whisker plot and has the advantage that it highlights the data distribution and indicates outliers.

As the creation of boxplots in R Studio requires definite dates, the variable `YEAR_DET` had to be used. As a reminder, if a timespan was ascertained for the variable `YEAR_ATT`, the variable `YEAR_DET` will show the mean of the timespan. The information provided under `YEAR_DET` is less certain than the information provided under `YEAR_ATT`, but is necessary, to make use of the time variable in R Studio.

In sum, a boxplot provides more information about the data distribution than the timeline approach. The boxes will indicate in which time period the majority of datapoints of a form were attested. The boxes, however, do not inform about frequencies and do not indicate that a certain form dominated. This has to be kept in mind as the visualisation might otherwise be misleading. The timeline has the benefit of not being restricted to a definite date. Therefore, both approaches

will be employed comparatively on the data. The differences between the approaches are summarised in Table 7.

Excel Timeline	Boxplot (R Studio)
→ based on YEAR_ATT	→ based on YEAR_DET
assigns either a definite or broader period (= thus, more certain date) to language samples	assigns a definite but in many cases uncertain date to language samples
ignores information on frequencies of forms	includes information on frequencies of forms but only regarding distribution of datapoints
obscures data distribution	highlights data distribution, showing outliers
clear visualisation of years in which forms were attested	sizes of boxes can be misleading and should not be equalised with the dominance of a form

Table 7: Contrasting juxtaposition of timeline vs. boxplot approach

4.3.4.3 Conditional inference trees

As the timeline and boxplot approach do not provide any degree of statistical rigor, the data was additionally tested for possible patterns with the help of *conditional inference trees* (*ctrees*) which are part of the *party* and *partykit* package in R (Hothorn, Hornik, & Zeileis 2006). The latter can be considered to be more stable since “the default in *party* is a maximum-type test statistic on the multidimensional test statistic when computing splits”, whereas *partykit* “employs a quadratic test statistic by default, because it was found to produce better splits empirically”, and thus, computes with log-p-values (Hothorn, Hornik, & Zeileis: 2015: 22). Thus, the tool *partykit* was applied where possible.⁸²

Although it is common in historical linguistics to work with *year-by-year* approaches or *predefined historical periods*, these methods were rejected. A year-by-year approach, which considers the frequencies of a variable and its variants per year, is not suitable for the present study since the individual years are too scarcely populated and since the data is unequally distributed among the years. The application of predefined historical periods was rejected as the “periodization into equidistant time periods may sometimes be misleading” (Gries & Hilpert 2012: 135). As Gries & Hilpert (2012: 136) have demonstrated, the decisions made regarding the number of periods as well as their length can lead to different results that may have an impact on the interpretation. Nonetheless, they state that “generalization by means of identifying stages constitutes the first step toward explaining a phenomenon” (2012: 136). They therefore suggest

⁸² Though according to the authors, the trees should be similar independent of whether *party* or *partykit* is used, this is not the case with my data. In the old implementation it could happen that from several highly significant variables, always the first is chosen because the p-values were essentially indistinguishable for the computer.

the application of a variability-based neighbour clustering (VNC) method to guarantee periods that are “data-driven and phenomenon-specific” (2012: 137) and, as such, objectively constructed. The VNC method was developed to cluster along one dimension for one or multiple features (cf. Gries & Hilpert 2008, 2012). Due to the complexity of the present study, involving three different varieties with several features, each of which showing several forms for how a feature can be realised, plus having measurements such as text type and author, the VNC algorithm was not applicable to the data. Instead, conditional inference trees proved to be most adequate for the analysis in the present study.

Conditional inference trees use recursive binary partitioning to create a visual tree structure, identifying “decision rules among the independent variables that best predict particular outcomes of the dependent variable” (Díaz-Campos & Dickinson 2019: 208). In the resulting tree structure, the outcome categories that are predicted are displayed on the bottom. A tree only displays splits if the data which are separated by a split are significantly different from one another. The tree structure will furthermore display the p-value of the split showing how confident one can be about a split.

The main advantage of using ctrees is that the algorithm allows a data-driven approach in terms of identifying whether time can be considered a predictable factor for the choice of the form. Ctrees is used to analyse whether the individual varieties show relevant temporal points in the usage of a specific form. In addition, ctrees allows for single as well as multiple predictors to be tested. Thus, next to the time variable, factors such as the impact of the author or text type can be analysed as well, which will help to evaluate the suitability of the early documents in regard to the research question. Ctrees can be applied for sparse or imbalanced data meaning “‘small n large p’, where n is the number of observations and p is the number of predictors” (Levshina 2015: 292).

For the analysis, the determined year of attestation (YEAR_DET) serves as the continuous predictor. Since features might have stabilised at different timepoints, each feature is analysed separately. In addition, the features and their developments are analysed in a variety-specific, non-comparative manner for a start, since the features in the varieties might have stabilised at different points in time. At the end of each feature chapter, a ctree based on all three varieties will be created to investigate whether two varieties are closer or further away from each other regarding the feature under investigation.⁸³

⁸³ The datapoints in the present study are too uneven and imbalanced for neighbour net phylogenetic methods. Since neighbour nets require stable datapoints to produce reliable results, running this very uneven data with neighbour nets is not appropriate.

4.3.5 Strategies in dealing with the reliability of sources

Although the focus on historical sources is not without its hazards, they remain the most accurate account of the spoken pidgin of that time. Section 2.2.2 referred to important aspects that need to be taken into consideration when collecting early language data on pidgin and creole varieties. Special focus was placed on the question of the reliability of historical written sources. In the following, I will outline how these factors were considered in the present study. Although the strategies listed do not warrant the identification of all unreliable texts (cf. Baker & Winer 1999: 105), they reduced the amount of them and helped to critically evaluate anomalies discovered in the data analysis.

I started with the factor of *authenticity* (see Section 2.2.2.1). As plagiarism was a common practice in early sources, it was necessary to ensure that only the original language examples were included into the database. Nonetheless, all language samples, independent of whether they were duplicates or not, had to initially be collected in order to identify who the original author of the language sample was and to which original year the sample refers. If a duplicate was already discovered during the compilation of the general databases, a note was made so that these could be removed in a second step. Next to those identified during the extraction of the data, further duplicates were found with the help of the filter function in Excel during the morpheme-by-morpheme analysis and were consequently deleted. Thus, it was ensured that duplicates were not included into the analysis.

Baker & Winer (1999: 105) argue that internal as well as external checks are necessary to prove the reliability of the early sources. The internal checks prove the consistency of features “within a particular publication or manuscript” by comparing the features found with those attested in earlier and later texts. Due to the number of sources which provide the basis for the present study, it is assumed that the data itself will reveal “phonological, lexical or grammatical anomalies” (Baker & Winer 1999: 105) for which explanations must be found. Those anomalies do not necessarily mean that a text is unreliable, but they “help[...] to identify those features for which explanations should be sought when making external checks” (1999: 105). External checks, according to the authors, “are those made on the experience, competence, attitudes and motivation of their authors” (1999: 105). As the data was coded for several of these external factors, they could be consulted to investigate to what extent they can explain discovered anomalies.

Coding of the year of publication (YEAR_PUB) and year of attestation (YEAR_ATT) was not only required for the reconstruction of the feature development, but also necessary in

order to learn about the *timespan between the documentation/publication and the actual language situation* (see Section 2.2.2.2).

To investigate *language attitudes* (see Section 2.2.2.3) and their impact on the pidgin portrayal, the contextual embeddings provided in the initial word documents were examined. In addition, the role of the author in the area (AUTH_ROLE) was useful as it may explain certain language attitudes.

Information about the *duration of sojourn and travel activities* (see Section 2.2.2.4), which Baker & Winer (1999) define as experience, was obtained either from the source text or additional investigations had to be pursued concerning the author. Moreover, the role of the author in the area (AUTH_ROLE) sometimes indicated how closely an author came into contact with the pidgin.

As shown by Huber & Velupillai (2016: 132) the competence or linguistic abilities of the authors cannot always be trusted (see Section 2.2.2.5). Therefore, language examples were extracted that were embedded in their original contexts, but translations provided by the early authors were not included in the spreadsheets. Instead, the morpheme-by-morpheme analysis was executed to ensure an analysis and translation of the data that are as neutral as possible. Nonetheless, the data might still contain filters based on the competence of the author. A qualitative comparison of language examples from different authors of (roughly) the same time period was made to identify a possible lack of linguistic abilities. The origin of the author (AUTH_ORIGIN), showing whether the text was written by an indigenous Islander or a European, further indicated the level of linguistic competence of the writer.

It is a common procedure in corpus linguistics and other database compilations to identify the text type of the sources included. It is assumed that “texts within each [text] type are maximally similar in their linguistic characteristics, regardless of their situational/register characteristics” (Biber 2009: 846). If a corpus should reflect the language of a complete population, it needs to involve various text types to describe the language at its best. Individual text types might have an impact on which form (variant) of a feature (variable) is used. The datapoints were therefore coded for text type (TXT_TYPE) to identify filters that may have been exposed on the data through *editor revisions* (see Section 2.2.2.6) and to differentiate between speech-related and written Pidgin English.

It needs to be noted that when the present study refers to ‘text type’, it is not the source type (monograph, newspaper article, etc.) that is meant. Since the language examples occurred in different source types by chance and texts differed in the amount of data that could be extracted (e.g. single words, phrases, sentences, whole texts), this classification was not considered helpful.

It is not of great importance whether the data was extracted from archive material or monographs. Instead, it is important to know about whether language data was found in a travel report, written by a European recalling a specific speech-situation, or whether the language data represents a complete letter written by a Pacific Islander. The data was thus classified into three broad categories, namely *speech-related attestations*, *written attestations* and *intermediate attestations* (cf. Table 8). The broad categories can be further subdivided into six sub-categories, which themselves consist of thirteen sub-classifications.⁸⁴

TEXT_TYPE_3		TEXT_TYPE_B		TEXT_TYPE_F	
1	speech-related attestations	1_1	“attempts to represent actual speech by specific individuals on specific occasion” (Roberts 2005: 42)	a	remarks and comments by travellers, missionaries and residents concerning their experiences in monographs, letters, diaries and reports
				b	scripted monologues (e.g. lecture)
				c	court testimonies
		1_2	“attempts to portray the typical manner in which members of various social and ethnic groups spoke” (Roberts 2005: 42)	d	dictionaries & grammars
				e	general descriptions of language use
				f	fiction
2	written attestations	2_1	natural writings produced in specific situations for interethnic communication	g	letters written by Pacific Islanders
				h	pamphlets
		2_2	permanent writings	i	translations
				j	other documents (e.g. PE newspapers)
3	intermediate attestations	3_1	documents based on speech-events but consciously transformed to fit the written medium	k	letters written on behalf of Pacific Islanders
				l	stories told by Pacific Islanders but edited to be written down
		3_2	unclear whether composed in written or spoken medium	m	e.g. songs by Pacific Islanders that were not translated

Table 8: Variable *TEXT_TYPE*

The first category, *speech-related attestations*, can be subdivided based on the distinction introduced by Roberts (2005: 42) who argues that language data can be categorised into “attempts to represent actual speech by specific individuals on specific occasions” and “attempts to portray

⁸⁴ Initially, a very fine-grained classification system was used to classify the texts, because it may also make a difference whether data was extracted out of a published vs. an unpublished diary. However, due to the limited amount of datapoints available, it was decided to work with the broader categories displayed in Table 8.

the typical manner in which members of various social and ethnic groups spoke”. A differentiation was thus made between those attestations that try to recall or record actual speech and those attestations that are not based on actual specific situations but nonetheless try to depict what the spoken language of a society was like. The former can involve (a) remarks and comments by missionaries, residents and travellers about encountered specific speech situations, (b) journalistic reports in which it is reported about ‘real’ speech situations and (c) court testimonies. The latter label can involve (d) dictionaries and grammars as well as (e) general language descriptions and (f) fictional writings.⁸⁵

The category *written attestations* can be subdivided as well. On the one hand, it contains written documents in PE that were naturally produced and used in specific single situations for interethnic communication, such as (g) letters written by Pacific Islanders and (h) pamphlets. On the other hand, the category includes rather ‘permanent’ writings which were produced for interethnic purposes that were enduring. Bible translations and law translations (i), as well as PE newspapers (j) fall into this latter category, as these documents were instruments that were made use of in interethnic communication, but they were not produced for a single written communicative interaction.

A third category named *intermediate attestations* was added as not all the texts can be easily defined as belonging to the speech-related or to the written category. An example represent PE letters that were written by Europeans on behalf of the islanders (k). It can be assumed that illiterate Pacific Islanders did not dictate the letters but that they rather narrated freely what the European should write into the letter. The writer will have used the spoken word as a basis but will have transformed the speech act to meet the formal standards that letters had to fulfil. Thus, the letters fit neither into the speech-related category (despite being based on a specific speech-situation), nor into the written category (despite being written down). Moreover, stories narrated by Pacific Islanders, which were intended to be written down by Europeans (l) belong to this category as well. Next to PE examples that were extracted out of documents based on speech-events which were transformed to fit the written medium, there are also data sources for which it is unclear whether they were composed in the written or spoken medium, e.g. songs by Pacific Islanders that were not translated (m). Since each text may include attestations of several of these categories, it needs to be noted that a thorough look into the texts is necessary.

⁸⁵ It needs to be noted that attestations which were introduced with sentences such as “and such a prayer can be heard” or “perhaps a man will come along and say” (Vance December 1946: 12) were classified as fictional writing as well. The use of the conjunctive may further be an indicator that fictional PE was documented.

The collected data is further characterised by an *anglicisation* of most of the examples (see Section 2.2.2.7), which is a problem that could not be opposed as easily. Therefore, especially those sources which make use of non-English orthographies were compared to the anglicised samples. Thus, if two or more authors provided language data for the same year or period, the examples were compared to uncover whether the anglicised examples obscure developments in the variety. Especially Pacific Islander-derived sources were considered to have a high potential to reveal the trustworthiness of anglicisations used in other sources. In addition, hints about the pronunciation of sounds and words were extracted with their dates of attestation and compared to the language data. Some attested words could be immediately classified as being anglicised, without comparing them to further examples. For instance, the database contains nouns with a plural -s ending and verbs with the 3SG morpheme -s or verbs in a past form. In most of these cases, the authors substituted an anglicised form for what in PE would have been an unmarked verb or an unmarked noun so that the verb or noun fits into the context of their writing. For example, the English sentence displayed below describes a situation that happened in the past. The author, Idriess, uses the anglicised form “made paper” for what in PE would have been the unmarked verb “make paper” to adjust the verb form to the past reading of the English sentence:

- (4) He brought twelve skinny, bushy-looking boys back to the beach, where they “made paper”– signed on for two years’ service. (Idriess 1941: 88)

Since the Excel files could be filtered for a specific morpheme, in some cases the comparison of the tokens could shed light on their reliability.

4.4 Summary and outlook

The present chapter outlined the empirical and methodological basis of the study. While Section 4.1 was devoted to the database compilation, the database was introduced in Section 4.2. Section 4.3 outlined the applied processes for the selection of features for the analysis (4.3.1), their coding and their extraction (4.3.2). The resulting data spreadsheets were introduced in Section 4.3.3 before the analytical procedure was explained in Section 4.3.4. Finally, Section 4.3.5 was devoted to demonstrating how the reliability of the sources was proven.

Several caveats of the present study were illustrated which are predominantly connected with the availability and trustworthiness of the data. It was shown that the major hazard lies in the determination of dates of attestation which is especially problematic in a study that investigates the development and change of pidgin varieties across time.

Nonetheless, it is important to recollect that despite the possible pitfalls, the historical data represent the only available sources for the reconstruction of early stages of contact languages

and have already “extended our knowledge of these languages” in the past (Baker & Winer 1999: 103). Working with historical data of this kind has its challenges but other types of data (for instance, modern day corpora) include challenges as well. The fragmented collection of early sources is the only possible way to analyse the development of individual features of the varieties in question and, thus, to gain insights into the individual development of the three varieties. Cautious and detailed attention to the data structure is a prerequisite for the analysis presented in the upcoming chapters. The next chapter will show the results of the analysis of meta-linguistic statements about the similarity of the varieties before Chapters 6-9 will focus on the analysis of the morphosyntactic features.

5 Meta-linguistic statements on the divergence of MPE varieties

A common method in measuring the similarity of two or more languages is to apply mutual intelligibility tests – a method which is not adequate for historical studies that try to investigate to what extent speakers of different but similar languages were able to comprehend each other. Nevertheless, in order to investigate the degree of similarity of SIP, BIS, and TP, extralinguistic socio-historical information on the varieties' resemblance was collected. The more the varieties diverged from each other, the more likely it is that authors have noticed this change and noted it down in their writings.

The analysis of the meta-linguistic statements provides evidence that the contact variety spoken in Melanesia was perceived as a *single* variety by pre-20th century writers. Serving as the lingua franca it is described as “the universal mode of communication” (Bridge 1886: 547) or as “the universal trade-tongue of all natives everywhere” (Mackellar 1912: 105; referring to the year 1900).

It is attested that Pidgin English continued to be used as a lingua franca or a “rough-and-ready method of intercommunication” in post-1900 Melanesia, enabling “boys from islands so distant from one another as the New Hebrides, Solomons, and New Guinea to work together or under white traders and employers and understand each other” (Alexander 1927: 213). What is striking is that Gilchrist Alexander, who from 1907 to 1925 acted inter alia as a “British Judge in the New Hebrides and a Lands Commissioner in the Solomon Islands” (Glasser 1999: 187), describes MPE as “the lingua franca of the Pacific” being “alike in British, French, or other possessions” (Alexander 1927: 213). The author, who, due to his role in the Pacific, came into contact with the variety as spoken in the Solomon Islands and Vanuatu, does not recognise any regional variation using the adjective “alike” to describe the varieties. Similarly, Rivers (1914: 466-467; referring to the years 1907-1914) states that:

the English language in a pidgin form is at present the almost universal language of this vast geographical area. [...] [A] Solomon Islander who wishes to converse with a native of the New Hebrides, finds his most convenient, and often his only, means of intercourse in pidgin English. At the present time, we can watch the process whereby a new and uniform language is displacing a condition of linguistic diversity.

Words such as *alike*, *universal* and *uniform* are used repetitively to describe Melanesian Pidgin English in the early descriptions. This gives rise to suspicion that the varieties did not differ to a great extent in the areas and that they were, in any case, mutual intelligible at least to a certain degree.

The earliest metalinguistic statement *at hand* declaring regional variation dates to the year 1914. Jacomb, who lived in Vanuatu from 1906-1914, claims that “[d]oubtless Pidgin-English varies in every Group, according to the state of development of the native language of the locality,

and the nature of the ordinary everyday life” (Jacomb 1914: 91). Referring to the variety as spoken in the New Hebrides he claims that “Pidgin-English is composed chiefly of English words with, in the New Hebrides, a very slight admixture of French” (1914: 91). Thus, differences are at least observed on a lexical level as early as in 1914.

Around 1926, travel accounts make notice of local variation but at the same time report the varieties to be mutual intelligible. Collinson (1926: 84) states that “with the exception of local variants [Pidgin English] is pretty much the same everywhere”. This conforms with Hogbin’s description who argues “that the pidgin spoken in that area [= New Guinea], despite many differences in vocabulary, is similar to that current in the Solomons” (1939: 163) and elsewhere he states that “Solomon Island pidgin is slightly, but not very markedly different [from Tok Pisin]” (1939: np). Cormack, who stayed in the Pacific from 1914-1944, writes at the end of his stay that “[w]ith its local variations this lingua-franca of the South Seas is universally used throughout the Western Pacific” (1944: 130). Also, Gordon points towards the lexical differences, stating that “[i]n each area changes occur in pidgin English & local idioms & words are used” (Gordon 1943, AU ANUA 584-1). Another early source pointing towards lexical differences is Johns’ *Pocket Vocabulary of Malay, Pidgin English, and Japanese Phrases*. Since according to Johns twenty percent of Pidgin English are of native origin, the varieties “vary with the locality” (1942: 78).

Nonetheless, the varieties are described to be used “in interdistrict, interisland, and commercial contacts throughout Melanesia” (cf. Krieger 1943: 48), so that mutual intelligibility can be expected. Interesting is also the description which Reed (1943: 270) provides:

The pidgin of New Guinea, however, is a sub-type of Beach-la-mar and is here to be distinguished as such. [...] [I]n the British Solomon Islands the ‘pidgin’ is far purer English than our own particular brand in this Territory. As with the Papuans you can address a Solomon Island work boy in the vernacular of Picadilly or Potts Point, and, in nine cases out of ten, he will understand your meaning. Further south, in the New Hebrides, there creep into the ‘pidgin’ certain French words [...] In view of the inadequacy of the term ‘pidgin English,’ and with the variations of Beach-la-mar in mind, it seems advisable that we seek a distinctive name for the *lingua franca* peculiar to the Mandated Territory.

Several remarks need to be attached to this quote. First, it becomes visible that Reed recognised variation between the three varieties. Secondly, early Solomon Islands Pijin seemed to be intensely influenced by its superstrate, since Reed describes it as “far purer English” and even claims that one can address the Solomon Islanders in Piccadilly or Potts Point vernaculars. A third important consideration is that Reed suggests a distinctive name for the variety spoken in the Mandated Territory of New Guinea based on the existing variation. In 1944, the three varieties are named *dialects* for the first time in a language guide produced by the Army Education Branch (cf. 1944: 1).

Summarising the metalinguistic statements, a movement away from adjectives such as *alike*, *universal* and *the same* to softer adjectives such as *similar* can be observed. From the attestations known to me, *local variation* was first mentioned in 1914, and from 1926 onwards, a growing number of authors observe differences between the varieties. It is interesting that all descriptions refer predominantly to lexical differences in the geographical areas and it is not referred to morphosyntactic variation. One possible reason for this is that lexical differences are more salient to a non-expert observer, i.e. someone who is not a linguist. Despite local variation, the varieties are described to be mutual intelligible in the timeframe from 1850 until 1950.

However, it should be kept in mind that the descriptions above were made by European non-linguists with a colonial attitude. Thus, it is possible that the European observers perceived greater similarities than there were. For example, it is likely that there were linguistic false friends. Because of the similarity of Pidgin English and English words, European observers may have assumed that they understood what was being said, but in fact they may have not. For example, the PE word *dai*, may have been translated as ‘to die’ due to its similarity to the StE verb *die* even though in PE it can have other meanings as well, such as ‘to faint’ or ‘to fall into a deep sleep’. False friends may have existed not only between PE and English, but also between Solomon Islands Pijin, Bislama and Tok Pisin. A contemporary example represents the word *inap/inaf* (< *enough*). While the word is used with the meaning ‘capable’ or ‘sufficient’ in all three MPE varieties, it can also function as an ability marker in TP. Moreover, while the word *ken* represents another marker to express ability in TP, the similar sounding marker *kan* encodes inability in SIP. Thus, the fact that early travellers, residents, etc. did not mention variation does not prove that variation did not exist. Only when they point towards variation, this might tell us something about the varieties’ divergence.

6 Case Study: Demonstratives

I then made signs (for we understood not a word of their language) that we wanted wood; and they made signs to us to cut down the trees.

(Cook 1842: 498)

Young children that have not acquired language yet make use of their eyes, gestures, signs and pointing to communicate with their caretakers. The pointing is said to mark the onset of “systematic intentional and referential communication, with shared attention focused on a third entity” (Levinson 2018: 2). As soon as children acquire language, they get to know additional linguistic means to refer to objects in their environment and to express a joint attentional focus (cf. Diessel 2006: 469; Levinson 2018: 2). According to Diessel (2006: 469), the best linguistic device to replace the initial pointing represent demonstratives.

Likewise, in adult speech situations in which speakers do not have a shared language at hand, the use of gestures and *pointing* serve as helpful devices to bridge a communication gap. In the introductory quote and in Section 3.1.1 it was shown that signs – including pointing – were a common communicative tool in the initial encounters between Pacific Islanders and Europeans. With more intensive contact between the groups, pointing became less important as linguistic solutions developed. It is thus possible that demonstratives were among the first words used by Europeans and Pacific Islanders during the development of the contact variety and that they replaced the initial pointing, similar as in first language acquisition. This, together with the fact that demonstratives exist in all languages around the world (cf. Diessel 1999a: 1), may explain why they are found with high frequencies in the early data.

The present chapter will focus on the development of demonstratives in Solomon Islands Pijin, Bislama and Tok Pisin. The case study starts with an outline of the theoretical background and will give an overview of the demonstratives in modern TP, BIS and SIP. This is followed by a survey of research on demonstratives in MPE with a focus on diachronic change. In Section 6.4 the methodological steps will be introduced. The chapter will close with a summary of the findings.

6.1 Theoretical background

The most extensive crosslinguistic studies on demonstratives have been conducted by Diessel (1999a) and Levinson (2018) and their frameworks provide the basis for this chapter. Demonstratives exist in all languages, but differences can be observed in their *morphological form*, *syntactic context*, *meaning* and *pragmatic use* (cf. Diessel 1999a: 1). While most languages encode demonstratives through free morphemes, there are also languages in which

demonstratives occur in form of clitics. The forms may be uninflected or marked for gender, number and/or case (cf. Diessel 1999a: 22-33).

Demonstratives can furthermore appear in different *syntactic contexts*. Usually four different types are distinguished, which have been introduced by Diessel (1999a: 57). According to his classification, demonstratives may occur:

- i. as independent pronouns in argument position of verbs and adpositions (pronominal)
- ii. together with a noun in a noun phrase (adnominal)
- iii. as verb modifiers (adverbial), or,
- iv. in copular and nonverbal clauses (identificational).

Those occurring as independent pronouns in argument position are referred to as *demonstrative pronouns* (e.g. English *That smells.*), while those being used with a noun in noun-phrases are called *demonstrative determiners* (e.g. English *this boy*). Verb modifiers are referred to as *demonstrative adverbs* and relate to manner and locative adverbs such as English *thus, here, and there*. *Demonstrative identifiers* is the name given to copular and nonverbal clauses (cf. Diessel 1999a: 57). Diessel (1999b: 3) remarks that demonstrative identifiers are “usually not distinguished from pronominal demonstratives, but since demonstratives in identificational sentences are often formally distinguished from (pronominal) demonstratives in other sentence types”, he keeps them apart. Levinson adds a fifth category which he names *presentationals*, referring to French expressions such as *voilà!* (cf. 2018: 4).

The analysis in this study is confined to pronominal, adnominal and identificational demonstratives. No differentiation is made between demonstratives occurring in identificational sentences and those occurring as independent pronouns in argument position of verbs and adpositions (see Section 6.4 below for details).

Languages around the world differ regarding whether they use a single form in pronominal and adnominal (or even all four) syntactic contexts or whether they make formal distinctions. English, for instance, uses the homonymous demonstrative pronouns *this* and *that* in both positions, whereas French makes use of different forms with *ce/cette* in adnominal and *celui/celle* in pronominal position. Typological studies have shown that it is more common for languages not to show any formal distinctions based on their syntactic context. Diessel, who compared demonstrative markers in a sample of 85 languages, identified 71.76% of the languages to not make any formal distinction between the different semantic contexts (cf. 1999a: 59). His results conform with the languages listed in the typological database *World Atlas of Language Structures* (WALS), in which 143 of the sample of 201 languages (= 71.14%) use the same form in pronominal and adnominal position (cf. Diessel 2013a: 174). Although this has also been confirmed for pidgins and creoles as well, it is proportionally less common. According to the

database *Atlas of Pidgin and Creole Language Structures* (APiCS), 54.67% of the listed 75 languages use the same form in adnominal and pronominal position (cf. Maurer 2013a: 122).

Demonstratives can also be classified according to their deictic and qualitative semantic characteristics (cf. Diessel 1999a: 35ff.). They are usually classified as place deictic expressions and “indicate the relative distance of an object, location or person vis-à-vis the deictic center” (Diessel 1999a: 36; cf. also Levinson 1983: 54).⁸⁶ The deictic centre usually refers to the speaker’s location and is also called *origo*.

Diessel argues that all languages have at least a dual demonstrative system, with “a proximal demonstrative referring to an entity near the deictic center, and a distal demonstrative indicating a referent that is located at some distance to the deictic center” (1999a: 36). This is, however, only valid for demonstrative adverbs. Crosslinguistic studies have shown that languages do not necessarily show a distance-contrast in pronominal and adnominal syntactic contexts. For instance, the languages listed in WALS show that even though the two-way distinction is the commonest approach (54.3%), seven (3%) of the 234 languages investigated do not show a deictic contrast and the remaining 37.6% of languages indicate a three-way distance contrast (cf. Diessel 2013b: 170). In APiCS the simple distance contrast also represents the dominant structure, occurring in 67% of the listed contact varieties but 19% of the listed pidgins and creoles are distance-neutral and 13% make use of three-way distinctions (cf. Maurer 2013b: 126). As the figures indicate, distance-neutral pronominal and adnominal demonstratives are cross-linguistically less frequent than demonstratives with a simple distance contrast. It can be said that distance-neutral demonstratives are nonetheless not uncommon in contact languages, since almost one in five pidgins and creoles does not show a distance contrast. Demonstrative adverbs, which tend to show a distance contrast, may therefore co-occur with distance-neutral adnominal demonstratives to differentiate between two referents (cf. Diessel 1999a: 38, 1999b: 8). In German, for instance, the adverbs *hier* (= proximal) and *da/dort* (= distal) can be added to express deictic contrast in a phrase such as *Das Mädchen da*. Similarly, the French distance-neutral demonstrative forms *ce/cette/ces* can obtain a deictic contrast by adding post-nominal *ci* (= proximal) or *là* (= distal), as in *cette fille-là*.

Languages that differentiate between more than two distance contrasts can further be classified into distance-oriented versus person-oriented systems. In the former “the deictic center is the only point of reference for the location of the referent”, while in the latter “the location of

⁸⁶ As will be shown in 6.4, a clear distinction between deictic place expressions and deictic time expressions (cf. Bühler 1934: 102) was not possible in the case of the early MPE data, as demonstratives were attested to be used together with the noun *time* in adnominal cases to make temporal references in relation to a speech event.

the hearer serves as another reference point” (Diessel 1999a: 50). What they all have in common is that they require contextual information to be able to “find the referent in the context” (Levinson 2018: 5).

Demonstratives may also encode “qualitative information about the referent” (Diessel 1999a: 47). They may indicate whether the referent is a location or an object/person (ontology), the animacy, humanness, sex and number of the referent, as well as whether they are “conceptualized as a restricted or extended entity” (= boundedness) (Diessel 1999a: 50).

Though demonstratives may serve as a device to encode deixis, they are not used for deictic purposes only but show a variety of *pragmatic uses* (see Figure 16).

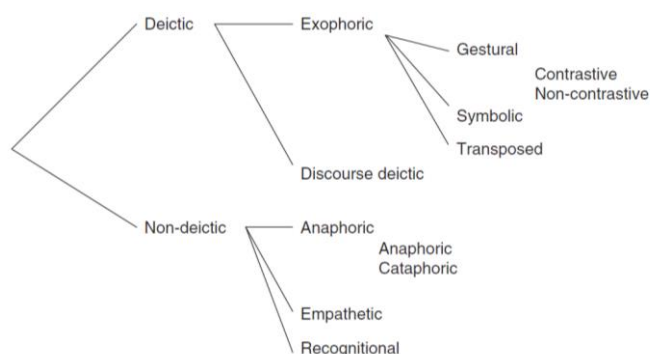


Figure 16: Different uses of demonstratives (Levinson 2018: 10)

According to Levinson (2006, 2018), non-deictic uses of demonstratives can be *anaphoric*, *empathetic* or *recognitional*. Anaphoric use occurs when reference is made to an entity that was introduced earlier, as in example (5). An example of an empathetic use is found in (6) and example (7) shows a recognitional use of demonstratives.

- (5) *Vanessa got up early in the morning though she did not sleep a lot. **This** woman is a drudge.*
 (6) ***This** daft cow!*
 (7) *Do you remember **that** nice bonfire?*
 (8) *Ahhhh! She screamed like **that**.*

Deictic demonstratives may be further classified according to their use as well. While exophoric use of deictic expressions occurs when the entity is available in the physical context, discourse deictic use is when a demonstrative is used to refer “to a chunk of discourse itself” (Levinson 2018: 10) as in (8). The former expressions may additionally be distinguished regarding whether they require a gesture or not (cf. Levinson 2018: 10). A sentence such as *This finger hurts* requires some pointing or indication, which is not the case in a sentence such as *This city is nice*.

The present study is restricted to the morphological realisation of demonstratives and will focus on surface phenomena only. Thus, while the syntactic context of demonstratives is considered, their deictic and qualitative semantic characteristics as well as their pragmatic

functions are not analysed. As SIP, BIS and TP are said to have developed out of MPE, and as differences in encoding demonstratives were observed during the morpheme-by-morpheme analysis, the focus is placed on their formal realisation. The aim is to investigate whether the early data shows when the three varieties made a choice for a specific form in adnominal and pronominal position and what factors may have had an impact on this choice of form. I will analyse from a typological perspective whether different forms were in use in the two syntactic contexts.

6.2 Demonstratives in MPE today

Modern Solomon Islands Pijin, Bislama and Tok Pisin show different preferences in the *forms* used to encode demonstratives but are quite similar from a typological perspective. In *Solomon Islands Pijin* the free forms *ia*, *disfala* and *desfala* are the most common forms (cf. Simons 1985: 59-60; Jourdan 2002: 36; Jourdan 2008: 470).⁸⁷ While *disfala* and *desfala* are pronominal (cf. example (9)), *ia* occurs post-nominally (cf. example (10)). Both forms can be used in adnominal and pronominal position so that a form-distinction based on the syntactic context does not exist. Additional demonstrative forms are *hemia* (cf. (11)), *diswan* (cf. (12)) and *diskaen*, with *hemia* more commonly found in pronominal contexts (cf. Jourdan 2002: 72; 37; 36). A distance contrast is not made so that contextual information is required to differentiate between proximate ‘this’ and distant ‘that’.

(9) disfala DEM ‘this/that girl’	gele girl	(10) <i>gele ia</i> girl DEM ‘this/that girl’
(11) hemia DEM ‘Only this one is left.’ (Jourdan 2002: 72)	<i>nomoa</i> <i>stap</i> only LOC	(12) Diswan hem blong me DEM 3SG POSS 1SG ‘This is mine.’ (Jourdan 2002: 23)

In *modern Bislama* the free form *ia* is used in pronominal and adnominal position. When *ia* functions as an adnominal demonstrative, it is postponed to the noun, as shown in example (13) (cf. Crowley 2004: 64; Meyerhoff 2013a: 225). Number is not marked on the form but through the noun preceding plural modifiers *ol* or *olgeta*. The demonstrative is not deictically contrastive so that the context in which the particle is used is necessary to comprehend whether it refers to ‘this/these’ or ‘that/those’ (cf. Crowley 2004: 65). Only occasionally *nao* or *lo(ng)we* are attached to the particle to express distance contrast, *ia nao* expressing the proximate ‘this/these’ and *ia*

⁸⁷ According to Simons & Young (1978: 159) “*desfala* functions as an article rather than as a demonstrative” (Siegel 2008: 182), but all other grammars list the form as a demonstrative.

lo(ng)we expressing the distant ‘that/those’ as exemplified in (14). According to Crowley (2008: 449), these additions are not necessary and are only used to “disambiguate the multiple senses of *ya*”. Moreover, the form *hemia* can be used in both syntactic contexts as a demonstrative (cf. sentences (15) and (16)).

(13) <i>gel ia</i> girl DEM ‘this/that girl’	(14) <i>Man ia nao o man ia longwe?</i> man DEM now or man DEM far.away ‘This man or that man?’ (Crowley 2004: 64-65)
(15) <i>gel hemia</i> girl DEM ‘this/that girl’	(16) <i>Hemia i mo gud, waet-wan ia.</i> DEM PM COMP good white-one DEM ‘That one is better, that white one.’ (Crowley 2004: 49)

Though the form *disfala* may also be heard among some Vanuatians, it is less commonly used. As Crowley states (2004: 65):

it is a little difficult to be certain about its status in modern Bislama. While some people certainly do use it, it is not nearly as common as *ia*, and some people regard *disfala* either as just random influence from English *this*, or as perhaps a borrowing from similar forms that are much more commonly used in Tok Pisin in Papua New Guinea or Solomons Pijin.

In most cases in which *disfala* is used, a postmodifying *ia* is used simultaneously.

In *Tok Pisin*, as exemplified in sentence (17), the free form *dispela* represents the prominent demonstrative particle, which can be pluralised through the preceding plural marker *ol* (cf. Smith 2002: 177-178). The form **datpela* is not used by Tok Pisin speakers. Reduced forms of *dispela*, such as *displa*, *disla* and *sla* have developed which are common in rapid speech (Smith & Siegel 2013b: online). An example for *disla* is found in (18). *Dispela* is used in pronominal as well as adnominal syntactic contexts and no distance contrast is made. In case of adnominal demonstratives, *dispela* precedes the noun.

(17) <i>dispela</i> DEM ‘this/that girl’	<i>gel</i> girl	(18) <i>Mi laik-im disla</i> 1SG like-TR DEM ‘I like this one’ (Smith & Siegel 2013b: online; Example 22-59)

The form *ia*, which serves as the major demonstrative particle in BIS and which is one of the possible particles in SIP, is only occasionally applied with a demonstrative meaning in Tok Pisin (cf. Smith 2002: 178). It is rather considered to fulfil the role of a focus marker.

From a typological perspective all three varieties make use of *free* forms that *do not show a distance contrast* and stay the same independent of their syntactic context (= adnominal vs. pronominal occurrence). However, while in SIP *disfala/desfala* and *ia* function as demonstrative pronouns and demonstrative determiners, BIS shows a preference for the form *ia* and TP for the form *dispela*. In SIP and BIS the additional demonstrative pronoun *hemia* is used which cannot

be found in TP. A co-occurrence of *disfala/dispela* and *ia* seems to be possible in all three varieties.

6.3 Previous diachronic studies on demonstratives in MPE

The development of demonstrative particles in the three MPE varieties and the concomitant stabilisation of a preferred form in TP and BIS has not yet been investigated from a diachronic perspective. Baker (1993: 45), whose studies focussed on earliest attestations, first attests the demonstrative forms *this fellow* and *that fellow* in QLD data from Aborigines and later in the MPE varieties, arguing in 1993 that the two forms had existed before the labour trade began but spread to the Melanesian area through returning labourers. Three years later, however, Baker (1996) argued that only *that fellow* is likely to have spread from QLD to Melanesia through returning labourers. Due to the fact that the first attestation of *this fellow* was found in the Northern Territory of QLD, which Baker described as “an area ill-placed for the diffusion of pidgin features to Melanesia” (1996: 534), he supposed that *this fellow* may have developed as an independent innovation in Vanuatu from where “it was carried [...] to all the other territories where it [was] subsequently attested” (1996: 534). As it was not the focus of Baker’s studies to investigate the further development of the demonstrative particles in the three Melanesian varieties, we do not have further information about the forms’ continuing development – whether *this fellow* and *that fellow* stabilised or disappeared. In addition, Baker’s feature list neither included the demonstrative particle *here* nor StE similar forms, as his feature list was based on forms that deviate from StE. Baker & Huber (2001) have successfully shown that *this fellow* NOUN and *that fellow* NOUN represent common Pacific features. In their analysis they do not differentiate between the individual MPE varieties and focus on first attestations. Therefore, information about the further development and use of the features in the three MPE varieties is not provided.

Siegel’s study on substrate reinforcement (2008: 183) devotes a small section to the stabilisation of demonstratives. Starting from differences observed in the three MPE varieties today, Siegel argues that the demonstrative forms *this fellow*, *that fellow* and *here* formed part of early MPE and had been present in the three areas under investigation. Next to the three forms, StE *this* would have been in use as well. To support this claim, he provides five historical sentences without indicating where these sentences were attested and if all forms were in fact attested in all three areas. As Siegel’s primary interest is to show that present day structures can be traced to substrate influence, he does not place the focus on the chronology of the forms’ attestations.

As most of the historical studies focus on earliest attestations and are predominantly interested in how features spread from one area to another, no studies have been conducted so far which attempt to show *when* the three MPE varieties diverged in their demonstrative usage. Although Siegel provides a possible explanation for their divergence, he does not focus on *when* differences stabilised. Thus, in the following an attempt will be made to fill this academic void.

6.4 Methodological considerations

The present study is confined to the morphological realisation of pronominal, adnominal and identificational demonstratives in the three MPE varieties. Although it was initially aimed at differentiating between pronominal and identificational demonstratives, the low number of early attestations made such a fine-grained classification and analysis impossible. Thus, identificational demonstratives are classified as pronominal demonstratives in the present study. As required contextual information is frequently insufficient or completely missing in the early collected data, the study only focusses on surface phenomena.

Factor	Levels
ST_FORM	<i>this</i> <i>that</i> <i>these</i> <i>those</i> <i>this fellow</i> <i>that fellow</i> <i>these fellow</i> <i>here</i> <i>this fellow + here</i> <i>this + here</i> <i>this one</i> <i>them</i> <i>them fellow</i> <i>that one</i> <i>him here</i>
WORD_ORDER	prenominal postnominal surrounding
FEATURE	dem_adn dem_pron dem_ambig
YEAR_DET	1832 ... 1950

Table 9: Linguistic coding for demonstratives

A primary step was to identify the individual forms (ST_FORM) used to encode demonstratives in the three varieties under investigation. The variable ST_FORM has 15 variants, which are displayed in Table 9, and will be introduced with an example sentence in the following. The examples were randomly selected, and their selection does not mean that a variant was solely

attested in one of the three varieties. Information regarding the distribution of the variants in the individual varieties will be provided in Section 6.5.

The first four forms listed in the table, namely *this*, *that*, *these* and *those*, look like their StE equivalents (see sentences (19)-(22)). It is probable that these forms were directly borrowed from the lexifier. Nonetheless, the similarity between the early MPE and StE forms does not mean that the actual forms were the same as in StE. From the early attestations it can be learned that *th*-stopping took place as *that* is occasionally orthographically realised as *dat* (cf., for instance, Cheesman 1933: 76) and *this* as *dis* (cf., for instance, Methodist Mission 1935: 5). In addition, there are attestations of *this* and *that* in contexts which in StE require a plural form. In the early PE attestations the forms are not necessarily overtly marked for plurality. In a few cases plurality is expressed through the demonstrative-preceding pluraliser *all*, as shown in sentence (23), *altogether* or through numerals.

- (19) *Master [...] that money no good along this country.*
 master [...] **DEM** money NEG good PREP **DEM** country
 ‘Master, that money is bad for this country.’
 (Ambrym 1892-1896; Lamb 1905: 136)
- (20) *Me no sleep last night me can ‘t forget ‘im that word*
 1SG NEG sleep last night 1SG INABIL forget-TR **DEM** word
 ‘I did not sleep last night, I couldn’t forget that word.’
 (Uru Island 1929; Young July 1929: 5; AU PMB DOC 439)
- (21) *me think about these two men they no ready for heaven*
 1SG think PREP **DEM** two man.PL 3PL NEG prepared PREP heaven
 ‘I thought about these two men who were not prepared for heaven.’
 (Malaita 1930; Waite 31.03.1930; AU PMB MS 1253)
- (22) *Me come here wait ‘im you, those word-s you speak cut ‘im heart belong*
 1SG come here wait-TR 2SG **DEM** word-PL.2SG speak cut-TR heart POSS

me all same knife.
 1SG like knife
 ‘I came here and waited for you, those words which you spoke cut me like a knife.’
 (Uru Island 1929; Young July 1929: 5; AU PMB DOC 439)
- (23) *me fella been do ‘im all this something you been tell ‘im*
 1PL PST do TR **PL** **DEM** thing 2SG PST tell TR
 ‘We did these things which you told us (to do).’
 (Malaita 1942; Deck January 1942; AU PMB DOC 440)

A further variant is the form *them* (cf. sentence (24)). *Them* is a common demonstrative pronoun and plural demonstrative determiner (= ‘those’) in contemporary non-standard and regional varieties of English (cf., for instance, Quinn 2009: 44). As pointed out in Section 4.3, it seems reasonable to assume that the colonisers and early sailors did not necessarily speak standard forms of the lexifier. Thus, the form may have entered PE by early traders or colonisers speaking non-standard varieties of English.

- (24) [...] **them** three Boy-s he poison this Boy call mallip [...]
 [...] **DEM** three boy-PL PM poison DEM boy called Mallip [...]
 ‘Those three boys poisoned this boy called Mallip.’

(Ambrym 1919; Dar 11.01.1919)

The forms *this*, *that*, *these* and *them* were additionally attested in combination with the modifier *fellow* to encode the demonstrative, as exemplified in (25)-(28). *This fellow* and *that fellow* were also in use on Samoan and Queensland plantations (cf., for instance, Eden 1872: 156 for QPPE and Schuchhardt 1889: 159 for SPPE). As mentioned above, while *that fellow* is assumed to have spread to Melanesia by returning labourers, two diffusion scenarios have been suggested for the spread of *this fellow* (cf. Baker 1993, 1996). The forms are likely to have their origin in the StE noun phrases *this fellow* and *that fellow*, in which *fellow* represents the head of the noun phrase. It is likely that ‘DEM *fellow*’ was a commonly used noun phrase in the English spoken by the European colonisers. As *fellow* was typically unstressed this may have led to it being reinterpreted as part of the demonstrative.

Although the example sentences in (25) and (26) give the impression that the forms were pronounced as in StE, some of the early sources indicate that this was not the case. *This fellow* in some sources is spelt as *dis pala*, *dispala*, *despela*, *dispela*, *disfelo*, *disfela*, etc., which supports the assumption that *fellow* was reinterpreted as part of the demonstrative particle. In addition, the spelling variants indicate that dental fricatives were realised as stops and that at least some speakers replaced the labiodental fricative [f] through the labial plosive [p].

- (25) you like **this fellow** papine?
 2SG like **DEM** girl
 ‘Do you like this girl?’

(Milne Bay 1884; Finsch 1888: 241)

- (26) **That fellow** woman Mary belong a me
DEM woman wife POSS 1SG
 ‘That woman is my wife.’

(Vanuatu 1882; Bridge 1886: 547)

These fellow and *them fellow* might represent author modifications and/or overextensions by Western observers since they occur with low frequencies and the sources show that the early writers often had a wrong understanding of the use of *fellow*. For instance, Collinson (1929: 21), writing about MPE as spoken in the Solomon Islands, claimed that the “word ‘fella’ (fellow) is always put in front of nouns” and also other authors give rise to suspicion that the use of the particle *fellow* was overextended (cf. also Daiber 1902: 254 and Stephan & Graebner 1907: 20).⁸⁸

⁸⁸ Cf. “[...] ‘fellow’ [...] wird überhaupt jedem Dinge vorgesetzt; es ist gleichsam die Bezeichnung eines Wortes als Substantiv” (Daiber 1902: 254); “Vor jedes Substantivum wird fellow (Bursche) gesetzt, sogar one fellow pain ein Schmerz” (Stephan & Graebner 1907: 20).

Thus, it is possible that the European observers used the demonstratives *these* or *them*, which themselves are likely to represent author modifications, and inserted a *fellow* between the English-deriving demonstrative and the noun, assuming that *fellow* “is always put in front of nouns” (Collinson 1929: 21).

- (27) *These feller* *boy* *belonga me*
DEM boy POSS 1SG
 ‘These boys of mine’

(New Guinea ~1941; Idriess 1941: 146)

- (28) *them fellow* *boat* *he* *no* *good*
DEM boat PM NEG good
 ‘these boats are bad’

(Espiritu Santo 1883; Colonial Office CO225/13 1882-83: 390)

Further early variants are the forms *here* (cf. sentence (29)), a combination of *this fellow* and *here* (sentence (30)), a combination of *this* and *here* (sentence (31)), and the forms *this one* (sentence (32)) and *that one* (sentence (33)).

- (29) *You* *no* *sabby* *picaninny* *‘ere* *‘e* *b’long* *white man?*
 2SG NEG know child **DEM** PM POSS European
 ‘You do not know that this child (is) of a European?’

(Vanuatu 1917; Lynch 1923: 195)

- (30) *em* *i-kič-em* *disfelə* *brum* *hir,* *naw* *em* *i-fajt-em* *hed* *biləŋ* *mi*
 3SG PM-catch-TR **DEM** broom **DEM** and 3SG PM-beat-TR head POSS 1SG
 ‘She took this broom and she hit my head.’

(New Guinea ~1943; Hall 1943: 63)

- (31) *All* *this* *something* *here* *me* *no savvy* *do* *him.*
 PL **DEM** thing **DEM** 1SG INABIL do TR
 ‘I am unable to do these things.’

(Tulagi; Aho, B. 23.12.1927)

- (32) *this one* *place* *belong* *people* *who* *die* *finish*
DEM place PREP people REL die COMPL
 ‘this place for the people who have died (i.e. cemetery)’

(German New Guinea 1921; Lambert 1942: 89)

- (33) *Debbil-debbil* *that one?*
 ghost **DEM**
 ‘Is that the ghost?’

(Solomon Islands ~ 1908; Abbott 1908: 69)

All variants represent free forms (at least in their orthographical representation) so that a distinction between *bound* vs. *free* forms was not considered useful. As the theoretical background has shown that demonstratives may differ depending on whether they appear in adnominal or pronominal position, a differentiation between demonstratives which determine a noun (*dem_adn*) and those replacing a noun (*dem_pron*) was made. While 13 of the above listed variants were attested in adnominal position, eleven occurred in pronominal position. An additional category called *dem_ambig* was necessary, since despite a thorough examination of the context in which the early

language examples occurred, the data contained cases which could not be clearly identified as pronominal or adnominal demonstratives.

Many of the datapoints which were challenging were those that contained the forms *this fellow* and *that fellow*. The problem can best be exemplified with a sentence from the data. The sentence *This fellow gammon* was extracted from trial data that was recorded in the Solomon Islands. In the sentence, *fellow* may function either as a noun as exemplified in (34) or as a part of the demonstrative (*this* + MODIF = DEM) as the glossing in (35) indicates. If *fellow* was used as a noun, *this* would have to be classified as an adnominal demonstrative. If *fellow* was used as a modifying particle which forms part of the demonstrative, *this fellow* should be interpreted as a pronominal demonstrative (= identificational demonstrative). The word *gammon* is not very helpful for the analysis as it itself is characterised by multifunctionality and can be used as a verb or as a noun.

(34) *This fellow gammon.*
 DEM fellow lie
 ‘This fellow lies.’

(35) *This fellow gammon.*
 DEM lie
 ‘This (is) a lie.’

(Tulagi; Talatova 09.01.1923)

The category *dem_ambig* contains not only those ambiguous datapoints for which a clear differentiation between pronominal and adnominal was not possible but also those tokens for which it could not be clearly decided whether the form is used to encode an adnominal/pronominal demonstrative at all. For instance, there are tokens in which the form *here* may either function as a deictic locational adverb or as a pronominal/adnominal demonstrative, as exemplified in sentence (36). All three varieties show the use of *here* as a locational adverb (cf. e.g. Dickinson 1927: 120 for SIP; MacFarlane 1873: 106 for BIS; Brown 1908: 141 for TP) but nowadays the particle can be used, as the section on present day demonstrative forms showed, with a demonstrative meaning as well. The particle *here* in sentence (36) may either form part of the prepositional phrase *here at One Pusu* or it may belong to the noun phrase *altogether man here*. As mentioned earlier, locational adverbs are sometimes referred to as adverbial demonstratives. They have a demonstrative function in that they refer to a place (cf. Diessel 2006: 473-474). It is due to the shared demonstrative function that spatial adverbs frequently grammaticalise into demonstratives. The classification of forms which occur in the early MPE dataset as either adverbial demonstratives or adnominal demonstratives is complicated by the fact that the forms are homonymous. In sentence (37) the form *here* may encode an adverbial demonstrative, an adnominal demonstrative or a demonstrative identifier.

(36) *Him He hear ‘im altogether man here at One Pusu [...]*
 3SG PM answer TR PL man here/DEM PREP One Pusu [...]
 ‘He answers the prayers of all men here at One Pusu./He answers the prayers of all these men at One Pusu.’
 (One Pusu 1943; Deck December 1943: 4; AU PMB DOC 439)

- (37) *here* *money* *me* *tell* *him* *along* *two* *fellow*
DEM/here money 1SG tell TR PREP 3DU
 ‘This money about which I told them./ This is the money about which I told them./ Here is the money about which I told them.’

(Tulagi; Puanimarau 24.01.1927)

Maurer (2013b: 126) has suggested three main criteria to distinguish between locational adverbs (= adverbial demonstratives) and adnominal/pronominal demonstratives. Elements could be classified as the latter if they differ in their “shape from [...] spatial adverbs”, if “the spatial adverbs are the only available demonstratives in the language” and if demonstratives and spatial adverbs need to co-occur together. The three criteria did, however, not serve as helpful criteria to analyse the ambiguous datapoints in the present study. The first criterion cannot be applied to the early attestations as most writers used an English orthography. Thus, even if differences in the shape existed, it is likely that these differences are obscured in most early attestations. The second and third criteria could not be applied as the focus is on the development of the contact varieties and thus the early data shows several competing forms. Fixed rules had not developed yet. Thus, it was decided that ambiguous cases should be excluded from the quantitative analysis. They will, however, not be completely ignored as they may be important indicators of grammaticalisation. With the remaining *here* tokens it might still occasionally be debatable whether *here* functions as a demonstrative adverb or pronoun.

The demonstrative tokens were furthermore coded for WORD_ORDER and thus, the examples were classified as *prenominal*, *postnominal* or *surrounding*, the latter in cases in which a demonstrative occurred before and after the noun, as shown in example (38). The coding of the linguistic variables is summarised in Table 9.

- (38) ... *na* *siɲawɬ-im* *disfelə* *pikɲimi* *hir* *kowi*
 ... and call-TR DEM child DEM Kowi
 ‘... and called this boy, Kowi.’

(New Guinea ~1943; Hall 1943: 48)

Ideally a classification regarding the various contexts of *usage* would have been desirable. However, the number of early attestations is too small to perform such a fine-grained classification and sentences were frequently quoted out of their context. In addition, it was decided not to differentiate between singular vs. plural contexts in the quantitative analysis. Due to the limited access to contextual information it was not always clear whether, for instance, nouns had a singular or plural reading. Nonetheless, some general remarks on how plurality was expressed in each variety will be provided.

It was almost impossible to investigate whether a distance contrast was made in the early use of demonstrative particles as contextual information is frequently missing. Thus, it is not clear whether forms were used to refer to a distal or proximal entity. Even if StE-derived forms were

in use, this does not necessarily imply that a distance contrast was made. Nonetheless, the qualitative data gave the impression that distance contrast was not common.

To trace the historical development of demonstratives, the factors YEAR_ATT and YEAR_DET represented the most important possible predictor variables in the analysis.⁸⁹ Furthermore, non-linguistic factors such as the text type and author were considered as possible predictor variables because they might have an impact on the results. As forms may have stabilised at different points in time in each of the three varieties, SIP, BIS and TP were first analysed separately before a comparative analysis was conducted.

6.5 Findings and discussion

In the following section the results of the diachronic analysis of adnominal and pronominal demonstratives will be presented. The section starts with an overview of the general data distribution (Section 6.5.1) before a closer look at demonstrative attestations and their development across time will be provided, starting with SIP in Section 6.5.2, continuing with BIS in Section 6.5.3 and focussing on TP in Section 6.5.4. The diachronic comparative analysis of adnominal and pronominal demonstrative particles in Section 6.5.5 will bring together the results of the individual analyses.

6.5.1 General data distribution

A total of 1,841 demonstrative tokens were attested in the early data. For the analysis in R, 38 ambiguous cases had to be excluded, leading to a total of 1,803 tokens. Figure 17 shows how the individual datapoints spread across time among the three varieties.

The first fact to note is that the attestations of demonstratives are unequally distributed across time. The earliest written attestations of demonstratives in the TP and SIP datasets date to the 1880s. The earliest written attestation of a demonstrative particle in BIS goes much further back in time to the late 1820s. This may reflect the fact that Vanuatu was earlier and more intensively involved in the early trading period than the other two areas but may also be due to differences in the surviving written records. The figure also shows that the years after 1930 are scarcely populated in the BIS dataset. This does not mean that demonstratives were not used in BIS after 1930, but simply reflects the general fact that only a small amount of early data could be obtained of BIS for the years 1930-1950. The present dataset thus cannot adequately reflect

⁸⁹ It is important to remember that while YEAR_ATT provides information about the time to which an utterance refers and can refer to dates, specific years or time periods, YEAR_DET refers to concrete determined dates, because statistical programs such as R Studio require definite dates.

the development of demonstratives in Bislama after 1930. From the distribution of datapoints it appears that demonstratives are best represented in the SIP dataset, at least in the years 1900 to 1950.

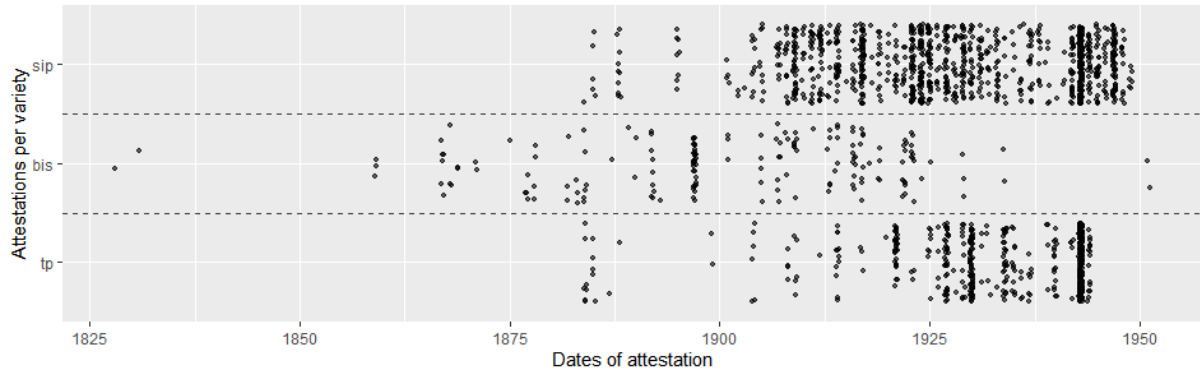


Figure 17: Distribution of demonstrative datapoints across time per variety (based on 1,803 tokens)

A second fact to note is that the number of demonstratives attested per variety differs to a great degree. The majority of demonstrative datapoints were found in the SIP dataset with 1,041 occurrences (= 57.74%). In the early TP data 586 tokens (= 32.50%) were identified and in the BIS dataset demonstratives amount to 176 tokens (= 9.76%).⁹⁰

Figure 18 shows the relative frequencies of attested forms per variety if the complete period under investigation is treated as a single period.

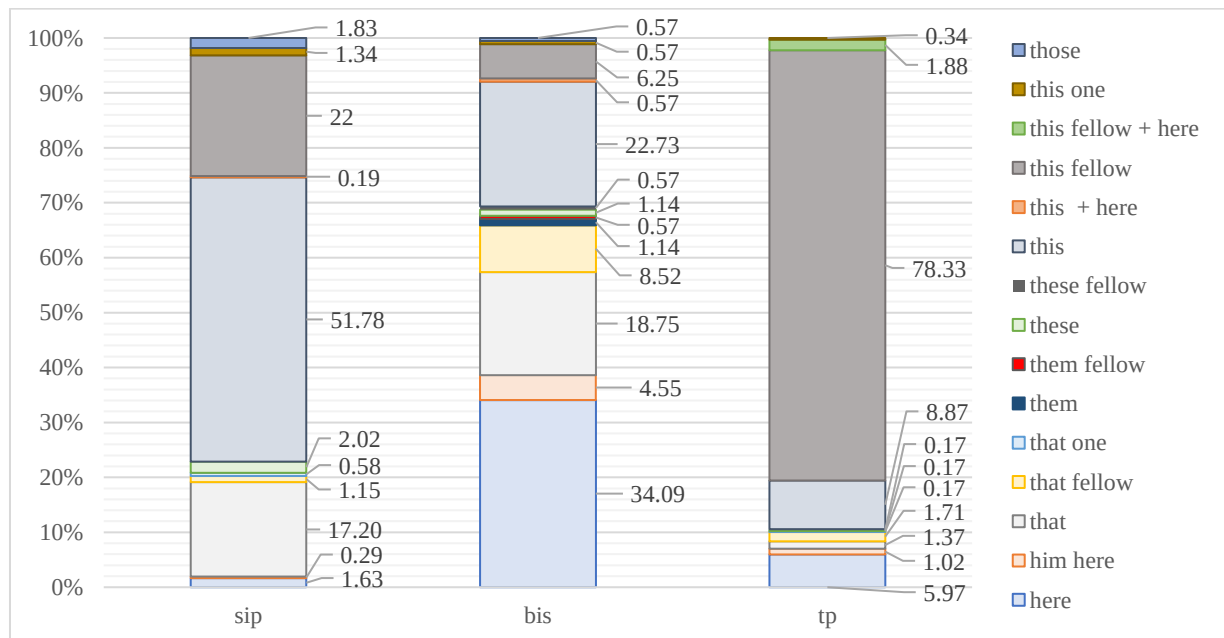


Figure 18: Relative frequencies of demonstrative variants per variety⁹¹

⁹⁰ If ambiguous cases are included, the numbers are as follows: Solomon Islands Pijin 1,060 (= 57.57%); Tok Pisin 594 (= 32.27%); Bislama 187 (= 10.16%).

⁹¹ Figure 34 displays a map showing the places where the individual forms were attested (see Section 6.6).

The figure shows that in all three varieties at least eleven of the 15 variants were attested between 1830 and 1950. In BIS even 13 variants were found. The following section provides a closer look at the individual attested forms, their frequencies and their attestation across time.

6.5.2 Diachronic analysis of demonstratives in Solomon Islands Pijin

In this section, demonstrative forms attested in the early SIP data will be investigated from a diachronic perspective. First, the attested forms will be introduced before the focus is placed on *when* the forms were attested. In a final step, the results of the ctree algorithm will be displayed to discuss whether there is statistical rigor to assume that the end of the labour trade resulted in the stabilisation of demonstrative particles in SIP.

6.5.2.1 Attested forms in SIP

As Figure 18 showed, 51.78% (= 539/1,041) of the demonstrative tokens in the SIP dataset are linguistically encoded with *this*. The form is predominantly used as an adnominal determiner (= 90.91%) and only serves to replace the noun in the remaining 9.09% of instances. While in at least 28 of the 539 attestations *this* is used in a plural context, only in four of these cases the overt plural markers *all* or *altogether* are used (cf., for instance, sentence (39)).

- (39) *Me onefella, which way me going to teach all this new people*
 1SG alone how 1SG FUT INF teach **PL** **DEM** new people
 ‘I am alone, how am I going to teach these new people?’
 (Solomon Islands 1947; Deck 1947: 3; AU PMB DOC 442)

The second most frequent form is *this fellow* (229/1,041 = 22%) which is dominantly attested in adnominal positions (216/229 = 94.32%) but was also identified in pronominal contexts:

- (40) *give me this feller picanninny*
 give 1SG **DEM** child
 ‘Give me this child.’
 (Utupua Island; Wari between 17.04.1925–25.06.1925)

- (41) *[...] more better you go look ‘im this fella close up*
 [...] should 2SG go watch TR **DEM** close up
 ‘You should go and watch this from close up.’
 (Sydney 1923; Collinson 1926: 229)

The third most frequent form is *that*, occurring 179 times in the 1,041 tokens (= 17.2%). In contrast, the form *that fellow* is only attested 12 times (12/1,041 = 1.15%).

The dominant usage of the superstrate forms *this* and *that* is striking. Though it was first assumed that the use of StE-deriving forms may be due to author or editor modifications, a closer examination of the data reveals that this was not necessarily the case. The qualitative analysis of collocations with the adnominal demonstrative *this* revealed that in 24.08% of the tokens (=

118/490) *this* co-occurs together with the word *time*. Similarly, the form *that* collocates together with *time* in 16.22% of the instances (= 24/148). While in sentences such as (42) *this time* has the meaning of ‘at this/that time/moment’, in sentences such as (43) the string *this time* is the temporal adverbial ‘now’. It is interesting to note that the noun *time* is never attested to co-occur with the forms *this fellow* or *that fellow* or other alternative demonstrative forms in the early SIP dataset. Therefore, it is likely that *this time* and *that time* had already grammaticalised into fixed deictic grammatical expressions.

- (42) *This time me no pay tax.*
 DEM time 1SG NEG pay tax
 ‘At this time, I did not pay the tax.’

(Tulagi; Ramichi 14.09.1927)

- (43) *Me no sorry along pipe this time; me no like'im any more*
 1SG NEG mourn PREP pipe now 1SG NEG like-TR any longer
 ‘I do not mourn after pipes now; I do not like them any longer.’

(Malaita 1937; Deck October 1937: 3, AU PMB DOC 439)

To test this assumption, a closer look in Jourdan’s *Pijin Trilingual Cultural Dictionary* (2002) was taken. The dictionary contains entries for *distaem* and its variants *destaem*, *tistaem*, the forms being translated as ‘now, at this time, at this moment’ (cf. 2002: 37). The variant *destaem* may have its origin in *that time*. Interestingly, also other words in Jourdan’s dictionary, such as *disaelan/desaelan* ‘island’ and *diskaen/deskaen* ‘this sort, this kind, this kind of’ (2002: 36) provide evidence that the forms *that* and especially *this* must have been used simultaneously with other competing demonstrative forms. Thus, the high number of attestations of *this* and *that* in early SIP seems to be reasonable and was a prerequisite for fixed deictic expressions to develop.⁹²

The attestation of further StE forms, such as *these* (21/1,041 = 2.02%) and *those* (19/1041 = 1.83%), which occurred in adnominal and pronominal position as well, are further indicators that Reed (1943: 270) may have been correct in his observation that “in the British Solomon Islands the ‘pidgin’ [was] far purer English”. However, *these* and *those* are attested with low frequencies and therefore the possibility remains that they were the result of author or editor modifications. For instance, in example (44), in which *these* was attested, another lexifier feature can be found. The use of plural -s may give rise to suspicion that author or editor revisions took place:

⁹² It may be debatable whether occurrences of *this time* and *that time* should be included in the data calculations if they had already grammaticalised into fixed deictic expressions. Unfortunately, however, it remains unclear whether, and if so when exactly, the strings grammaticalised. Moreover, it raises the question in how far strings such as *this island* ‘island’ and *this kind*, which also grammaticalised into fixed deictic expressions from today’s perspective, should be in- or excluded. Nonetheless, the following calculations were done twice, once including the strings and once excluding them. While those including the strings are displayed in text, those excluding the strings are referred to in the footnote.

- (44) *Me find these good lessons help me more*
 1SG find DEM.PL good lesson-PL help 1SG more
 ‘I realised these good lessons help me more.’

(Malaita 1933; Deck July 1933: 4; AU PMB DOC 439)

The form *here*, as exemplified in (45), occurs in 1.63% (= 17/1,041) of all demonstrative tokens and was attested in adnominal as well as pronominal contexts. The forms *this one* (14/1,041 = 1.34%), *that one* (6/1,041 = 0.58%) and *him here* (3/1,041 = 0.29%) were attested in pronominal contexts only. An example for *him here* is found in (46) in which the deictic usage is supported by the speaker pointing to the dock.

- (45) *Now me think, what something here missionary he tell-im*
 and 1SG think what thing DEM missionary PM tell-TR
 ‘and I thought, what was this thing, the missionary told’

(Malaita 1926; Sullivan August 1926; PMB 1150)

- (46) *Me savvy Talatova him here [points to the dock]*
 1SG know Talatova DEM
 ‘I know Talatova, this one!’

(Tulagi; Garrie 09.01.1929)

6.5.2.2 Timeline and boxplot approach

Figure 19 shows the timeline of demonstrative attestations in Solomon Islands Pijin. The earliest forms are *that* and *this* with the former being first attested in 1880 and the latter in 1881. In addition, the form *this fellow* is already attested in 1881 for the first time. All three forms are present throughout the complete time-period from 1880 until the 1940s.

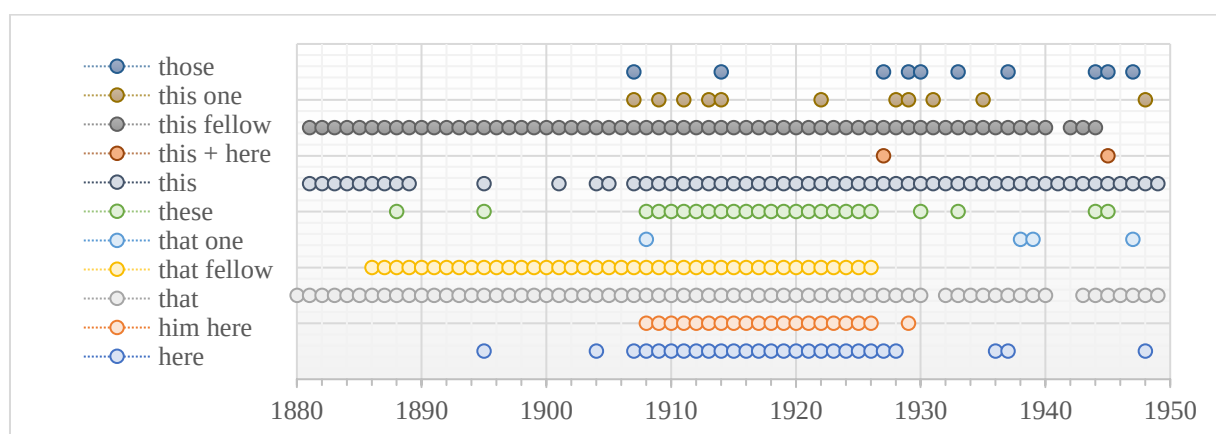


Figure 19: Timeline of demonstrative variants in SIP

The form *that fellow* occurs for the first time in 1886 but is only attested until 1926 in the written data. It is striking that the forms *this*, *that*, *these* and *those* are observed until at least 1945, if not even longer. A possible explanation for the frequent occurrences of *this* and *that* was provided above. The form *here* is first attested in 1895 and continues to be used until the end of the covered timeframe. The form *him here* is first attested in 1908 which suggests that *here* had stabilised by

that time so that a new form could develop out of the personal pronoun *him* and the form *here*. Attestations can be found in the written data from 1908 until 1929. The earliest evidence of the forms *this one* and *that one* dates to the years 1907 and 1908 respectively and both forms seem to have survived until the end of the here-portrayed timeframe though they are less frequently attested than other forms.

Figure 20 displays the distribution of forms (ST_FORM) across time in a boxplot visualisation. Though the timeline created in Excel and the boxplot created in R look similar, it should be kept in mind that the boxplot approach is based on the variable YEAR_DET. Thus, if no concrete date of attestation was identified but the sample was assigned to a timespan, the middle date of this period was chosen as the year of attestation, and therefore differences in the starting dates are possible.

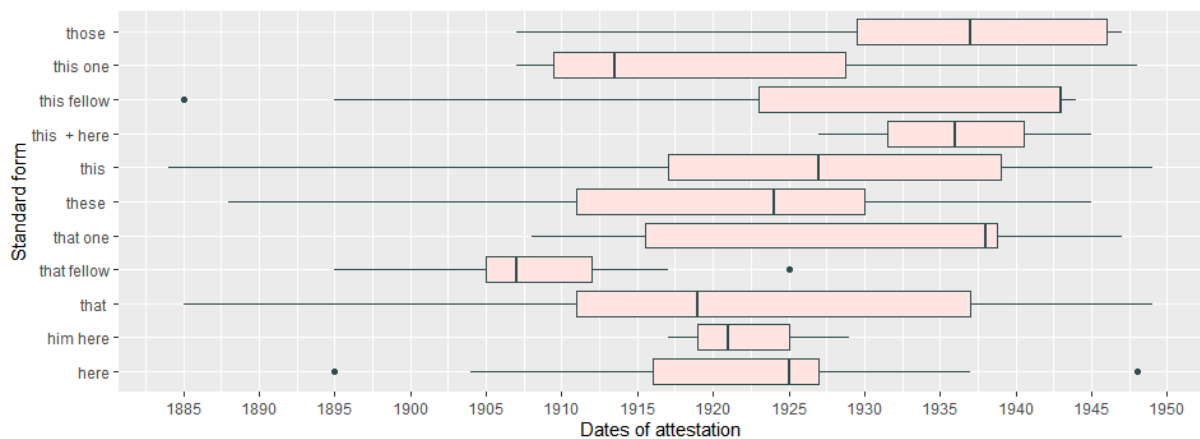


Figure 20: Boxplot of demonstrative variants in SIP

The boxplot illustrates what already became apparent in Figure 17, namely that most of the datapoints for SIP were attested post 1905. In addition, the boxes and whiskers of *this*, *that*, *this fellow*, *that one*, *this one*, *those* and *these* are stretched out far horizontally, implying that these forms were still in use by the end of the time period covered. The median of *this fellow* dates latest, lying in 1943. This might be an indicator for *this fellow* turning into the dominant form. However, as all other forms are still present, it is not visible whether, and if so *when*, competing forms were eliminated by solely focussing on the boxplot and timeline approach.

As the choice of demonstrative forms might be dependent on the syntactic context in which they occur, Figure 21 shows a boxplot in which it is differentiated between pronominal or adnominal demonstratives. The boxplots of *this one* and *that one* show that the forms were attested in pronominal position only. The figure furthermore shows that *this fellow* was only attested rarely in pronominal contexts, in which *here* was predominantly attested.

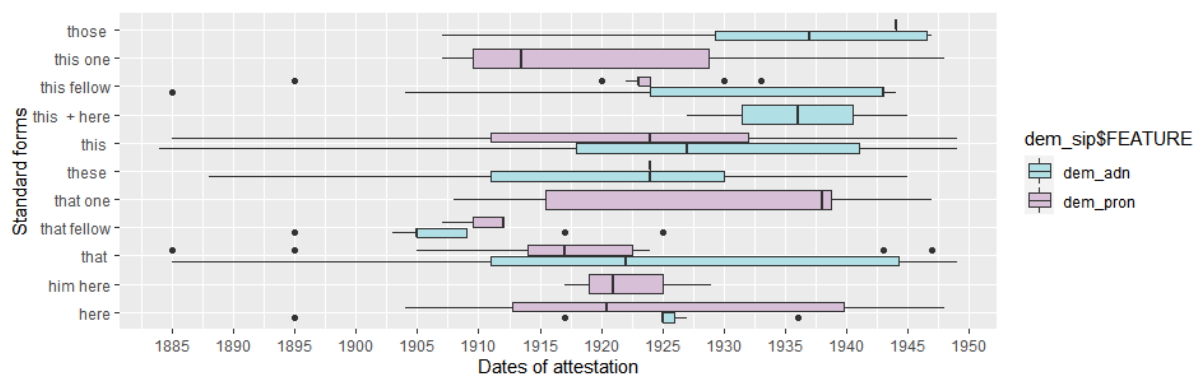


Figure 21: Boxplot of demonstrative variants in SIP depending on their syntactic context

6.5.2.3 Testing for the impact of the years of attestations on the choice of form

In the boxplot and timeline approach, frequencies and authorship were not taken into consideration. In addition, due to the presence of coexisting forms throughout the complete period under investigation, the application of conditional inference trees was necessary to test whether, and if so when, the data shows differences regarding the forms used to encode demonstratives. Figure 22 shows the tree obtained when analysing the impact of the determined year of attestation (= YEAR_DET) on the formal realisation of demonstratives (= ST_FORM). The recursive algorithm detected six highly significant splits ($p < 0.001^{***}$). The highest-order split is in the year 1919. The data before and in 1919 is most likely to make use of the forms *this* and *that*. A second split is observed in the year 1946. The data after 1946 seems to behave in a similar way to the data before and in 1919 in that the forms *this* and *that* are most likely to occur. Nodes 4, 5, 6 and 10 further subdivide the data into five periods (1920-1922, 1923, 1924-1942, 1943, 1944-1946). The end nodes seem to take turns in whether they show *this* (Nodes 7, 9, 12) or *this fellow* (Nodes 8, 11) as the most likely form to encode demonstratives. It needs to be noted that the periods created by the splits are very small (cf., for instance, Node 8, which only refers to the year 1923). Thus, the changes in the end nodes might represent preferences of individual authors or texts. Nonetheless, the dominant forms that can be observed to have been in use after 1919 are *this* and *this fellow*.⁹³

As the use of the form might be dependent on the syntactic context in which the demonstrative occurs, a conditional inference tree was created in which the syntactic context (= FEATURE) was considered as well. As Figure 23 shows, the algorithm has detected five highly significant splits ($p < 0.001^{***}$).

⁹³ If possible grammaticalised fixed deictic expressions were excluded from the calculations, a tree consisting of nine splits is obtained. The splits in the years 1919, 1922, 1923, 1942 and 1943 remain the same but further splits are observed in the years 1936, 1937, 1940 and 1945. Independent of whether the strings are included or excluded, *this fellow* only represents the most likely form in the period 1922-1923 and in the year 1943.

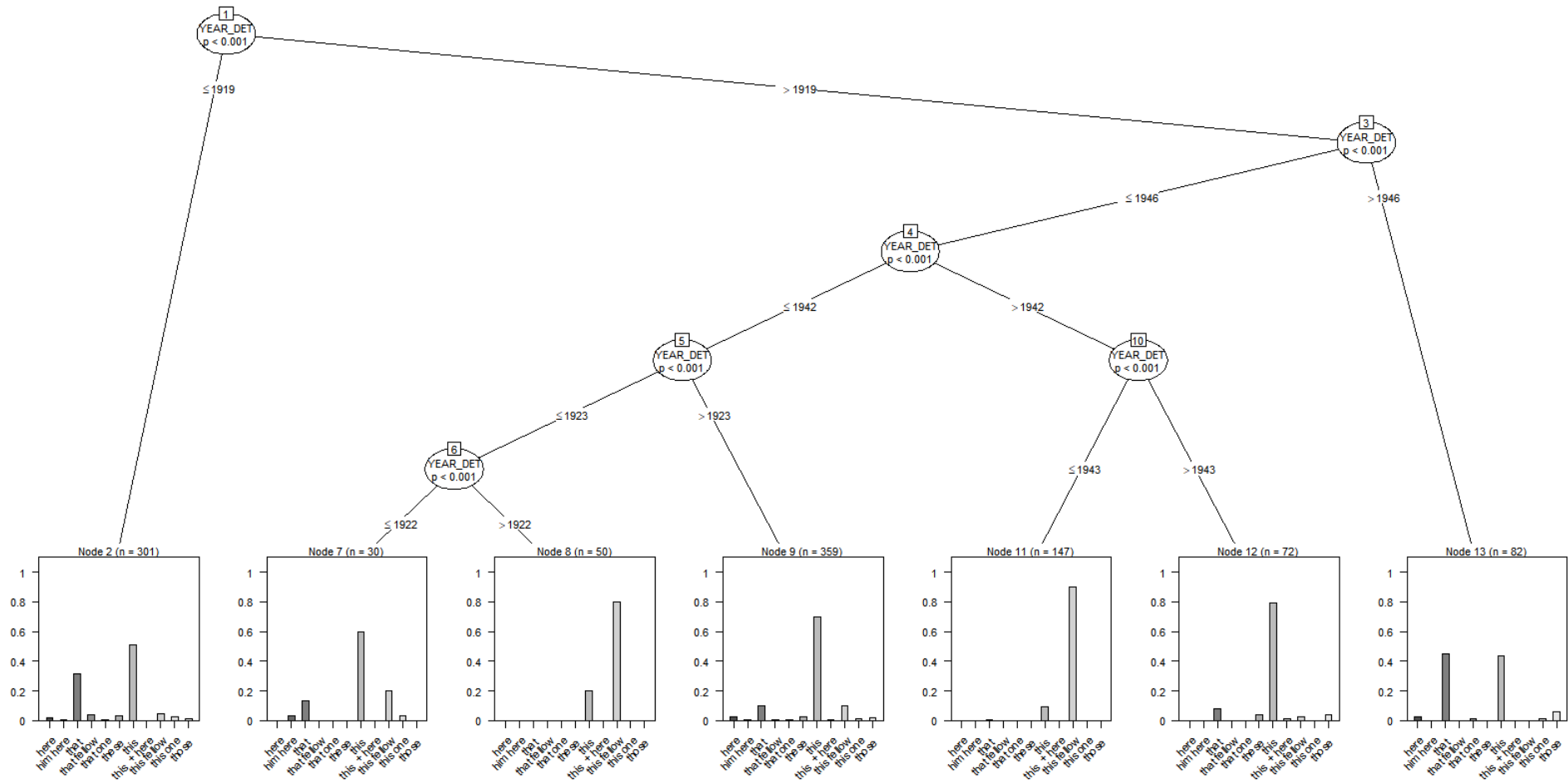


Figure 22: Conditional inference tree for demonstratives in SIP $ST\_FORM \sim YEAR\_DET$ ⁹⁴

⁹⁴ The conditional inference trees are based on the package *partykit*. I used the package *party* as well and compared the resulting tree structures to investigate whether the different algorithms lead to similar results. In the present case, the *party* package shows the same splits but solely in a different order.

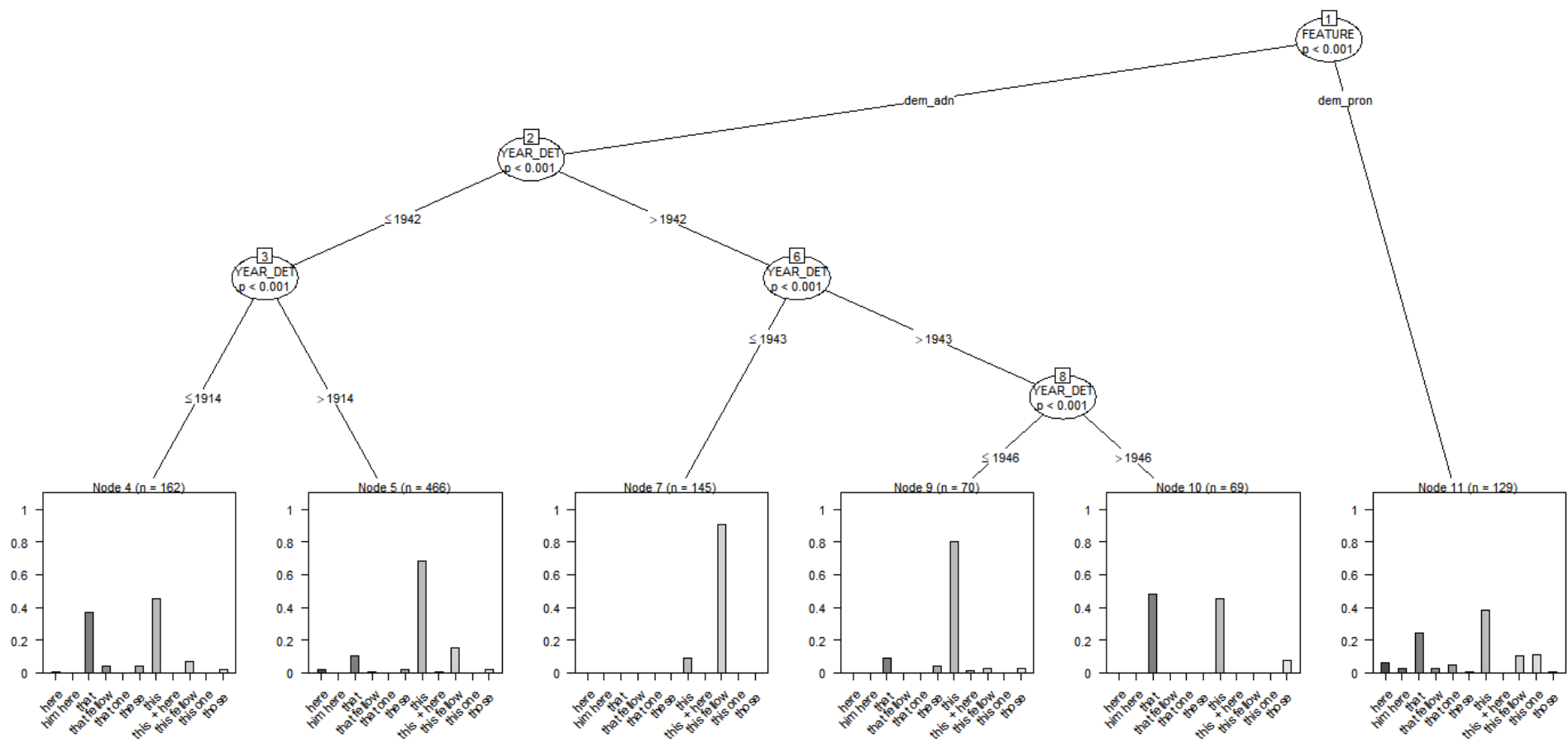


Figure 23: Conditional inference tree for demonstratives in $SIP\ ST\_FORM \sim YEAR\_DET + FEATURE$ ⁹⁵

⁹⁵ The conditional inference tree created with the *party* package led to the same tree structure and splits.

The highest-level split is based on the syntactic context, whereas the remaining four splits are based on the year of attestation. While the choice of the form to encode pronominal demonstratives seems to be independent of the year of attestation, the recursive algorithm has detected four time splits with adnominal demonstratives. It is important to note that a split in 1919 is no longer observed. Instead, the first time split is displayed in the year 1942. Furthermore, the years 1914, 1943 and 1946 are detected as split-evoking years. While adnominal demonstratives attested before 1914 and after 1946 are most likely to be encoded with the forms *this* and *that* (cf. Nodes 4 and 10), those attested in the years 1915-1942 and 1944-1946 are predominantly encoded with *this* (Nodes 5 and 9). Only in 1943 *this fellow* is detected as the most probable variant (cf. Node 7).⁹⁶

To test whether the text type, in which the demonstrative was attested, has an impact on the results, the text type category was added as a further predictor variable.⁹⁷ The highest-level split is still based on the syntactic context in which the demonstrative occurs. The time splits in the years 1942, 1914, and 1943 also remain the same. The split in the year 1946 is no longer observed. Instead, Nodes 5 and 9 indicate that in the period 1915-1942 and in 1943 the choice of adnominal demonstrative forms is dependent on whether they occur in speech-related or written and intermediate attestations. Especially for the year 1943 extreme differences are attested with speech-related attestations showing a predominant use of *this*, and written and intermediate attestations indicating a dominant use of *this fellow*. Post-1943 attestations are most likely to show the demonstrative *this* in speech-like and written attestations and *that* in intermediate attestations.

As the author may have an impact on the choice of form as well, a conditional inference tree based on the year of attestation, the syntactic context, the text type and the author was produced.⁹⁸ The resulting tree consists of 15 splits whereby the first split is based on the author variable.⁹⁹ In the writings of the authors Woodford, Norden, Collinson, Muspratt, Davies,

⁹⁶ If *this/that*-strings were excluded from the data that may have grammaticalized by that time, a tree with eleven splits is obtained. Independent of whether the strings are included or not, the highest-level split is based on the syntactic context and shows that the choice of form to encode pronominal demonstratives seems to be independent of the year of attestation. Ten time-based splits (= 1907, 1919, 1922, 1923, 1924, 1926, 1940, 1942, 1943, 1945) can be observed with adnominal demonstratives, whereby *this* represents the most likely form in six of the end nodes. *This fellow* remains the most likely form in the period 1922-1923 and in the year 1943.

⁹⁷ The tree can be found online at

https://www.dropbox.com/s/jhylvx18ft2a3gt1/Ctree_Demonstratives_SIP_TT.png?dl=0 (last access 29 September 2021). The same tree structure was obtained when doing the analysis with the package *party*.

⁹⁸ Including the author variable into the algorithm using the package *partykit* led to the error message “it cannot be searched for unordered splits in 31 levels”. As this is a bug of the *partykit* tool for which no solution exists so far, the tool developers suggest using the package *party* in those instances where the error message occurs (<https://cran.r-project.org/web/packages/partykit/news.html>, last access 13 May 2021). Therefore, ctree's including the author variable were built with the package *party*.

⁹⁹ The tree is too large to be presented here, but can be viewed online at

https://www.dropbox.com/s/1ktiieducy7i9o7/Ctree_Demonstratives_SIP.png?dl=0 (last access 29 September 2021).

Hogbin, Luke and Cormack *this fellow* represents the dominant attested form. While data of the first two authors is most likely to encode demonstratives with *this fellow*, *this*, *that fellow* and *that* (Node 3), the latter authors use *this* and *this fellow* as variants, with a split based on the text type (Node 4). The occurrences of *this* are higher in speech-related attestations than in written or intermediate ones.

The remaining authors of the first split show two further author-based splits (Node 7, 9) before a split can be observed based on the syntactic context (= FEATURE; Node 11). While pronominal demonstrative attestations show five further splits based on the author (Nodes 19, 21, 23, 25, 27), adnominal demonstrative attestations show differences in the choice of the demonstrative form depending on the text type (Nodes 12). Written and intermediate attestations show a further time-split in the year 1914. After 1914 a further text type and author-based split are observable though the end nodes show that *this* and *that* represent the forms which are most likely to occur. The trees suggest that the author is crucial in the choice of the demonstrative forms attested.¹⁰⁰

6.5.3 Diachronic analysis of demonstratives in Bislama

The present section focusses on the diachronic development of demonstrative particles in BIS. First, an overview of the attested forms will be provided before the focus will be placed on *when* the features were attested across time. The ctree algorithm will be applied in Section 6.5.3.3 to test whether the data at hand provides useful information regarding *when* a stabilisation or change in the use of demonstrative particles occurred in BIS.

6.5.3.1 Attested forms in BIS

A total of 13 different forms were attested in the Bislama dataset. *Here* (60/176 = 34.09%) represents the form which is attested with highest frequencies and occurs in adnominal as well as pronominal contexts:

- (47) 'oo 'ere boy where Mis Collins 'e bin kill 'im?
 who DEM boy REL Miss Collins PM PST kill TR
 'Who is this boy whom Miss Collins killed?'

(Paama 1916; Lynch 1923: 326)

¹⁰⁰ If strings that may have grammaticalised were excluded from the calculations, a very similar tree is obtained, consisting of 15 splits as well. There are eleven author-based splits, one text type split, one syntactic split and two year-based splits. The author-based splits on the first two levels are identical to the ones obtained when the strings are included. However, while the time-based split in 1914 remains the same, an additional split is attested in 1924 with demonstratives attested in the works of Collinson, Muspratt, Davies, Luke, Cormack and Hogbin. While *this fellow* represents the most probable form before and after 1924, *this* only represents an alternative form before 1924 for the first five authors. With Hogbin, both forms occur prior to and post 1924, but after 1924, the use of *this* becomes less likely and the use of *this fellow* increases.

The second most frequent form is *this* (40/176 = 22.73%), which is directly followed by *that* (33/176 = 18.75%) (cf., for instance, example (19)). As both forms are attested to occur in travel reports, court proceedings as well as letters written by Pacific Islanders, it can be assumed that the StE forms were used by Europeans as well as non-Europeans as variants to encode demonstratives. What is interesting is that in contrast to SIP only two tokens were identified in which the demonstratives collocate with the noun *time* and that a fixed adverbial expression has not grammaticalised.

Next to these forms, *that fellow* occurred 15 times (= 8.52%) and *this fellow* 11 times (= 6.25%) in the data (cf. (26) for an example of *that fellow*). *Him here*, being used as a pronominal demonstrative, is attested in 8 out of 176 tokens (= 4.55%):

- (48) *him ere* *which way* *you* *no* *come* *look* *me feller* *long* *place* *here*
 DEM why 2SG NEG come see 1PL PREP place DEM
 ‘... this is why you do not come to see us at this place.’
 (Espiritu Santo 1951; Letter by A.M. in Tyron & Charpentier 2004: 365-366)

Other attested forms are *these*, *these fellow*, *them*, *them fellow*, *this one*, *those*, and a form combining *this* and *here*, all occurring with a frequency of less than 1.5%. For instance, *them* is attested twice in the early data. The occurrences are found in two distinct letters by the islander Albert Dar dating to the years 1916 and 1919 (cf. NHBS 1/I Vol 1, 17/1914). In these letters *them* is used instead of the demonstrative *those*, which, as explained earlier, is a feature which can also be found in non-standard varieties of (Australian) English. A closer look at the sentences in which *them* was attested (cf. Sentence (24) and (49)) makes clear, however, that though there are no indicators that the letters were written *on behalf of* Albert Dar, the writer’s language is closer to StE than to PE. For instance, in (49) the author uses the 1SG pronoun *I* and the adverb *quickly* which are features that were not common in the PE of that time. In addition, the sentence may allow for a different analysis of *them* as well: *them* might not represent the demonstrative particle but may represent the transitive marker of the verb *take out*. It is possible that we find *them* instead of *him* due to author modification. However, as the form *them fellow* was also attested once in a position where StE would make use of the demonstrative particle *these* (cf. sentence (50)), it is likely that *them* and *them fellow* were occasionally used by some speakers as demonstratives. Pacific Islanders were in close contact with Australians and colonisers might not have spoken StE. The form may thus have entered early MPE through these contacts. Still, the fact that there are only three attestations which include *them* is an indication that the morpheme did not represent a common demonstrative particle.

- (49) *I want you take out them three Boys quickly.*
 1SG want you take out **DEM** three boy-PL quick-ADV
 ‘I want you to take those three boys out quickly.’

(Ambrym 1916; Dar 13.08.1916)

- (50) *them fellow boat he no good*
DEM MODIF boat PM NEG good
 ‘these boats are bad’

(Espiritu Santo 1882; Colonial Office CO225/13 1882-83: 390)

Though *these*, *those*, *these fellow* were only attested seldomly, it is possible that the forms were occasionally used. However, the small frequencies of occurrence indicate that the particles did not represent dominant demonstrative forms used in the variety.

6.5.3.2 Timeline and boxplot approach

So far, the demonstrative tokens have only been examined without considering their dates of attestation. Figure 24 shows the occurrences of the attested forms across time, though ignoring the data distribution and frequencies of occurrence. Based on the timeline it can be observed that *this* is the earliest demonstrative form attested in 1831. It is an isolated attestation, as the next documented occurrences of demonstratives date 34 years later. In the late 1860s a variety of forms including *this*, *that*, *that fellow* and *here* were attested, giving rise to suspicion that the forms coexisted and were used interchangeably by that time. It is also interesting that the first occurrence of the form *this fellow* is in 1871 and thus, the form occurs two years after the first attestation of *here*. While the timeline visualises the coexistence of competing features until 1929, only the forms *here* and *him here* are attested after 1930. Due to the BIS data gap in the years 1930 until 1950 (cf. Figure 17), this observation is not very meaningful.

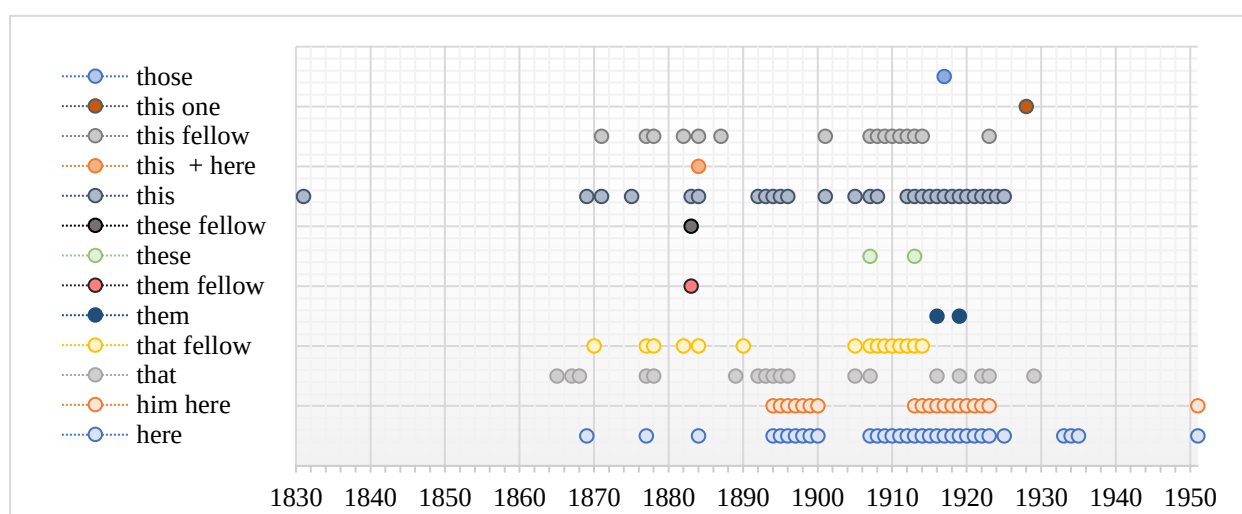


Figure 24: Timeline of demonstrative variants in Bislama

As the timeline ignores the frequency and distribution of datapoints across time, the boxplot in Figure 25 visualises the distribution of the various demonstrative forms across time based on the

continuous variable YEAR_DET, and thus on concrete years of attestations. The first thing to note is that the boxplot contains vertical lines without a box, which means that the forms were attested in a single year only. The forms *those*, *this one*, *these fellow* and *them fellow* are thus only attested in a single year in the BIS data. *These fellow* and *them fellow* are attested in 1883 which speaks for the previously mentioned hypothesis that the forms represent overextensions. Although *them* and *these* occur more than once, they are attested with low frequencies and do not spread over a long period of time.

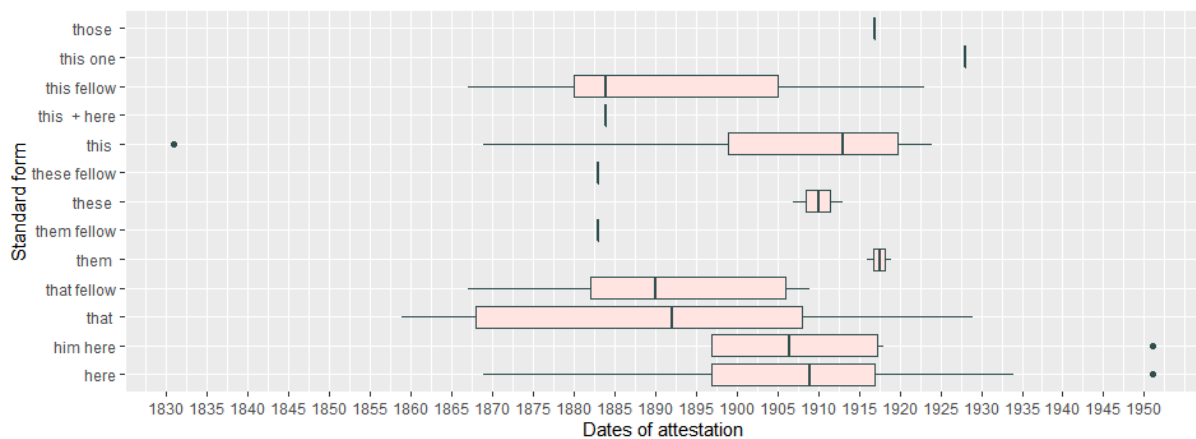


Figure 25: Boxplot of demonstrative variants in Bislama

The boxes and whiskers of *this*, *this fellow*, *that* and *here* stretch out rather far horizontally, almost all showing whiskers that reach the year 1923. This shows that the four forms were present in Bislama at least until 1923. Though *this* represents the earliest form attested, *this fellow* occurs already as early as in 1857 in the written dataset. The forms *this* and *that* seem to have coexisted with *this fellow* and *that fellow* for a long period, although the use of *that fellow* declined. It is interesting to note that *here* is already attested with a demonstrative reading before 1870, the box starting around 1897 and its median lying around 1909. Despite being an outlier, there is also a datapoint in 1951 and the whisker stretches out until 1935. The form *him here* is attested from 1897 onwards. Although its box covers the same time period as the box of the form *here*, the latter shows a whisker on the left side and was attested earlier. It makes sense to assume that the form *here* had to stabilise before it merged with the pronoun *him* so that the new form *him here* could develop. Since the data for Bislama does not represent the variety well enough in the years 1930-1950, it is unfortunately not possible to show how the forms developed in these twenty years. Due to the two outliers attested in 1951, it can only be said that the forms *here* and *him here* were still present.

While the medians of *this fellow*, *that* and *that fellow* lie before 1900, the medians of *this*, *him here* and *here* lie after 1900 and, therefore, the latter seem to represent the dominant attested

forms after 1900. A change in the choice of the demonstrative form may thus be observable around 1897.

As the use of forms might be dependent on whether the demonstrative is used in adnominal or pronominal contexts, Figure 26 shows a boxplot which differentiates whether the forms are used as adnominal determiners or as pronominal pronouns. What is interesting to note is that *here* is predominantly attested in adnominal positions and it is only occasionally found as a pronominal pronoun. In contrast, *him here* is only attested in pronominal contexts. Still, the boxplots give the impression that around 1897 some changes occur.

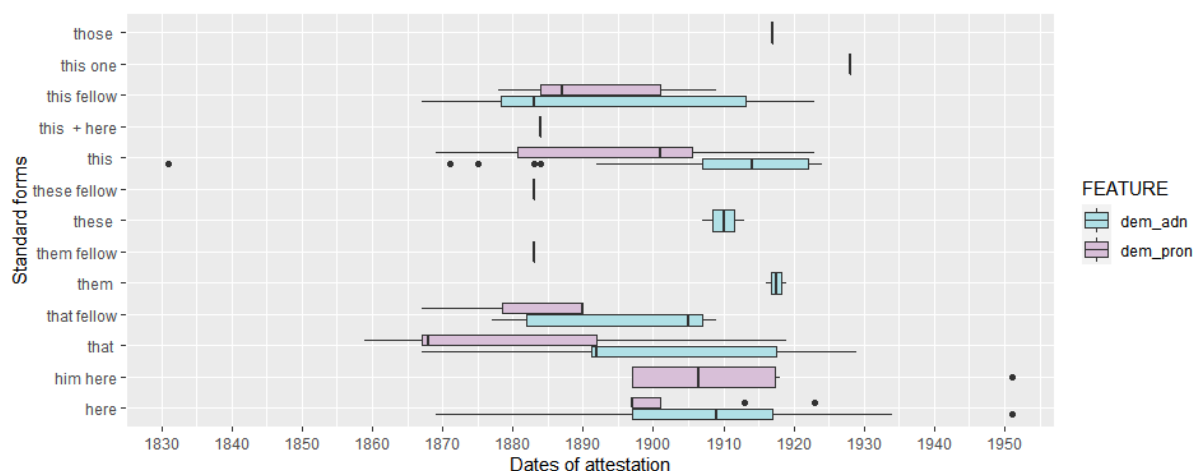


Figure 26: Boxplot of demonstrative variants in BIS depending on their syntactic context

6.5.3.3 Testing for the impact of the years of attestations on the choice of form

Conditional inference trees were created to test whether the year of attestation has an impact on the encoding of demonstrative forms. Figure 27 shows whether the year of attestation has a significant impact on the formal realisation of demonstratives in BIS.

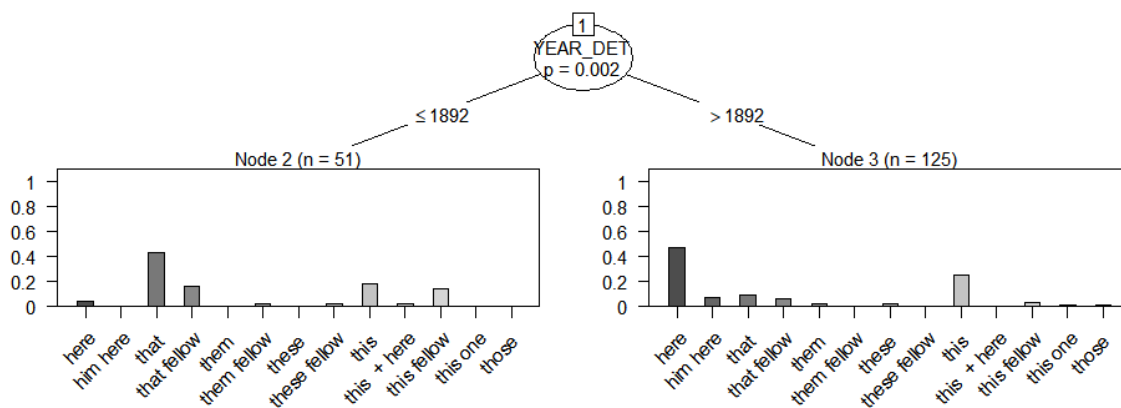


Figure 27: Conditional inference tree for demonstratives in BIS YEAR_DET~ST_FORM¹⁰¹

¹⁰¹ The package *party* yields two splits, one in 1868 ($p=0.002^{**}$) and one in 1893 ($p=0.031^{**}$).

The tree shows a single split in the year 1892, which is very significant ($p < 0.002^{**}$). While demonstratives before and in 1892 have a significantly higher probability to be encoded with *that*, or, with *that fellow*, *this* and *this fellow* etc. (albeit less frequently), demonstratives occurring in the data after 1892 are most likely to be encoded with *here*. The form *this* represents the second most likely form in post-1892 attestations.

Considering the choice of form may be influenced by the syntactic context in which a form is used, the variable FEATURE was added resulting in a tree with two splits:

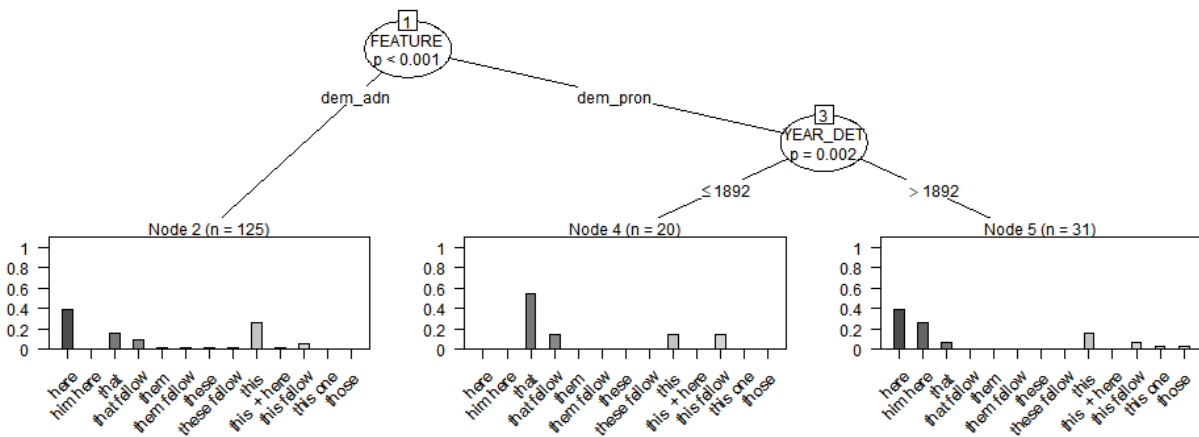


Figure 28: Conditional inference tree for demonstratives in BIS YEAR_DET~ST_FORM+FEATURE¹⁰²

The first split is highly significant ($p < 0.001^{***}$) and shows that adnominal demonstratives behave differently from pronominal demonstratives. With adnominal demonstratives the forms *here* and *this* represent the most likely forms independent of the year of attestation. With pronominal demonstratives, there is a very significant split after the year 1892 ($p = 0.002^{**}$). While in post-1892 data *here* and *him here* represent the dominant pronominal forms, these forms are not attested in pronominal contexts before 1892 in which *that* is attested with the highest frequencies and *this*, *this fellow* and *that fellow* represent less frequently used variants.

A tree including the text type was constructed to investigate the impact of the source on the choice of form. The resulting tree structure in Figure 29 shows that the first and most important split remains the one based on the syntactic context ($p = 0.001^{***}$). The choice of form is dependent on whether the demonstrative replaces a noun/noun phrase or whether it co-occurs with it. In the latter case, the choice of form is further dependent on whether the demonstrative occurs in speech-related attestations (variant 1) or in written and intermediate attestations (variants 2 and 3). While in speech-related attestations *here* represents the dominant form, demonstratives in written and intermediate attestations are most likely to be encoded with *this*.

¹⁰² The package *party* yields two splits as well. It also splits the data in terms of adnominal vs. pronominal and then shows a split in the year 1868.

This can be explained by the fact that *here* was originally a spatial adverb which usually makes the more sense in face-to-face interactions. The significant split in Node 4 ($p=0.012^*$) highlights that written and intermediate attestations further show differences in whether they were attested before or after 1914. The written and intermediate datapoints dating before 1914 illustrate the tendency to make use of the forms *this* and *these* with *this* being the dominant variant. Post-1914 datapoints are most likely to be encoded with the form *this* but the forms *here*, *that* and *them* are also possible. For pronominal demonstratives the text type does not seem to be of importance and the previously observed split in the year 1892 (cf. Node 7) remains very significant ($p=0.005^{**}$).

The choice of form may also be dependent on the author of the source document which is why a ctree including the year of attestation, the syntactic context, the text type and the author was created.¹⁰³ The first three splits of the resulting tree are with respect to the author. Node 6 splits the data with respect to the syntactic context and Node 8 according to the text type. It turns out that the year of attestation does not seem to be crucial in terms of the choice of the forms used to express demonstratives if the author is included into the ctree algorithm.

Even though the author represents the most dominant predictor variable, it needs to be noted that this result does not invalidate the data but shows that given the nature of the sources, information about *the author* is important. The aim of using statistical tools, such as ctree, is to identify patterns in the data. If in a genre-based discourse analysis using statical methods the genre turns out to be the most important predictor variable, it does not mean that the *genre* is not reliable, but it only indicates that the genre is one pattern based on which differences in the data can be observed. Similarly, if the author turns out to be the most important predictor variable in the present study, this does not mean that the author is not reliable. It only shows us that the linguistic forms/features we identify should always be evaluated in the light of the author. In other words, the ctree of demonstratives in BIS shows us that the author as well as the text type and the syntactic context in which the demonstratives occur have some predictive power, but this does not mean that the data or the author is not reliable.

¹⁰³ The tree was built with the algorithms of the *party* package. It can be viewed online at https://www.dropbox.com/s/fv0nvzbw9bk7om9/Ctree_Demonstratives_BIS.png?dl=0 (last access 29 September 2021).

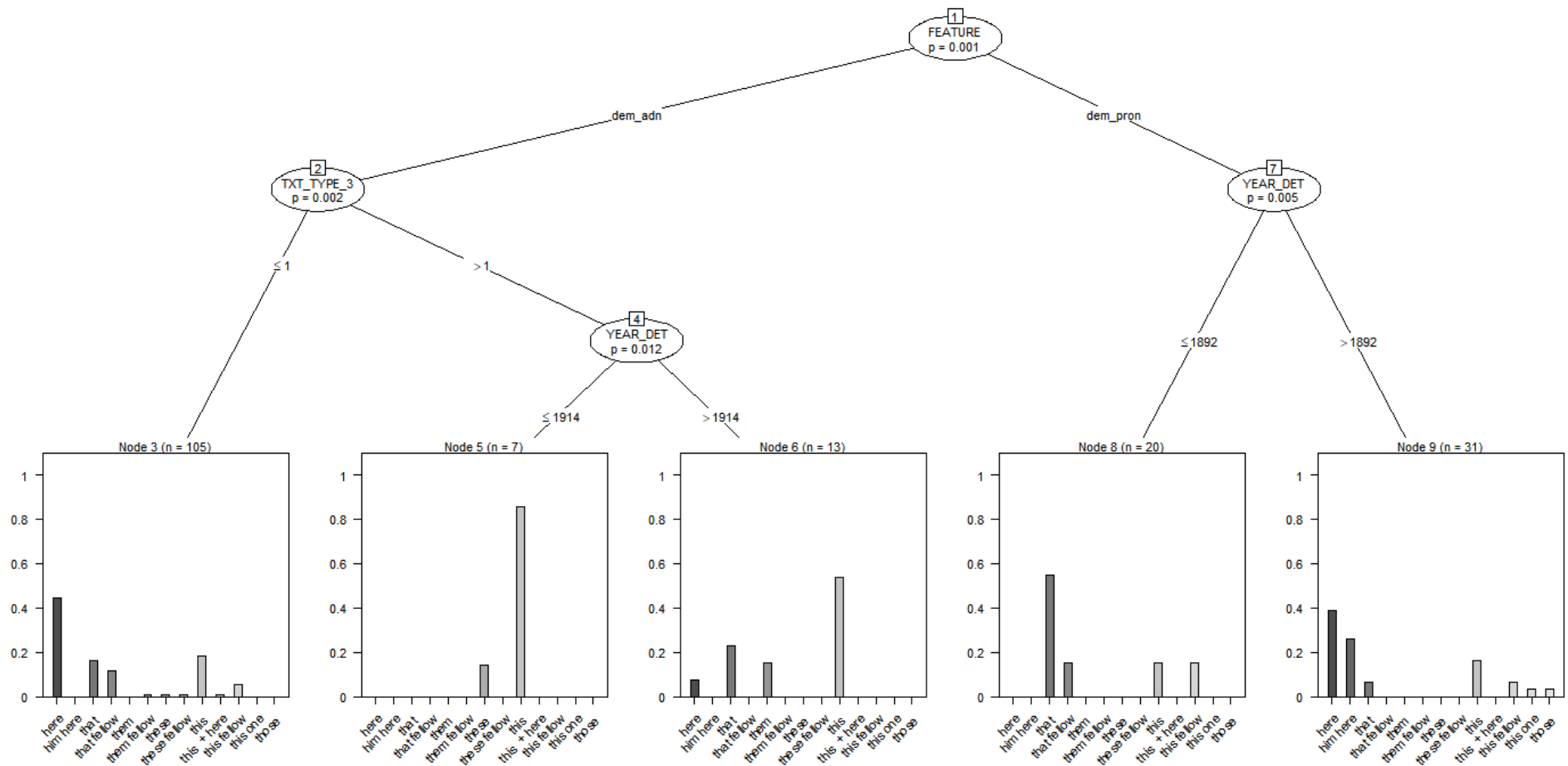


Figure 29: Conditional inference tree for demonstratives in $BIS\ ST\_FORM \sim YEAR\_DET + FEATURE + TEXT\_TYPE\_3$ ¹⁰⁴

¹⁰⁴ The number of splits is the same in the tree received with the *party* package. All splits are the same, despite the year split (7) which is 1868 in *party*.

6.5.4 Diachronic analysis of demonstratives in Tok Pisin

This section focusses on the diachronic development of demonstrative particles in TP based on pre-1950 attestations. The attested forms will first be introduced before the timeline and boxplot approach will be applied. The results of the ctree analysis in Section 6.5.4.3 will show whether time represents a significant predictor variable in the choice of forms used to encode demonstratives.

6.5.4.1 Attested forms in TP

This fellow is the demonstrative form that occurs with highest frequencies in the TP data (459/586 = 78.33%). It is used as a demonstrative pronoun and as a demonstrative determiner:

- (51) *snek i kam raus long disfelo hul*
 snake PM come out PREP DEM hole
 ‘The snake came out of this hole.’

(Mugil District ~1930; van Baar 1930: 17)

- (52) *Me no like this fella, masta*
 1SG NEG like DEM master
 ‘I do not like this, master.’

(New Guinea 1940, Forsyth 1942: 3; AU ANUA 305-66)

At least 32 of these 459 attestations have a plural meaning, yet they are not overtly marked for plurality. It is also interesting to note that in almost 28.76% of the 459 attestations *fellow* was orthographically realised with the letter <p>. This indicates that at least some speakers replaced the labiodental fricative [f] with [p].¹⁰⁵

The second most frequently attested form is *this* (52/586 = 8.87%), which is attested in singular and plural contexts. In plural contexts it may directly precede a numeral:

- (53) [...] *na kaun-im ologeda liklik kopra i stap long dis-tu-pela bag*
 [...] and count-TR all little copra PM LOC PREP DEM-two-MODIF bag
 ‘and count all the little copra which is inside of the two bags’

(New Guinea War pamphlet 1943-1945; Kerr 1985; NLA MS9002 P106)

It is interesting to note that the forms *that* and *that fellow* only occur 10 (= 1.37%) and 8 (= 1.17%) times in the Tok Pisin dataset. Thus, a clear preference of *this* and *this fellow* can be observed. In 5.97% of all tokens (= 35/586) the form *here* is attested. Other forms that were present in the early

¹⁰⁵ In general, almost all attestations in which *fellow* was orthographically realised as *pela*, *pella*, *pala*, etc. were attested in the TP data. Only a few attestations were found in SIP. Interestingly, all these attestations are from a single source, namely, a war pamphlet written by the Japanese. As explained in Section 3.1.6, the allied troops usually learned a single variety of MPE. Because New Guinea was the area that was most extensively involved in the war, the variety of the region was taught and learned. Thus, the attestations might rather indicate and reflect the pronunciation that was common in early TP. It is also possible that the attestations indicate the Japanese’ perception of how the modifier was pronounced.

TP data are *this one*, *that one*, *these*, *these fellow*, *him here* and *this fellow* + *here* which all occur with a frequency of less than 2%. The forms *these*, *these fellow* and *this one* represent once attestations. The former two occur in the novel of the Australian writer Idriess. As the forms are only attested in a single source, they seem to have their origin in author modification and cannot be considered typical demonstrative markers of early Tok Pisin.

6.5.4.2 Timeline and boxplot approach

The timeline in Figure 30 shows how the occurrences of the attested forms spread across time.

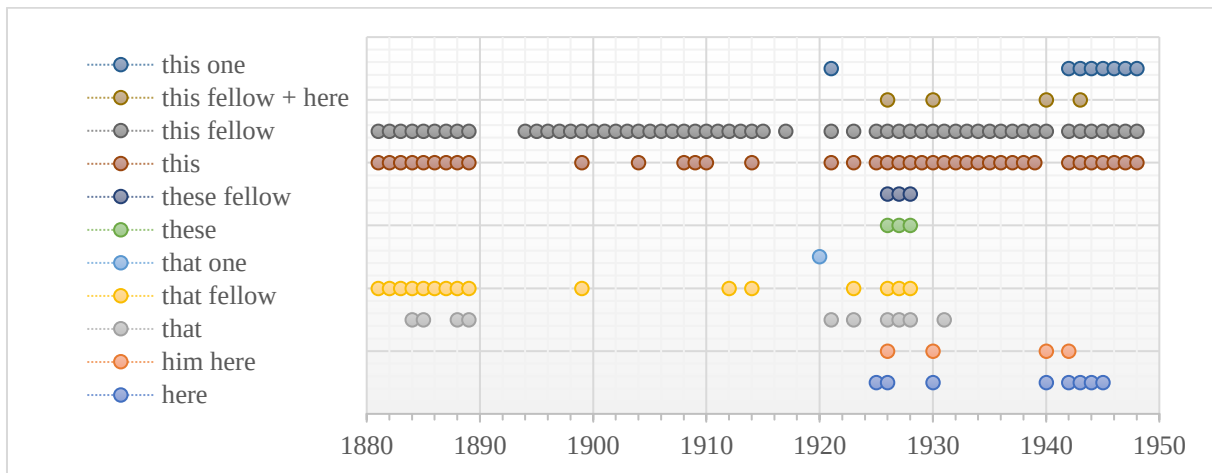


Figure 30: Timeline of demonstrative variants in TP

This, *this fellow* and *that fellow* represent the earliest used forms. They are first attested in the early 1880s but only the former two forms are present throughout the time period presented here. *That fellow*, as well as the StE-derived form *that*, are first attested in the 1880s, but after 1931 written attestations are not available. It is interesting to note that after 1920 new forms such as *this one*, *here*, *him here* are first attested in the written data and their usage seems to continue until the end of the covered period. The attestations of *these fellow* and *these* refer to 1926–1928 and, as pointed out earlier, seem to reflect idiosyncrasies.

As the timeline does not consider the frequency and distribution of demonstrative datapoints, the boxplot in Figure 31 was created. The boxplot visualisation reconfirms that the forms *these*, *these fellow* and *this one* were each attested in a single year only. The median of *this fellow* lies in the year 1943 and the median of *this* in 1927, which shows that although the forms are present from the 1880s onwards, most of the attestations date after 1920. The majority of datapoints of *that* and *that fellow* are attested earlier. Based on the boxplot visualisation it appears as if around 1921 major changes can be observed in the written data. The use of forms such as *that* and *that fellow* seems to decrease, while the use of forms such as *this* and *this fellow* seems to increase. Additional new forms such as *this one*, *here* and *him here* are attested as well.

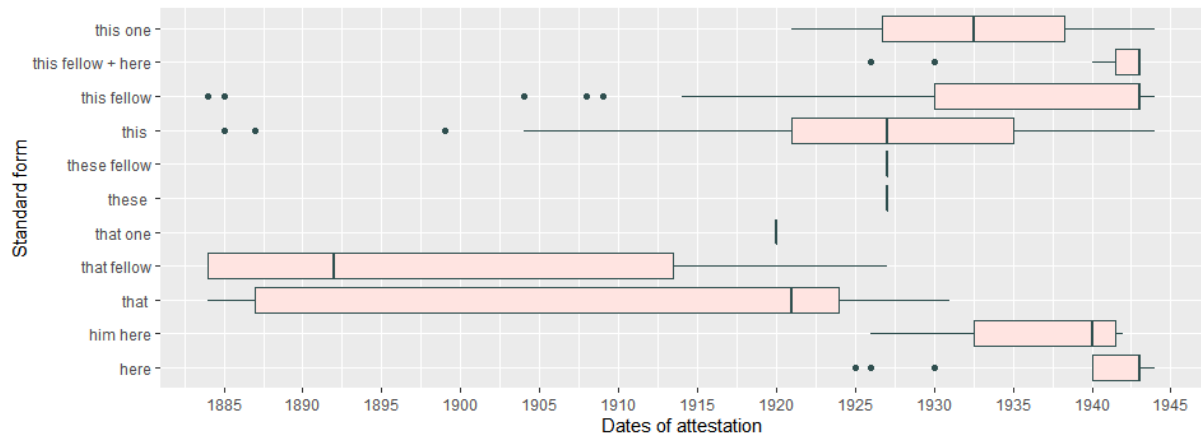


Figure 31: Boxplot of demonstrative variants in TP

Figure 32 differentiates between forms used to encode demonstrative pronouns and those used to encode demonstrative determiners. It becomes evident that only the forms *this fellow*, *this*, *that fellow*, *that* and *here* encode adnominal demonstratives in the early data. In terms of pronominal pronouns *this* represents the dominant form until around 1920. *This fellow*, *that*, *him here* and *this fellow + here* turn into the dominant forms post 1920.



Figure 32: Boxplot of demonstrative variants in TP depending on their syntactic context

6.5.4.3 Testing for the impact of the years of attestations on the choice of form

The assumptions made so far did not consider statistical significance. As with SIP and BIS, the recursive algorithm of ctree was used to investigate whether the year represents a predictor variable for the choice of demonstrative forms. Figure 33 shows the resulting tree based on the year of attestation and the forms used to encode demonstratives. The tree consists of seven time-splits, the first split being in the year 1899 (Node 1, $p < 0.001^{***}$). Before and in 1899 the forms *that*, *this*, *that fellow* and *this fellow* were used simultaneously, with *that fellow* being most likely to occur. The second split is detected in 1927 (Node 2, $p < 0.001^{***}$). Attestations dating from

1900 until 1927, in contrast, show a higher likelihood of *this fellow* and *this*, the probability for *that* and *that fellow* is very low. The next time split is observed in 1943 (Node 5; $p=0.045^*$). The data attested in the years 1928-1943 is further subdivided in the year 1935 (Node 6; $p=0.024^*$). Although there are further time splits (Nodes 7, 10, 11), these splits produce very small periods, sometimes covering a single year and may thus represent idiosyncrasies of an author or text. For instance, in Node 8 *this fellow* is shown as the dominant form whereas in Node 9 *this* is considered as the dominant form. Since the data in Node 9 refers to a single year only, namely 1935, and since the number of n is very low as well ($n=12$), the differences do not seem to reflect realistic changes in the language. After 1927, the form *this fellow* represents the most dominant attested form, but the occurrence of forms such as *here* and *him here* represent further post-1927 data observations.

Because the use of the form might be dependent on the syntactic context in which the demonstrative occurs, a tree resulting from using both the time and syntactic context predictor variables was created and can be found online.¹⁰⁶ The algorithm detected several splits, with the first split being the same as in Figure 33. Thus, the year 1899 remains the most significant factor independent of whether the syntactic context is considered or not. However, the split on the next level is based on the syntactic context. Only post-1899 demonstratives attested in adnominal contexts still show a highly significant time split in the year 1927 (Node 4; $p<0.001^{***}$). The further subdivisions by the year of attestation in Nodes 6 and 7 are assumed to again represent idiosyncrasies of individual texts or authors. A comparison of the Nodes 8, 9 and 10 shows that while in Nodes 8 and 10 *this fellow* is the dominant form, in Node 9 *this* represents the most probable form. As Node 9 again only covers a single year (= 1935) and is only based on 12 tokens, it is likely to present idiosyncrasies of a single text or author. Pronominal demonstratives are split in the year 1921 (Node 11, $p=0.017^*$). Pronominal demonstratives occurring between 1900 until 1921 show an equal amount of attestations of *this* and *this fellow* (*that one* and *that* represent less frequent variants. The data attested after 1921 indicates a dominant use of *this fellow*.

Applying the ctree algorithm again and including the text type as a further possible predictor variable, the splits on the first three levels (Nodes 1, 3, 4, 21) remain the same.¹⁰⁷

¹⁰⁶https://www.dropbox.com/s/l698kjfp0slmqhq/Ctree_Dem_TP_ST_FORM~YEAR_DET%2BFEATURE.png?dl=0 (last access 29 September 2021).

¹⁰⁷ Due to the size of the resulting tree, it is only accessible online at https://www.dropbox.com/s/9oo8hx0uhci5qes/dem_TP_ST_FORM~YEAR_DET%2BFEATURE%2BTXT_TYPE_3.png?dl=0 (last access 29 September 2021).

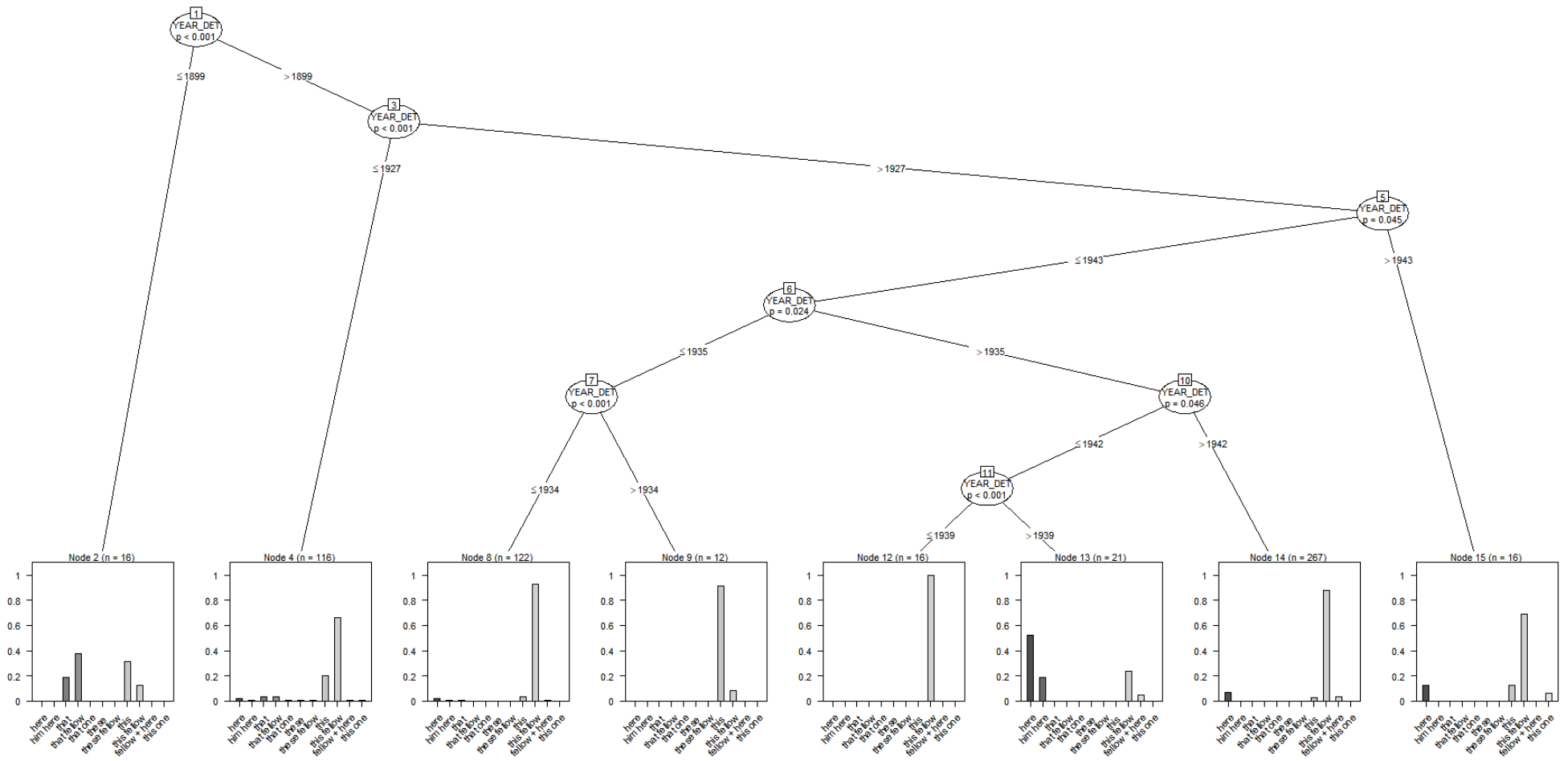


Figure 33: Conditional inference tree for demonstratives in $TP\ ST\_FORM \sim YEAR\_DET$ ¹⁰⁸

¹⁰⁸ The package *party* yields splits in 1884, 1921 and 1925.

Thus, a highly significant split occurs based on the time predictor variable in the year 1899. The second split is based on the syntactic context and on the third level, adnominal determiners show a further time split in 1927, while demonstrative pronouns show a time split in the year 1921. Only for adnominal attestations between 1928 and 1950 the text type seems to be a further predictor variable for the choice of the demonstrative forms. Since the years 1899, 1921 and 1927 are not dependent on the text type, they may reflect important dates in terms of changes in the choice of demonstratives in TP.

Next to the text type, the author may be a predictor variable that influences the choice of the demonstrative form. Therefore, the impact of the year of attestation, the syntactic context, the text type and the author on the demonstrative form was tested. The resulting tree consists of eight splits, of which six are highly significant.¹⁰⁹ The first four splits are with respect to the author. The two splits on the fifth level are based on the syntactic context and Node 9 and 11 split the data based on the author and text type again. It turns out that the author represents the strongest predictor variable and that no significant results regarding the year of attestation are obtained.

6.5.5 Diachronic comparative analysis of demonstratives in MPE

Based on the observations of the individual analyses, the following assumptions can be made about the diachronic development of demonstrative markers in early MPE. All three varieties show attestations of *this* and *that*. The forms *this fellow* and *that fellow* could be attested in all three regions as well.

Though there is evidence that the forms *this* and *that* were in use in the early period of the varieties' development, the forms differed from StE in that they were also used in plural contexts. The early sources give further rise to suspicion that phonological adaptations took place. Orthographic realisations such as *dis* and *dat* show that *th*-stopping seems to have been common. Realisations of *fellow* as *pela/pala/fela* indicate that *fellow* differed from its StE pronunciation as well. For instance, the labiodental fricative [f] was replaced with [p] by some TP speakers. As the analysis has shown, *this* and *this fellow* are more frequent than *that* and *that fellow* in all three varieties.

The form *here* entered the varieties as a locative adverb. It was first attested as a demonstrative particle in the BIS dataset in 1869. In SIP it was first attested in 1895 and in TP in 1924. The grammaticalisation of the locative adverb *here* into a demonstrative marker can also

¹⁰⁹ The tree was built with the algorithms of the *party* package. The tree is available at https://www.dropbox.com/s/mzsq7w007gxuan/Ctree_Demonstratives_TP.png?dl=0 (last access 29 September 2021).

be observed in other pidgins and creoles (cf. Heine & Kuteva 2004: 173). What is interesting is that the form is not attested to fulfil the function of a demonstrative marker in the early Samoan and Queensland plantation pidgin data in which it is only attested as a demonstrative adverb. It may thus have its origin in Vanuatu. It may be of importance that in French the locative adverbs *ici* ‘here’ and *là* ‘there’ have grammaticalised into postnominal elements that form part of the demonstratives (see Section 6.1). As the French were present in Vanuatu, it is possible to imagine that French language structures had an impact on the development and stabilisation of the postnominal marker. In addition, Camden (1979: 76) and Siegel (2008: 183-184) have successfully shown that Bislama’s substrate languages make use of postnominal demonstratives as well. This is of interest as the postnominal demonstrative *here* dominates in Bislama, whereas prenominal forms dominate in Tok Pisin and Solomon Islands Pijin. Thus, the postnominal form *here* may have been reinforced in Bislama by similar structures being present in substrate languages.

The ctree analysis showed that the author represents the greatest predictor variable for the choice of demonstrative forms in all three varieties. If the author was not considered, the trees constructed with ctree indicated that in TP a change in the data can be seen in 1899. Data attested before this year consisted of the four mentioned forms, whereas data attested from 1900 onwards predominantly contained the form *this fellow*. In Bislama, the year 1892 proved to be a significant year regarding the choice of form, with *here* turning into the most likely form in pronominal syntactic contexts. Nonetheless, competitive forms do not fall out of usage yet and due to insufficient data covering the years 1935 until 1950, the results are not meaningful. In contrast to the BIS and TP data, a tendency towards specific forms is not observable in the early SIP data due to a high degree of variation attested until the end of the observed period. This may be explained by the fact that the variety ended up being used on the home plantations much later than in the other two areas.

Based on the data at hand, the analysis does not provide an answer regarding the question when the stabilisation and choice for one (or several) of the demonstrative particles took place, as different forms coexisted until the end of the time period considered in this study.

6.6 Summary

The present chapter discussed the diachronic development of demonstratives in SIP, BIS and TP. It started with an introduction of the theoretical background (6.1) before insight into contemporary demonstrative forms in SIP, BIS and TP was provided (6.2). The section on previous diachronic research (6.3) demonstrated the necessity for further research on

demonstrative particles and was followed by an outline of the methodological consideration (6.4). The subsequent Section 6.5 focussed on demonstrative forms and their dates of attestations in early SIP, BIS and TP.

The analysis has shown that the early MPE datasets were characterised by a high degree of variability with StE similar forms being used next to new forms such as, for instance, *that fellow*, *this fellow*, *here* and *him here*. The latter four forms are of special interest as these are the most common forms in the contemporary varieties. The analysis has shown that although in BIS and TP the amount of competing forms was reduced by the end of the observed period (or at least a preferred form existed), competitive forms did not yet fall out of use. Although the year 1899 in TP and 1892 in pronominal syntactic contexts in BIS proved to be significant years regarding the choice of form, the author turned into the dominant predictor variable as soon as the variable was considered in the algorithm as well. Nonetheless, it is of interest that these split evoking years date prior to the end of the labour trade.

The coding of the data proved difficult. The multifunctionality of morphemes, an anglicised spelling and missing contextual information made the analysis of some datapoints impossible. In addition, the different sizes of the databases represented the challenge of receiving wrong results or making wrong judgements. The data for the analysis could not be collected according to the same standards but is based on the surviving written record. This leads to imbalanced datasets. A major problem of the BIS data is that insufficient data for the time period 1930 to 1950 is available and the SIP data consists predominantly of material that is connected to the South Sea Evangelical Mission (SSEM). For future studies it will be of importance to fill the BIS data gap and to find SIP attestations in further non-religious sources in order to obtain more balanced datasets.

Post-1950 data is required to reconstruct the development and stabilisation of demonstrative particles in the three varieties and to see whether the lectal diversity is reduced or whether those differences can be observed until today. Corpora covering the time period from the varieties' earliest attestations until today are necessary to understand the complete grammaticalisation process of demonstratives in SIP, BIS and TP and to identify, when the varieties diverged from each other.

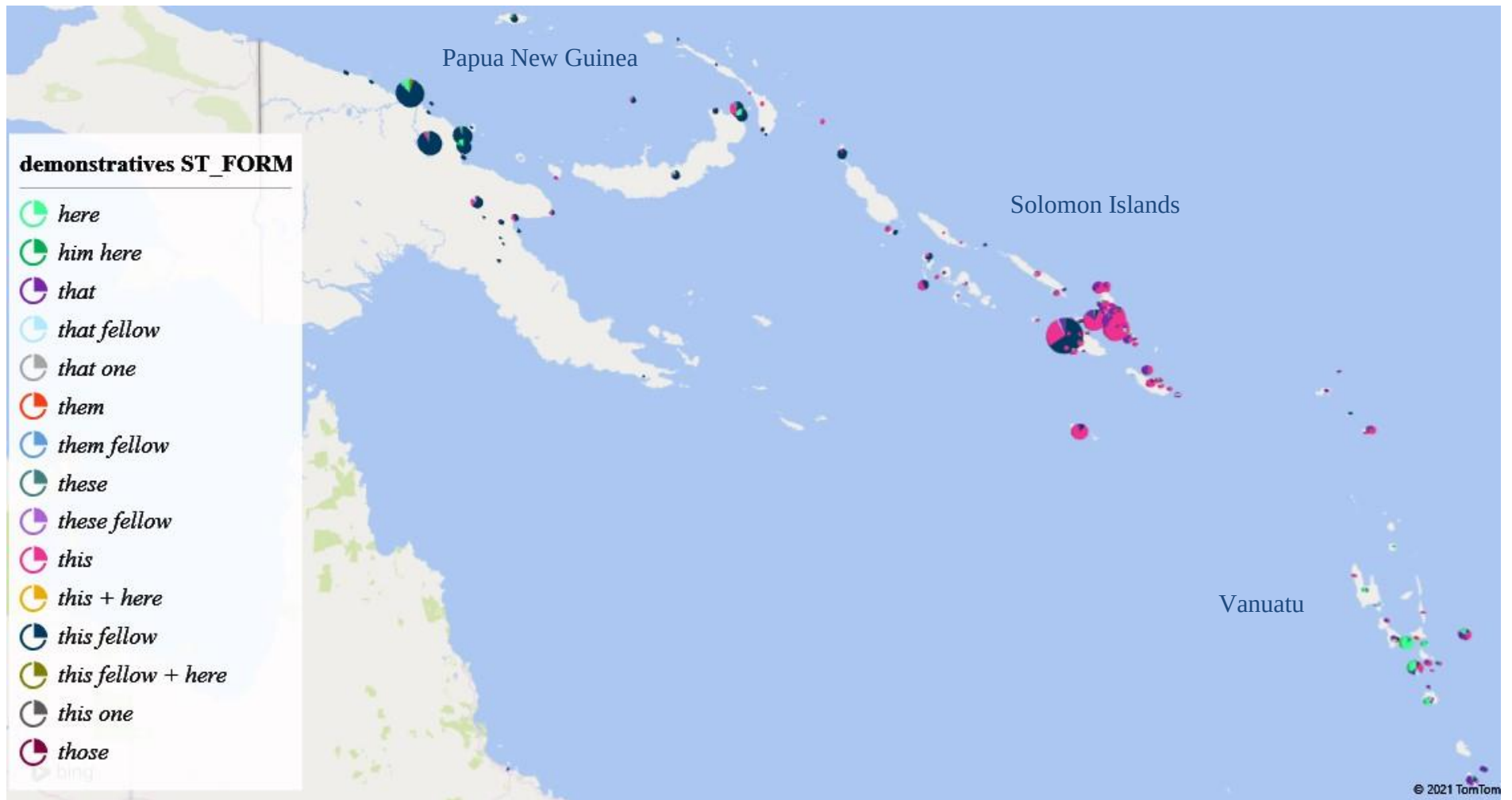


Figure 34: Map of Melanesian Islands, showing attestations of demonstrative ST_FORM variants

7 Case Study: Relative clauses

This chapter explores the development of relativisation strategies in SIP, BIS and TP. Since it has been assumed that “[d]istinctive marking of relative clauses comes later in the stabilisation and expansion phase of the pidgin life cycle” (Romaine 1985: 11), innovative changes may be observed in the development from Melanesian Pidgin English into SIP, BIS and TP. This chapter will start by providing a brief overview of the theoretical background of relative clauses (7.1) before it looks at relative clauses in MPE varieties today (7.2). This will be followed by a discussion of previous diachronic studies on relative clauses (7.3) before the methodological considerations will be introduced (7.4) and an analysis will be given (7.5). Finally, the chapter will conclude with a summary of the results (7.6).

7.1 Theoretical background

Relative clauses (RCs) have been studied extensively and various definitions have been proposed. In this study, following the definition by Lehmann (1986: 664), relative constructions are defined as “consisting of a nominal [...] and a subordinate clause interpreted as attributively modifying the nominal”. The subordinate clause, which is the relative clause, modifies a noun or noun phrase “by expressing a proposition one of whose arguments is coreferential with that noun” (Bruyn 1995b: 149). In English grammars, *restrictive* relative clauses are distinguished from *non-restrictive* ones. While the former delimit the noun or noun phrase they are referring to (cf. Aitchison 1992: 298), non-restrictive relative clauses add additional information about the noun phrase and are therefore also referred to as *appositive* relative clauses. Sentence (54) shows a restrictive relative clause as the noun *boy* could refer to all boys living around the world, but the relative clause restricts the meaning to the specific boy whom the speaker saw yesterday. In contrast, in sentence (55) the relative clause is not restricting the noun but adds additional information that is not required. The sentence could be rephrased with the conjunction *and* into *This is Tom and he is my best friend* indicating that the information is appositive only.

(54) *What is the name of the boy*[whom we saw yesterday?]

(55) *This is Tom*,[who is my best friend.]

However, a formal distinction between restrictive versus non-restrictive relative clauses is typologically rare. As shown by Comrie (1989: 139), it is therefore not necessary to distinguish between restrictive versus non-restrictive relative clauses. In addition, it is not always very straightforward to classify relative clauses as instances of one or the other. From a diachronic point of view the two types of relative clauses develop in tandem (cf. Bruyn 1995b: 150). This study will thus henceforth not differentiate between the two.

Relative clauses have been studied from various typological perspectives. Keenan & Comrie (1977: 66) analysed relative clauses based on their syntactic function, showing that languages differ in terms of the noun phrase positions that can be relativised. Comparing the relativisation of noun phrase positions in 50 languages, they introduced the following *Accessibility Hierarchy*:

SUBJECT>DIRECT OBJECT>INDIRECT OBJECT>OBLIQUE>GENITIVE>OBJECT OF COMPARISON

The Noun Phrase Accessibility Hierarchy is an implicational hierarchy and suggests that “the relativizability of certain positions is dependent on that of others” (1977: 66). All languages that have relative clauses will relativise subjects but do not necessarily relativise the grammatical functions listed to the right. Languages that relativise indirect objects will also relativise subjects and direct objects but not necessarily the syntactic positions to the right of the hierarchy.

Lehmann (1984) investigated relative clauses based on the order of occurrence of the relative clause and the noun. He identified that the relative clause can be postnominal, prenominal or circumnominal in embedded relative clauses, whereas in adjoined relative clauses preposed and postposed RCs can be differentiated.

Typological studies have furthermore shown that languages may use different morphosyntactic strategies to represent the noun in the relative clause. Comrie & Kuteva (2013a: 494-495) have distinguished four main strategies. The *relative pronoun strategy* refers to relative clauses in which a case-marked pronominal component is present and is used to refer to the subject or object of the main clause. For example, the Portuguese-lexified creole Angolar, spoken in São Tomé and Príncipe, has a relative pronoun *ki*, which serves as a relativiser that at the same time fulfils the function of the subject in the relative clause (see example (56)).

- (56) *ome si ki ba tamba*
 man DEM REL.SBJ go fish
 ‘the man who went fishing’

(Maurer 1995: 55)

The second strategy is called *non-reduction strategy*, which may further be subdivided into *correlative* and *head-internal* relative clauses. While in the former “the head noun appears as a full-fledged noun phrase in the relative clause and is taken up again at least by a pronoun or other pronominal element in the main clause” (Comrie 1998: 62), in the latter type the noun phrase occurs in the relative clause as well but is not explicitly stated in the main clause (cf. 1998: 62). The former strategy is, for example, found in Pidgin Hindustani (Pidgin (Fiji Hindi-lexified): Fiji), as shown in (57). The latter strategy occurs marginally in Ternate Chabacano (Pidgin (Spanish-lexified): The Philippines) as shown in (58).

- (57) *Jon kempa u-lon baito, u maila*
 REL camp 3-PL COP 3SG dirty
 ‘The camp that they stayed at was dirty.’

(Siegel 2013: online; Example 70-11)

- (58) *Kel a-konosé bo ómbri agóra mi ermáno*
 REL PFV-know you man today my brother

‘The man you met today is my brother.’ (Lit. ‘The – you met the man today – is my brother.’)

(Sippola 2013: online; Example 44-12)

The third strategy is called *pronoun retention strategy* and refers to relative clauses in which “the position relativized is explicitly indicated by means of a resumptive personal pronoun” (Comrie & Kuteva 2005: 211):

- (59) *Dispela pik em sa kaikai ol man em sa raun.*
 DEM pig 3SG HAB eat PL man 3SG HAB go.round
 ‘The pig which eats people was going around.’

(Smith & Siegel 2013b: online; Example 22-158)

In contrast, the fourth strategy, named *gap relativisation strategy*, describes relative clauses in which the head noun is not overtly referred to:

- (60) *Mi save long wan-pela meri _ i gat twenti pikinini*
 1SG know PREP one-MODIF woman _ PM have twenty children
 ‘I know a woman who has twenty children.’

(Smith & Siegel 2013b: online; Example 22-160)

A strategy not referred to by Comrie & Kuteva (2005, 2013a, 2013b, 2013c) is what I will call the *relative particle strategy*. Like the *relative pronoun strategy*, the *relative particle strategy* is characterised by an invariant morpheme which occurs in the relative clause as a relativiser. However, in contrast to the *relative pronoun strategy*, the relative particle only introduces the relative clause but does not indicate the head’s role. A relative particle may co-occur with a *resumptive pronoun* indicating the head’s role or with a *gap*, as the following example from Nigerian Pidgin English (Pidgin (English-lexified): Nigeria) shows:

- (61) *dì pikín wé (ì) dè sidón fòr dyar*
 ART.DEF child REL (3SG.SBJ) NCOMPL sit LOC there
 ‘The child who sots over there.’

(Faraclas 2013: online; Examples 17-157 and 17-158)

It is a common assumption that pidgins lack overt markers to express relativisation (cf. Sebba 1997: 39). As pidgins are frequently described as characterised by “[r]elative grammatical simplicity” and to be “less complex than the grammars of their source languages” (Sebba 1997: 37), overt marking of relative clauses is only assumed to develop when the pidgin starts to stabilise or creolise (cf. Romaine 1985: 11).

This idea is supported by Bickerton who assumes that pidgins are the result of contact situations in which speakers aimed at acquiring a native-like competence in the lexifier language but failed. Since from a structural perspective there is no need to overtly mark relative clauses, the speakers of the contact variety are assumed to have initially started without the use of relative pronouns (cf. Bickerton 2016: 58). However, as zero-marked relative clauses are difficult to process, contact varieties may come up with other strategies for *functional* reasons (cf. Bickerton 1991: 28). Keeping in mind that Bickerton assumed that contact-situation participants aimed at acquiring the lexifier language, he regards the use of resumptive pronouns as an intermediate stage only (cf. Bickerton 1977: 274) that will come to be replaced by the “full range of English relative pronouns” (Romaine 1985: 12). Bickerton’s theory ignores the possibility that pidgin speakers may not have aimed at acquiring a native-like command of the lexifier language and it does not consider the role which lexifier language speakers have played in the development process of the pidgin variety. In addition, research has shown that the zero-strategy can also be found in some creolised contact languages, such as, for instance, Sri Lankan Malay¹¹⁰ (Slomanson 2013) or Sri Lanka Portuguese¹¹¹ (Smith 2013). Moreover, there are contact languages in which the “relative markers of the lexifier language were lost during pidginization or creolization, or do not function in the same way in the creole as they do in the source language” (Bruyn 1995b: 149).

Thus, innovative changes can be expected to occur in the development from an early contact variety into an expanding pidgin or creole (cf. Bruyn 1995b: 149). In cases in which the relative markers of the lexifier language get lost completely it is thus of interest whether other words (either from the superstrate or substrate languages) took over this function. A study focussing on relative clauses in the languages of the world has shown that demonstratives, locative adverbs such as *here*, and *wh*-question words serve as the most frequent basis in the grammaticalisation process of relativisers (cf. Heine & Kuteva 2004: 115, 174).

Though there are different typological approaches, the focus of the present study will be on the *morphosyntactic strategies* used in the three varieties to realise relative clauses. The study will investigate whether early MPE started with zero-marking as well and whether the strategy was retained or replaced during the development from MPE to SIP, BIS and TP. As the focus of the study is on *when* the varieties diverged, the analysis will not only look at *if* changes in the choice of relativisation strategies can be observed but also *at what point in time* these changes happened in the varieties. A closer look will be taken at the individual *forms* used to encode relativisation since the morpheme-by-morpheme analysis provided evidence that different forms

¹¹⁰ (Mixed Language (Malay-lexified): Sri Lanka)

¹¹¹ (Creole (Portuguese-lexified): Sri Lanka)

were in use in the three varieties. I will only consider subject and object relative clauses in this case study.

7.2 Relative clauses in MPE today

The contemporary MPE varieties show different preferences regarding the forms and strategies used to encode relativisation. As the focus of the study is on subject and object relative clauses, the present section will focus on how relative clauses (RCs) are encoded in these syntactic contexts in modern SIP, BIS and TP.

In *contemporary SIP* various pronoun strategies are in use. Relative clauses may be zero-marked with either a gap or a pronoun in subject/object position as exemplified in sentence (62). This corresponds to Keenan & Comrie's (1977) gap-relativisation and pronoun retention strategies. According to Jourdan, this is "the most important way of building relative clauses in Pijin" (1985: 160). Alternatively, the clause may contain the relative particle *wea* which may be followed by a gap or a pronoun as exemplified in sentence (63). As *wea* can be used in subject and object position, it is referred to as a *relative particle*. In addition, the relative pronoun *hu* may be used in subject relative clauses only (cf. Huebner & Horoi 1979: 173).

(62) *Mi luk-im wan-fala man Ø(hem) i gar-em wan-fala leg nomoa*
 1SG look-TR one-MODIF man Ø (3SG.RES) PM have-TR one-MODIF leg only
 'I saw a man who had only one leg.'

(63) *Mi luk-im wan-fala man wea(hem) i gar-em wan-fala leg nomoa*
 1SG look-TR one-MODIF man REL (3SG.RES) PM have-TR one-MODIF leg only
 'I saw a man who had only one leg.'

(adjusted from Huebner & Horoi 1979: 172)

In *Bislama* relative clauses are usually introduced by the relative particle *we* (cf. Crowley 2004: 66). While constructions in which *we* is followed by a gap seem to be the most prominent, it is also possible that *we* is followed by a resumptive pronoun, as exemplified in sentence (64), in which the second person singular pronoun is reiterated after the relative particle (cf. Meyerhoff 2013b).¹¹²

Next to these strategies, the form *ia* is sometimes classified as fulfilling a relative clause indicating function. Tryon (1987: 119) argues that *ia* "is very widely used in Bislama following and modifying the object of the main clause to signal the introduction of a following relative clause". As sentence (65) shows, the *we* particle is nonetheless needed and thus remains the

¹¹²According to APiCS, the *relative particle and gap* strategy is more common than the *relative particle and resumptive pronoun* strategy in both subject and object relative clauses. In subject relative clauses the ratio is 75% / 25% and in object relative clauses it is 90% / 10% (cf. Michaelis & Haspelmath 2013a: 366, 2013b: 370).

primary relative clause marker. Due to the multifunctionality of *ia*, there is no common agreement about the reanalysis of the demonstrative marker *ia* as a relative clause marker.

- (64) *yu*, *we* *yu* *sapot-em* *gavman*, *yu* *save* *kas-em* *wok*
 2SG **REL** **2SG.RES** support-TR government 2SG can get-TR work
 ‘You who support the government, you’ll get work.’
 (Meyerhoff 2013b: online; Example 23-144)

- (65) *Em* *i* *save* *man* ***ia*** *we* *—* *i* *ded.*
 3SG PM know man **DEM** **REL** *—* PM died
 ‘He knew the man who died.’
 (Tryon 1987: 118)

In *contemporary Tok Pisin* relative clauses can be realised in a variety of ways. According to Smith & Siegel (2013b), it is most common that both subject and object relative clauses are not overtly marked but indicated by juxtaposition only (see sentence (66)). It is also possible that the relative clause is introduced with a resumptive pronoun. A strategy which is less commonly used in comparison with the other two strategies is the use of a relative particle which is followed by a gap.¹¹³ As sentence (67) demonstrates, the relative particle used in TP is *we*, which seems to be similar to *wea* in SIP and *we* in BIS. It is only used in some varieties of TP (cf. Mühlhäusler 1985b: 153). Other studies, such as the one conducted by Smith (2002), have demonstrated that the relative pronoun *husat* provides a further way to encode relative clauses. Example (68) shows that the pronoun *husat* can only be used with human referents (cf. Smith 2002: 151). It is usually associated with a form used in formal and written contexts (cf. Siegel 1981) and lects which are closer to Standard English (cf. Mühlhäusler 1985c: 416). Furthermore, it is only occasionally attested in spoken TP (cf. Smith 2002: 151).

- (66) *Mi* *save* *long* *wan-pela* *meri* *∅* *—* *i* *gat* *twenti* *pikinini*
 1SG know PREP ART-MODIF woman *∅* *—* PM have twenty children
 ‘I know a woman who has twenty children’
 (Smith & Siegel 2013b: online; Example 22-160)

- (67) *Em* *stori* *blo* *wanpla* *sneik* *we* *—* *i* *bin* *stap* *lo* *Bali*
 3SG story POSS one snake **REL** *—* PM PST stay PREP Bali
 ‘It’s the story of a snake that lived in Bali [...].’
 (Smith & Siegel 2013b: online; Example 22-159)

- (68) *Brata* *blo* *JP* *usat* *i* *bin* *stap* *wantem* *ol* *manggi* *hia* *lo* *vilidj*
 brother POSS JP **REL** PM PST LOC with PL boy here PREP village
 ‘JP’s brother who stayed with the boys here in the village’
 (Smith 2002: 152)

¹¹³ Zero marking followed by a gap is attested in 55.6% of subject relative clauses and in 63.6% of object relative clauses. Zero marking followed by a resumptive pronoun is attested in 33.3% of subject relative clauses and in 27.3% of object relative clauses in APiCS. A relative particle (+gap) is only attested in 11.1% of subject relative clauses and in 9.1% of object relative clauses (cf. Michaelis & Haspelmath 2013a: 366, 2013b: 370).

Similar to BIS, the particle *ia* has been brought into connection with relativisation strategies in TP. The idea that *ia* may function as a way to encode relative clauses was first articulated by Sankoff & Brown (1976). *Ia-bracketed relative clauses* are however difficult to identify and do not seem to be commonly used. For instance, Smith's data contains only a small amount of the latter, and he concludes that "[o]ften it is difficult to assign an exact function to the particle, which seems to be highly variable and may merely mark the end of a section of discourse or stress a particular feature" (Smith 2002: 156). This conforms with Siegel (1981: 27-28) who did not attest the use of *ia* as a relative clause indicating device.

A further means of marking subject and object relative clauses in Tok Pisin is the use of clause-final *longen/long em*. The form has its origin in locative relative clauses, which are frequently attested to co-occur with the locative relative particle *we* (cf. Mundhenk 1990: 354; Smith 2002: 156-157). However, according to Smith (2002: 157), it is "undergoing semantic bleaching, i.e. losing its semantic content and being grammaticalized as a relative marker". This means that the application of *longen* or *long em* has expanded to subject and object relative clauses in which they have a "clause-delimiting role", as the following example demonstrates:

- (69) [...] *em* *smel-im* *pik* (*tu-pla* *bin* *kil-im* ***longen*** [...])
 [...] 3SG smell-TR pig (two-MODIF PST kill-TR **REL** [...])
 '...he smelled the pig, which the two had killed...'

(Smith 2002: 157)

7.3 Previous diachronic studies on relative clauses in MPE

The development of relative clauses in pidgins and creoles has long been of interest to creolists, which is why several studies on relative clauses in Tok Pisin can be found (cf. Sankoff & Brown 1980; Siegel 1985; Aitchison 1992; Romaine 1992; Levey 2001). Although the studies *inter alia* aim to reconstruct the diachronic development of relative clause markers, they are primarily based on data after 1950. For instance, Sankoff & Brown (1980) try to reconstruct the origin and spread of relativisation strategies based on recordings made in Lae and the Buang area in 1971 and only make reference to two earlier sources by Hall (1943) and Churchill (1911). Romaine's analysis is based on a corpus consisting of spoken data collected in 1986/1987 and written data referring to the years 1966, 1982, 1986/87. The study conducted by Levey (2001) compares relative clause markers attested in *Wantok* newspaper articles dating to the years 1971-1974 with those attested in the years 1990, 1994 and 1998.

It is of special interest that the existing studies show many discrepancies regarding the development and spread of the relative particle *where*. Sankoff & Brown (1976: 637), for instance, do not identify *where* as a relativiser in their TP data and argue that "we is not used even in the formation of place relatives", whereas Aitchison (1992: 309) and Mühlhäusler (1997: 174)

assume that *where* was first used in locative adverbial relative clauses and grammaticalised into a subject and object relative clause marker through a reanalysis of the marker. This is supported by Romaine's findings who attests the *where* particle primarily in locative relative clauses but also finds attestations in which the particle occurs in subject and object relative clauses. This lets her assume "that we spread from locative to subject and then direct object relatives" (1992: 161).

Mühlhäusler considers *where* to be a recent innovation in Tok Pisin since he does not attest the marker in his early data collection but confirms the existence of the particle in his post-1950 TP data (cf., for instance, Mühlhäusler 1985b: 153, 1997: 174; Mühlhäusler et al. 2003: 11). According to his findings, *where* remains a marginal marker which is regionally restricted and "encountered in some second-language varieties of Tok Pisin though it is only documented where significant numbers of creole speakers are found" (1985b: 153). While Siegel (1981: 30) supports the assumption of *where* being a feature of creolised TP, he assumes in a later publication that the relative particle *where* was used as an optional relative clause marker in all regions in which early MPE was spoken but that it never established in TP as it "was not reinforced by the substrate languages in the New Guinea Islands" (2008: 187).

The TP particle *husat*, which has its origin in the bimorphemic question particle *who's that* has been claimed to represent a 1970s development, yet the particle is not very frequently used and more common in written than in spoken language. Romaine identifies only three instances of the marker (1992: 159) and Siegel shows that *husat* only started to be used in *Wantok* newspaper articles from April/May 1979 onwards (1981: 31) claiming that it "is not a feature of either commonly spoken or creolized Tok Pisin" (1981: 30). Levey's study concludes that from the beginning of the 70s until the 90s a "trend towards more explicit marking of relative clauses" with particles such as *we* and *husat* can be observed in TP (2001: 265-266). The late attestation of *husat* is of interest since bimorphemic question particles were common in the *early* development stages of contact varieties. Thus, one would assume that the grammaticalisation of the question particle *who's that* into the relative pronoun *husat* would have taken place at an earlier time period.

Studies conducted on relative clauses in BIS and SIP are less common. Though relative clause constructions are referred to in grammars and language descriptions (cf. Crowley 2004; Meyerhoff 2013a), little research has been done focussing on them from a historical perspective. An exception provides the study conducted by Crowley (1990a).

Crowley (1990a: 330) first attests the form *we* to encode relative clauses in early 20th century Bislama. Since *we* is a common marker in contemporary SIP, he assumes that the particle must have developed into a relative marker before the end of the 19th century and that it would

therefore be present in contemporary SIP and BIS. In addition, Crowley (1990a: 330) assumes that *where* has been present in all three varieties claiming that:

[a]lthough *we* is not found today in many varieties of Tok Pisin, it is fairly widespread in the Tok Pisin of New Ireland speakers. This suggests that this was originally a feature of all varieties of Melanesian Pidgin by the mid-1880s and that it subsequently became reduced in its distribution in Tok Pisin.

However, Crowley is unable to provide early evidence to support his assumption. There remains the possibility that *where* entered SIP much later through renewed contact with English. Moreover, as was shown above, opposing views regarding the existence and development of *where* in TP exist and, for instance, Mühlhäusler claims *we* to be a recent innovation in Tok Pisin. Since Crowley does not have examples of early SIP and TP to support his claims, these are assumptions that are still in need of proof. The existence of *where* in early Solomon Islands Pijin is thus also disputed.

Studies focussing on the interrelationship of MPE varieties that touch upon the topic of RCs are, for instance, Keesing (1888), Baker (1993) and Siegel (2008). Siegel (2008: 187), citing to Crowley (1990a), assumes that the relative particle *where* existed in all regions in which early MPE was spoken and thus suggests *where* to have been a feature of SIP as well. He assumes that substrate influence was decisive in whether the varieties retained the particle or not. As substrate languages spoken in the Solomon Islands and New Guinea do not use relativisers, the particle would not have been reinforced in these areas. For SIP, he bases his assumptions on data extracted from an interview conducted by Keesing & Jourdan in the 1980s with a Solomon Islander who learnt Pijin on the plantations in the 1930s. As the Solomon Islander had not used Pijin frequently since the war, he assumes the pidgin to reflect what the variety was like in the 1930s and 1940s. The speaker only used the particle *where* in two out of 19 relative clause constructions, which is why Siegel (2008: 188) assumes that the *where* particle was not reinforced during dialect differentiation, but instead represents a recent consequence of renewed contact with English:

the widespread use of *wea* as a relativizer in Pijin appears to be a fairly recent consequence of renewed contact with English or language-internal expansion. In the earlier stage of dialect differentiation, however, the use of *we* was not reinforced by the Southeast Solomonian substrate languages [...] which do not have relativizers.

Baker (1993: 30), who compares features and their first attestations in Chinese Pidgin English, New South Wales Pidgin English and QPPE with their dates of earliest attestation in the MPE varieties, attests the relative particle *where* only in his Bislama dataset, assuming it to be a modern development in SIP. He does not attest the form in TP and comes to the conclusion that *where* ('who/which/that') represents a Vanuatu innovation (cf. 1993: 52).

Keesing (1988: 145-148) claims that the resumptive pronoun *he* took over the function of embedding relative clauses due to substrate influence. He argues that the use of the form *he* is the result of a “collusion between English speakers’ resumptive pronouns and Oceanic speakers’ SRPs (= Subject-Referencing-Pronoun)” (Keesing 1988: 144). The latter is the designation referring to distinct pronominal morphemes which are used to “recapitulate subjects and [to] introduce predicates” (Hall 1966: 83). As SRPs in Oceanic languages can additionally serve to embed relative clauses without any further overt marker, Keesing assumes that this functionality was also taken over during the development of the Melanesian Pidgin varieties. Sentences (70) and (71) shows examples of SRPs in Kwaio (Austronesian (Southeast-Solomonic): Malaita) which embed relative clauses. According to Keesing, early MPE sentences such as the one exemplified in (72), represent a direct calque from Oceanic languages.

- (70) *rua wane gala nigi naaboni ta-gala suga-a*
 two man SRP(they) arrive yesterday FUT-SRP(they) buy-it
 ‘The two men who came yesterday, will buy it.’

- (71) *wane ku aga-si-a naaboni te-’e suga-a*
 man SRP(I) see-TR-him yesterday FUT-SRP(they) buy-it
 ‘The man I saw yesterday will buy it.’

(Keesing 1988: 146)

- (72) *All boy he stop long place belong sick, no can limlimbur long nother fella place*
 PL boy PM/SRP LOC PREP place POSS sick PROH walk PREP other MODIF place
 ‘The boys who are in the hospital are not allowed to walk to other places’
 (Kokopo 1927; McCarthy 1926-1952; PMB 616)

Since he assumes that the MPE varieties were alike at this early stage of their development, *he* would have fulfilled a relative clause embedding function in all three regions.

To sum up, although research on the individual varieties has been conducted and suggestions were made about how and when the relative particle *where* spread, there are still obscurities regarding the development of relative clause markers in the three varieties. The form *husat* has been claimed to represent a post-1950 development. However, new data may lead to new insights, and StE similar forms have not been considered at all. In addition, no diachronic comparative analysis on the development of relative clause strategies in early SIP, BIS and TP has been conducted so far that includes statistical measurements to identify *when* the varieties developed their individual RC marking strategies.

7.4 Methodological considerations

Based on the differences observed during the morpheme-by-morpheme analysis, I decided to investigate the diachronic development of relative clauses by approaching them from two

linguistic perspectives, namely, by investigating not only the relativisation strategies (STRATEGY) as is common in typological studies, but also the concrete relative pronoun and relative particle forms used to encode the relative clauses (ST_FORM). The two variables were analysed independently from each other. The analysis of each variety first concentrates on the relativisation strategy before the concrete particle forms are focussed on.

The study is, as mentioned above, only based on subject and object relative clauses due to the fact that these represented the dominant types in the data.¹¹⁴ Subject relative clauses were coded as *rel_sub* and object relative clauses as *rel_obj* (FEATURE). Due to the limited amount of early data at hand, the case study focusses on surface phenomena only.

The variables and variants for which the data was coded are summarised in Table 10. The individual variants of the variables STRATEGY and ST_FORM will be further exemplified below. It should be noted that the examples were randomly selected out of the early data. Reference regarding the presence or absence of the forms in each variety will be given in the subsequent sections.

Five different relativisation strategies were identified in relative clauses in the early data. The data was coded as *zero+gap* if the utterance contained a zero marked relative clause which has a gap in subject or object position, as in sentence (73). *Him* in *me look him* does not equate StE 3SG *him* but represents the form to encode transitivity. Zero marked relative clauses may also explicitly refer to the relativised position with a resumptive pronoun, which was coded as *zero+res* in the data and is exemplified in sentence (74). *Zero+gap* is in line with what was earlier defined as the gap relativisation strategy, whereas the coding *zero+res* equates with Keenan & Comrie's (1977) *pronoun retention strategy*.

- (73) *Me think mark [Ø me look him –] he mark belong Talatova*
 1SG think mark [Ø 1SG see TR PM mark POSS Talatova
 'I think the sign which I saw was the sign of Talatova.'
 (Tulagi; Tor Tor 09.01.1923)

- (74) *He tambu for tell out anyway something [Ø him he no true.]*
 PM forbidden PREP say anything [Ø 3SG.RES PM NEG true]
 'It is prohibited to tell lies.'
 (Solomon Islands; Hogbin 1944: 282)

The identification of zero-marked relative clauses proved to be difficult and it may be that zero marked relative clauses were not recognised due to author modifications. In sentences such as example (75), it was difficult to decide whether the clauses are simply concatenated or whether

¹¹⁴ Genitive, temporal, instrumental and adverbial clauses introduced through relativisation were not included since a comparative analysis was impossible due to their low frequencies of occurrence. For example, the number of instrumental clauses amounted to less than five sentences and the number of possessive relative clauses was below 50 sentences.

they represent two relative clauses that are not overtly marked. It is also likely that authors used full stops when transcribing past speech events although the speaker may have produced a single sentence containing a non-overtly marked relative clause. This may have resulted in clauses not being identified as relative clauses.

(75) *Master, man belong longlong, walk about long place,*
master man PREP foolish walk PREP place

$\left(\begin{array}{l} \emptyset \\ \emptyset \end{array} \right. \begin{array}{l} he \\ 3SG.RES/PM \end{array} \begin{array}{l} no \\ NEG \end{array} \begin{array}{l} get \\ have \end{array} \begin{array}{l} cloth \\ cloth \end{array} \begin{array}{l} no \\ NEG \end{array} \begin{array}{l} labalaba, \\ lava-lava \end{array} \right)$

$\left(\begin{array}{l} \emptyset \\ \emptyset \end{array} \right. \begin{array}{l} he \\ 3SG.RES/PM \end{array} \begin{array}{l} no \\ NEG \end{array} \begin{array}{l} belong \\ PREP \end{array} \begin{array}{l} this fellow \\ DEM \end{array} \begin{array}{l} place, \\ place \end{array} \right)$

he belong other fellow side belong Sâng.
3SG/PM PREP other MODIF side PREP Sang

‘Master, a maniac walks through the village. He neither wears a loincloth nor other cloth, he is not from here, he belongs to the area of Sang./Master, a maniac walks through the village, who neither wears a loincloth nor other cloth and who is not from here but belongs to the area of Sang.’

(former Kaiser Wilhelmsland 1917; Detzner 1920: 200)

Another problem highlighted by sentence (75) is that sometimes it is not clear from the context whether the form *he* was used as a resumptive 3SG pronoun or whether it represented the predicate marker. The form *i*, which has grammaticalised out of the Standard English 3SG pronoun *he* into what Sankoff (1993: 120) and Crowley (2000) define as a predicate marker and Meyerhoff (2013a: 226) calls an agreement marker, is present in all three contemporary varieties. When *he* occurs in the early data, it is not always clear whether the form has already grammaticalised but is covered by its Standard English orthographical portrayal, or whether it was used as a resumptive pronoun. In all three varieties a derivative of the 3SG pronoun *him* has grammaticalised into the 3SG pronoun.¹¹⁵ In a sentence such as (76), we find a relative clause in which the *zero+res* strategy is used, and in which we can be sure that *he* is used as a predicate marker as it appears after the third person singular pronoun *him*. Thus, *him* represents the resumptive pronoun and *he* serves as the predicate marker.

(76) *This fellow something* $\left(\begin{array}{l} \emptyset \\ \emptyset \end{array} \right. \begin{array}{l} him \\ 3SG \end{array} \begin{array}{l} he \\ PM \end{array} \begin{array}{l} been \\ PST \end{array} \begin{array}{l} do-im \\ do-TR \end{array} \begin{array}{l} - \\ - \end{array} \right)$
DEM thing
 ‘This thing, which he has done’

(Solomon Islands 1943; Hogbin 1944: 280)

If *he* occurs after a relative particle or zero without a preceding personal pronoun, such a clear classification is not possible. One attempt to identify the function which the form fulfilled in the individual relative clauses was to prove whether *him* was attested in other sentences of the same source to encode the 3SG subject pronoun. If this was the case, it appeared to be unlikely that *he*

¹¹⁵ In Bislama and Solomon Islands Pijin the form developed into *hem*, in Tok Pisin into *em*.

was used in the relative clause to fulfil this function. Instead, this would suggest that *he* was used as a predicate or agreement marker.

Sometimes the capitalisation of *he* provided evidence that the marker was used by the speaker, or rather, interpreted by the writer as a 3SG pronoun and not as a predicate marker. In a sentence such as *me want to tell you myself about Jesus He been save me* (Deck 27.09.1932; PMB 1150), one finds referential capitalisation in which the form *He* is interpreted as referring to a divine being.¹¹⁶ Nonetheless, it cannot be assumed that *He* represents a resumptive pronoun in this clause since we do not find capitalisation in speech. Thus, the writer may have reinterpreted the occurrence of the form as a pronoun, although the speaker might have used it as a predicate marker. Thus, a clear classification of all RC tokens was not possible. Classification problems arose especially with tokens that were extracted out of texts in which a mesolectal or acrolectal form of Pidgin English or code mixing was attested, since in these sources the number of StE similar pronouns was generally higher.

Since *him* may have grammaticalised at different points in time in various regions, it could also not be taken for granted that attestations of *he* dating after the first occurrence of *him* *he* can be interpreted as predicate markers. Ambiguous datapoints were included into the analysis as well but were coded separately.

Besides zero-marking, *relative particles* were attested as a further means to encode relative clauses in the early data. Similar to zero-marked RCs, relative clauses containing a relative particle either explicitly indicated the relativised position with a resumptive pronoun or left the extraction site empty. The former type of relative clauses was coded as *rel.part+res* and the latter as *rel.part+gap*. An example of both types can be found in sentence (77) and (78):

- (77) *You me* **[where we sign long Jesus,]** *no eat im this pigpig*
 1PL.INCL **[REL 1PL.RES sign PREP Jesus]** NEG eat TR DEM pig
 ‘We who follow/sign up for Jesus, do not eat this pig.’
 (Inakona 1932; Cowie 01.08.1932; PMB 1150)
- (78) [...] *you go cut head belong Suimbinalosum off and throw along one*
 [...] 2SG go cut head POSS Suimbinalosum off and throw PREP ART.INDEF

big nabanga tree **[where _ he stop long long way]**
big nabanga tree **[REL _ PM LOC far-away]**
 ‘Go and cut Suimbinalosum’s head off and throw it at a nabanga tree which is very far away.’
 (Hog Harbour 1931; Salisbury 03.07.1931)

¹¹⁶ It should be noted that an optional past tense marker *bin* deriving from StE *been* has grammaticalised in TP and BIS. According to Jourdan, the past auxiliary can also be found in SIP but is “predominantly [used] by people from the west of the country, and by people who have been to school in Papua New Guinea” (2002: 21). Huebner & Horoi (1979: 57) observed in the 1980s that although *bin* did not occur very frequently, it was used by some speakers to encode past tense. When *bin* is used as a past tense marker, it is usually preceded by the predicate marker *i* if the subject of the sentence requires it. (For example, the 1SG pronoun *mi* does not take a PM).

Moreover, ambiguous relative clauses in which it could not be clearly determined whether *he* served as a resumptive pronoun or a predicate marker were encoded separately as *rel.part(+res)* + *gap*.

Sentences in which a relative pronoun was used, as exemplified in sentence (79), were coded as *rel.pro*. Following the classification system used in APiCS, relative particles and relative pronouns were thus differentiated. Relative clause markers were coded as relative pronouns if they were only applicable for either subject or object references.

- (79) [...] something (**which** me want 'im -)
 [...] something (**REL** 1SG want-TR -)
 'something I want'

(One Pusu 1943; Deck December 1943: 4; AU PMB DOC 439)

When focussing on the **form** (ST_FORM), six different variants were attested. Next to *zero*, as was exemplified in sentences (73-74), the relative particles *where* (cf. (77-78)) and *that* (cf. (80)) and the relative pronouns *who* (cf. (81)), *which* (cf. (79)), and the form *who's that* (cf. (82)) were attested in various orthographic forms.

- (80) small things (**that** stop along bush)
 small thing-PL (**REL** LOC PREP bush)
 'small things which/that are in the bush (= insects)'

(Malakula 1929; Cheesman 1933: 159)

- (81) this one place belong people (**who** die finish)
 DEM place POSS people (**REL** die COMPL)
 'this place of the dead people' (= here: heaven)

(German New Guinea 1921; Lambert 1942: 89)

- (82) balus (**hu sat** i fain-im dis fala man) i sal-im tok log waitman.
 aircraft (**REL** PM find-TR DEM man) PM send-TR talk PREP white-man
 '(The) aircraftman, who finds these men, will inform the white men.'

(Solomon Islands 1942; Luke 1945: 95)

The most interesting forms represent the relative particle *where* and the relative pronoun *who's that* as these forms seem to deviate most from Standard English. The former (= *we*) is a common RC particle in English-based pidgins and creoles around the world. Its usage is attested, for instance, in Krio (cf. Finney 2013), Ghanaian Pidgin English (cf. Huber 2013)¹¹⁷ and Nigerian Pidgin English as well (cf. Faraclas 2013).¹¹⁸

¹¹⁷ Krio (Creole (English-lexified): Sierra Leone); Ghanaian Pidgin English (Pidgin (English-lexified): Ghana).

¹¹⁸ It is also a common phenomenon in other informal or non-standard languages. See, for example Ballarè & Inglese (2021: 20) who provide examples that Italian *dove*, English *where*, French *où* and German *wo* are not "confined to expressing location" in RCs, showing that the markers can also encode subjects and/or objects in non-standard varieties of these language. The following is an example from regional/informal German:

Kennst du den Film wo er mitspielt?
 kennen.2SG 2SG ART.DEF.MASC Film REL 3SG.MASC mitspielen
 "Do you know the movie in which he played?"

Some further specifications are necessary to understand which clauses were considered RC constructions. Although relative clauses usually appear in sentences consisting of a main clause and the relative clause, I decided to also include circumlocutions which do not represent a complete sentence but nonetheless contain a relative clause. Consider the examples (83) and (84).

- (83) *Senake* $\left(\begin{array}{ccccc} \emptyset & - & E & Got & Hand \\ snake & \emptyset & PM & got & hand \end{array} \right)$
 ‘snake which has hands’ = centipede

(Mandated Territory of New Guinea ~1943; Helton 1943: 60)

- (84) *snek* $\left(\begin{array}{ccccccc} \emptyset & - & i & stap & long & bel & bolong & man \\ snake & \emptyset & PM & LOC & PREP & belly & POSS & man \end{array} \right)$
 ‘snake which is in the stomach’ = roundworm

(Vunapope ~1940; Kutscher ~1940: 99)

Although they are not representing full sentences but are circumlocutions to describe missing vocabulary, they exemplify restrictive relative clauses which delimit the noun or noun phrase they are referring to. As the item *snake* could refer to different kinds of insects and snakes, the relative clause is necessary so that the listener can restrict the contemplable objects to the one the speaker is referring to, namely a ‘centipede’ in example (83) and a ‘roundworm’ in example (84).

Factor	Levels
ST_FORM	zero which who that who’s that where
FEATURE	rel_sub rel_obj
STRATEGY	zero+gap zero+res zero(+res)+gap rel.pro rel.part+gap rel.part+res rel.part(+res)+gap
YEAR_DET	1832 ... 1950

Table 10: Linguistic coding for relative clauses

7.5 Findings and discussion

In the following the results of the analysis will be discussed. First, some remarks regarding the general data distribution will be made (7.5.1) before the developments in each variety will be treated separately (7.5.2-7.5.4). In Section 7.5.5 the developments in the three varieties will be observed from a comparative perspective before a summary of the results will be given (7.6).

7.5.1 General data distribution

In total, 815 tokens of relative clauses were found in the early data. The majority of RC constructions were attested in early TP with 439 RCs (53.87%), the SIP data collection contained 279 RCs (34.23%), and in the BIS data only 97 RC constructions (11.90%) were identified. It is necessary to point out that in all three varieties subject relative clauses were more common than object relative clauses, the ratio being 395/44 (90/10%) in TP, 200/79 (72/28%) in SIP and 75/22 (77/23%) in BIS.

The diverging number of datapoints attested in each variety also becomes visible in Figure 35, which highlights how the attested relative clauses in each variety spread across time. The bulk of the TP datapoints dates after 1920, with the highest amount of attestations in 1940 and 1943. The majority of datapoints for SIP are attested after 1905 and spread relatively even. The distribution of the BIS datapoints shows that in contrast to the other two varieties, more pre-1900 attestations exist. However, it is visible that after 1935 only three further RCs were attested, which reflects the generally low number of written data that could be collected for BIS covering the 1930s and 1940s. It needs to be noted that relative clauses might have been used with higher frequencies in the three varieties before these years but might not have survived in the written record.

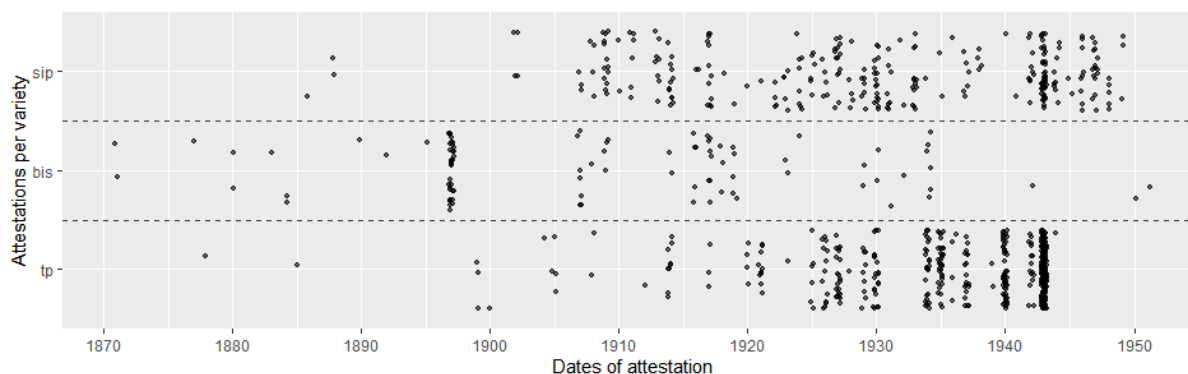


Figure 35: Distribution of relative clause datapoints across time per variety

Figure 36 shows the relative frequencies of strategies attested per variety if the timespan from the earliest attestation until 1950 is regarded as a single period only. The figure shows that *zero+gap* represents the strategy that occurs with highest frequencies in all three datasets. In Tok Pisin 86.10% (= 378/439) of the attested RCs are based on the *zero+gap* strategy, whereas in 6.15% (= 27/439) of the tokens the extraction side is indicated with a resumptive pronoun. In addition, 7.52% (= 33/439) of the TP tokens represent ambiguous cases in which it remained unclear whether the role of the head inside the relative clause was indicated by an overt expression or not. In 0.23% (= 1/439) of the TP tokens a *relative pronoun* was used.

Zero+gap also represents the dominant attested strategy in BIS, being identified in 73.20% (= 71/97) of the tokens. In 9.28% (= 9/97) of the attestations, *he* occurred in the RCs but could not be clearly identified as representing either a resumptive pronoun or the predicate marker. In contrast to TP, early BIS showed the use of the relative particle strategy *rel.part+gap* in 17.53% (= 17/97) of the attestations.

Though *zero+gap* represents the dominant strategy attested in SIP as well, a higher degree of variation can be observed in contrast to BIS and TP. *Zero+gap* was only identified in 36.20% (= 101/ 279) of the tokens, *zero+res* was attested in 13.62% (= 38/289) and *zero+(res)+gap* in 3.94% (= 11/279) of the RCs. Strategies based on a relative particle were found in more than 35% of the tokens. While it could not be clearly identified whether an overt form was used to indicate the head inside the *where*-based relative clause in 2.51% (= 7/279) of the tokens, an overt form was clearly identified in 8.96% (= 25/279) and a gap in 24.37% (= 68/279) of the *where*-based RCs. The high percentage of the relative pronoun strategy, which was attested in 10.39% (= 29/279) of the tokens, is worth noting because in early TP the strategy is less common and in BIS it is not attested at all.

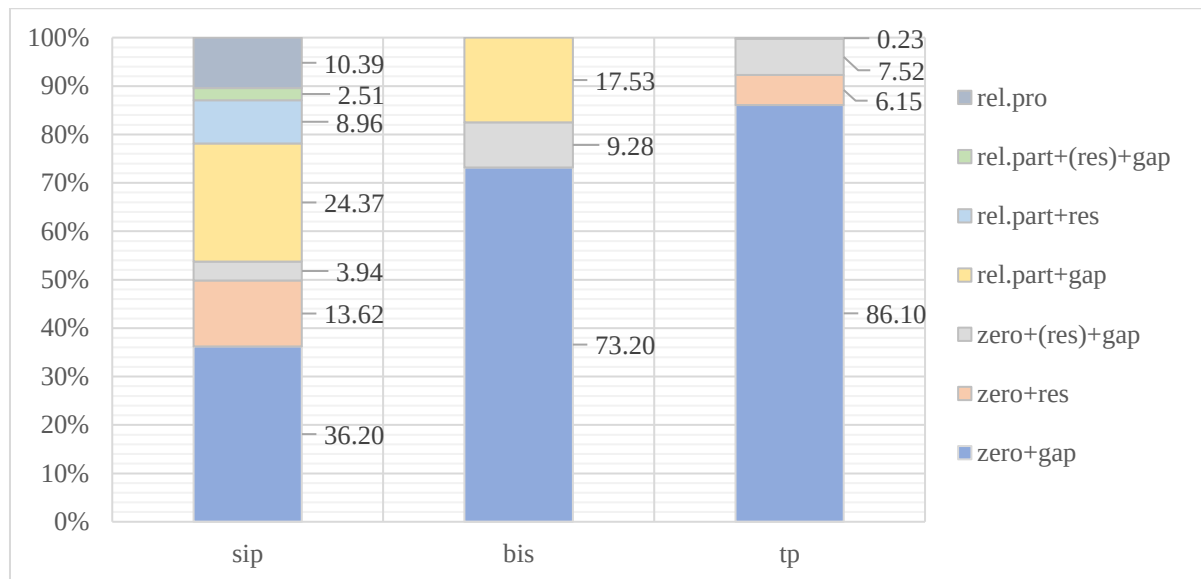


Figure 36: Relative frequencies of RC STRATEGY variants per variety

When focussing on the relative frequencies of the form (ST_FORM), the frequencies of zero-marking are in line with what became visible in Figure 36. Zero-marking is the most common form in all three varieties being attested in more than half of the datapoints in each variety (cf. Figure 37).

The frequencies of the individual relative particles and relative pronouns attested in the varieties are of special interest. For instance, it appears striking that in SIP and BIS the particle *where* is the second most common marker, but it is absent in the TP data. Moreover, the figure

reveals that the early SIP data shows the greatest amount of variability regarding relative clause markers, whereas in the TP dataset a clear preference of the zero-marker can be observed.¹¹⁹

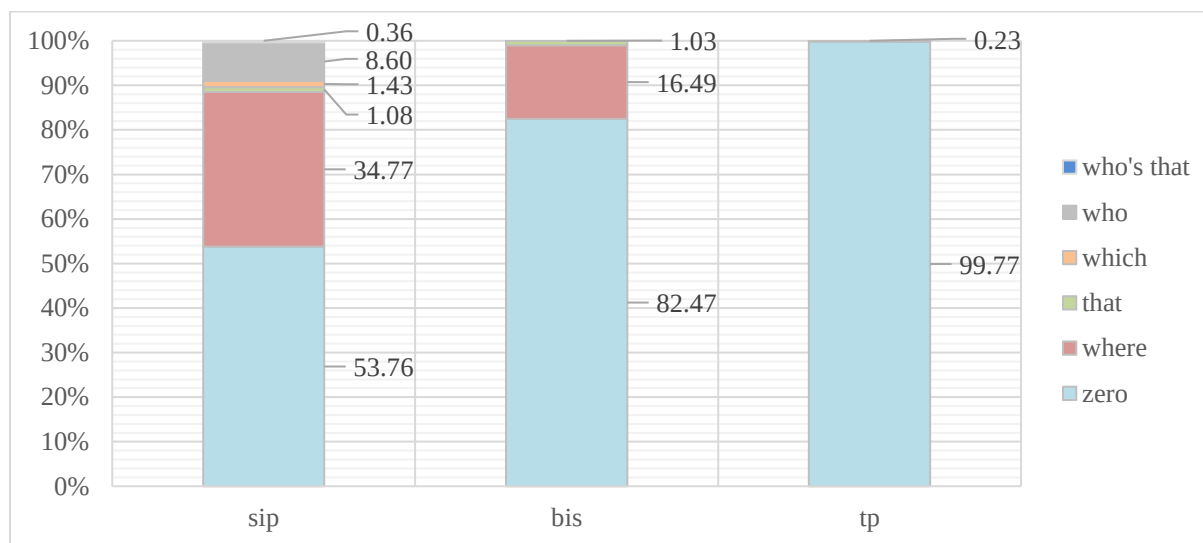


Figure 37: Relative frequencies of RC ST_FORM variants per variety¹²⁰

7.5.2 Diachronic analysis of relative clauses in Solomon Islands Pijin

Figure 36 and 37 show that relative clauses in SIP, as attested in written sources referring to the years 1870-1950, were characterised by a high variability. In this section, a closer look at the various attested *strategies* and *forms* will be taken.

7.5.2.1 Attested strategies and forms in SIP

Solomon Islands Pijin represents the variety that shows the greatest amount of variation regarding the strategies and concrete forms used to encode relative clauses. The form occurring with highest frequency is the zero-relative which is attested in 150 of 279 sentences (= 53.76%). Zeros in early SIP can be followed by a gap or by a resumptive pronoun (cf. sentence (73-74)). The form *where* is attested in 34.76% of the sentences (= 97/279) and, thus, represents the second most frequently attested form. Like the zero-relative, *where* occurs optionally with a gap or with a resumptive pronoun, as demonstrated in the examples (85) and (77). Next to *where*, the particle *that* is attested as a RC marker in three of the tokens (= 1.08%). In contrast to the particle *where*, it was not attested to co-occur with a resumptive pronoun (cf. sentence (86)).

¹¹⁹ A map showing *where* and *which* relative clause markers were attested is displayed in Figure 63.

¹²⁰ It should be noted that *who* and *which* represent in fact relative pronouns, which means that the use of the relative pronoun *who* vs. *which* agrees with the animacy of the head in the noun phrase.

- (85) *Man* [*where* _ *he* *blind-eye*,¹²¹] *he* *make im* *this one?*
 man [*REL* _ PM blind] PM make TR DEM
 ‘Did the man, who is blind, make this?’

(Solomon Islands 1948; Deck 1948: 2; AU PMB DOC 442)

- (86) *Me* *help* *long* *that* *regional school* [*that* *Jason* *start* _ *long* *Bina*] *me*
 1SG help PREP DEM regional school [*REL* Jason start _ PREP Bina] 1SG
alsame *father* *for* *ten* *young* *boy* [*who* *stop* *in* *my* *house*.]
 like father PREP ten young boy [*REL* stay PREP POSS.1SG house]
 ‘I helped at that regional school which Jason started at Bina – I was like a father for ten young boys who
 stayed in my house.’

(*The Stories of the Crew*; NIV June 1947: 8; AU PMB DOC 439)

Next to the attestation of relative particles, relative pronouns were observed in the data. Sentence (86), for instance, shows an occurrence of *who*, which was found in 8.60% (= 24/279) of the attested relative clauses, and sentence (79) shows an example of *which*, attested in 1.43% (= 4/279) of the relative clause tokens. The relatively high occurrence of the relative pronoun *who*, together with the fact that the choice of the relative pronouns *who* vs. *which* agrees with the animacy of the head in the noun phrase, seemed striking at first. A closer look into the data, however, revealed that *who* appeared in various text types, such as travel reports, court proceedings and in religious documents of the *SSEM*. Its appearance in a variety of source types may be an indicator that SIP was exposed to a greater degree to its lexifier than the other two MPE varieties or that it adopted those forms (which just happen to be considered ‘correct’ StE).¹²² It would thus support Reed’s (1943: 270) claim that “in the British Solomon Islands the ‘pidgin’ is far purer English” than in the other regions. Moreover, it needs to be kept in mind that the relative pronoun *hu*, which is an orthographic variant of *who*, is used in modern Solomon Islands Pijin to refer to human referents as well. It is imaginable that the form already found its way into SIP at this early period. Post-1950 data is required to investigate those claims and to explore whether *who* persisted throughout time or whether it disappeared and entered through renewed contact with the lexifier language at a later point in time.

The form *who*’s *that* was only attested once. Its orthographical realisation is *hu sat* and it was attested in a war pamphlet produced by the Japanese. The form *who*’s *that*, as exemplified in example (82), is interpreted as a subject relative clause marker. Although the word *balus* is normally used to express ‘aircraft’, the context makes clear that it does not refer to the aircraft as

¹²¹ A ‘blind man’ was commonly circumscribed as *man i no save lukluk* or *man i no gat ai* (cf. van Baar ~1930: 14; Borchardt 1926: 6). Similar as in these circumscriptions, in example (85) the noun *man* is followed by the predicate marker *he* which introduces a verb. The lexical item *blind-eye* might function as an adjective, or as an adjectival verb, as it is quite common for languages in the world to have stative verbs as adjectives (cf. Velupillai 2012: 127). *Blind-eye* was only attested once in the early data but it can nonetheless be assumed that the lexical item represented an alternative to the other circumscriptions as the stative verb *blaenae* is still used in SIP today (Jourdan 2002: 23).

¹²² StE acrolectal forms were generally more common in the SIP data.

such but to the aircraftsman. Keeping in mind that relative particles usually derive from *wh*-question words, it needs to be noted that *husat* was not attested as a question particle in SIP. The occurrence of the form thus appears to be striking. Instead, *who's that* is attested as a question particle in early Tok Pisin. For instance, in the booklet for soldiers produced by the Army Education Branch (1944: 20) an entry for the question particle “hoosat?” (‘who’) can be found. As pointed out in Section 3.1.6, the Australian and American governments produced war pamphlets in MPE, while Japanese, Australian, and American forces were taught pidgin in order to get along in the Melanesian area during World War II. As the MPE varieties were claimed to be mutual intelligible, only a single variety was taught, which was usually the variety that was spoken in New Guinea (cf., for instance, Army Education Branch 1944: 1-2). Thus, it seems likely that the form *hu sat*, despite being attested in the early SIP data, was not a SIP feature. Since in 8.60% of the SIP tokens the form *who* is used, which is also the relative pronoun attested today, it is more likely that *hu sat* represented an innovation of TP rather than of SIP and that it was ‘falsely’ used by the non-pidgin speaking Japanese producers of the pamphlet. It should be kept in mind that the Japanese (as well as the American and Australian soldiers) produced pamphlets in a language that they themselves were not used to. Most of the pamphlet writers learned the contact variety with the help of textbooks or vocabulary lists and were in contact with PE, if at all, for a short period only. Therefore, it is likely that the form *hu sat* was overused and extended to contexts in which it was not used before.

7.5.2.2 Relativisation in SIP with a focus on strategies

7.5.2.2.1 Timeline and boxplot approach

In order to investigate when the individual strategies were attested in the data, the timeline and boxplot approaches were used. Figure 38 shows the timeline of the attested relative clause strategies (STRATEGY) across time based on the year of attestation (=YEAR_ATT). The attestation of some forms could be traced to concrete dates, others to specific years and again others to time periods.

The timeline shows that *zero+gap* represents the earliest attested strategy in the data being first attested in 1886. Two years later the first attestation of a *relative pronoun* can be observed. The *zero+res* strategy is first attested in 1914, although some ambiguous cases are already attested in 1908. Relativisation strategies that include a relative particle are not attested before 1908. As the figure shows, the pronoun strategy, the relative particle and the zero strategy persist until the end of the observed period in the 1940s.

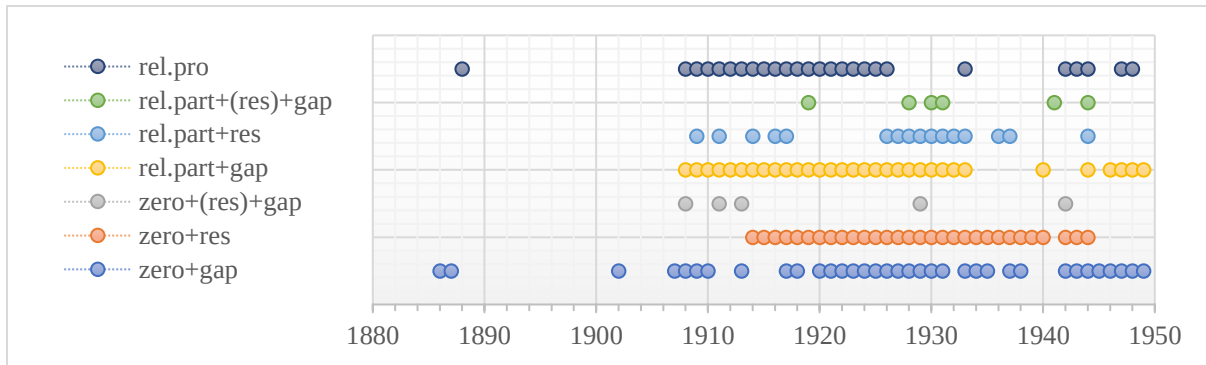


Figure 38: Timeline of RC STRATEGIES in SIP

The comparison of the timeline with the boxplot in Figure 39, which is based on a single determined year for each attestation (=YEAR_DET), shows that the results are quite similar. As almost all boxes and whiskers stretch out to years at the end of the observed period, most of the strategies seem to have been in use until 1950. Almost all strategies are already attested in 1910. *Zero+res* represents an exception. However, due to the amount of ambiguous *zero+(res)+gap* attestations, it can be assumed that the strategy was already used as a variant by 1910 as well. A difference between Figures 38 and 39 can be observed regarding the strategy *zero+res*, which in the timeline starts in 1914 and in the boxplot in 1925. This is because for some datapoints no clear dates of attestation, but timespans, were ascertained.

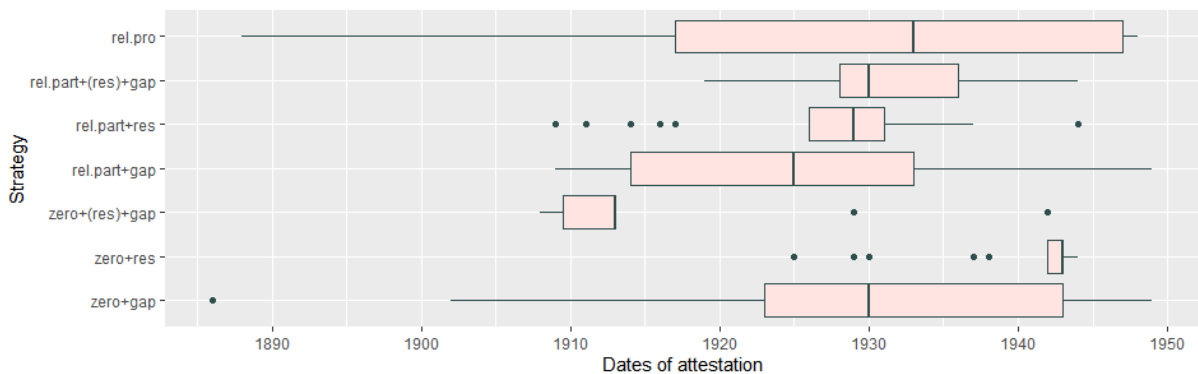


Figure 39: Boxplot of RC STRATEGIES in SIP

Even if the timeline and boxplot show that early SIP was characterised by a great degree of variation, it should be pointed out that strategies based on a relative particle, a relative pronoun or a *zero* marker also represent the common relativisation strategies in contemporary SIP. Thus, the observed variation does not exclude the possibility that these strategies already began to stabilise in the sense of some being selected as the dominant strategies.

7.5.2.2.2 Testing for the impact of the years of attestations on the relativisation strategy

To test whether the data shows significant changes in terms of the choice of strategy (STRATEGY) depending on the year of attestation (YEAR_DET), the ctree algorithm was

applied, which in contrast to the timeline approach also takes frequencies of occurrence into account. The resulting tree, as shown in Figure 40, yields four time splits on three levels. The first three splits are highly significant ($p < 0.001^{***}$) and indicate that a change in the strategy is observable based on whether RCs were attested before or after the years 1936, 1924 and 1943. The last split in the year 1944 is significant with $p = 0.043^*$. Through the identification of these dates, the data is split into five periods (beginning-1924, 1925-1936, 1937-1943, 1944, 1945-1950). However, a closer look at the terminal nodes reveals that in each of the resulting periods a high degree of variation exists. An exception forms the period 1937-1943 in which a clear preference of the zero strategies *zero+gap* and *zero+res* is visible. As in the fourth and fifth constructed period, however, the likelihood of relative particle-based forms increases again, it can be assumed that the zero preference is likely to represent particularities that can be traced to individual authors or texts. Author idiosyncrasies may also explain the sudden high occurrence of the relative pronoun strategy in the Nodes 8 and 9.

Before testing whether the author (= AUTH_NAME) and text type (= TXT_TYPE_3) have an impact on the chosen strategy, it was analysed if the choice of the relativisation strategy is dependent on whether the clause is a subject or object relative clause (= FEATURE). The resulting tree, which is displayed in Figure 41, shows five splits on three levels. The highest-level split shows that the choice of strategy is dependent on whether the head noun is a subject or object ($p < 0.001^{***}$). The second split indicates that even though in object relative clauses both the *zero+gap* and *rel.part+gap* strategy were attested across time, the *rel.part* strategy is more likely to occur in RCs attested before 1927, and the *zero+gap* strategy is more likely after 1926. Subject relative clauses, according to the ctree algorithm, show differences depending on whether they occur before or after the years 1936, 1913, and 1944.¹²³

A great amount of variation is visible in the Nodes 7 and 8 and there is a high probability of the *rel.pro* strategy. As in Figure 40, these observations appear to be striking. If the text type variable is added, the ctree shown in Figure 42 is obtained. The tree consists of five splits on four levels. What is interesting to observe is that the highest-level split is still based on whether the clause is a subject or object relative clause (= FEATURE). However, while in the previous figure a time-dependent split was identified with object relative clauses, the year 1926 no longer seems to be significant if the text type is included as a possible predictor variable because no significant time split is obtained. Thus, according to Figure 42, object relative clauses are more likely to make use of the *zero+gap* strategy than of the *relative particle+gap* strategy, although both forms were in use in early SIP. Subject relative clauses show differences depending on the text type

¹²³ Node 5 and Node 9 are highly significant with $p < 0.001^{***}$; Node 6 is very significant with $p = 0.002^{**}$.

(Nodes 3 and 7) and the year of attestation (Nodes 4 and 8). In speech-like attestations *zero* strategies were attested with highest frequencies before 1914. After 1913, relative particle-based strategies represent the dominant strategy in speech-like attestations.

Node 7 indicates that a highly significant difference can be observed depending on whether the data was extracted from written or intermediate attestations. In the latter type, only the relative pronoun strategy was attested. As there are only seven tokens that are considered in the node, the high usage of the relative pronoun strategy seems to represent a particularity of a specific text. A closer look into the data confirms that all seven attestations were extracted from a single source in which acrolectal features were mixed with basilectal features. With written attestations further differences are observed depending on whether the data was attested before or after the year 1937 ($p < 0.001^{***}$). While before 1938 the strategy *rel.part+res* dominated, after 1937 *zero+res* was identified as the dominant strategy. If the author is considered as a further predictor variable, no time splits remain. Four of six observed splits are dependent on the author, one on the feature and on the text type.¹²⁴ This shows us that the author, whether the clause is a subject or object relative clause, as well as the text type have some predictive power for the choice of a RC strategy and that the year of attestation cannot be used to predict the preferred strategy.¹²⁵

To summarise these observations, the timelines have shown that the *zero+gap* strategy represents the earliest attested strategy in the SIP data. By 1908, however, the relative particle strategy has developed into an alternative strategy. The application of the ctree algorithm has shown, however, that the author, the position relativised and the text type are stronger predictor variables than the year of attestation. This indicates that extralinguistic factors have a strong influence on the choice of form. Since differences may be observable on the morpheme-level, the following sections will focus on the concrete forms used to encode relative clauses.

¹²⁴ The figure is too large to be presented here but can be viewed online at https://www.dropbox.com/s/zskyvfvczvr3c/Ctree_RC_SIP.png?dl=0 (last access 29 September 2021).

¹²⁵ This does not mean that the authors are not reliable (see explanation in Section 6.5.3.3).

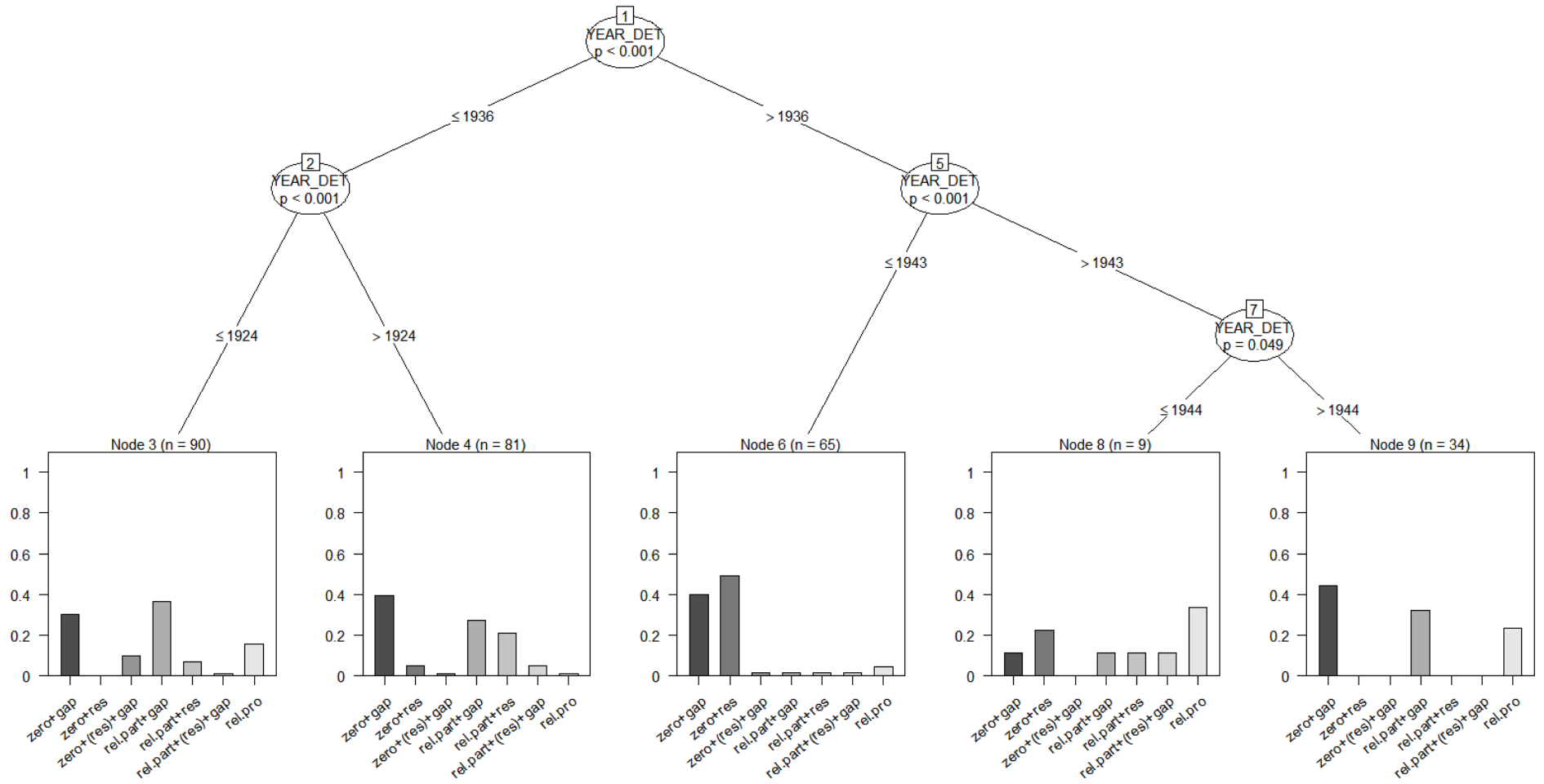


Figure 40: Conditional inference tree for relative clauses in SIP STRATEGY~YEAR_DET

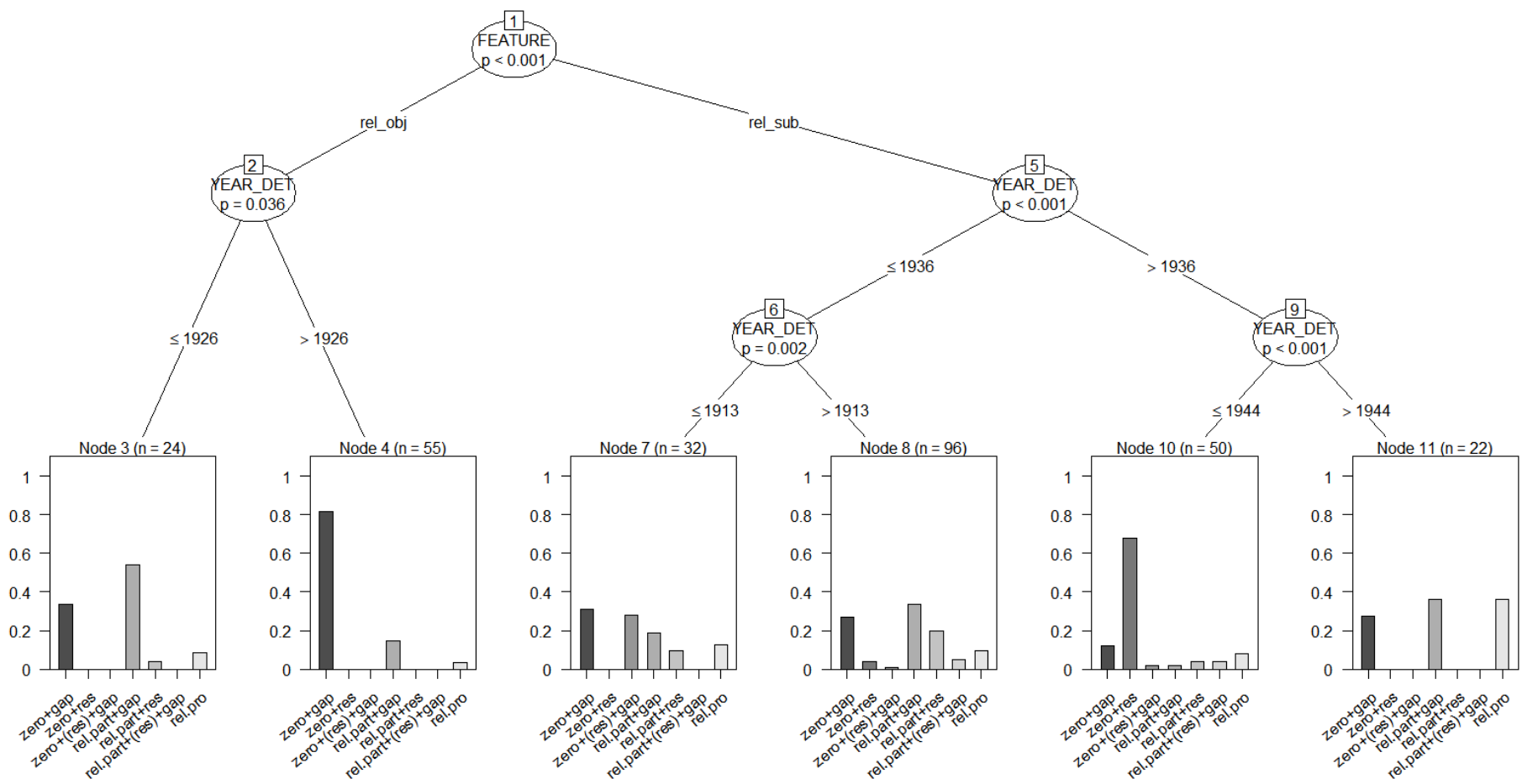


Figure 41: Conditional inference tree for relative clauses in SIP STRATEGY~YEAR_DET+FEATURE

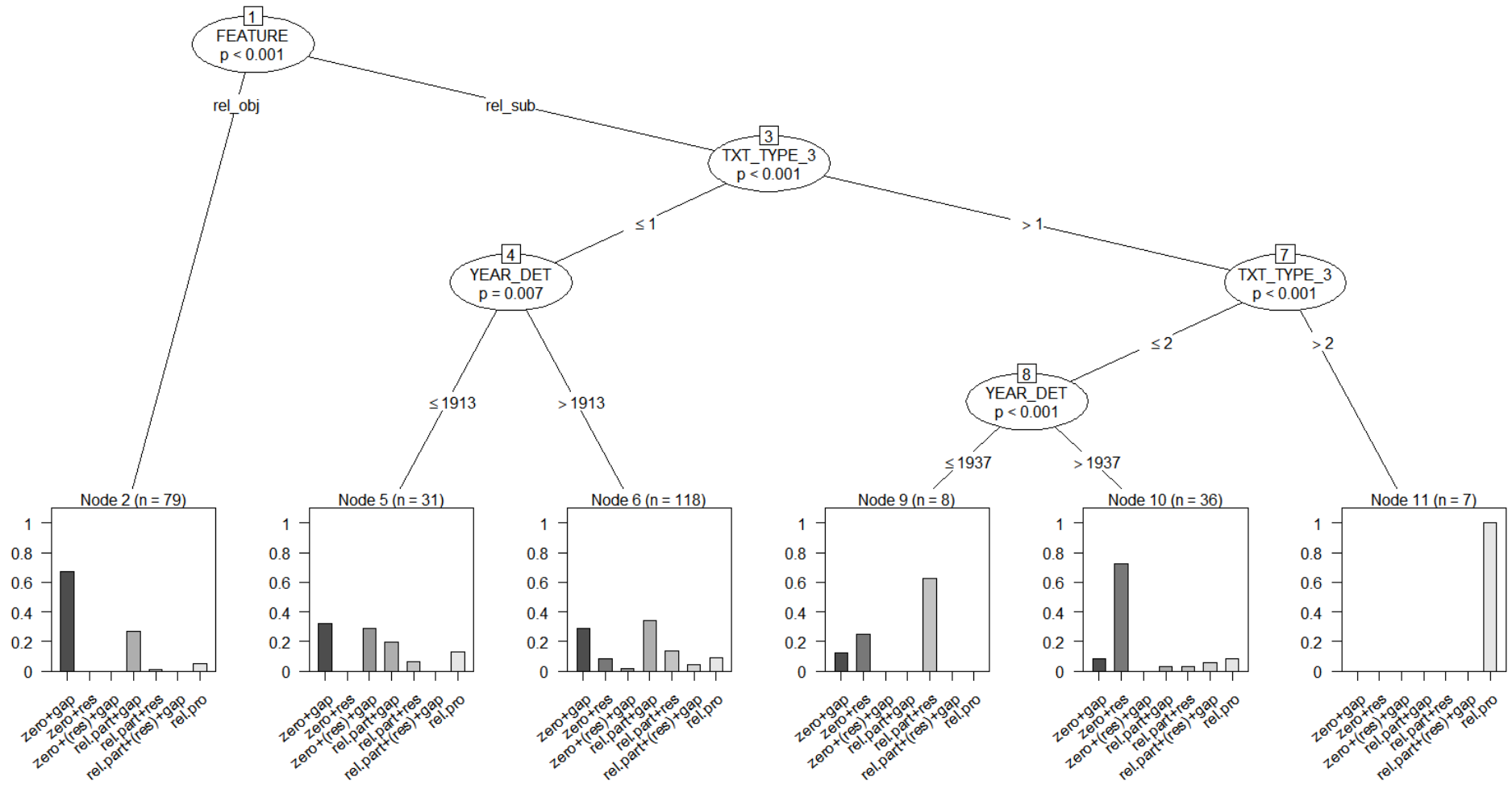


Figure 42: Conditional inference tree for relative clauses in SIP STRATEGY~YEAR_DET+FEATURE+TXT_TYPE_3

7.5.2.3 Relativisation in SIP with a focus on concrete forms used to mark RCs

7.5.2.3.1 Timeline and boxplot approach

In a second step, the analysis was repeated by focusing on the concrete forms used to mark relative clauses (= ST_FORM). Figure 43 shows the forms attested in relative clauses of early SIP across time. In comparison to Figure 38, the relative pronoun strategy is now replaced by an individual timeline for each of the individual relative pronouns *who's that*, *who* and *which*. Moreover, it can be observed that the timeline of the three relative particle strategies is reflected by the timelines of the particles *that* and *where* in the figure below. The datapoints of the three zero strategies displayed in Figure 38 are summarised in the timeline of the form *zero* in Figure 43.

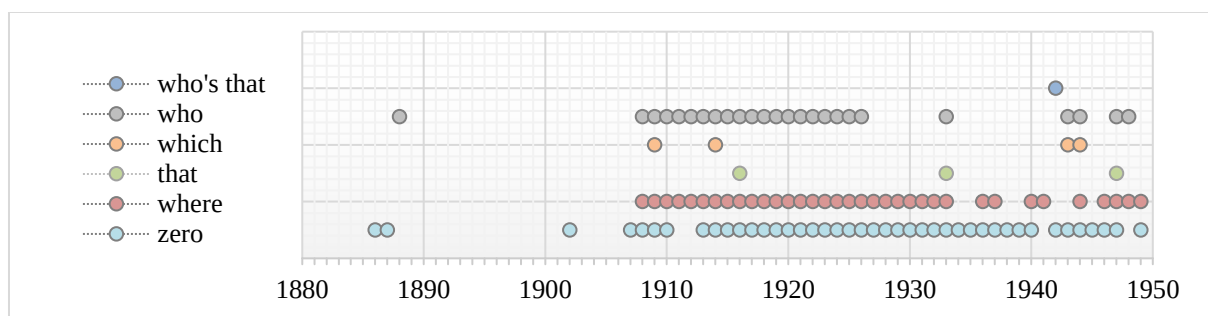


Figure 43: Timeline of RC ST_FORM variants in SIP

The timeline indicates that SIP started without overtly marking relative clauses. The first zero-attestation in the written data stems from the year 1886. Two years later the form *who* is attested for the first time. Both forms seem to have been in use until at least the late 1940s. A higher degree of variation can be observed in the post-1900 written data. From 1908 onwards the relative particle *where* is attested and remains present until the end of the investigated period. In addition, the forms *which* and *that* are occasionally attested. The single attestation of the form *who's that* dates to the year 1942. Solely based on the timeline approach, a preference for a single form cannot be observed but it appears as if *zero*, *where* and *who* developed into the dominant forms.

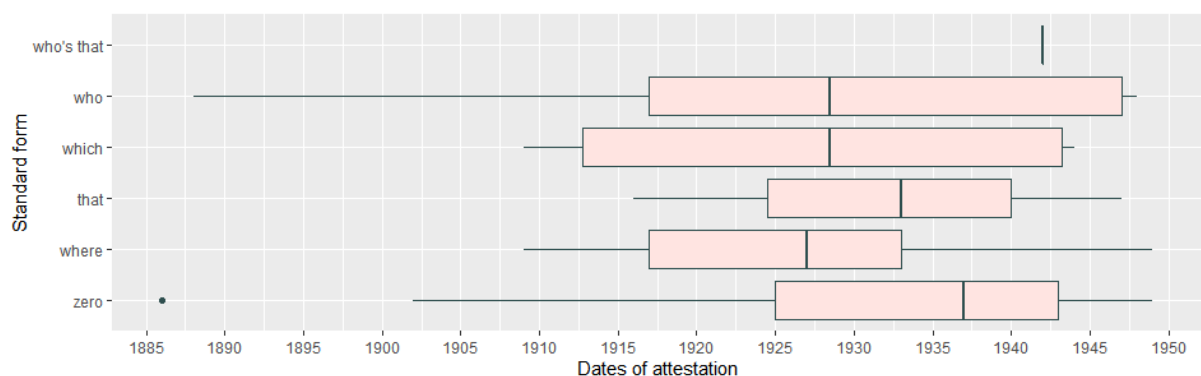


Figure 44: Boxplot of RC ST_FORM variants in SIP

The same results are obtained when creating a boxplot in which the overall distribution of datapoints is considered. Figure 44 shows that a clear preference of a specific RC marker cannot be observed as the boxes and whiskers of all forms stretch out into the 1940s.

7.5.2.3.2 Testing for the impact of the years of attestations on the choice of form

Even though the timeline and boxplot approach do not suggest that a stabilisation of forms occurred before 1950 since several forms coexisted, the ctree functionality was used to test whether the determined year of attestation (= YEAR_DET) influences the choice of the form (= ST_FORM). Figure 45 shows that a ctree analysis yields three time-splits if the year of attestation is considered as the only possible predictor variable.

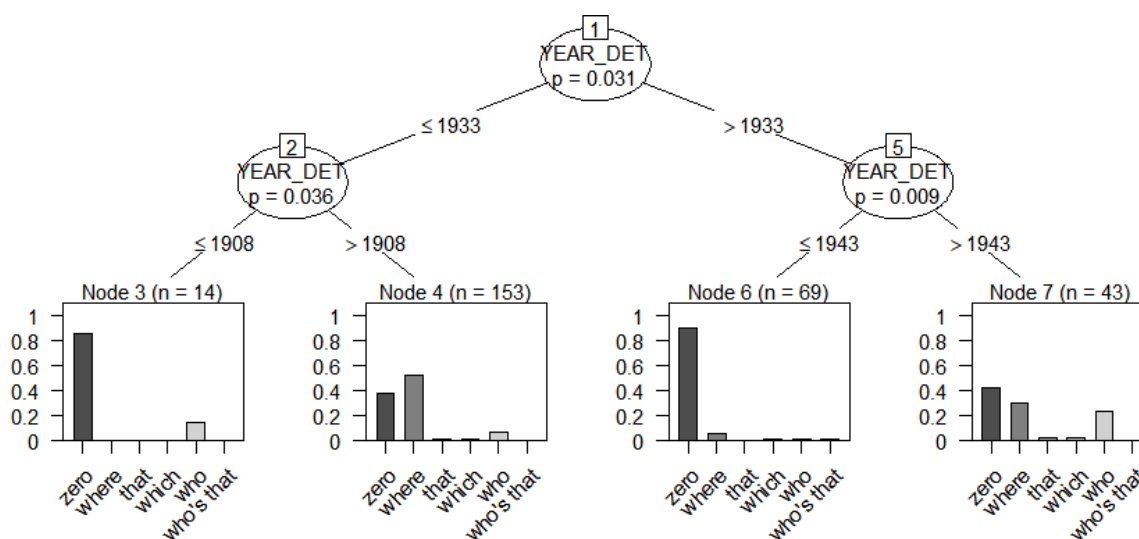


Figure 45: Conditional inference tree for relative clauses in SIP ST_FORM~YEAR_DET

The splits 1 and 2 are significant with $p=0.031^*$ and $p=0.036^*$; the third split is very significant with $p=0.009^{**}$. The splits separate the data into four time periods. In the first period from 1886-1908, zero-relatives represent the dominant form. From 1909-1933, in contrast, *where* is the dominant particle attested in the written data despite the zero strategy being still present and commonly used. Attestations referring to the years 1933-1943 predominantly show the use of a zero-particle, while *where*, *which*, *who* and *who's that* are less likely to occur. RCs belonging to the fourth period, which dates from 1943 onwards, still are most likely to make use of the zero structure but the particle *where* and the relative pronoun *who* almost lie level with zero-particles.

To test if the choice of form is dependent on whether the relative clause refers to a subject (*rel_sub*) or object (*rel_obj*), the variable FEATURE was added to the analysis, as shown in Figure 46. What is remarkable to note is that the tree yields two splits with the highest-level split being based on the variable FEATURE ($p<0.001^{***}$). The second level split is based on the year of attestation ($p<0.01^{**}$). According to the tree, object relative clauses are most likely to use the

where particle if attested before 1926 and a zero particle if attested after 1926. In each period, the other form is also attested but less likely to occur. In subject relative clauses the forms *zero*, *where* and *who* occur independent of their determined year of attestation, with *zero* being the most probable form that is closely followed by *where*.

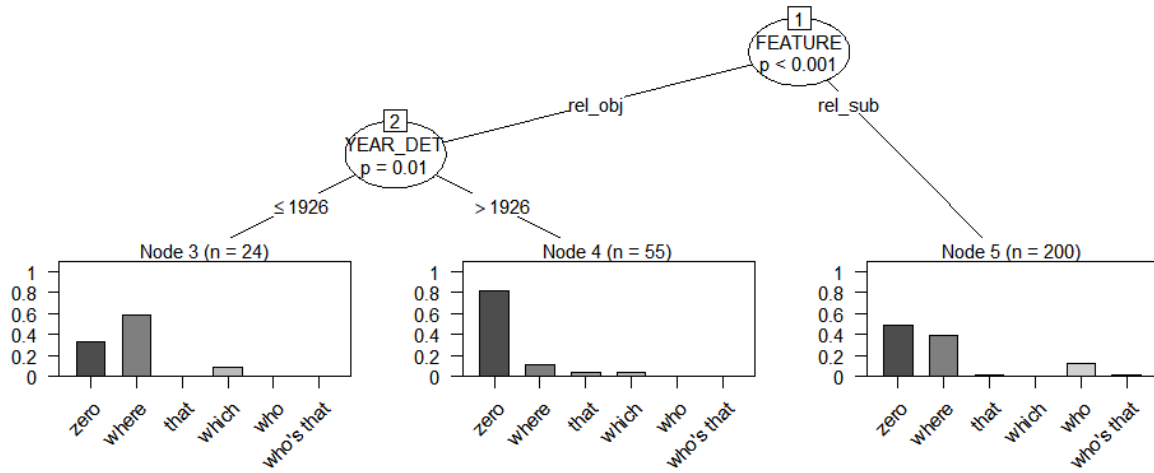


Figure 46: Conditional inference tree for relative clauses in $SIP\ ST\_FORM \sim YEAR\_DET + FEATURE$

Figure 47 shows the tree which is obtained when the text type is added as a further possible predictor variable and reveals that the most dominant splits are caused by the text type variable (cf. Nodes 1 and 2). Although there are also splits based on the year of attestation (Nodes 6, 7, and 9), these create end nodes that only refer to single years and, therefore, seem to reflect particularities of individual texts. If the more fine-grained text type classifications were used, which consist of five or 13 sub-categories (see Table 8), no time-splits were observed at all. If the author (AUTH_NAME) is considered in addition to the text type, the resulting ctree consists of three splits.¹²⁶ While the highest-level split is based on the author ($p < 0.001^{***}$), the second-level split is based on the relativised element ($p < 0.001^{***}$), and the third on the text type ($p < 0.001^{***}$). Due to these results, a closer look at the authors was taken. What is remarkable is that the particle *where* was almost exclusively found to occur in sources that were connected to the South Sea Evangelical Mission (SSEM). Only 8 of 97 *where* attestations were attested in (five) other sources.

Summarising the observations of all constructed trees, the greatest change in the SIP written data is observed in 1908 when the relative particle *where* was attested for the first time. This is supported in that the year 1908 was yielded as one split-evoking variable, as shown in Figure 45. However, as the remaining ctree analyses showed, the author, text type and the relativised element of the relative clause turned out to be the dominant predictor variables.

¹²⁶ See https://www.dropbox.com/s/djbero0q5gqmw9o/Ctree_RCF_SIP.png?dl=0 (last access 29 September 2021).

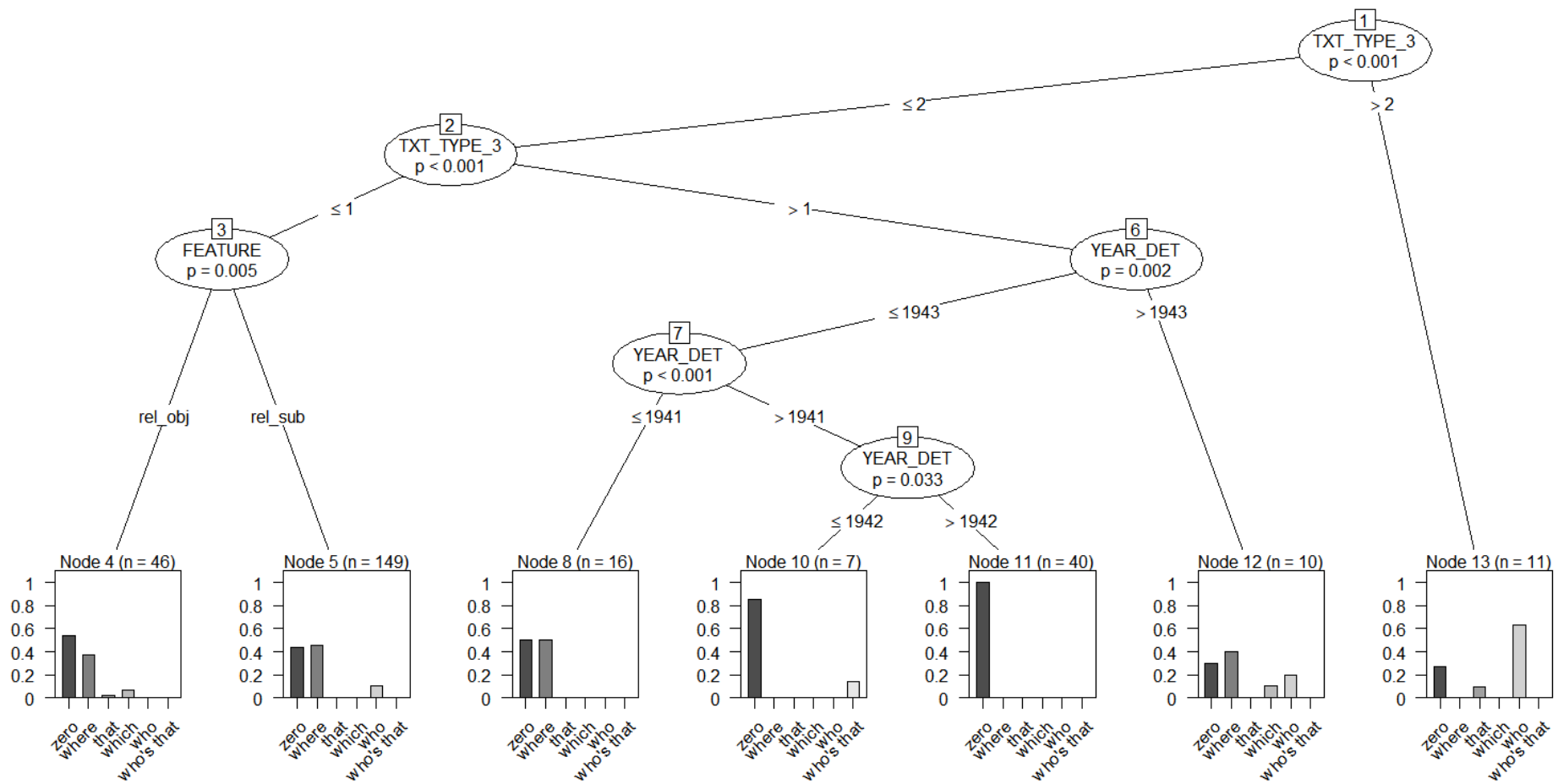


Figure 47: Conditional inference tree for relative clauses in SIP ST_FORM ~ YEAR_DET + FEATURE + TXT_TYPE_3

7.5.3 Diachronic analysis of relative clauses in Bislama

After having analysed the development of relative clauses in SIP, the present section will focus on strategies and forms attested in BIS. The attested forms will be first introduced before the diachronic development of the strategies and forms will be explored.

7.5.3.1 Attested strategies and forms in BIS

In Bislama, three of the five relative forms that were introduced in the methodology section were attested. The form with the highest frequencies of occurrence is the zero-form. It is found in 82.47% of the 97 attested relative clauses. Zeros were only attested with a gap in subject position (87-88), although ambiguous cases that could not be clearly classified existed as well.

- (87) [...] *me want to work along one man* $\left[\begin{smallmatrix} \emptyset & - & he & speak & true \\ \emptyset & - & PM & speak & truth \end{smallmatrix} \right]$
 [...] 1SG VOL to work PREP ART man
 ‘I want to work for a man who speaks the truth’
 (Atchin, New Hebrides 1917; Stewart 28.04.1917)

- (88) *You no sabby picaninny ‘ere* $\left[\begin{smallmatrix} \emptyset & - & ‘e & b’long & white\ man? \\ \emptyset & - & PM & POSS & white\ man \end{smallmatrix} \right]$
 2SG NEG know child DEM
 ‘Don’t you know the/this child, who is of the white man?’
 (Vanuatu 1917; Lynch 1923: 195)

It is interesting to note that in sentence (88) the object of the main clause, namely the child (= *picaninny*), is followed by the demonstrative pronoun *‘ere* (>*here*). As outlined in Section 7.2, the use of an orthographic variant of *here* following and modifying the object of the main clause is common in contemporary Bislama and has led some researchers to assume that it might serve as a relative clause indicating device. It occurs only in five of the attested sentences (four times in a zero-relative clause, and once in co-occurrence with the particle *where*).

The second most frequently attested form is the relative particle *where*, which occurs in 16 of 97 sentences (= 16.49%) (cf. sentence (89)).

- (89) *White man* $\left[\begin{smallmatrix} where & - & ‘e & look\ out & store & long & Liro \\ REL & - & PM & look.after & store & PREP & Liro \end{smallmatrix} \right]$
 white man
 ‘The Europeans who look after the store at Liro.’
 (Lehili 1916; Lynch 1923: 326)

Although the particle *that* is attested as well, it occurs only once in the complete dataset as was shown in sentence (80). What is remarkable is that it is characteristic for this author, Evelyn Cheesman, to regularly anglify her Pidgin English examples to fit the flow of her text. For instance, she writes “he ought to have ‘talked small’ to the missus because ‘she no like big talk’” (Cheesman 1933: 120) so that the boundaries between Pidgin English and English items are sometimes blurred. The sentence in which the form *that* occurs is introduced as “The natives began by searching for ‘small things that stop along bush’” (Cheesman 1933: 159). Although

Cheesman uses quotation marks in this context to clearly indicate the PE example, it is likely that the relative particle *that* was inserted by her to make the PE example fit into her sentence. This is supported by the fact that no other author makes use of the particle.

7.5.3.2 Relativisation in BIS with a focus on strategies

7.5.3.2.1 Timeline and boxplot approach

In early BIS three different RC strategies were identified to have been in use. Figure 48 shows their attestations across time. What can be seen is that the earliest attested strategy is *zero+gap* with its first attestation in 1871. It is likely that the relativised position could also be explicitly indicated with a resumptive pronoun as the ambiguous strategy *zero+(res)+gap* is found in 1883 for the first time. A change occurs in the data around 1913 when the *rel.part+gap* strategy is first attested. Although the zero strategies remain present, a relative particle strategy seems to establish itself as an alternative means to encode relative clauses.

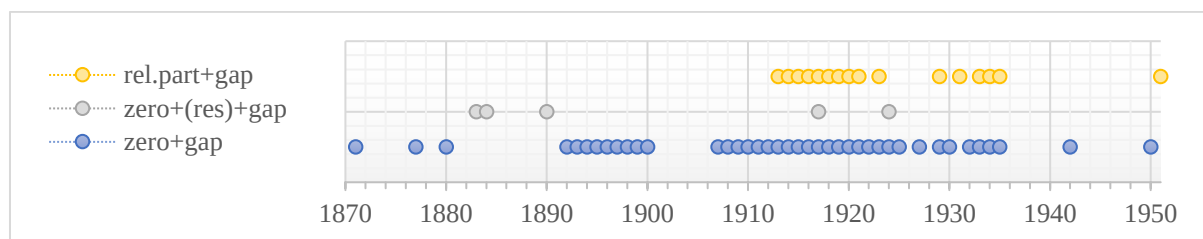


Figure 48: Timeline of RC STRATEGIES in BIS

A comparison of the timeline in Figure 48 with the boxplot in Figure 49 shows that the later attestations of *zero+gap* and *rel.part+gap* are indicated as outliers, which has to do with the fact that, in general, only few datapoints covering the years 1930-1950 could be collected for BIS. Nonetheless, the outliers show that both strategies were in use in the late 1940s. Furthermore, it is interesting to compare the medians. The median of *zero+gap* lies in 1897, whereas the median of the relative particle strategy lies in 1918. The difference in the starting date of attestations of the *rel.part+gap* strategy in Figure 48 and 49 can be explained by the fact that in Figure 48 the variable YEAR_ATT served as the basis for the analysis, whereas in Figure 49 the variable YEAR_DET was used.¹²⁷

¹²⁷ Thus, while in Figure 48 either specific dates, years or periods were used to create the timelines of attestations, in Figure 49 concrete years had to be determined for each attestation as statistical programs such as R Studio require definite dates. If periods were documented under YEAR_ATT, the year in the midst of the period was chosen as YEAR_DET.

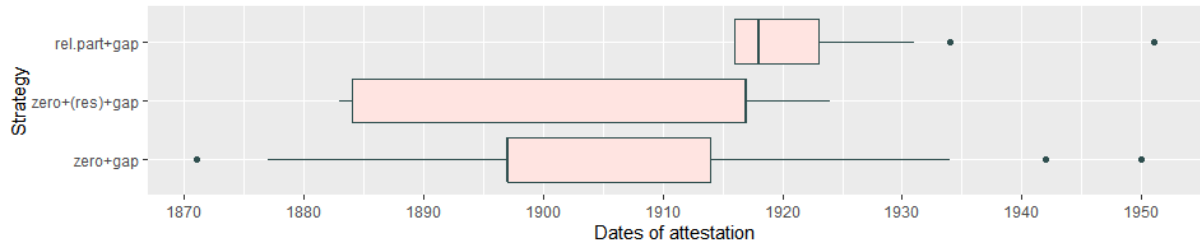


Figure 49: Boxplot of RC STRATEGIES in BIS

7.5.3.2.2 Testing for the impact of the years of attestations on the relativisation strategy

To investigate whether time represents an important factor in the choice of the relative clause strategy in BIS, the ctree algorithm was applied. First, only the year of attestation (= YEAR_DET) was considered as a possible predictor variable. Figure 50 shows the resulting tree, which consists of three time splits.

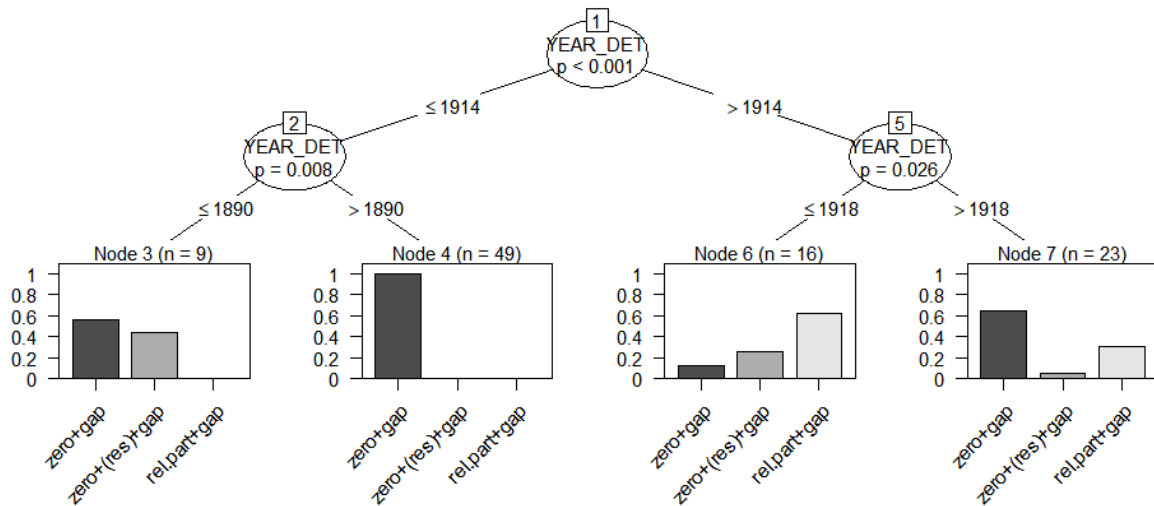


Figure 50: Conditional inference tree for relative clauses in BIS STRATEGY~YEAR_DET

The highest-level split is highly significant ($p < 0.001^{***}$) and indicates that relative clauses attested before 1915 use zero strategies only, whereas relative clauses attested after 1914 additionally use relative particle strategies. Node 2 and 5 further split the data in the years 1890 and 1918, creating four periods. The first period starts with the earliest attestations and lasts until 1890, the second period is from 1891-1914, the third from 1915-1918 and the fourth from 1919-1951. Especially the latter two periods are of interest as in the third period the relative particle strategy outnumbers the zero strategy. In the fourth period, however, the *zero+gap* strategy returns to be the dominant strategy despite the relative particle strategy remaining as well.

If the syntactic role of the head noun is considered as a further possible influencing factor (= FEATURE), the decision tree displayed in Figure 51 is obtained. What is of major interest is that the time split in 1918, which was observed in Figure 50, no longer evokes a significant split. Only the time splits in 1914 ($p = 0.001^{***}$) and 1890 ($p = 0.017^*$) remain significant split-evoking

years. Even if the text type is added as a further predictor variable, the significant splits stay the same.¹²⁸

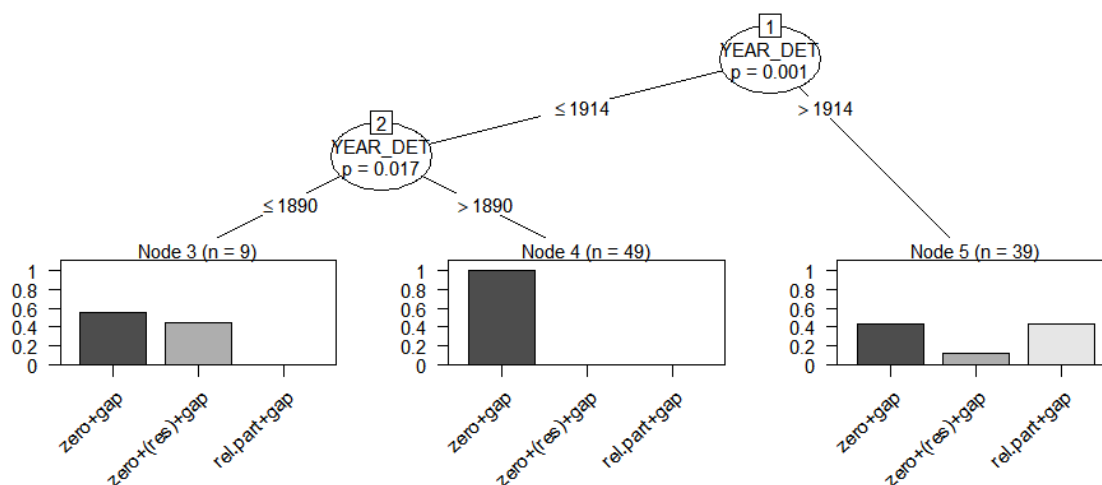


Figure 51: Conditional inference tree for relative clauses in BIS STRATEGY~YEAR_DET+FEATURE

Thus, the text type does not represent a significant predictor variable, and the years 1890 and 1914 seem to be change-evoking years. However, if the author is considered when applying the algorithm, the year of attestation no longer represents a significant predictor variable (see Appendix II). Instead, the author turns into the most significant predictor variable, evoking two splits. This indicates that the author has some predictive power for the choice of a RC strategy.¹²⁹

To sum up, the analysis of strategies in early BIS shows that the zero strategy was the earliest to be attested but that from 1913 onwards a relative particle strategy became used simultaneously. In fact, the ctree analysis revealed that the year 1914 (next to 1890) evoked changes in the choice of RC strategies if the author was not considered. Including the latter, the author variable proved to be the dominant predictor variable.

7.5.3.3 Relativisation in BIS with a focus on concrete forms used to mark RCs

7.5.3.3.1 Timeline and boxplot approach

Next, the focus was placed on the concrete relative pronoun and relative particle forms (= ST_FORM) used to encode the relative clauses. The timeline in Figure 52 shows the concrete forms and their distribution across time.

It becomes visible that zero-relatives represent the earliest attested form in the dataset of early written Bislama. Its earliest attestation dates to the year 1871 and the timeline shows that the form is persistent throughout the period that was observed in the present study. The relative

¹²⁸ The p-level changes in Node 1 to $p=0.002^{**}$ and in Node 2 to $p=0.025^{*}$ if the predictor variable TXT_TYPE_3 is added.

¹²⁹ This does not mean that the authors are not reliable (see explanation in Section 6.5.3.3).

particle *that* is only attested in the year 1929 and a possible explanation for its occurrence was provided earlier. From 1913 onwards the particle *where* is attested to occur in the early Bislama data and seems to establish as an alternative way of marking relative clauses. The form, albeit with time gaps, is attested until the end of the observed period.

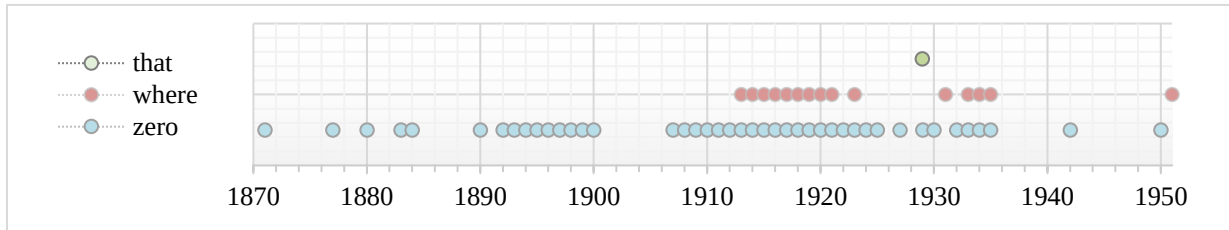


Figure 52: Timeline of RC ST_FORM variants in BIS

Comparing the observations with the boxplot results displayed in Figure 53, it can be seen that although both the zero-relative as well as the particle *where* are attested until the end of the time period under investigation, the bulk of the zero attestations dates pre-1900, the median being in the year 1897. The median of the particle *where* is in the year 1918 instead. Since there is only a vertical line for the form *that*, the boxplot shows that the form was attested in a single year only. Although it can be assumed that the particle *where* was used in spoken Bislama already prior to 1910, based on the timeline and boxplot approach, a change in the written data can be observed after 1913.

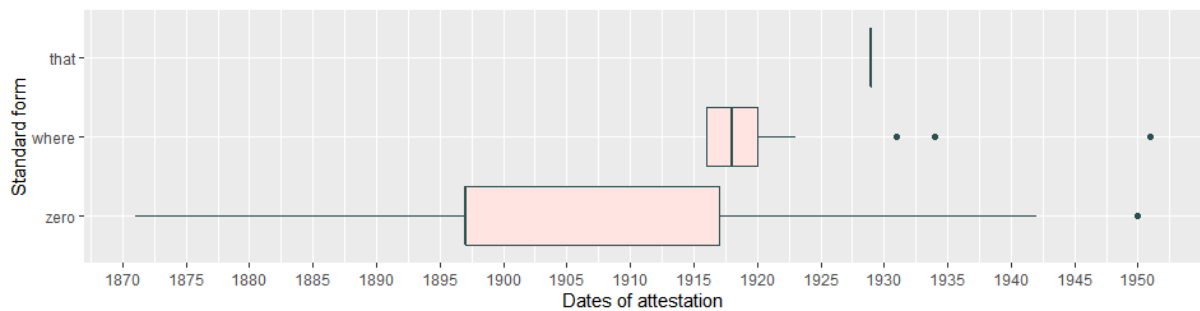


Figure 53: Boxplot of RC ST_FORM variants in BIS¹³⁰

7.5.3.3.2 Testing for the impact of the years of attestations on the choice of form

The ctree in Figure 54, which analyses the impact of the year of attestation on the choice of form, confirms the assumption that changes occur around 1913. According to Node 1, a highly significant split with $p < 0.001^{***}$ is observed in the year 1914, with pre-1914 RCs showing zero-

¹³⁰ Note that in the boxplot approach, the first attestation of *where* dates to 1916, while in the timeline approach the first attestation dates to 1913. This is due to the different variables YEAR_ATT and YEAR_DET being taken into consideration in the two approaches.

relatives only and post-1914 data showing both zero-relatives and the use of the *where* particle as an alternative marker.

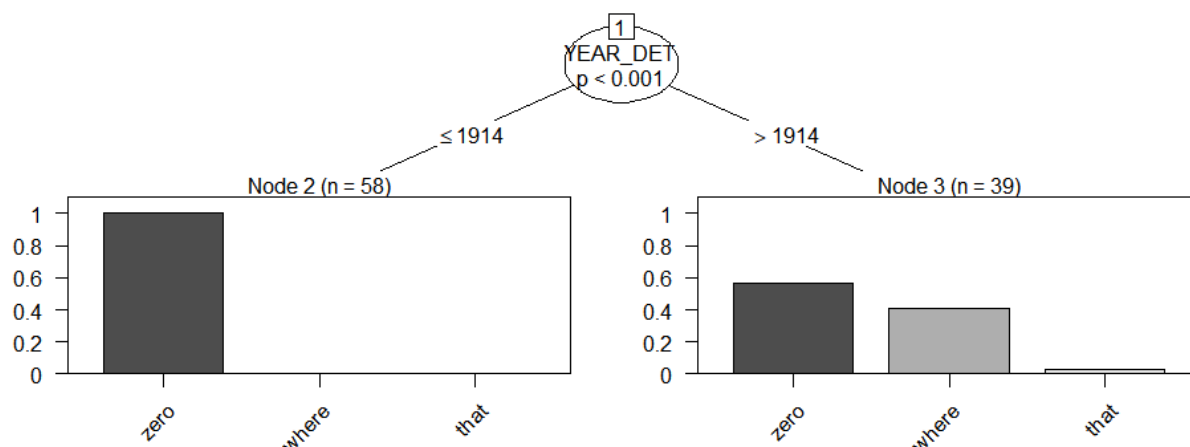


Figure 54: Conditional inference tree for relative clauses in BIS ST_FORM~YEAR_DET

Even if the variable FEATURE is added, which distinguishes between relativisation on subjects and relativisation on objects, the resulting tree does not change. This indicates that the choice of form for RCs in BIS is independent of whether the noun in the main clause is a subject or an object. The year 1914 also remains the only highly significant split if the text type is included as a further possible predictor variable. However, as soon as the author is included as well, a single split based on the variable AUTH_NAME ($p=0.001^{***}$) is obtained which shows that the author seems to be the most significant predictor variable in the choice of form (see Appendix II).

7.5.4 Diachronic analysis of relative clauses in Tok Pisin

This section focusses on the diachronic development of relative clause strategies and forms attested in TP until 1950. The strategies and forms will be first introduced before the focus will be placed on their dates of attestation.

7.5.4.1 Attested strategies and forms in TP

Almost all attested relative clauses in the Tok Pisin dataset lacked overt marking. The zero form, either followed by a gap as in sentence (90), or by a resumptive pronoun as in sentence (91), represented the dominant strategies.

- (90) All place, $\left[\begin{array}{c} \emptyset \_ \\ \emptyset \_ \end{array} \right]$ he got 'red cross' he tabu belong fight
 PL place $\left[\begin{array}{c} \emptyset \_ \\ \emptyset \_ \end{array} \right]$ PM have red cross PM prohibited PREP fight
 'The places, which have a red cross sign, are forbidden for fighting'
 (Rabaul 1914; Dempwolff 1914 in Mühlhäusler et al. 2003: 53)

- (91) Plenti man $\left[\begin{array}{c} \emptyset \text{ ol} \\ \emptyset \text{ 3PL.RES} \end{array} \right]$ i bihaind-im, ol i sin-daun long nabich
 many man $\left[\begin{array}{c} \emptyset \text{ ol} \\ \emptyset \text{ 3PL.RES} \end{array} \right]$ PM follow-TR 3PL PM sit-down PREP beach
 'Many men who were following him sat down at the beach.'

(Alexis Harbour ~1935; Wolf 1935:10)

In addition, the relative pronoun *who*, as shown in example (81), was attested only once. The single occurrence of the form *who* seems dubious and is likely to represent an author modification.

7.5.4.2 Relativisation in TP with a focus on strategies

7.5.4.2.1 Timeline and boxplot approach

Four strategies were attested in the early TP dataset. As the timeline in Figure 55 shows, three of these strategies are based on *zero*. The difference between the choice of the strategy, if one is observed at all, is only based on whether a resumptive pronoun was used or not. Both, *zero+res* and *zero+gap* are attested until the end of the observed period. The single attestation of the relative pronoun strategy in 1921 seems to be an exception rather than the norm. Even though the author uses direct speech to report what his New Guinean interpreter Jerope said, it needs to be kept in mind that the author probably did not record the words of Jerope right away. Moreover, he writes himself that he had “been studying Pidgin English for nearly a year, but had not reached the point where [he] could use it in [his] lectures” (Lambert 1942: 83). Thus, the use of the relative pronoun strategy might represent a result of author modification.

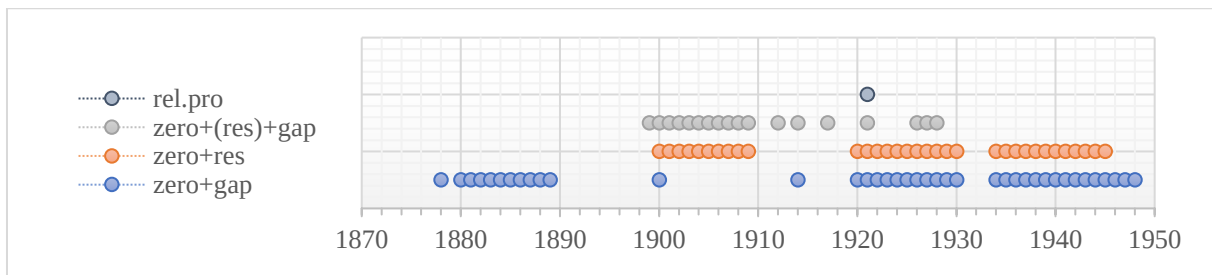


Figure 55: Timeline of RC STRATEGIES in TP

The boxplot in Figure 56 conforms with these results. As the medians of the strategy *zero+res* and *zero+gap* are both in 1940, it can be assumed that both strategies were in use by the end of the observed period.

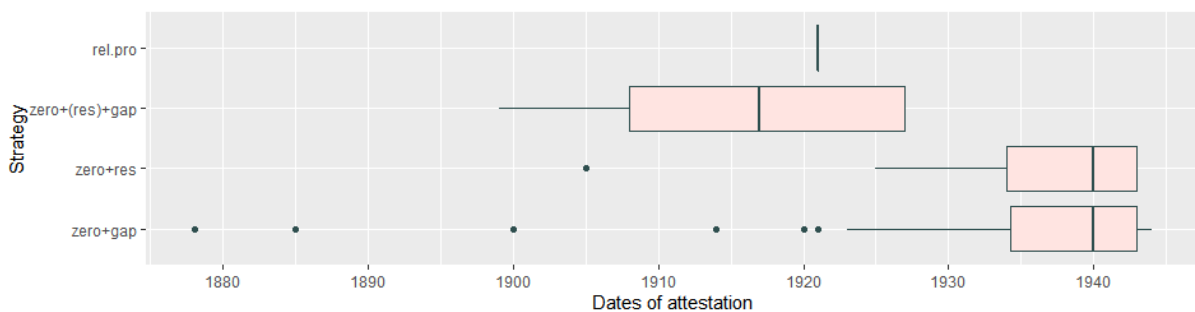


Figure 56: Boxplot of RC STRATEGIES in TP

7.5.4.2.2 Testing for the impact of the years of attestations on the relativisation strategy

Although a clear preference for zero-strategies seems to have existed in the early TP variety, the ctree algorithm was applied for the sake of completeness. Despite expectations, a ctree, analysing the impact of the year of attestation on the choice of the relativisation strategy, results in a decision tree consisting of four splits. The highest-level split in Figure 57 shows that differences in the choice of the relativisation strategy are observable depending on whether RCs were attested in and before or after 1917 ($p < 0.001^{***}$). Nodes 2, 5 and 6 indicate that the years 1904, 1927 and 1926 are further split-evoking years.

However, it is important to take the number of tokens per end node into consideration. While the first four end nodes consist of a small amount of attestations only, RCs attested after 1927 make up 357 of all datapoints. According to the tree, the *zero+gap* strategy has stabilised as the preferred option from 1918 onwards. At least in the Nodes 7 and 9 it is shown that from 1918 until 1926 and from 1928 onwards it is the preferred option. Although in Node 8 a higher amount of *zero+(res)+gap* is identified, it is striking that all 25 attestations of Node 8 refer to the year 1927 and may thus reflect particularities of a specific author. Furthermore, the ambiguous strategy does not exclude the possibility that *zero+gap* may have dominated in Node 8.

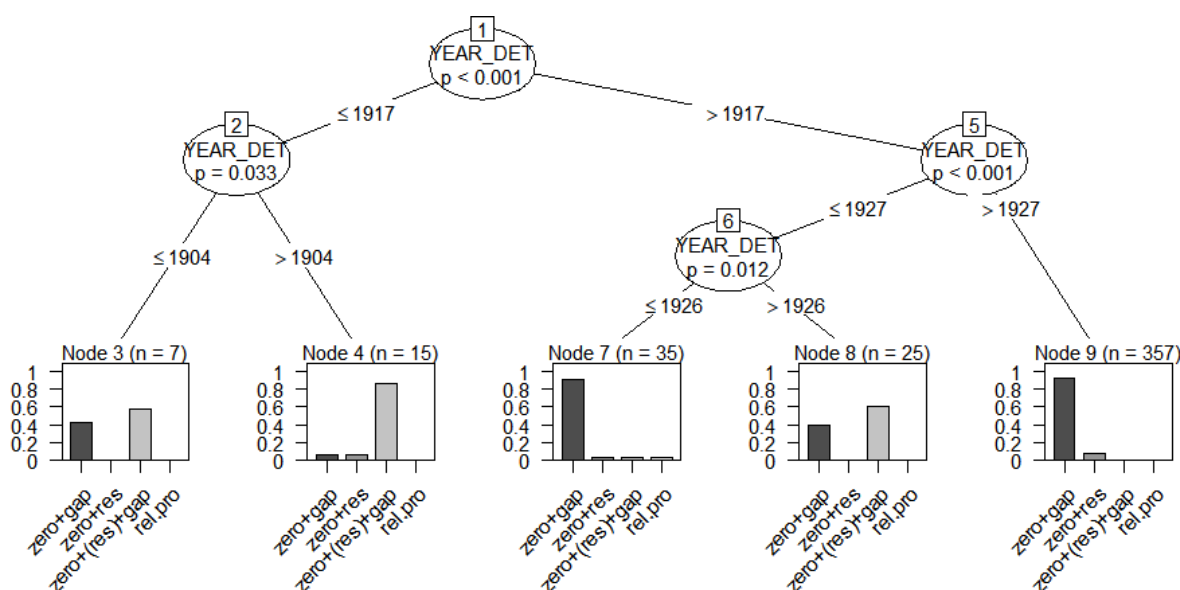


Figure 57: Conditional inference tree for relative clauses in TP STRATEGY~YEAR_DET

If the possible predictor variable FEATURE is included, which tests whether the syntactic role of the head noun has an impact on the choice of strategy, the decision tree in Figure 58 is obtained. The tree consists of three time-based splits. RCs attested before 1918 are most likely to be encoded with the ambiguous strategy. In the data after 1917 the *zero+gap* strategy is the preferred identified option. An exception forms the year 1927 in which the *zero+(res)+gap* strategy

dominates again. As explained above, this may represent a particularity of a specific text or author.

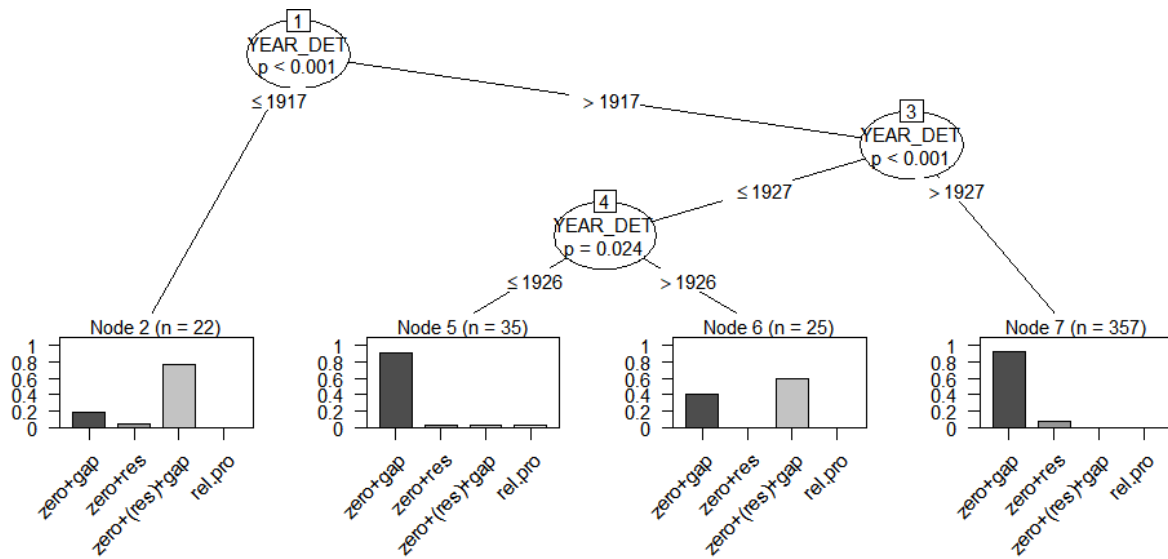


Figure 58: Conditional inference tree for relative clauses in TP STRATEGY~YEAR_DET+FEATURE

Figure 59 shows that the text type has an impact on the choice of strategy as well. While the years 1917 and 1927 remain the dominant split-evoking years, both the data between 1918 and 1927, as well as between 1928 and 1950 show further differences depending on whether RCs were extracted from speech-like attestations or written and intermediate attestations.

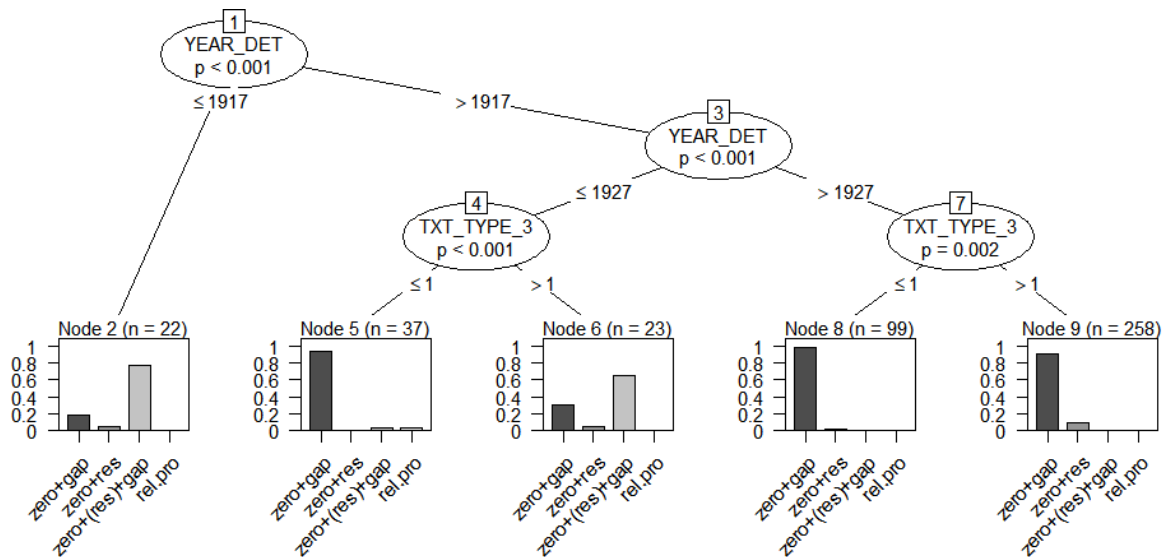


Figure 59: Conditional inference tree for relative clauses in TP STRATEGY~YEAR_DET+FEATURE+TXT_TYPE_3

If the author is considered as a further predictor variable, the resulting tree consists of four significant splits, whereby the first three are based on the author and only the fourth split is time-based.¹³¹ Again, the author seems to represent the dominant predictor variable.

7.5.4.3 Relativisation in TP with a focus on concrete forms used to mark RCs

7.5.4.3.1 Timeline and boxplot approach

Both a timeline as well as a boxplot were also created for the concrete relative pronoun and relative particle forms (ST_FORM) used to encode the relative clauses. Figures 60 and 61 indicate that a clear preference of form is observed throughout time. Zero relatives are attested from 1878 onwards and there is only a single attestation of the form *who* in 1921.

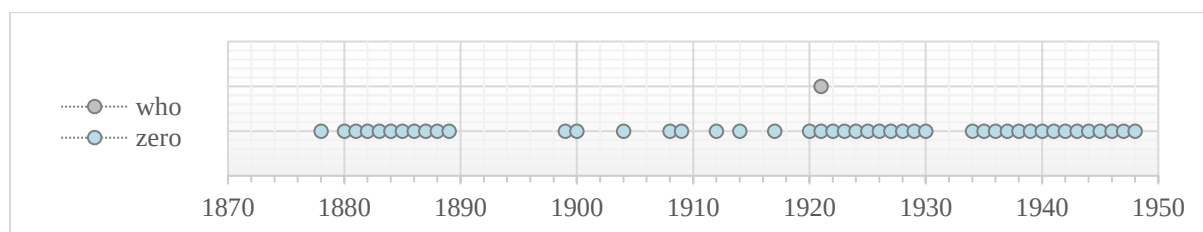


Figure 60: Timeline of RC ST_FORM variants in TP

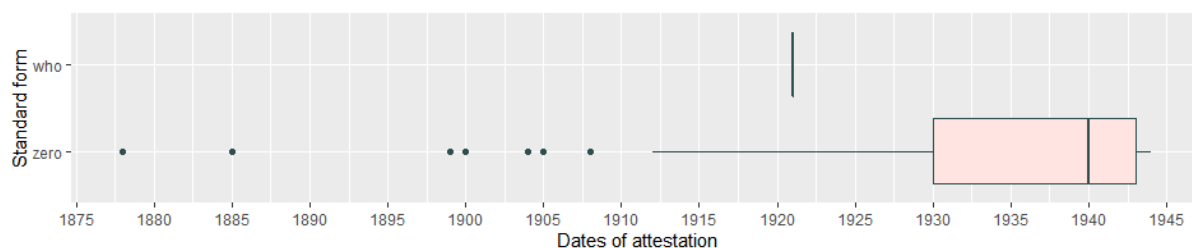


Figure 61: Boxplot of RC ST_FORMs in TP

7.5.4.3.2 Testing for the impact of the years of attestations on the choice of form

As only zero-relatives are attested except for the single, dubious occurrence of the form *who*, no time splits could be observed in the data. The algorithm results in a stacked bar blot.

7.5.5 Diachronic comparative analysis of relative clauses in MPE

Comparing relative clause strategies in the three varieties SIP, BIS and TP, it can be observed that different forms were in use to varying degrees in the three varieties across time. All three varieties are similar in that the earliest forms attested represent zero-relatives. This aligns with Bickerton's assumption that creoles start without overtly marked relative clauses (cf. 2016: 58). The *zero+gap* strategy seems to have been in use prior to the *zero+res* strategy in all three

¹³¹ See https://www.dropbox.com/s/psni2b8mxhjytjl/Ctree_RC_TP.png?dl=0 (last access 29 September 2021).

In addition, although there is this single sentence only, it gives rise to suspicion that the particle was in use in Queensland. Returning converts may thus have spread the feature to the Solomon Islands and New Hebrides. When the QKM turned into the SSEM and transferred to the BSIP, this may have contributed to the further spread and establishment of the *where* particle in SIP. Due to the fact that the mission was not active in New Guinea prior to World War II, this may explain why the feature is not found in the early TP variety. Even after World War II, the pidgin as spoken by the mission had probably little impact on TP as by then the latter mentioned variety had already established.

The early attestation of the relative particle *where* is also of interest since it contradicts earlier findings which were outlined in Section 7.3. Earlier studies do not find the *where* particle in RC constructions in SIP, but in the present data the particle is attested to occur in SIP even earlier than in BIS. While Baker (1993) assumed *where* to have its origin in Vanuatu, the present chapter outlined the possibility of its origin and spread through the QKM. As New Guineans were no longer recruited for Queensland plantations, they were not in contact with the QKM which might explain why the *where* particle was not attested in the early TP dataset.

Based on the data at hand, the assumption by Crowley (1990a: 330) that the relative particle *where* was a widespread feature in all Melanesian Pidgin English varieties is called into question. The relative particle *where* was not attested in the Tok Pisin data. Although non-attestation does not necessarily mean that a feature was not present in a variety, it needs to be considered that the Europeans usually (over-) emphasised peculiarities if they deviated from StE use when documenting the – for them – foreign language. Thus, it is reasonable to assume that they would have pointed out the use of *where* as a relative particle if it had been in use. Therefore, the results support Mühlhäusler's assumption that the *where* particle, which can be occasionally observed in contemporary TP, represents a recent (or at least a post-1950) innovation. In addition, other contemporary TP forms, such as the relative pronoun *husat* and constructions with *ia*, were not attested in the TP dataset. However, since *husat* occurred in a war pamphlet that was dropped over the Solomon Islands and usually the New Guinea variant of MPE was used in these pamphlets produced by non-speakers, it can be assumed that the relative pronoun was developing into a relative clause marker in TP as early as in the first half of the 20th century (cf. Section 7.5.2.1). Based on these observations, SIP and BIS seem to have been in closer contact.

Building a tree based on all three varieties to analyse the impact of the year of attestation, the syntactic role of the head noun and the variety on the form, the output tree in Figure 62 is obtained. The tree shows that the highest-level split is based on the variety, with $p < 0.001^{***}$. What appears to be rather striking is that the first split separates SIP from TP and BIS even though

zero-relatives and the *where* particle can be found in SIP and BIS. However, as the SIP data showed the greatest variability in terms of whether standard *lexifier features* were included into utterances, this may be one of the reasons why it is separated from BIS and TP on the first level. BIS and TP are split on the second level ($p < 0.001^{***}$). While in BIS the year 1914 is displayed as the split-evoking year, the choice of form in TP is time-independent.

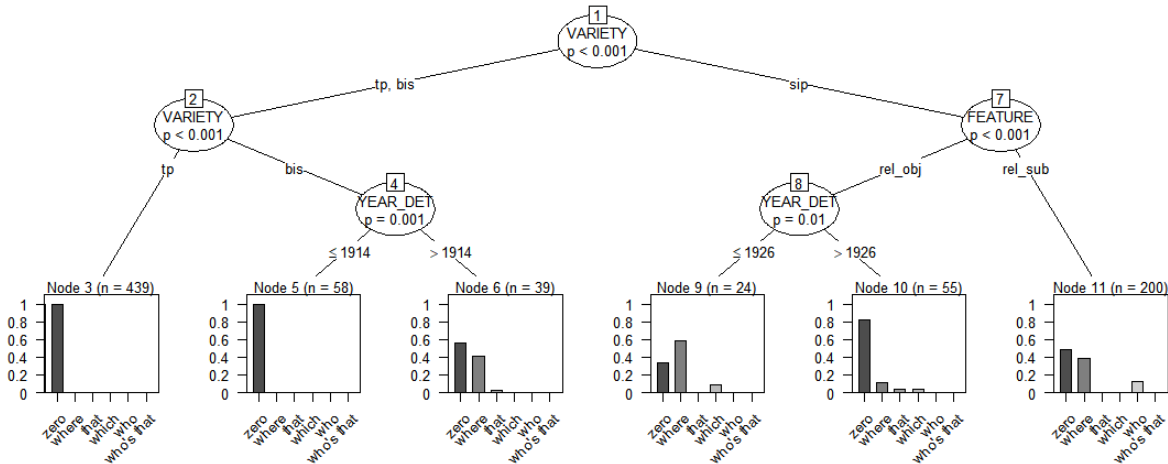


Figure 62: Conditional inference tree for relative clauses in TP ST_FORM~VARIETY+YEAR_DET+FEATURE

Regarding the question when the varieties diverged from each other, no concrete answer can be given. The individual analysis with the help of ctree clarified that the time splits observed in SIP are dependent on the text type as well as the author. In the BIS data the year 1914 proved to evoke a split even if the variable text type was considered. However, as soon as the author was considered as a further possible predictor variable, the latter turned into the dominant predictor variable for the choice of the form. In regard to TP, a preference for *zero* was attested independent of the year of attestation.

7.6 Summary

The present chapter traced the development of strategies and forms used to encode subject and object relative clauses in SIP, BIS and TP. It started with a short introduction of the theoretical background (7.1) before it provided an insight into subject and object relative clauses in contemporary SIP, BIS and TP (7.2). Previous research was reviewed in Section 7.3 before the methodological considerations were explained (7.4). The analysis of attested forms and their change across time was presented in Section 7.5.

The analysis of concrete *forms* used to encode relative clauses did not confirm the general claim that the end of the labour trade resulted in diverging forms used in the varieties. It is true that the first attestations of *where* in subject and object relative clauses were traced to the years 1908 in SIP and 1913 in BIS and were thus after the end of the labour trade. However, it needs to

be taken into consideration that the first attestation in the written data is not equatable with the first attestation in spoken language. It can be assumed that a time gap exists which suggests that *where* was already in use in the varieties before the end of the labour trade. This is also supported by the fact that the relative particle was not attested in the early TP data but occurred in a letter written by a Pacific Islander in Queensland around 1898-1899. Although it is likely that the relative particle *where* will have further stabilised in SIP and BIS after the end of the labour trade, the differences between SIP, BIS and TP in the *choice of form* are more likely to have their origin in the labour movement and plantation histories of the areas. It should be kept in mind that despite the *where* particle being attested in SIP and BIS, the present study does not provide information about post-1950 developments. Nonetheless, it seems likely that the particle persisted until today, since *where* is also found in contemporary BIS and SIP.

The high number of StE originating forms in SIP demonstrate the meso- and acrolectal character of the early SIP data. In addition, the closeness to StE forms might have its origin in author and editor modifications. However, as even early writers such as Reed (1943: 270) claimed Solomon Islands Pijin to be closer to StE than the remaining varieties, it is likely that the variety in fact showed acrolectal characteristics. The variability of attested forms may further be an indicator that the grammaticalisation of relative clauses was not finalised yet.

That the stabilisation of relative clauses was still ongoing can be further observed in the BIS dataset. Contemporary BIS only makes use of the relative particle strategy in subject and object relative clauses, but the early data still shows the use of both relative particle and zero strategies. Thus, the written data material is not sufficient to make claims about when the relative particle strategy developed into the only grammatical option. The splits obtained by ctree indicated changes in the forms and strategies used in the varieties, but most of the time splits were no longer observable as soon as the author was included as a possible predictor variable.

The analysis presented several methodological challenges. Multifunctional forms such as *he* led to a difficulty in classifying RCs according to the strategies. This resulted in the construction of categories for ambiguous cases which might have an impact on the results. In addition, the datasets could not be collected according to the same standards and are based on the surviving written record. Hence, differences in the quantity and quality of the data were observed. The SIP data contains a vast amount of material written down by the SSEM, for BIS little to no data is available for the years 1935 until 1950 and the TP data consists primarily of World War II material. Taken this into account, there is the possibility that the data availability has an impact on the outcome of this study. Future contributions may change the results and post-1950 data needs to be collected to obtain a better understanding of the varieties' development from 1950.

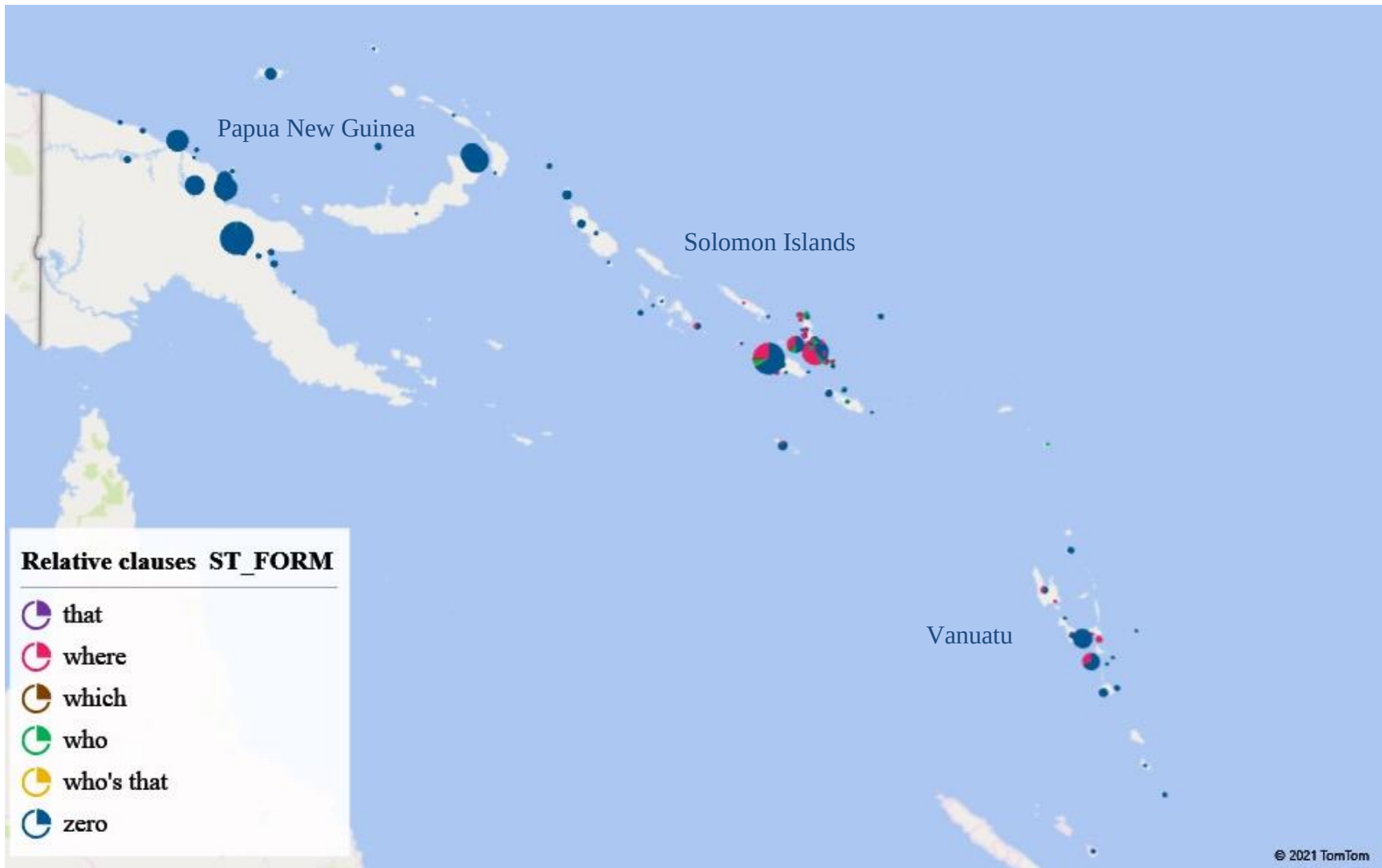


Figure 63: Map of Melanesian Islands, showing attestations of relative clause ST_FORM variants

8 Case Study: Modality

There is no doubt that the overall picture of the modals is extremely ‘messy’ and untidy and that the most the linguist can do is impose some order, point out some regularities, correspondences, parallelisms.

(Palmer 1990: 49)

During the morpheme-by-morpheme analysis of the data it turned out that the varieties showed differences in the forms used to encode certain semantic notions of modality. In this section I will therefore focus on markers used to encode *volition*, *ability*, *permission* and *speculation* in the three varieties. Since it is often said of pidgins that they are unlikely to show modality distinctions by overt mood markers, it is of special interest *when* the varieties developed verbal markers because this may be an indicator for their stabilisation into pidgincreoles. In addition, diverging forms may bring to light the history of the varieties and the impact which the end of the labour trade had on the feature development. The chapter starts with an outline of the underlying framework (Section 8.1) which is followed by a short focus on modal verbs in the contemporary varieties (Section 8.2). Section 8.3 deals with previous diachronic research on the development of mood markers in the MPE varieties. Sections 8.4 – 8.7 will focus on the diachronic analysis of volition, ability, permission and speculation markers in the three varieties. Each section starts with an outline of the methodological considerations before the findings are presented. The chapter will close with a summary and some concluding remarks.

8.1 Theoretical background

Modality is the concept which describes the variety of semantic notions which can be distinguished when “dealing with speakers’ judgements expressing their world view, and not a reality outside language” (Narrog 2012: 7). The notion of *modality* is frequently mentioned together with *mood*. The two labels have often been used interchangeably as a “category that codes a speaker’s attitude toward a situation or statement” (Velupillai 2012: 214).¹³² If they tend to be distinguished, however, *mood* is usually considered a category of grammar and modality a category of meaning (cf. Huddleston & Pullum 2016: 172). As Velupillai (2012: 214) points out, “mood tends to denote a higher level distinction for the whole clause of realis [...] versus irrealis [...]”. In other words, it is a grammatical category which codes whether a speaker asserts an event to be true (factual) or not. *Modality*, by contrast, represents a semantic sub-category of mood which tries to distinguish between different types of speaker attitudes towards possible or

¹³² The term *mode* is sometimes used as a cover term for both mood and modality (cf. Velupillai 2012: 214).

imagined events and includes a variety of semantic notions such as ability, permission, volition, prohibition, obligation and hypotheticality.¹³³

The focus of the present chapter is on *modality*. Several attempts have been made to distinguish the various types of modality that can be distinguished (cf., for instance, Bybee et al. 1994; van der Auwera & Plungian 1998; Palmer 2001; Aikhenwald 2010). Like the general use of the labels mood and modality, however, there is also no agreement on the number and types of modalities that linguists have distinguished. The present chapter follows the classification system introduced by Palmer (2001):

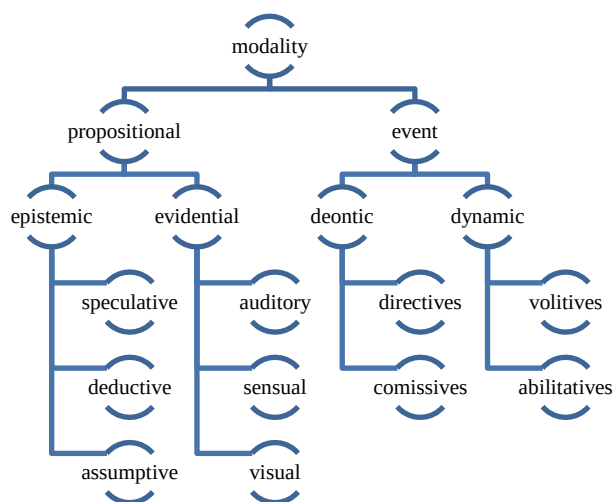


Figure 64: Classification of modality¹³⁴

Palmer (2001) distinguishes between *propositional* modality and *event* modality. The former is described as “the speaker’s attitude to the truth-value or factual status of the proposition” and is thus concerned with the degree to which speakers consider a proposition to be true or factual (Palmer 2001: 24). In contrast, event modality refers to a speaker’s judgement regarding potential future events (cf. Palmer 2001: 8).

Propositional modality can be conceptually subdivided into *epistemic* and *evidential* modality. *Epistemic* modality refers to speaker’s judgements concerning the factual status of a proposition. To put it in the words of Nuyts (2001: 21), a speaker evaluates “the chances that a certain hypothetical state of affairs under consideration [...] will occur, is occurring, or has

¹³³ It should be noted that this is just one way of using the notions *mood* and *modality*. In the literature several controversial definitions of mood and modality can be found (cf. Nuyts 2016a; van der Auwera & Zamorano Aguilar 2016). While modality is usually understood as the notion “covering semantic domains such as abilities/needs, potentials/inevabilities, deontics, and epistemics” (though controversies concerning what to include exist as well), mood has been inter alia used to refer to the “grammatical coding of [these] meanings on the verb”, to refer to “basic sentence types and the illocutionary categories expressed by them” or to refer to “indicative vs subjunctive or realis vs irrealis coding and its semantics” (Nuyts 2016a: 1-2).

¹³⁴ Categories such as future, negative, interrogative, conditional, purposive, presupposed, conditional, purposive, imperative-jussive, wishes and fears are, according to Palmer (2001: 22), usually found with mood and therefore do not appear in Figure 64.

occurred in a possible world”. Thus, in a sentence such as *Jenny must have gone to work*, the speaker makes the judgement based on evidence, e.g. Jenny told the speaker earlier that she will soon go to work. Epistemic statements can range in their degree of certainty, possibility and likelihood. Palmer (2001: 24-25) distinguishes between three main levels, which he refers to as *speculative*, *deductive* and *assumptive*. Speculative, as the designation indicates, refers to statements in which the speaker is uncertain about the factual status, resulting in a sentence such as *Jenny may have gone to work*.¹³⁵ In deductive statements, the speaker makes a judgement based on evidence as in *Jenny must have gone to work*. Assumptive refers to judgements that are made in terms of what is generally known about the proposition. Thus, if it is generally known that Jenny goes to work from 8 am to 3 pm every day, the example sentence might be changed into *Jenny will have gone to work*.

Evidential modality refers to evidence which is available to support the factual status. An example sentence for evidentiality represents *Jenny claims to have gone to work*. The speaker of the utterance indicates as evidence for the factual status that *Jenny has claimed it*. Evidential modality can be expressed through visual, sensual and auditory input. While English does not have grammaticalised evidentials, grammatical markers of evidentiality are common in the languages around the world (cf. de Haan 2013).

Event modality is also subdivided into two categories, namely into *deontic* and *dynamic* modality (cf. Palmer 2001: 70). While “with deontic modality the conditioning factors are external to the relevant individual, [...] with dynamic modality they are internal” (Palmer 2001: 9). The most common type of *deontic* modality are directives which have the aim to cause a person to perform an action. They include, for instance, obligation (*You must eat it*) and permission (*You may eat it*).¹³⁶ Another type of deontic modality are commissives, through which a speaker ensures that an action will take place by implying a threat or promise (*You shall eat it, or you do not get a dessert*).

Dynamic modality relates to the internal ability and willingness of a subject to perform an action. For instance, in a sentence such as *He can eat five apples in two minutes* the physical and mental abilities enable the subject of the sentence to perform the action. As Palmer (2001: 10) indicates by bringing in the example sentence *He can escape*, the category may also refer to “possibility in a more general sense”. In that sentence, it is not the physical or mental ability which enables the subject to escape, but rather the general circumstance that the door is not locked.

¹³⁵ According to Palmer (2001: 25), speculative modality may also be named dubitative modality. He prefers the former designation as “the forms do not generally indicate positive doubt”.

¹³⁶ It could be argued that imperatives (*Eat it*) and jussives (*Let me eat it*) also belong into this category.

Therefore, we may differentiate between neutral/circumstantial and subject-oriented dynamic modality (cf., for instance, Palmer 1990: 83). While subject-oriented ability markers can be paraphrased with ‘has the ability/competence to’, neutral/circumstantial possibility has to be rephrased with ‘it is possible for’ (cf. Palmer 1990: 84). As Palmer clarifies, although “[o]nly animate creatures may have ability [...] subject orientation is possible with inanimates, where it indicates that they have the necessary qualities or ‘power’ [...] to cause the event to take place” (Palmer 1990: 85). A differentiation between possibility and ability is, according to Palmer (1990: 85), not always possible.

In addition to the distinction problem discussed above, it may also be difficult to distinguish between neutral possibility and deontic modality. By providing the example sentence “*In the library you can take a book out and keep it out for a whole year unless it is recalled*” (Palmer 1990: 103), Palmer successfully shows that the auxiliary may encode either deontic modality or “what is dynamically possible or necessary” (Palmer 1990: 104).

Palmer argues that the major difference between deontic and dynamic modality is that deontic modality emanates from an external source, whereas dynamic modality “comes from the individual concerned” (Palmer 2001: 10). Therefore, he classifies *volition*, which can be defined as the willingness of an individual, as dynamic modality as well. However, it needs to be noted that the classification of volition has been highly debated. One reason for classification difficulties is that volition “may extend its scope over another participant in the projection than the subject of the volition itself, so that it becomes an indication of deontic necessity or obligation (e.g. I want YOU to help)” (Verplaetse 2003: 155; see also Palmer 2001: 134-135). Yet another position is that volition should not be considered a category of modality at all (cf., for example, van der Auwera & Plungian 1998; Nuyts 2016b).

Languages around the world differ in the grammatical strategies used to encode the various semantic notions of modality. Modality can be encoded through modal verbs, verbal inflections or by suffixes, clitics or particles (cf. Palmer 2001: 19). In addition, languages may differ in whether they show modality differences at all.

In terms of contact languages, it has frequently been claimed that pidgins lack modality distinctions or that they at least have no overt grammatical markers to express modality. Instead, it is assumed that contextual information is required to interpret sentences in terms of their modality (cf. Parkvall & Bakker 2013: 42). Creoles, by contrast, were said to possess a single tense, a single mood and a single aspect marker only, namely the anterior tense, the punctual aspect and the irrealis mood, following the prototype theory introduced by Bickerton (1981). Research in the last two decades has shown, however, that the TMA of many creoles is much

more diverse than the prototype theory suggests (cf. Winford 2018). For instance, Velupillai (2015: 399) demonstrates that most creoles have more than one mood marker and that pidgincreoles show mood distinctions as well. According to her analysis, only pidgins “are unlikely to have modal marking” (Velupillai 2015: 393, 395, 399). The varieties under investigation support the results of Velupillai in that they all have a well-developed TMA system by now, which consists of several tense, mood and aspect markers.

8.2 Modality in MPE today

Several *modality* markers are used in the contemporary varieties of MPE. While there are some modality markers which can be found in all three varieties, such as preverbal *mas* to indicate obligation (cf. Jourdan 2002: 132; Smith 2002: 136; Crowley 2004: 97),¹³⁷ differences can be inter alia observed in regard to the encoding of *volition*, *ability* and *permission*.

To express volition contemporary BIS uses the auxiliary verb *wantem*, whereas TP uses *laik*. In SIP the four verbs *laek*, *laekem (fo)*, *wande* and *wandem (fo)* represent alternative possible markers (cf., for instance, Crowley 2004: 100 for BIS; Smith 2002: 128 for TP and Jourdan 2002: 115, 256 and Huebner & Horoi 1979: 132-133 for SIP). The auxiliaries precede the full verb.

Ability is encoded with the form *save* in SIP and BIS. As the SIP sentence (93) and the BIS example (94) illustrate, the form is negated with the preverbal negative marker *no* to express inability. In addition, both varieties are reported to share the form *kanduit* to express inability. With regard to BIS, Crowley claims the form to be archaic and the meaning to be more specific than the marker *save*, as *kanduit* expresses ‘not manage to’ (2004: 101). Besides the similarities, contemporary SIP shows a range of further *ability* markers not commonly used in BIS. To denote inability, the forms *kan* and *kanot* can be used in preverbal position. The forms *fitim fo* and *inaf fo* represent additional means to express ability though their meaning is more restricted than the one of *save*. The form *fitim fo* refers to physical, mental or emotional competence, whereas *inaf fo* refers to the “skill at performing an action” (Huebner & Horoi 1979: 117).

Contemporary TP deviates most from the other two varieties in that *save* is usually not used to encode ability. Mühlhäusler (1985c: 387) argues that even though *save* can be used “to express competence in the sense of knowing how to do something”, its usage in abilitative contexts is usually avoided due to its ambiguity with habitual action. Smith (2002) does not attest

¹³⁷ In all three varieties *sud* is used as an alternative to *mas*. The form has its origin in the English word ‘should’ but is less commonly used than *mas*.

save as a means to encode ability in his TP corpus. Instead, the form *inap* is used to encode ability and its negated counterpart *no inap* to encode inability, as exemplified in (95).

Solomon Islands Pijin:

- (93) *Sapos mi siki bae mi no save go miting*
 if 1SG sick FUT 1SG NEG ABIL go go meeting
 ‘If I am sick, I will not be able to go to the meeting.’

(Jourdan 2002: 199)

Bislama:

- (94) *Joel i no save klaem kasem top blong hil*
 Joel PM NEG ABIL climb TERM top POSS hill
 ‘Joel couldn’t climb to the top of the hill.’

(Crowley 2004: 101)

Tok Pisin:

- (95) *Ol no inap kuk-im kaikai bilong yu.*
 3PL NEG ABIL cook-TR food POSS 2SG
 ‘They are not able to cook your food.’

(Smith & Siegel 2013a: 220)

Differences in the varieties can also be observed in the encoding of permissive and prohibitive modality. Contemporary SIP and BIS encode the notions with *save* and its negated form *no save*, as shown in sentence (96) and (97). In both varieties, *save* thus represents a multifunctional item which, next to being used as a full verb and as a marker of habituality, is used to encode ability and permission. It also fulfils the former two functions in contemporary TP, and may be used to encode competence, but to express permission and prohibition, *ken* and negated *no ken* are used instead (cf. sentence (98)).

Solomon Islands Pijin:

- (96) *Boe ia save kaekae kek tu?*
 boy DEM PERM eat cake too
 ‘Is the boy allowed to eat cake as well?’

(Jourdan 2002: 199)

Bislama:

- (97) *Long saed blong kastom ol woman oli no save dring kava.*
 PREP side POSS custom PL woman 3PL NEG PERM drink kava
 ‘Traditionally women were not allowed to drink kava.’

(Crowley 2004: 99)

Tok Pisin:

- (98) *Taim yupela go long ples no ken stori long ol mama [...]*
 when 2PL go PREP village NEG PERM tell PREP PL mother [...]
 ‘When you go back to the village you are not allowed to tell your mothers [...]

(Smith & Siegel 2013a: 220)

Though *ken* usually indicates permission and *inap* usually expresses ability, Smith (2002: 137) remarks that the two forms and their meanings are sometimes used interchangeably and cannot always be distinguished clearly.

In speculative contexts, the varieties make use of one or several of the dubitative markers *maet*, *mebi* and *ating*. *Maet* derives from the StE auxiliary verb *might*, which is used in the lexifier language to express speculative modality. The form is attested in SIP and BIS in dubitative

contexts as well, as shown in (99) and (100), but has, however, changed its syntactic position. Instead of directly preceding the verb, the dubitative marker occurs in clause-initial position. The form *mebi*, which is attested in contemporary BIS and SIP, derives from StE *maybe*. As *maet*, it occurs in utterance-initial position, precedes the subject and “can only ever be embedded within an utterance” (Crowley 1990a: 207). The third form *ating* derives from StE *I think* and is used in all three varieties (cf. Mühlhäusler 1985c: 377; Crowley 2004: 143; Jourdan 2008: 480). Sentence (101) shows a contemporary TP example. While *I think* occurs in clause-initial position in SIP (cf. Simons & Young 1978: 161), it can occur in various positions in BIS sentences (cf. Crowley 2004: 143).

Solomon Islands Pijin:

- (99) *Fren bilong mi maet hem-i save*
Fren bilong mi ating hem-i save
 friend POSS 1SG **might/DUB** 3SG-PM know
 ‘Maybe my friend knows./My friend might know.’

(Jourdan & Keesing 1997: 408)

Bislama:

- (100) *Maet hem i kam*
Mebi hem i kam
Ating hem i kam
maybe/DUB 3SG PM come
 ‘Perhaps he comes/He might come.’

(Crowley 2004: 143)

Tok Pisin:

- (101) *Ating em bai help-im mi*
probably 3SG FUT help-TR 1SG
 ‘It will probably help me./It might help me.’

(Smith & Siegel 2013a: 218)

Although in BIS all three forms are present and can be used interchangeably, research by Tryon (1991: 517) and Tryon & Charpentier (2004: 292) has shown that *maet* is the main and most frequently found form, while the forms *mebi* and *ating* represent regionalism that nowadays can only be attested in older Bislama speakers in Tanna. In Tok Pisin only the form *ating* is used and the use of *might* is not observed (cf. Smith 2002: 137).

It may be arguable whether the three markers warrant inclusion in a section that focusses on modality markers, since it is not clear to what extent the forms have already grammaticalised into epistemic markers. Researchers express different opinions regarding the functionality of the forms *maet*, *mebi* and *ating*. In research focussing on SIP, the forms are classified as markers used to encode dubitative modality (cf., for instance, Jourdan & Keesing 1997). Referring to BIS, Crowley describes the markers as “words performing a diverse range of functions that can also be included within the overall class of adverbs” (2004: 142), and Yakpo (1996: 18) designates them as ‘epistemic particles’ but does not list them into his table of TMA markers, because he does not consider them sufficiently grammaticalised. Focussing on TP, Mühlhäusler (1985c: 377) describes the form *ating* as a manner adverbial, and Smith (2002) and Smith & Siegel (2013a)

also do not explicitly define *ating* as a modality marker. However, the three forms are used in sentences in which speakers speculate on possibilities. They are treated as epistemic markers in this work because they are on the path to grammaticalising just like other features discussed in previous chapters.

Although there is more that can be said and explored regarding modality in the three MPE varieties, the present study is restricted to the phenomena outlined above and will not focus on further notions of modality. A summary and comparison of the modality markers that are discussed in this section can be found in Table 11 below.

	SIP	BIS	TP
volition	(no) <i>laek/laekem/wande/wandem (fo)</i>	(no) <i>wantem</i>	(no) <i>laik</i>
abilitative	<i>save; fitim fo; inaf fo</i>	<i>save</i>	<i>inap; (save)</i>
inabilitative	<i>no save, kanduit, kan, kanot</i>	<i>no save, kanduit</i>	<i>no inap</i>
prohibition	<i>no save</i>	<i>no save</i>	<i>no ken</i>
permission	<i>save</i>	<i>save</i>	<i>ken (inap)</i>
dubitative	<i>maet, mebi, ating</i>	<i>maet, ating</i>	<i>ating</i>

Table 11: Tabular comparison of selected modality markers

8.3 Previous diachronic studies on modality in MPE

As Winford (2018: 202) points out, “mood and modality remain the most neglected aspects of the study of creole TMA systems”. Since they are neglected from a present-day perspective, there are also almost no studies available that focus on modality from a diachronic perspective.

Despite TP representing the most extensively studied variety of MPE, studies that focus on the diachronic development of mood and modality are rare. As pointed out above, Bickerton (1981) proposed that creoles have a single mood marker only, namely the irrealis mood. This may be one reason why most studies have so far focused on the marking of realis versus irrealis. Romaine (1995), for instance, investigated the grammaticalisation path of the future marker *baimbai* in Tok Pisin by comparing historical sources with language material collected prior to her publication in 1995. In addition, she analysed the grammaticalisation of *laik* ‘want’ and *klostu* ‘near’ into proximate markers (Romaine 1999). Studying the external and internal history of TP, Mühlhäusler (1985a, 1985b) refers to developments regarding tense and aspect but does not focus on modality markers. He states that “[t]he need to express possibilities, contingencies and similar ideas is met by the emergence of sentence qualifiers” (1985b: 94) but does not refer to the forms discussed here. In one of the example sentences, Mühlhäusler (1985b: 116) shows how volition was encoded, but he does not discuss volition in the context of mood and modality.

As far as Bislama is concerned, Crowley's data demonstrated that by the end of World War I the pre-verbal auxiliaries *kan* 'inabilitative', *save* 'abilitative', *wan* and *laek* 'volition' were in use (cf. 1990a: 189). He does not find any semantically positive form of *kan* but assumes it to have existed from the 1880s onwards.

Comparative studies focusing on the diachronic development of modality are rare as well. Keesing (1991: 325) states that English derived lexical items, such as *by and by*, *just*, *might*, and *now* grammaticalised into aspect and modality markers, based on the models of the Oceanic substrate languages by the 1880s. The assumption that substrate systems led to the grammaticalisation of modality markers is also supported by Siegel (2008) who focusses on the comparison of abilitative markers in SIP, TP and BIS. Emphasising the role of substrate influence in contemporary MPE varieties, he provides only two historical sentences in which the form *save* was used. Based on the present day observations that *inap* is used as an auxiliary in contemporary TP and as a subordinate conjunction in contemporary BIS, Siegel assumes that a form of *inap* was already in use around the 1880s in the various regions in the Pacific, although he does not have historical data to prove this assumption (Siegel 2008: 191). He disregards the possibility that forms may have entered at a later stage into the varieties due to renewed contact with the other MPE varieties or renewed contact with the lexifier language.

Baker (1993: 25) lists the marker *might* 'perhaps' in his list of earliest attestation, assuming that the marker spread from QLD, where it was attested in 1906, to Vanuatu and the Solomon Islands, for which he finds the earliest attestations in 1914 and 1937.

None of the studies has focused extensively on the diachronic development of mood markers to encode various types of modality in the three varieties or on when the varieties diverged. Therefore, there is a need for further research on the development of modality markers in MPE across time. The following sections will trace the development of *volitive*, *abilitative*, *permissive* and *speculative* markers in SIP, BIS and TP.

8.4 Volition

The present section targets volition markers attested in the early MPE varieties. It starts with some general remarks on the methodology before the findings will be presented. As in the previous chapters, volitive markers will first be analysed individually in each variety before the results are comparatively discussed.

8.4.1 Methodological considerations

As Table 12 shows, eight different forms (ST_FORM) were identified during the morpheme-by-morpheme analysis to have been in use in MPE to encode volition.

Factor	Levels
ST_FORM	like like him like to want want for want him want to wish to
FEATURE	mod_volit mod_negvolit
M_MORPH	want like wish
YEAR_DET	1832 ... 1950

Table 12: Linguistic coding for volition

Although only three main morphemes provide the basis for these forms, i.e. *like*, *want* and *wish*, several combinations were evidenced due to the possibility of the root form occurring with or without the transitive marker *him* and due to the possibility of occurring with or without the infinitive markers *to* and *for*.

The marker *like* indicating volition appears in three different manifestations in the early data. It is attested in its bare form as in sentence (102), it may occur with the transitive marker *him* as exemplified in (103), or it may occur together with the infinitive marker *to* as known from StE and shown in (104). The commonality between these various markers is that they appear in preverbal position.

- (102) *Master, mi tufelo laik help-im yu*
 master 1DU.EXCL **VOL** help-TR 2SG
 ‘Master, we want to help you’

(New Guinea ~1935; Wolf 1935)

- (103) *Me laik-im see Duk-Duk dance*
 1SG **VOL-TR** see Duk-Duk dance
 ‘I want to see the Duk-Duk dance.’

(New Guinea 1942-48; Barrett 1954: 51)

- (104) *Me like to go along OnePusu*
 1SG **VOL** **INF** go PREP One Pusu
 ‘I want to go to One Pusu.’

(Solomon Islands 1934; Deck June 1934; PMB 1150)

The verb *want* also occurs as a preverbal marker in the early data. It is attested in its bare form as in sentence (105), may appear with a transitive marker as in (106), or it may occur as in StE with

the infinitive marker *to* (cf. sentence (107)). As exemplified in sentence (108), a further attested form is *want for*. We see that *for* seems to fulfil a similar function as the infinitive marker *to* in the verb phrase. A further variant identified during the morpheme-by-morpheme analysis represents *wish to*, which only occurred in very acrolectal PE clauses (see Section 8.4.2.3.1).

- (105) *Me no want kaikai you!*
 1SG NEG VOL eat 2SG
 ‘I do not want to eat you.’

(Malakula 1929, Cheesman 1933: 173)

- (106) *JAPANI WANT-IM SIMAS-IM EVERI AILAN PASIFIKI*
 Japanese VOL-TR smash-TR every island Pacific
 ‘The Japanese want to destroy every Pacific island.’

(Solomon Islands 1943, Tedder 16.11.1943; AU ANUA 445-140)

- (107) *Man, he want to take ship!*
 man PM VOL INF take ship
 ‘The man wants to take the ship.’

(New Guinea 1878, Wawn 1893: 297)

- (108) *He true me want for kill-em along knife*
 PM true 1SG VOL INF kill-TR PREP knife
 ‘(It is) true, I wanted to kill (him) with a knife’¹³⁸

(Lingatu, Santa Isabel; Takwafala 22.03.1938)

The morpheme-by-morpheme analysis further revealed that the form of the auxiliary verb stays the same in positive and negative volition utterances. Volitive utterances are negated by a prefixed negative marker. The negative marker was not included into the field ST_FORM but instead, the data was coded as *mod_volit* and *mod_negvolit* (FEATURE) to differentiate between volition in non-negated and negated contexts.

Since the analysis of eight different variants proved to be difficult due to the small amount of data available, a further column called M_MORPH was added in which it was solely differentiated between three variants, based on whether the marker derives from the Standard English verbs *wish*, *like* or *want*.

Semantically ambiguous structures presented a problem. There were several tokens in which it could not be clearly decided whether the form *like* was used as a volition marker or whether it functioned as an aspect marker indicating proximity. For instance, in sentence (102) introduced above, *laik* can be interpreted with the help of the contextual information as a clear case of volition. Similarly, in example (109) the form *laik* is a marker of proximity. However,

¹³⁸ *-em* is interpreted as a transitive marker in the sentence, as the remaining sentences by the speaker show that *em* had established as a transitive marker (e.g. *me killem you fellow*). Even if there is no overt object in the sentence, the preceding sentence indicates the object:

He true me chase-m boss along knife.
 PM true 1SG chase-TR boss PREP knife
 ‘(It is) true I was chasing the boss with the knife.’

Either there is no need to refer to the object (= the boss) again, or the object got lost through author modification.

sentence (110) represents an ambiguous case in which both readings are possible. The sentence is extracted out of a war pamphlet. Both a volitive and aspectual reading are possible considering the Americans *wanted to* chase away the Japanese, but they also claimed that they *were about to* throw bombs on them soon. The context does not provide enough information for a clear classification. Due to the grammaticalisation of *like* from a full verb into an auxiliary of volition and further into an aspect marker of proximity (see Romaine 1999), the form can fulfil several functions in TP so that sufficient contextual information is required for an unambiguous assignment. Such ambiguous cases were excluded from the analysis.

- (109) *Mastah, me laik die.*
 master 1SG ASP die
 ‘Master I am about to die.’

(New Guinea 1927; Matches 1930: 65)

- (110) *Nau mipela i laik raus-im Japan long olgeta ples [...]*
 now 1PL PM VOL/ASP throw.out-TR Japanese PREP all ples [...]
 ‘Now we want to/are about to throw out the Japanese from all places’

(New Guinea War pamphlet 1943-1945; Kerr 1985; NLA MS9002 PM97)

In addition, it had to be ensured that only those sentences were included in which the markers express internal willingness and desire to do something. Sentences in which the verbs *want* and *like* appeared as main verbs or were used to request someone else to do something, as in *You ouandème mi ouach héd bilong you?* (Pionnier 1913: 196), were not classified as cases of volition. Though there is a wish or desire expressed, the target of the modal verb is not the subject but the object of the sentence.

8.4.2 Findings and discussion

In this section the results of the analysis of volition markers will be presented. The section starts by providing an overview of the general data distribution before taking a closer look at the attestations of volition markers and their development across time, starting with SIP in Section 8.4.2.2, continuing with BIS in Section 8.4.2.3 and focussing on TP in Section 8.4.2.4. The diachronic comparative analysis of volition markers in Section 8.4.2.5 will bring together the results of the individual analyses.

8.4.2.1 General data distribution

In total, 539 tokens expressing volition are found in the early data. Half of them occur in SIP (= 269/540), 33.14% in TP (= 178/540) and 17.04% in BIS (= 92/540). In all three varieties only

around one fifth of the tokens represent instances of negated volition.¹³⁹ The datapoints are distributed unequally across time, as Figure 65 indicates.

As with the previously analysed features, the TP and SIP datapoints predominantly date after 1900. The scarcity of volition markers in the 19th century is compatible with the general assumption that modal markers only develop when a contact variety stabilises. At the same time the small number of volition tokens might be based in what has survived of the written record. The BIS dataset, by contrast, contains a higher amount of pre-1900 attestations. What is important to note is that there is a data gap in the BIS data from 1926 onwards in which no volition cases are found. The data gap does not indicate that volition markers were not used in this timeslot. As pointed out in Section 4.2, only a few sources could be collected for BIS that covered the later time period under investigation which explains the data gap.

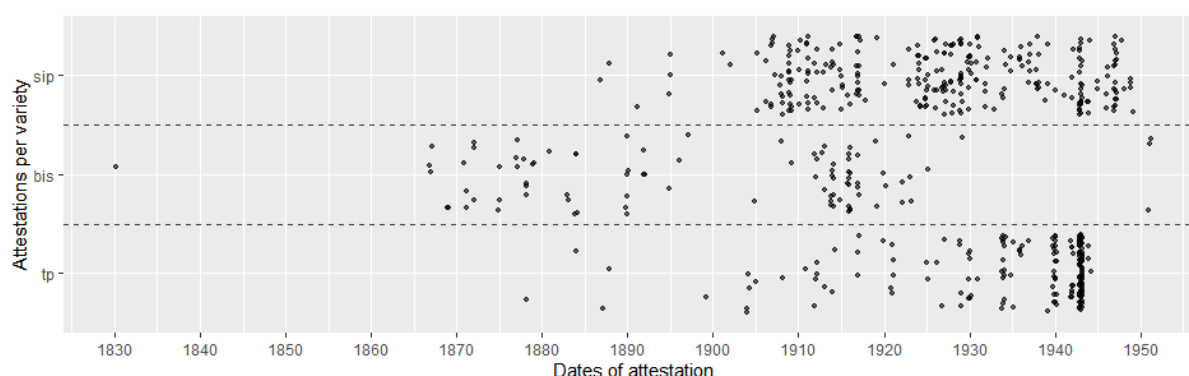


Figure 65: Distribution of volition datapoints across time per variety

If the period under investigation is treated as a single period and the focus is placed on whether volition markers derived from the StE verbs *wish*, *want* or *like*, the early data indicates that each of the three varieties shows a preference for a specific form (see Figure 66). While in TP a form based on *like* seems to be the preferred option, both SIP and BIS show a significantly higher use of a form based on *want*.¹⁴⁰

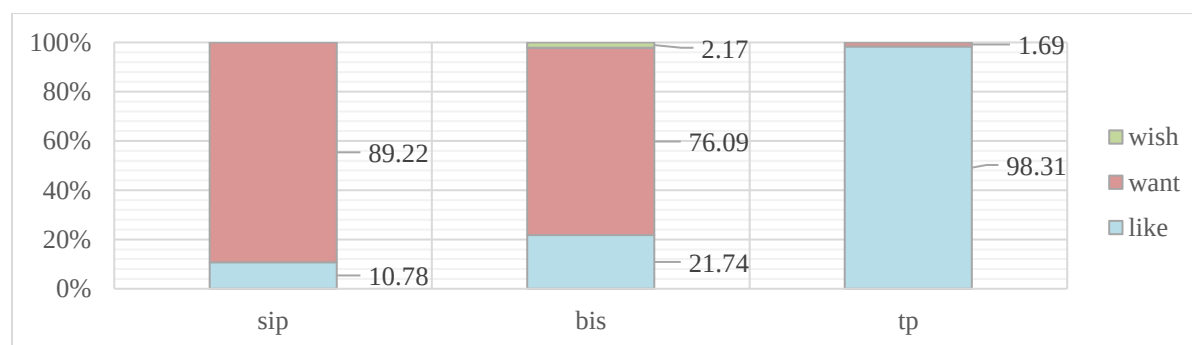


Figure 66: Relative frequencies of volition *M_MORPH* variants per variety (want~like $p < 0.001^{***}$)

¹³⁹ SIP: 48/269 = 17.84%; BIS: 20/92 = 21.74%; TP: 37/178 = 20.78%.

¹⁴⁰ A map showing where the volition markers were attested is displayed in Figure 78.

8.4.2.2 Diachronic analysis of volition in SIP

In this section the concrete forms used to encode volition in early SIP will be investigated from a diachronic perspective. After introducing the forms used, I will focus on when the forms were attested across time and what this might tell us about their development.

8.4.2.2.1 Attested forms

In SIP seven of eight different markers were attested to encode volition (see Table 13). The form which occurs with the highest frequency in the early data is *want to*, as exemplified in sentence (111). The form was found in affirmative (82.78%) as well as negated contexts (17.22%). The dominant marker which is attested as a negation for *want to* is preverbal *no*. However, the form *not want to* was attested twice, while the form *no more want to* was attested once. *No more* has grammaticalised as an alternative marker to negate verbs in early SIP, which is why it may have been occasionally used to encode negated volition semantics as well. The occurrence of *not* is likely to represent an author modification or, even likelier, a typographical error as *not* did not serve as a negative marker in the variety. In 4.83% of all tokens, the volition auxiliary *want* was followed directly by the verb without an additional morpheme, as example (112) shows. The transitive form *want him*, as shown in sentence (106), was only attested seven times.

It appears that the dominant form is the one in which the *to*-infinitive is preserved and that there are only 20 instances in total in which *(no) want(him)* is not followed by an infinitive marker. Most contact languages do not keep the infinitive marker from their lexifier when taking over constructions such as *want to*. Only pidgins and creoles that are very close to the lexifier are observed to preserve the infinitival *to* (cf. Michaelis & Haspelmath 2013c: 386). The question arising out of this observation is whether *want to* was used as shown in the example, or whether *to* represents a result of author or editor modification.

	negated	non-negated	total
<i>want</i>	5	8	13 (4.83%)
<i>want him</i>	0	7	7 (2.60%)
<i>want to</i>	36	173	209 (77.70%)
<i>want for</i>	6	5	11 (4.09%)
<i>like</i>	0	20	20 (7.43%)
<i>like him</i>	1	1	2 (0.74%)
<i>like to</i>	0	7	7 (2.60%)
Total	48	221	269

Table 13: Frequencies of attested volition forms in SIP

There is a reason to suppose that *to* was used in the early years of the variety's development, not only because of the dominant occurrence of the marker. A further indicator may be that not only *to* but also *for* is attested to co-occur with *want*. *Want for* appears five times and *no want for* six

times in the early data. The StE preposition *for* grammaticalised in SIP into a marker sharing inter alia the functions of the StE infinitive marker *to* (cf. Schäfer in progress a).¹⁴¹ Sentence (113) shows that *for* is used equivalently to Standard English *to* in infinitive constructions following the volition verb. Although there are only 11 tokens of *(no) want for*, their occurrence gives rise to the assumption that the form *want to* derived from StE but that *to* was replaced with *for* by some speakers when the latter grammaticalised into a marker fulfilling similar functions as the StE infinitive marker *to*. Moreover, a similarity between the forms *want to* and *wande* can be observed. Thus, there is a possibility that *wande* derived from *want to*.

- (111) *No, because me cross along Funansua me want to help-em you*
 NEG because 1SG angry PREP Funansua 1SG VOL INF help-TR 2SG
 ‘No, I wanted to help you, because I am angry with Funansua.’
 (Tulagi; Viti, B. 19.09.1927)

- (112) *Me want kill ‘em Mary*
 1SG VOL kill TR woman
 ‘I wanted to kill the woman.’
 (Tulagi; Okea, H. 29.12.1911)

- (113) *He true me want for kill-em along knife*
 PM true 1SG VOL INF kill-TR PREP knife
 ‘It is true, I wanted to kill him with a knife.’
 (Lingatu, Santa Isabel; Takwafala 22.03.1938)

The second most dominant form attested in the early data is *like*, which occurs in 7.43% of all tokens but was only found in affirmative contexts:

- (114) *You like look-im along shippie?*
 2SG VOL look-TR PREP ship
 ‘Do you want to take a look at the ship?’
 (Rennell Island 1933-40; Lambert 1942: 337)

- (115) *Me like ‘im fight and kill ‘im man*
 1SG VOL TR fight and kill TR man
 ‘I want to fight and kill the man.’
 (Fiu Bay 1925; Deck October 1925: 8; PMB 1253)

There is a single attestation of both *like him* and *no like him* which may indicate that an optional transitive marker could be used with the verb *like* in SIP (see sentence (115)). The infinitive marker *to* co-occurred with *like* in 2.60% of the tokens, as shown in example (104).

In summary, volition markers in early SIP have derived from the StE verbs *want* and *like*, whereby *want* is the dominant form attested. Both morphemes were attested with and without the transitive marker and with and without an infinite marker. *To* was attested with both morphemes, whereas *for* only occurred with *want*.

¹⁴¹ It should be noted, however, that an infinitive verb form does not exist in the same way as it does in StE. In general, the idea of infinitive verb forms is fairly Indo-European centric.

8.4.2.2.2 Timeline and boxplot approach

The timeline in Figure 67 shows the forms (ST_FORM) across time (by YEAR_ATT). The earliest attested form is *want to* which occurs for the first time in 1886. The form exists throughout the complete period under observation. The form *want* occurs for the first time in 1895 in the data. Although there are only datapoints available for some of the covered years, the form is attested in the 1940s as well.

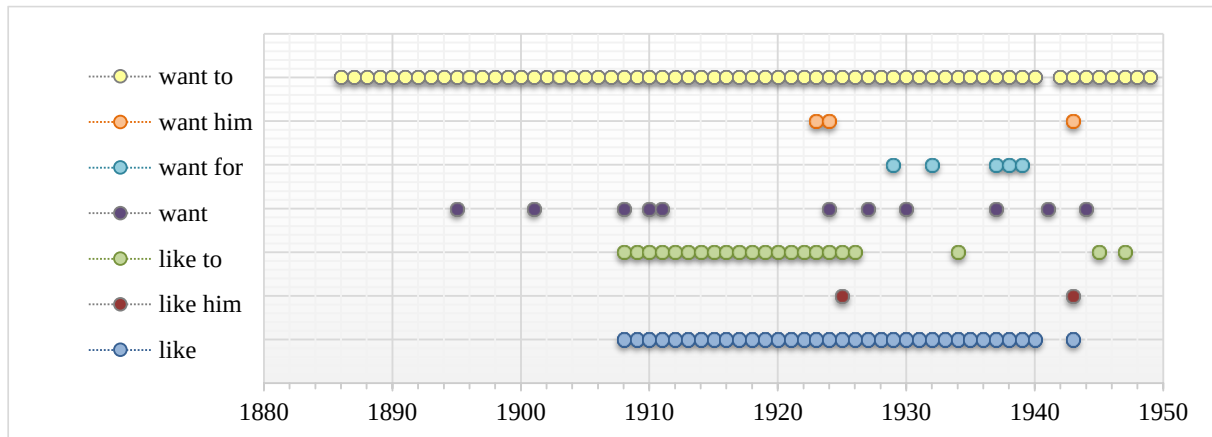


Figure 67: Timeline of volition variants in SIP

The forms *want him*, first attested in 1923 and *want for*, first attested in 1929, seem to have been in use at a later stage in the development of the varieties. For the two forms to develop it was necessary that *him* had grammaticalised into a transitive marker and that *for* had grammaticalised into a preposition fulfilling similar functions as StE *to*, which may explain their late occurrence. Forms built with the morpheme *like* are attested in the written data collection from 1908 onwards. The first attestations of *like to* and *like* date to the year 1908. Both forms spread over the complete period and seem to represent variants of the forms build with *want*. The two attestations of the form *like him* do not occur before 1925.

A similar picture is obtained when constructing a boxplot based on concrete years of attestations which were determined for each of the datapoints (YEAR_DET), as shown in Figure 68. Although volition markers built with *like* are attested later than volition markers based on *want*, both *want*- and *like*-based forms are widespread in the post-1900 data. The boxes and whiskers spread into the 1940s, which might mean that the forms have been used interchangeably. A change in the written data may be observed shortly before 1925. This is the time in which forms built with the transitive marker start to occur and the form *want for* is attested.

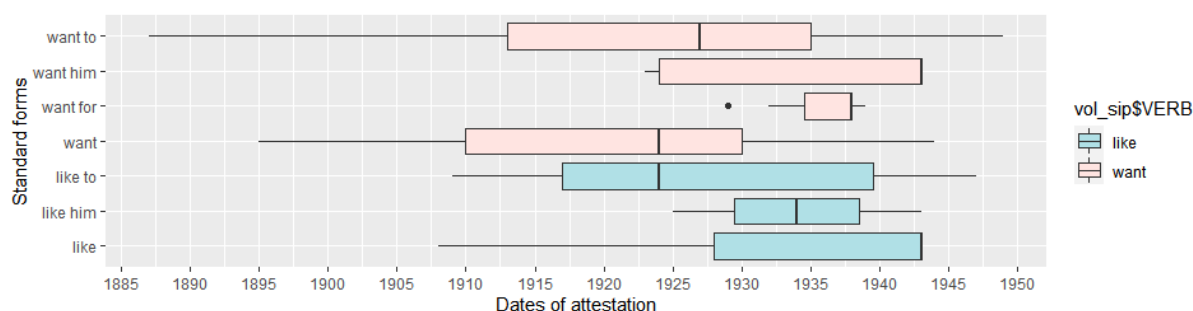
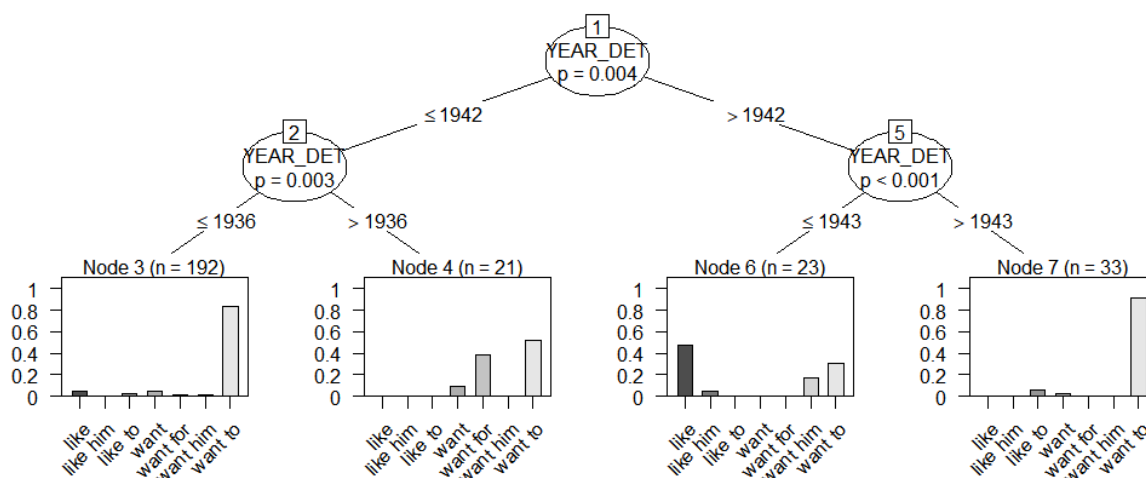


Figure 68: Boxplot of volition variants in SIP

8.4.2.2.3 Testing for the impact of the years of attestations on the choice of form

The function `ctrees` in R was used to observe whether time can be regarded as a factor influencing the choice of the volition marker. If only the year of attestation (= `YEAR_DET`) and the formal realisation (`ST_FORM`) is taken into consideration, the resulting tree shows three time-splits on two levels (see Figure 69). The highest-level split is very significant with $p=0.004^{**}$ and splits the data between before 1942 and after 1942. The second-level splits indicate the years 1936 (Node 2; $p=0.003^{**}$) and 1943 (Node 5; $p<0.001^{***}$) as further dates where differences in the choice of form can be observed. Data attested before 1936 shows a higher than 80% likelihood to be of the form *want to*. In the data dating from 1937 to 1942 *want to* still represents the most likely form but the variant *want for* shows a 40% chance to occur. *Want*, by comparison, is only attested in 10% of the data between 1937 and 1942. Based on Nodes 1 and 5, the data attested in the year 1943 is said to behave differently from the rest of the data, showing a high likelihood of the form *like*. The fact that almost all occurrences of *like* can be attributed to the year 1943 may indicate that the particularities have their origin in an individual text. The fourth period created by the splits, which clusters the datapoints from 1943 to 1950 into one period, is most likely to show the form *want to*.

Figure 69: Conditional inference tree for volition in SIP `ST_FORM~YEAR_DET`

In a second step, it was analysed if the choice of the form is dependent on whether the volition marker is negated or not (FEATURE). Figure 70 shows the resulting tree which consists of four splits on three levels. The first level split shows that negated volition tokens seem to differ from non-negated volition tokens ($p=0.002^{**}$). As can be learned from Node 2 ($p=0.031^*$), negated volition verbs that were attested before and in 1930 were most likely of the form *want to* and less likely of the forms *want* or *want for*. In negated volition utterances that date after 1930 *want to* and *want for* have almost the same frequencies and *like him* is attested as a less likely variant. It is also interesting to take a closer look at the further time-based splits created by Node 5 ($p=0.049^*$) and Node 7 ($p<0.001^{***}$). In non-negated volition utterances, the years 1942 and 1943 were identified as split-evoking years. Tokens attested before 1943 and those dating after 1943 (cf. Node 6 vs. Node 9) behave similarly in that *want to* represents the preferred option. Again, it is the year 1943 in which the data behaves differently with *like* representing the preferred volition marker. Thus, it is very likely that the occurrence of the form may be dependent on a specific text or specific author.

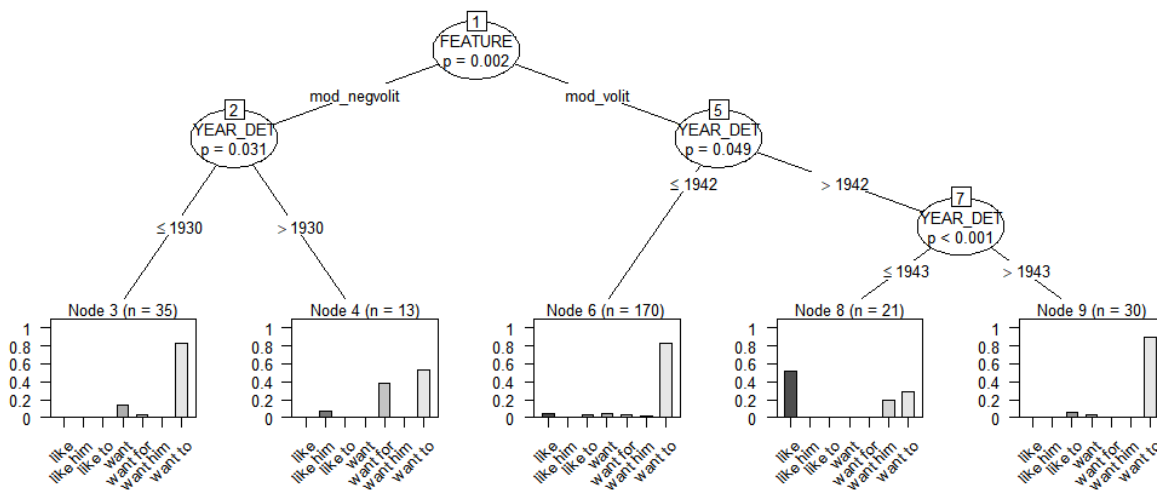


Figure 70: Conditional inference tree for volition in SIP ST_FORM~YEAR_DET+FEATURE

Adding the factor text type into the analysis, the resulting tree shows that from three significant splits only a single split is based on the year of attestation (see Figure 71). The first level-split, which is highly significant ($p<0.001^{***}$), is based on the text type, indicating a different behaviour of volition forms in speech-related versus written/intermediate attestations. The second-level split indicates that in speech-related attestations differences in the choice of form are dependent on whether the volition marker is negated or not ($p=0.005^{**}$). The third split is based on the time variable ($p=0.023^*$) and indicates that negated volition markers in speech-related attestations attested before 1931 are most likely to be of the form *want to*, whereas negated speech-related attestations after 1930 encode volition most likely with the form *want to* or with the form *want for*. The occurrence of *want to* is only slightly more probable. In non-negated

speech-like attestations a clear time-independent preference of the form *want to* can be observed. Similarly, *want to* represents the form that is most likely to be attested in written and intermediate attestations, although the chance for the use of the *like* marker is higher than it is in speech-related attestations (cf. Node 6 and 7).

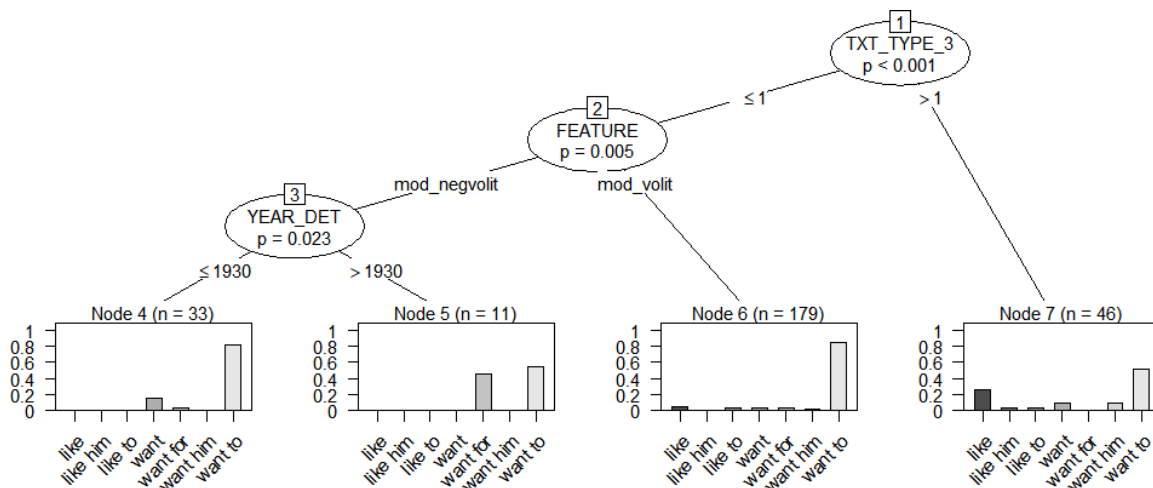


Figure 71: Conditional inference tree for volition in SIP $ST\_FORM \sim YEAR\_DET + FEATURE + TEXT\_TYPE\_3$

When the author (AUTH_NAME) is added as a further possible predictor variable, the resulting ctree consists of four splits on three levels, whereby three of the splits are based on the author and one split is based on whether the volition marker is negated or not (FEATURE).¹⁴² Especially Node 7 is of importance as the split separates the author Herbert Ian Hogbin from all other authors and shows that the high use of *like* in the year 1943 can be traced back, as suggested, to a single author. If the author and text type are considered, the year of attestation no longer acts as a significant predictor variable for the choice of volition forms.

8.4.2.3 Diachronic analysis of volition in BIS

The focus of this section is on volition markers attested in early BIS and how they develop across time. First, the attested forms will be introduced before the timeline and boxplot will be focused on. In a last step, the results of the analysis with ctree will be displayed.

8.4.2.3.1 Attested forms

In the early data of BIS, seven of the eight introduced forms were attested (cf. Table 14). Similar as in SIP, *want to* is attested with the highest frequencies and appears in both negated and non-negated contexts (see, for instance, sentence (116)). As exemplified in the BIS originating

¹⁴² The tree is too large to displayed here but can be found at https://www.dropbox.com/s/u2a98r9im1llqis/Ctree_Volition_SIP.png?dl=0 (last access 29 September 2021).

sentence (105), the morpheme *want* is attested without the infinitive marker in 23.91% of the datapoints.

	negated	non-negated	total
<i>want</i>	2	20	22 (23.91%)
<i>want him</i>	1	3	4 (4.35%)
<i>want to</i>	13	31	44 (47.83%)
<i>like</i>	3	11	14 (15.22%)
<i>like him</i>	0	1	1 (1.09%)
<i>like to</i>	1	4	5 (5.43%)
<i>wish to</i>	2	0	2 (2.17%)
Total	22	70	92

Table 14: Frequencies of attested volition forms in BIS

The early data contains only four instances of the transitive verb form *want him*, as exemplified in (117). Although the morpheme *row* was treated as a verb in the example, it is also possible to have functioned as a noun. The dual functionality may indicate how and why *transitive* volition markers have developed over time. It is likely that lexical morphemes such as *row*, following the volition verb *want*, were interpreted as nouns by some speakers and as verbs by others. If considered a noun, the verb requires a transitive marker. The construction *volition verb* + *transitive marker* + *noun* may have been reinterpreted as *volition verb* + *transitive marker* + *verb*.

- (116) *You want to steal me feller, Captain?*
 2SG VOL INF steal 1PL captain
 ‘Do you want to steal us, captain?’
 (Port Vila, New Hebrides; Wartevioch 13.11.1912)
- (117) *We fellow no want him row now.*
 1PL.EXCL NEG VOL TR row now
 ‘We do not want to argue now.’
 (Perturtsis 1914; Clarker 02.03.1914)

Forms based on the morpheme *like* are less frequent than forms based on the morpheme *want*. In its bare form, as exemplified in sentence (118), *like* was attested eleven times in non-negated and three times in negated contexts, making up 15.22% of the datapoints. Five times *like to*, as demonstrated in sentence (119), was attested. The form *like him* was only attested once, the sentence being displayed in (120). Similar as with example sentence (117), the morpheme following the volition marker was treated as a verb, based on the contextual information of the utterance. There is, however, the possibility that *smoke* was reinterpreted as a noun by the speaker or author.

- (118) *Boy he like go.*
 boy PM VOL go
 ‘The boy wants to go.’
 (Tanna Island 1875; Wawn 1893: 15)

- (119) *Suppose some man like to come close up along missionary [...]*
 if some man **VOL** **INF** come PREP missionary [...]
 ‘If some men want to come close to the missionary ...’
 (Orlip 1914; Clarker 02.03.1914)

- (120) *me too much like-em-smoke.*
 1SG really **VOL-TR**-smoke/tobacco
 ‘I really want to smoke.’
 (Tanna 1877; Giles 1968: 37)

BIS is the only variety which makes use of *wish to*, as shown in (121). It occurs only twice in a letter written by the Vanuatuan Supabo. When taking a closer look at the sentences in which the form occurs, it becomes clear that the writer of the letter uses either a very acrolectal form of Bislama or that the writer uses English and we can just observe a few Pidgin English elements. Standard English elements are, for instance, the 1SG pronoun, the 3SG possessive pronoun, plural -s, the prepositions *in* and *about*, and lexical items such as *wife*.

- (121) *I wish to tell you about some trouble-s in here in*
 1SG **VOL** **INF** tell 2SG about some trouble-PL PREP here PREP
last year one man he kill his wife
 last year ART man PM beat 3SG.POSS wife
 ‘I wish to tell you about some troubles which happened here; last year a man killed his wife’
 (Epi Island 1923; Supabo 04.06.1923)

8.4.2.3.2 Timeline and boxplot approach

To add diachronic information to the forms, the timeline in Figure 72 displays the attestations of volition markers in BIS across time. The earliest form attested in the written data is *like* in 1830. No further volition markers are attested until 1866. In the year 1867 the next attestation of *like* is found and the marker *like to* is attested for the first time as well. Although the latest attestation of *like to* dates to the early 1910s and the latest attestations of *like* dates to 1923, one can barely draw inferences about the occurrences of the forms from the data. As pointed out earlier, the data covering the years 1930 until 1950 was scarce so that the absence of the forms does not necessarily mean that the forms were no longer present.

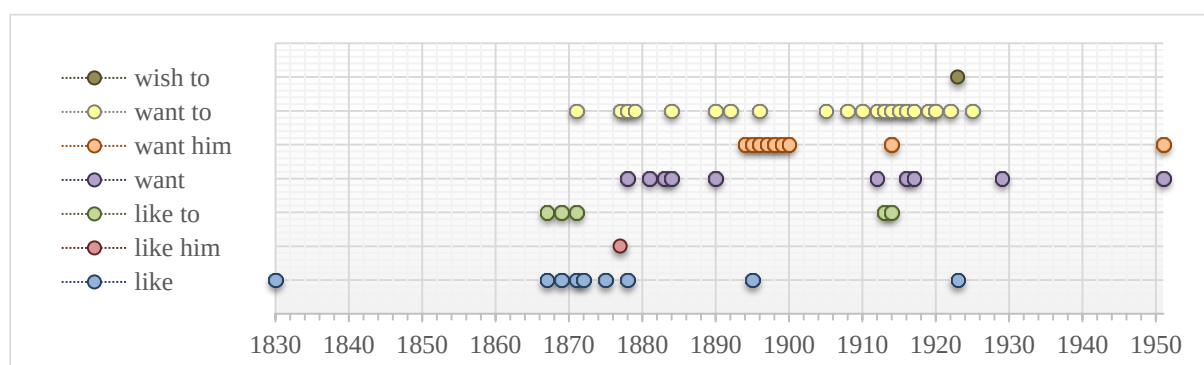


Figure 72: Timeline of volition variants in BIS

The first occurrence of *want to* dates from the year 1871 and the form is attested in the written data until 1925. The forms *want* and *want him* occur in the data in 1878 and 1895 for the first time. Both forms were also attested in 1951. As frequencies are not considered in the timeline and due to the data gap in the 1930s and 1940s, the timeline is not very conclusive concerning the choice of volition markers.

Although the boxplot displayed in Figure 73, in which periods of attestation are replaced by determined years of attestation (= YEAR_DET), offers a similar visualisation, it has more explanatory power. For instance, it becomes visible that the medians of the forms based on the verb *like* date before 1877, whereas the medians of the forms based on the verb *want* date after 1910. This indicates that forms built with *want* may have turned into the preferred option. The data gap, however, remains a problem in interpreting the results.

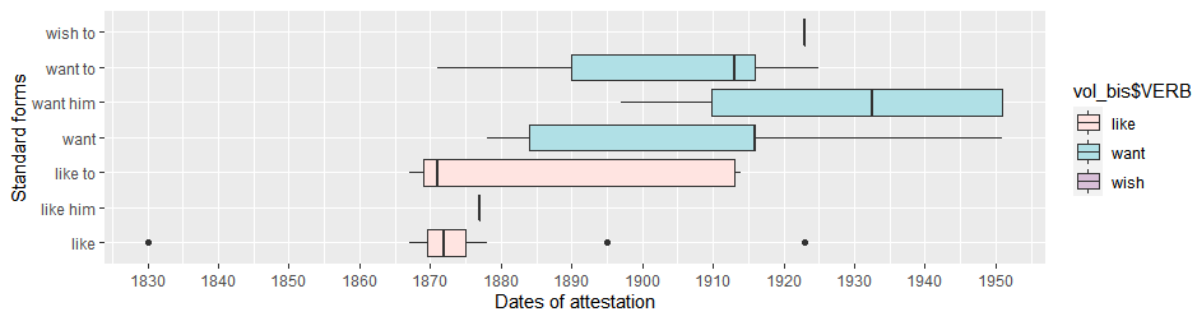


Figure 73: Boxplot of volition variants in BIS

8.4.2.3.3 Testing for the impact of the years of attestations on the choice of form

In order to test whether forms built with *want* turn into the preferred form in the written attestations after 1876, the ctree algorithm was applied since it takes frequencies of occurrence into consideration and is based on statistical rigor. Independent of whether only the year of attestation (YEAR_DET) was considered as a possible predictor variable or whether it was combined with the variable FEATURE (= measuring the impact of negation vs. non-negation) and the text type variable, all three analyses resulted in the same tree structure which is displayed in Figure 74.

The resulting trees show a single split based on the year of attestation which is highly significant in all three cases ($p < 0.001^{***}$). Independent of whether the volition verb is negated or not and independent of the text type, volition verbs attested before 1876 are most likely to be encoded with *like* (with *like to* and *want to* representing less common alternative forms), whereas volition verbs attested after 1875 are most likely to be of the form *want to* or *want*.

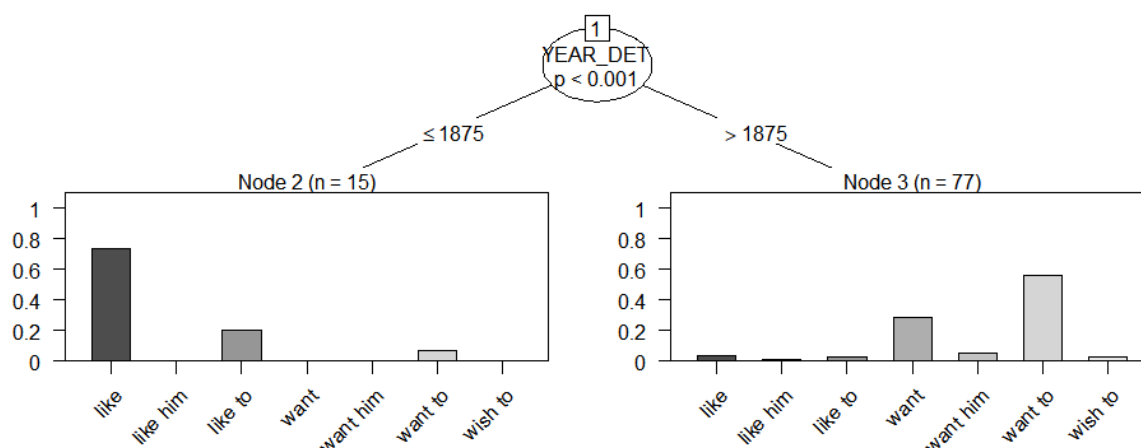


Figure 74: Conditional inference tree for volition in BIS ST_FORM~YEAR_DET+FEATURE+TXT_TYPE_3

If the author (AUTH_NAME) is added to the analysis, the resulting tree contains three splits, whereby the first and third split are based on the author and the second one on the year of attestation.¹⁴³ Although four authors are said to use volition markers differently, the time split in the year 1875 prevails an indicative factor for the remaining 22 authors in the choice of form ($p < 0.001^{***}$). Even though the results show that there is a significant year in regard to the preferred form, it should be kept in mind that the analysis is based on *written data* and that the available data does not cover the years 1930 to 1950 sufficiently.

8.4.2.4 Diachronic analysis of volition in TP

The present chapter focusses on volition markers attested in early TP. First, a look at the different attested variants is taken before their occurrence across time is visualised and further analysed with the help of the ctree algorithm.

8.4.2.4.1 Attested forms

In TP only four of the ST_FORMS were attested:

	negated	non-negated	total
want to	0	3	3 (1.69%)
like	136	37	173 (97.19%)
like him	0	1	1 (0.56%)
like to	0	1	1 (0.56%)
Total	136	42	178

Table 15: Frequencies of attested volition forms in TP

The most dominant attested marker represents *like*, which in its bare form occurs in 97.19% of the 178 attested volition tokens (see (102) for an example sentence). Sentences (103) and (122)

¹⁴³ See https://www.dropbox.com/s/yw5e3cypfxsr60a/Ctree_Volition_BIS.png?dl=0 (last access 29 September 2021).

show the single occurrences of the forms *like him* and *like to*. The form *want to*, as exemplified in (107), was only attested three times.

- (122) *Now me like to make work b'long you*
 now 1SG VOL INF make work POSS 2SG
 'Now I want to work for you'

(Finschafen 1914; Detzner 1920: 96)

8.4.2.4.2 Timeline and boxplot approach

As the frequencies indicated, *like* represents the by far dominant form in the TP dataset. Frequencies alone, however, do not provide information about the timeframe in which the forms were attested. Figure 75 shows the timeline which is based on the specific dates, years, or periods that were identified for each attestation (YEAR_ATT).

As Figure 75 shows, the form *want to* represents the earliest attested form. It occurs in 1878 for the first time but is only attested in two further years with the latest attestation dating to the year 1920. The marker *like* first appears in the TP data in 1886 and is attested throughout the complete time period under investigation. The transitive volition marker *like him* represents a late-attested token in the TP dataset; its first occurrence is in 1942. As previously pointed out, the transitive form *like him* presupposes the condition that *him* grammaticalised into the transitive marker, which explains why transitive forms occur later in the timeline.

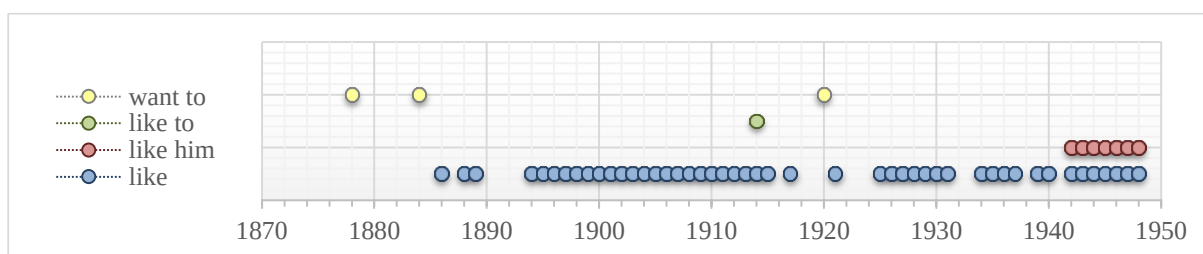


Figure 75: Timeline of volition variants in TP

Similar results are obtained when definite dates had to be determined for each attestation to create a boxplot in R (YEAR_DET), as shown in Figure 76.

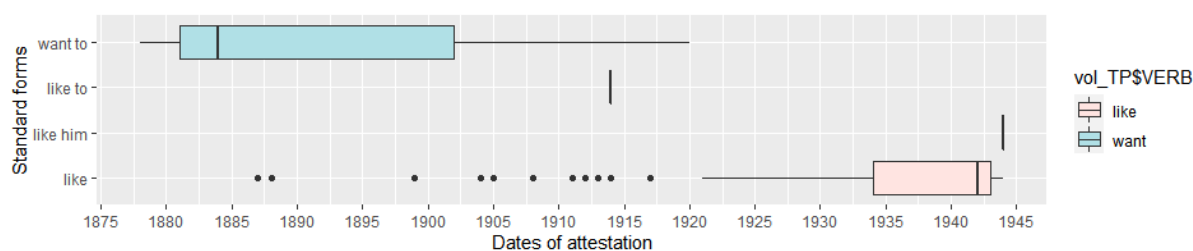


Figure 76: Boxplot of volition variants in TP

The median of *want to* indicates that the form might have been used in initial encounters with Europeans; the median of *like* shows that this is the dominant form at the end of the observed

period. *Like him* only occurs in a single year in the data and is therefore displayed as a vertical line. In Figure 75 it was portrayed as several datapoints since the attestation could not be clearly referred to an exact year of attestation.

8.4.2.4.3 Testing for the impact of the years of attestations on the choice of form

Because the timelines do not take frequencies into consideration and do not provide information regarding statistical significance, the data was further tested in terms of whether time represents a significant factor for the choice of the volition marker. One and the same tree structure was obtained independent of whether only the year of attestation, the year of attestation and the variable FEATURE (= impact of negation vs. non-negation), or the year of attestation, the variable FEATURE and the text type variable were taken into consideration.

As Figure 77 shows, the tree consists of a single split which is based on the year of attestation and is highly significant ($p < 0.001^{***}$). The year which is said to split the attested forms is the year 1920. While data after 1920 shows a clear preference of the form *like*, attestations dating before 1921 show next to the dominant form *like* the use of *like to* and *want to*.

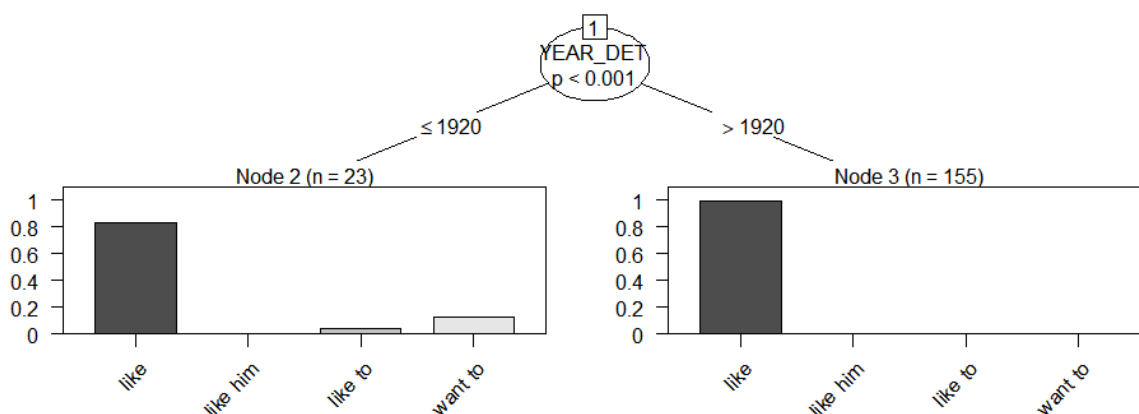


Figure 77: Conditional inference tree for volition in TP ST_FORM~YEAR_DET+FEATURE+TXT_TYPE_3

However, if the author is included as a further predictor variable, the year of attestation no longer evokes a significant split. Instead, the resulting tree consists of a single split which is based on the author variable.¹⁴⁴ It is interesting to note that only four authors make use of the forms *like him*, *like to* and *want to*, whereas the remaining 27 authors show a clear preference of the form *like*. This shows that there is no significant time split in the data. By contrast, based on the data at hand, it seems as if *like* already represented the dominant form at an early stage in the development of the variety.

¹⁴⁴ See https://www.dropbox.com/s/6p5pm7qqa8pomsm/Ctree_Volition_TP.png?dl=0 (last access 29 September 2021).

8.4.2.5 Diachronic comparative analysis of volition in MPE

Comparing the development of volition markers in SIP, BIS and TP across time, it can be observed that based on the written attestations, no generalising year of change can be attested which is valid for all three varieties. A ctree including data of all three varieties and testing the impact of the year of attestation, the variety, the text type, as well as negation on the choice of form, supports these observations. The tree consists of seven splits on five levels.¹⁴⁵ The first two splits (Nodes 1 and 2) are based on the variable VARIETY and are highly significant ($p < 0.001^{***}$). While the first split separates TP from SIP and BIS, Node 2 splits SIP and BIS from each other. According to the tree, SIP and BIS show a greater similarity regarding the choice of the volition marker. Furthermore, it is important to note that SIP and BIS split independent of the year of attestation. Only after they have been separated, further time-based splits can be observed. The remaining splits for each variety are those which were already discussed above in the individual variety sections.

In SIP, the forms *like* and *want* were attested across time and no clear choice of form could be observed. Negated volition markers attested before 1931 were most likely to be of the form *want to* and less likely of the form *want*. Post-1930 negated volition markers were either of the form *want to* or *want for*, with the former showing a higher likelihood. These differences, however, could not be observed if the author was considered as a predictor variable as well. Non-negated volition markers showed a remarkable likelihood of the form *like* in the year 1943, whereas in the remaining periods the form *want to* was preferred. The high occurrence of the form *like* could be referred to a single author and source and thus, represented idiosyncrasies of a single author/source.

In BIS, the analyses revealed that the form *like* was the likeliest form in pre-1875 data, but afterwards forms with *want* dominated. It is the only one of the three varieties in which the time-split prevails if the author is added as a further predictor variable. However, volition markers dating between 1930 and 1950 are only scarcely represented in the data.

In TP, a clear preference of the form *like* can be observed. The occurrence of other markers before and in the year 1920 is only a significant factor if the author is not considered as a predictor variable. It is thus likely to assume that the form *like* had already dominated in volition contexts in TP before the end of the labour trade.

Since SIP and BIS predominantly make use of forms based on the morpheme *want to* express volition and since forms based on *want* scarcely occur in early TP, this gives rise to

¹⁴⁵ The tree is too large to be displayed here but can be viewed online at https://www.dropbox.com/s/tqkccy6n9uhm9ml/Ctree_Volition_MPE.png?dl=0 (last access 29 September 2021).

assumption that the *want*-based forms may have been in use on QLD plantations but not on the Samoan ones. Although there are fewer surviving records of SPPE than of QPPE, I searched for *want*- and *like*-based auxiliary forms in early SPPE and QPPE. While in the early SPPE data only *like* was attested (cf. (123)), QPPE shows auxiliary volition markers based on *want* and *like*, as shown in example (124) and (125).

- (123) *Me no like go place belong me*
 1SG NEG VOL go place POSS 1SG
 ‘I do not want to go to my home village.’

(Samoa 18833; Schuchardt 1889: 159)

- (124) *Oh me want to tell you something*
 oh 1SG VOL INF tell 2SG something
 ‘Oh I want to tell you something.’

(Queensland 1880; Fussell 1903: 56)

- (125) *Suppose me like to go me walk along ground*
 if 1SG VOL INF go 1SG walk PREP ground
 ‘If I want to go, I walk by foot.’

(Queensland 1892; Queensland Department of Justice 1892: 1180)

The available surviving written attestations thus support the assumption that the morpheme *want* was only in use on QLD plantations, whereas the morpheme *like* was in use on both QLD and Samoan plantations. This might be the reason why in early TP a clear preference of the morpheme *like* is observed. A comparison of the forms attested in the early data of SIP and BIS with the forms which are used in the varieties today shows that the grammaticalisation of the volition markers had not been completed in SIP and BIS by the end of 1950. Further data is needed that covers the time after 1950 to understand the complete stabilisation path of volition markers in SIP and BIS. In addition, more Bislama data covering the years 1930 until 1950 is required.

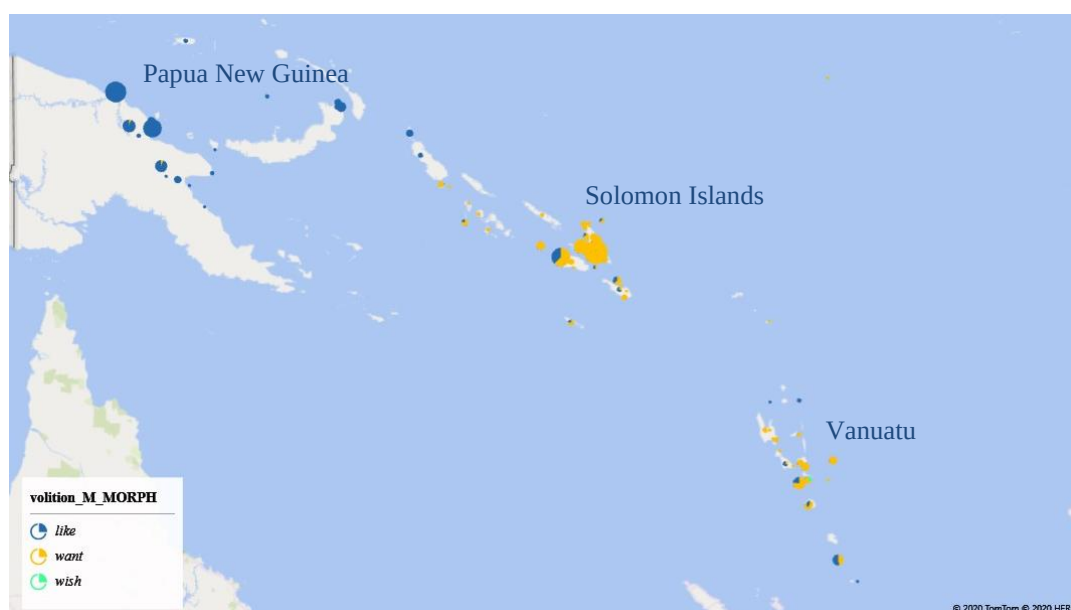


Figure 78: Map of Melanesian Islands, showing attestations of volition *M_MORPH* variants

8.5 Abilitative

In this section I will focus on the development of markers to encode *ability* in SIP, TP and BIS. After providing some remarks on the methodology, the findings of the analysis will be introduced and discussed. The varieties will first be analysed individually before a contrastive approach is undertaken.

8.5.1 Methodological considerations

Ability, as pointed out in Section 8.1, may be subcategorised into *subject-oriented* and *neutral/circumstantial* possibility. Although the different subcategories may lead to the choice of different forms to encode ability, the two types are not differentiated in this section. The reason is that for many of the datapoints a clear differentiation of the functionality was not possible because required contextual information was missing. Mental and physical ability was also not differentiated. Again, although the different types of ability may lead to the choice of different forms, the attempt to work with a very fine-grained classification system had to be abandoned. Since the number of *ability* tokens is quite low and contextual information is limited, a broad classification system had to be applied.

As Table 16 shows, seven different forms (ST_FORM) to encode ability were identified during the morpheme-by-morpheme analysis. In the following, the forms will be introduced with example sentences. The examples were selected randomly, and their selection does not mean that a variant was solely attested in one of the three varieties.

One of the possible modal markers to encode ability represents *able to*, as exemplified in sentence (126). Despite deriving from StE, it differs from its occurrence in the lexifier language in that the copula is omitted.

- (126) *I not able to see what men Manoba do along of Government.*
 1SG NEG ABIL INF see what man.PL Manoba do PREP government
 ‘I could not see what the Manoba men did with the government.’
 (Bundaberg; Enow 24.12.1888 [Australian Station 1888: 20])

Another form which derived from the lexifier language is *can* which is orthographically realised as *can*, *kan* and *ken*, as, for instance, in sentence (127). The circumlocution used to describe a swivel chair is literally translated a chair that can be turned around.¹⁴⁶

¹⁴⁶ During the morpheme-by-morpheme analysis it became evident that many abilitative markers could be extracted from early circumlocutionary dictionary entries. Especially adjectives which in StE contain the morpheme *-ible/-able* and thus are used to encode the meaning ‘capable of X’ are expressed through circumlocutions in the early MPE varieties. For instance, *ol i no ken daunim em* is the circumlocution for

- (127) *sia ol i ken tan-im nambaut*
 chair 3PL PM **ABIL** turn-TR around
 'swivel chair'

(Vunapope ~1940; Kutscher 1940: 37)

The next three forms listed in Table 16 have their origin in the English adjective *enough* and only differ in whether, and if so which, infinitive markers are inserted between the form and the following verb. Sentence (128) shows, for instance, a sentence in which the form *enough* is followed by the morpheme *long*, which in abilitative contexts has a similar function as the English infinitive marker *to*. *Long* was glossed as a separate item because it did not always co-occur with *enough*. Thus, it is assumed to have been optional. *For* was attested as a variant marker which could co-occur with *enough*.

The sixth and seventh form listed in the table are *save* and *save for*, the former being exemplified in sentence (129). *For* in the phrase *save for*+*VERB* seems to fulfill a similar function as the English infinitive marker *to* (cf. Schäfer in progress a).

- (128) *Sapos masta i-no inap long wokabaut yupala mek-im bet nau kar-im*
 if master PM-NEG **ABIL** INF walk 2PL make-TR stretcher and carry-TR
 'If the master cannot walk, you will make a stretcher for him and carry him.'

(New Guinea War pamphlet 1943-1945; Kerr 1985; NLA MS9002 PM13)

- (129) *suppose I go along school I savey write all the same*
 if 1SG go PREP school 1SG **ABIL** write likewise
 'If I go to school, I will also be able to write'

(Baunani 1912; Deck February 1912: 4-5; AU PMB MS 1253)

The factor ST_FORM does not provide information on whether the marker expresses ability or inability. Therefore, the data was additionally coded for FEATURE, in which the variants *mod_abil* and *mod_inabil* were distinguished. To express inability the aforementioned forms were found to be negated by the MPE preverbal negative markers *no* or *no more*, the former being exemplified in (130). However, not all negated forms are constructed with the common MPE negative markers but forms such as *can't* and *cannot*, which are based on StE *not*, were attested as well (cf., for instance, (131)).

- (130) *JAPAN I NO-GAT SIP NAU I NO-KAN KIS-IM NU-PELA*
 Japanese PM NEG-have ship now PM **NEG-ABIL** catch-TR new-MODIF
KAGO BILOG OL
 good POSS 3PL

'The Japanese do not have ships so (they) cannot catch new goods.'

(New Guinea War pamphlet 1943-1945; Kerr 1985; NLA MS9002 PK4)

'invincible'. Therefore, circumlocutions offer a chance to learn more about the ways in which ability is expressed in the early forms of the varieties.

- (131) *Me* *say,* *No!* *ship* ***can't*** *break* [...]
 1SG *say* NEG *ship* **INABIL** *break* [...]
 'I said: "No, the ship cannot break."'

(*The Stories of the Crew*; NIV June 1947: 8; AU PMB DOC 439)

It is possible that the three varieties show differences in how they negate ability. The coding that was used so far under ST_FORM obfuscates such differences. Therefore, the data was further coded for ST_FORM_F(ine-grained) which lists not only the main morpheme but also the negative marker and the infinitive marker if applicable. This allows for an investigation of whether the varieties show differences on a more fine-grained level. In addition, forms may have been reinterpreted and taken over as monomorphemic items which can be better identified the more precise the analysis level is.

A very broad classification factor called M_MORPH was additionally added, which provides information about the main morpheme but does not inform about whether the main morpheme is negated or not or whether a marker is following the main *ability-expressing* morpheme.

FACTOR	VARIANTS
ST_FORM	<i>able to</i> <i>can</i> <i>enough</i> <i>enough for</i> <i>enough long</i> <i>save</i> <i>save for</i>
FEATURE	mod_abil mod_inabil
ST_FORM_F	<i>not able to</i> <i>can</i> <i>no can</i> <i>no more can</i> <i>cannot</i> <i>can't</i> <i>enough</i> <i>no enough</i> <i>enough for</i> <i>enough long</i> <i>no enough long</i> <i>save</i> <i>no save</i> <i>save for</i>
M_MORPH	<i>able</i> <i>can</i> <i>save</i> <i>enough</i>
YEAR_DET	1832 ... 1950

Table 16: Linguistic coding for ability

It is crucial to mention that a clear classification of forms as markers of ability was not always possible. Several of the identified forms have served as multifunctional items in the early data. As such, the classification of *save* proved to be difficult. *Save* does not only occur as an auxiliary indicating ability but it is also attested as a full verb, as a habitual-aspect marker and as a permissive marker. In some cases a differentiation between HAB(itual), ABIL(ity) and PERM(ission) was not possible. This reflects a difficulty which Palmer (1990, 2001) himself identified in his modality model and to which the introductory quote of this section refers. For instance, in a sentence such as (132), it is not clear whether *no savvy* encodes ability in that the child does not have the knowledge to mix a cocktail, or whether it encodes permission, in that children are not allowed to mix cocktails. A similar classification problem can be observed in Standard English which becomes clear when taking a look at *can't* in the English translation. The form can encode the meaning 'are not allowed to' or 'do not have the capacity'. The problem of distinguishing between ability and permission is not restricted to the marker *save* but can also be observed with other markers. In sentence (133), for instance, the verb *cant* may encode inability or prohibition. The restricted availability of contextual information further hampers a proper analysis of some early tokens.

- (132) *Small fella piccaninny no savvy make 'im cocktail!*
 small MODIF child NEG SAVVY make TR cocktail
 'Small children can't make cocktails.'

(Simbo Islands 1920-26; Collinson 1926: 159)

- (133) *Suppose pig-pig he make trouble, me savvey sell him or eat him; but this*
 if pig PM make trouble 1SG POSS sell TR/3SG or eat TR but DEM
piccinini he born long me, me cant sell him, what me do now?
 child PM born PREP 1SG 1SG CANT sell TR/3SG what 1SG do now
 'If a pig causes trouble, I can sell or eat it, but this child of mine, I can't sell, what do I do now?'

(Nongosila 1928; Deck 13.05.1928; PMB 1150)

The marker *can* was occasionally used to encode futurity in the early data. Sometimes a clear differentiation between ability and futurity was not possible. Example (134) represents such an ambiguous case. The sentence was extracted from a war pamphlet, which was written by the Australian Government during World War II.

- (134) *No long time now kiap can come along Mortlock and*
 NEG long time and officer ABIL/FUT come PREP Mortlock and
Green along one ship now look-im you fella [...]
 Green PREP ART ship and see-TR 2PL [...]
 'In a short time, the government officer will (be able to) come to Mortlock and Green by ship and meet you ...'

(Green Islands 1943-1945; Kerr 1985; NLA MS9002 PK18)

The marker *can* may express *circumstantial* possibility. At the same time, the sentence refers to a future event that is not only possible, but that will definitely happen according to the writer of the pamphlet. This future reading is intensified by the words *no long time* at the beginning of the sentence. Examples such as this make clear that ability and futurity are not completely separate areas. As Bybee et al. (1994: 266) have shown, the grammaticalisation path ABILITY > POSSIBILITY > FUTURITY, although not very common, exists in the languages around the world. It is based on the premise that if something is possible, it will occur. In light of these issues, ambiguous tokens were not included into the analysis.

8.5.2 Findings and discussion

In this section, the results of the analysis of ability markers will be presented. The section starts by providing an overview of the general data distribution before taking a closer look at the attestations of ability markers and their development across time. While the varieties will first be considered individually, starting with SIP in Section 8.5.2.2, continuing with BIS in Section 8.5.2.3 and focussing on TP in Section 8.5.2.4, a comparative analysis of ability markers in the three MPE varieties will follow in Section 8.5.2.5.

8.5.2.1 General data distribution

In total, 474 sentences were attested in the collected data material in which an overt marker was used to express ability. Of these, 183 sentences could be ascribed to the Solomon Islands (= 38.61%), 262 to New Guinea (= 55.27%) and only 29 to Vanuatu (= 6.12%). The mere enumeration alone elucidates the problematic data situation. As the amount of datapoints available per variety differs to a high degree, it makes a comparison difficult. Figure 79 shows how the individual collected datapoints spread across time in the three varieties.

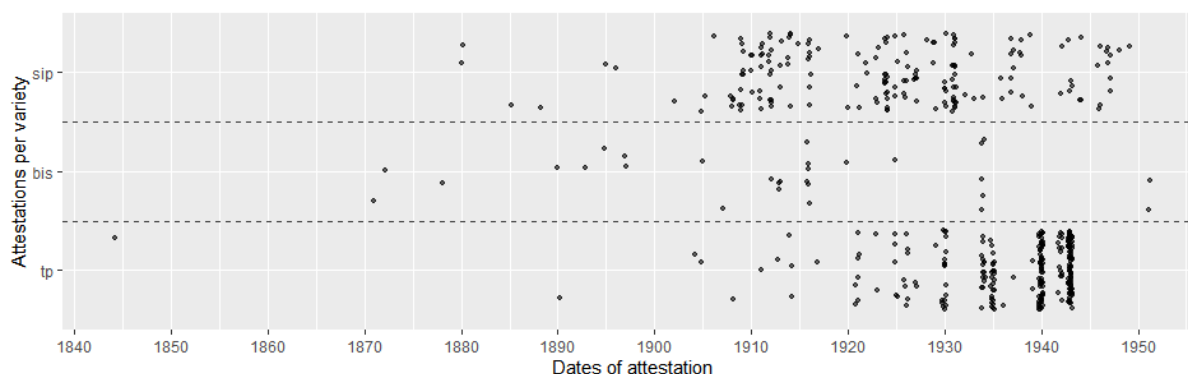


Figure 79: Distribution of ability datapoints across time per variety

What is interesting to note when focusing on the TP and SIP datapoints is that although there are scattered attestations before 1900, the majority of datapoints is dated after 1905. While in SIP the

datapoints spread rather equally, in TP most of the datapoints can be localised around 1930-1945. The small amount of datapoints that are available for BIS are scattered. Again, no attestations are available covering the time from 1935 to 1950 in the BIS dataset.

If the time from 1840 to 1951 is treated as a single period, the three varieties show different preferences regarding the choice of the marker used to encode ability, as Figure 80 illustrates.¹⁴⁷ Focussing solely on the main morphemes (M_MORPH) of the ability markers, the forms built with *can* and *save* represent the dominant attested morphemes. While in SIP and TP most forms seem to be constructed with the morpheme *can*, BIS shows a preference for forms constructed with the morpheme *save*. SIP represents the only variety that used forms constructed with *able*, whereas TP is the only variety in which forms based on the morpheme *enough* were attested. It should be noted that the figure does not include information about how often a variant was used by individual authors.

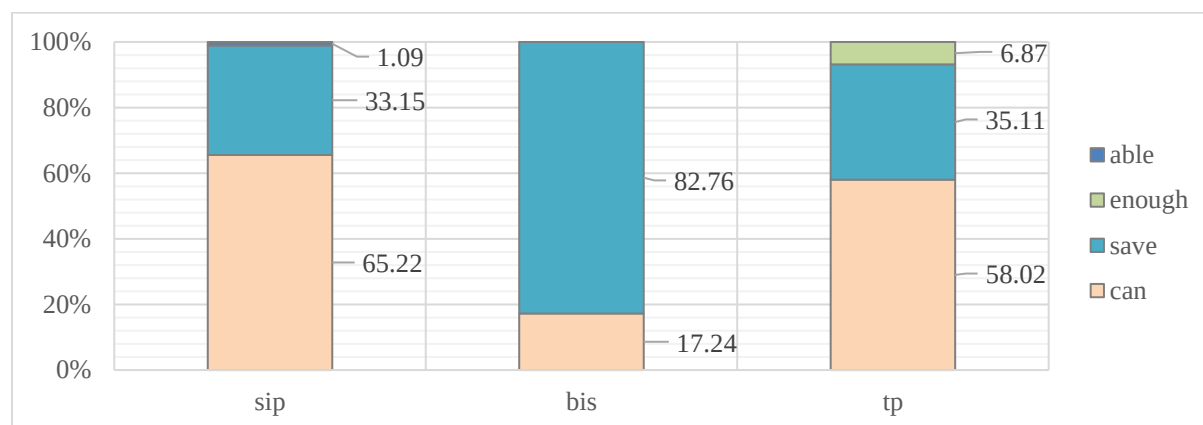


Figure 80: Relative frequencies of abilitative M_MORPH variants per variety (can ~ save: $p < 0.001^{***}$)

By solely focussing on the main morphemes of the forms, and by not including more specific information about, for instance, negated vs. non-negated markers and infinitive markers, information gets lost. If one compares Figure 81, which contains the specific forms used to encode ability in affirmative contexts with Figure 82, which contains the concrete forms used in negated contexts, it is the negated forms that seem to be of special interest. Although both SIP and TP show high frequencies of forms based on the morpheme *can* (see Figure 80), they differ in the concrete forms used to express inability (see Figure 82). In TP almost all negated forms that contain the marker *can* are of the form *no can* while in SIP the form *no can* is not attested. Instead, the form *can't* occurs in 69.34% and *cannot* in 5.15% of the negated tokens. Thus, in the comparative chapter it will be necessary to not only focus on the main morphemes, but to also look at the more fine-grained forms.

¹⁴⁷ A map showing where the ability markers were attested is displayed in Figure 94.

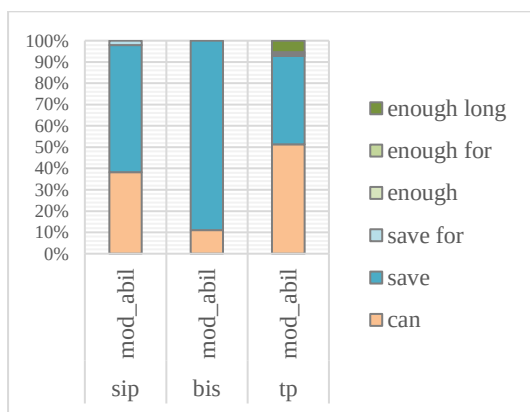


Figure 81: Relative frequencies of abilitative variants in affirmative contexts per variety

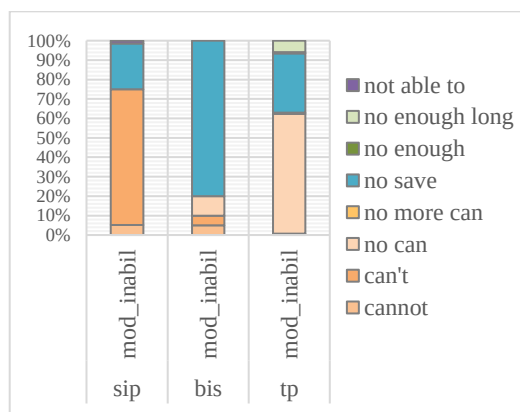


Figure 82: Relative frequencies of abilitative variants in negated contexts per variety

8.5.2.2 Diachronic analysis of abilitative in SIP

In the following section, a closer look at the various forms used in early SIP to encode abilitative modality will be taken. First, an overview of the attested forms will be provided before the focus will be on their temporal development. In the end, the results of the ctree algorithm will be presented and discussed in order to answer the question whether the data at hand provides useful information regarding when a stabilisation or change in the use of abilitative markers took place in SIP.

8.5.2.2.1 Attested forms

In SIP the forms *save*, *save for* and *can* were attested in affirmative contexts, whereby *save* occurred with highest frequencies ($28/47 = 59.57\%$). *Save* derives from the Portuguese word *saber* and was already used by the early traders in the Pacific area as a full verb meaning ‘to know’. The word also entered with this meaning into the plantation pidgins that were spoken in Queensland and Samoa.¹⁴⁸ There is evidence of an early use of *save* as a full verb in early SIP as well, as sentence (135) illustrates. At the same time, around 1880, there are already first attestations in the written data in which a form of *save* directly precedes a full verb. Sentence (136) shows such an instance, i.e. the full verb *save* no longer acts as the only verb in the clause and can be translated as ‘know how to/be able to’. Similar attestations are also found in SPPE (cf. Governor Solf 1895 in Mühlhäusler 1978) and QPPE (cf. for instance QLD Supreme Court 1884). Therefore, the preverbal use of the form may have been brought to the area by returning labourers. The verb continued to grammaticalise and stabilised into an ability marker in SIP as sentence (137) shows. The overall development of *save* in SIP follows a common grammaticalisation path since it is very common that verbs which have the meaning ‘to know’ form the primary source

¹⁴⁸ Cf., for instance, Schuchhardt (1889: 159) for evidence in Samoan Plantation Pidgin English and Fussell (1903: 48) for evidence in Queensland around 1883.

for the grammaticalisation of ability markers (Bybee et al. 1994: 190). From a cognitive-semantic perspective this is replicable since the ability to do something postulates that one knows how to do something and vice versa.

- (135) *White man, he savez too much*
 white man PM **know** a.lot
 ‘The white man knows a lot.’

(Alu Island 1881; Guppy 1887: 22)

- (136) *White man allsame woman, he no savee fight*
 white man like woman PM/3SG **NEG** **know** fight
 ‘The white man is like a woman who/he does not know how to fight.’

(Kolombangara 1880; Coote 1882: 206)

- (137) *To’ouna, you think Jesus savey save you?*
 To’ouna 2SG think Jesus **ABIL** save 2SG
 ‘To’ouna, do you think that Jesus can save you?’

(One Pusu 1943; Deck December 1943: 4; AU PMB DOC 439)

The form *save for* was only attested once in the written data. A closer look at sentence (138) shows that *for* takes the place of the English infinitive marker *to*. The combination of *save* ‘know’ with the infinitive-like particle *for* ‘to’ results in the literal translation ‘know how to’ and thus, expresses ability as well. The form *can* originates from StE and was only attested in direct preverbal position without additional infinitive markers (cf., for instance, (139)).

- (138) *me savvy for look out longa im.*
 1SG **ABIL** **INF** take.care PREP 3SG
 ‘I can take care of him.’

(Tulagi Island 1937; Horton 1965: 17)

- (139) *We man, we all right, we can kai-kai banana.*
 1PL man 1PL all right 1PL **ABIL** eat banana
 ‘We are men, we are all right, we can eat bananas.’

(Ugi Island 1909; Dickinson 1927: 107)

The morphemes *save* and *can* do also provide the basis for some of the attested forms to express inability. What is interesting is that although the common manner of verbal negation in SIP is through the markers *no* and *no more*, the marker *no* only negates the form *save*, but does not combine with *can*. Instead, the forms *can’t* and *cannot* were attested in the SIP data, which are based on the StE negative marker *not*. Although one might assume that the forms result from author or editor modifications, sentences such as (140) show that the form *cant* indeed occurred in basilectal MPE as well. In the example, *can’t* is realised as *cant* which shows that the form is likely to have been reinterpreted as a monomorphemic marker and to have not been used as an ability marker that is negated. Therefore, it is glossed in the example as INABIL. Lexifier and substrate speakers may have used the form simultaneously but with different grammatical interpretations. It is also possible that it entered SIP through acrolectal forms of the variety. The

form *cannot* was predominantly attested in sentences which contain several features that are closer to StE. For instance, the speaker of the utterance displayed in (141) starts the sentence with *It is very bad for him* and ends the sentence with *with all his shell*. The beginning and end are clearly StE, whereas the part in the middle, in which we find *cannot*, represents SIP. As was shown in Section 8.2, *kanot* as well as *kant* are forms that are still attested nowadays so that the occurrence of the forms in the early data is not unrealistic.

- (140) *me no savey, me cant look inside heart belong man*
 1SG NEG know 1SG **INABIL** look inside heart POSS man
 ‘I do not know, I am unable to look into a man’s heart.’
 (One Pusu 1931; Sullivan February 1931; PMB 1150)

- (141) *me think for now he cannot go back to Noumea quick time*
 1SG think PREP now PM **INABIL** return PREP Noumea quickly
 ‘... I think, for now, he can’t return quickly to Noumea ...’
 (Solomon Islands 1924; MacQuarrie 1948: 90)

What appears to be striking is that *can/can’t* is followed nine times by the verb *do* and the 3SG pronoun *it* (= *can’t do it*) as in (142). Due to the use of *do* instead of *make* and *it* instead of *him*, this may either be an indicator that a very acrolectal form of SIP was used, or it may already be the precursor of the form *kandoit* which is listed by Jourdan (2002: 89) as an ability marker. A closer look into the sentences in which the verbal phrase occurred reveals that almost all of the attestations are found in data of the SSEM. At least one further acrolectal feature (e.g., 1SG *I*, or 3SG *she*) was attested in the sentential contexts from which the verb phrases were extracted. The attested instances are not considered as grammaticalised *kanduit* here, since in the nine tokens no full verb is following the phrase so that *do* might represent a lexical verb.¹⁴⁹ Nonetheless, the form is used to express inability so that it is possible that the contemporary inability marker *can’t do it* has its origin in acrolectal PE speech as used by some devotees of the SSEM. It might be that *kandoit* represents an early stage of grammaticalisation.

- (142) *Me can’t do it, me fright*
 1SG **INABIL** **do** **3SG** 1SG afraid
 ‘I couldn’t do it, I was afraid!’
 (Solomon Islands 1946; Deck December 1946: 5; AU PMB DOC 439)

Another acrolectal form which was only attested in negated contexts is *not able to*. The form only occurs twice in the dataset:

¹⁴⁹ Cf. in contrast the following example in which the lexical verb *waka* follows the auxiliary form *kanduit*:

Hem kanduit waka
 3SG **INABIL** work
 ‘He cannot work.’ (Jourdan 2002: 89)

- (143) *I not able to see what men Manoba do along of Government.*
 1SG NEG ABIL INF see what man.PL Manoba do PREP government
 ‘I could not see what the Manoba men did with the government.’
 (Bundaberg; Enow 24.12.1888 [Australian Station 1888: 20])

8.5.2.2.2 Timeline and boxplot approach

To investigate whether the use of a form is restricted to a certain time period, a timeline based on the year of attestation (YEAR_ATT) was created, which is displayed in Figure 83. The horizontal dashed line separates the forms used in negated contexts from forms used in unnegated contexts.

Focusing on the timeline only, no change in the choice of form can be observed. The two verbs *save* and *can* are both attested for the first time in non-negated contexts in 1895 and both forms persist until the late 1940s. Especially the attestation of *can* is of interest because the form is not used any longer in contemporary SIP. As *save* and *can* are both attested in the period under investigation, this might be an indicator that *can* only disappeared after 1950.

In a similar way, the two main forms that are attested in negative contexts, *no save* and *can't* share their year of earliest attestation and both forms are attested until shortly before the end of the observed period. The two attestations of *not able to* occur in 1888 and 1947 and it could thus be assumed that they persisted throughout the period or might just have been scattered oddities. However, the fact that only two occurrences in acrolectal contexts were observed suggests that the two occurrences might just represent Standard English. The form *cannot* is only attested in the years 1924, 1927 and 1938.

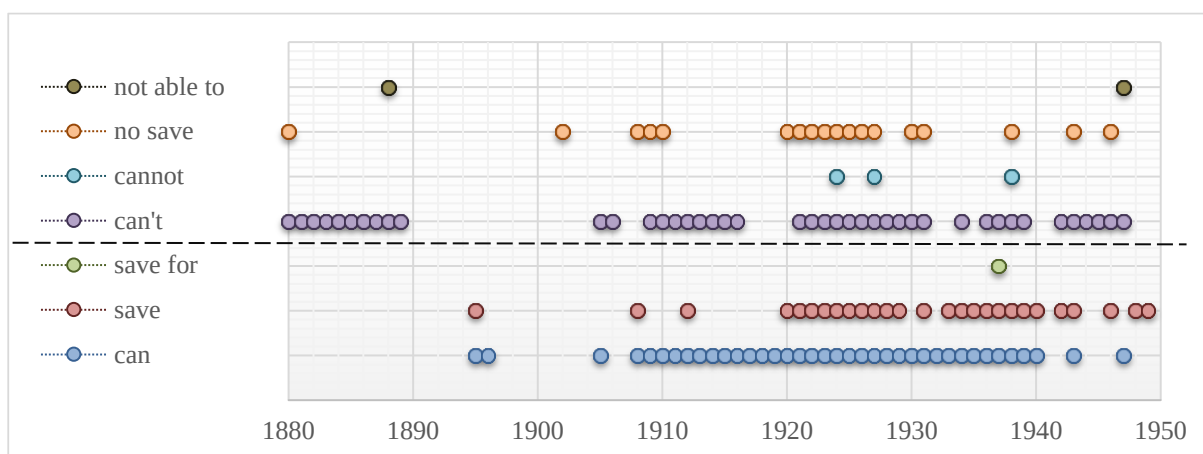


Figure 83: Timeline of ability variants in SIP

The boxplot in Figure 84 visualises the attestations across time if the focus is placed on the main morpheme used to encode the abilitative mood (M_MORPH). A comparison of the medians shows that the quantitative peak of *save* dates latest, namely to the year 1926. As the whiskers of the three main morphemes stretch out until the end of the observed period, all three morphemes

save, *can* and *able* are attested across time.¹⁵⁰ The visualisation is misleading since it seems as if the morpheme *able* was consistently attested despite the fact that there were only two attestations in total. In general, the results should be treated carefully due to the small number of datapoints.

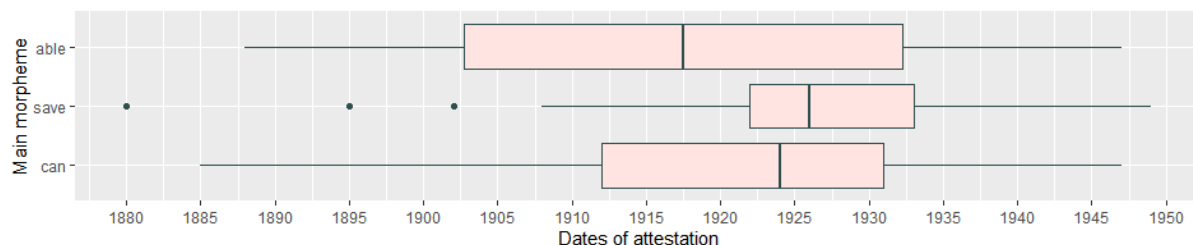


Figure 84: Boxplot of ability variants in SIP based on *M_MORPH*

8.5.2.2.3 Testing for the impact of the years of attestations on the choice of form

The timeline and boxplot approach do not include frequencies of attestations that may serve as indicators of ongoing change. In addition, it should be investigated whether the author and text type have an impact on the choice of form. Focussing on the timeline and boxplot, it seems as if the year of attestation has no impact on the choice of form. Applying the ctree algorithm to the data, testing the impact of the year of attestation on the forms used to encode abilitative modality, these observations are confirmed. The algorithm does not identify any splits independent of whether the focus is on the morphemes, the broad or fine-grained classification (*M_MORPH*, *ST_FORM* or *ST_FORM_F*).

If the factor *FEATURE* is added, which indicates whether the form is negated or not and the impact of both predictor variables on the form is analysed, the resulting tree consists of a single split only, which is based on the variable *FEATURE* (cf. Figure 85; $p < 0.001^{***}$). This shows that based on the data at hand, the year of attestation has no impact on the form used to encode ability in SIP but instead, the forms differ regarding whether they are negated or not. If the form is used to express ability, *save* is attested with higher frequencies. In contrast, to express inability, forms based on the morpheme *can* are most likely in the early sources.

¹⁵⁰ A boxplot based on the fine-grained classification of forms (= variable *ST_FORM_F*) can be found in Appendix I. The median of the forms *can* and *can't* date earlier than the medians of the forms *no save* and *save*, which may be a first indicator that *save* develops into the dominant form. However, due to the amount of data available per year, this cannot be claimed for certain.

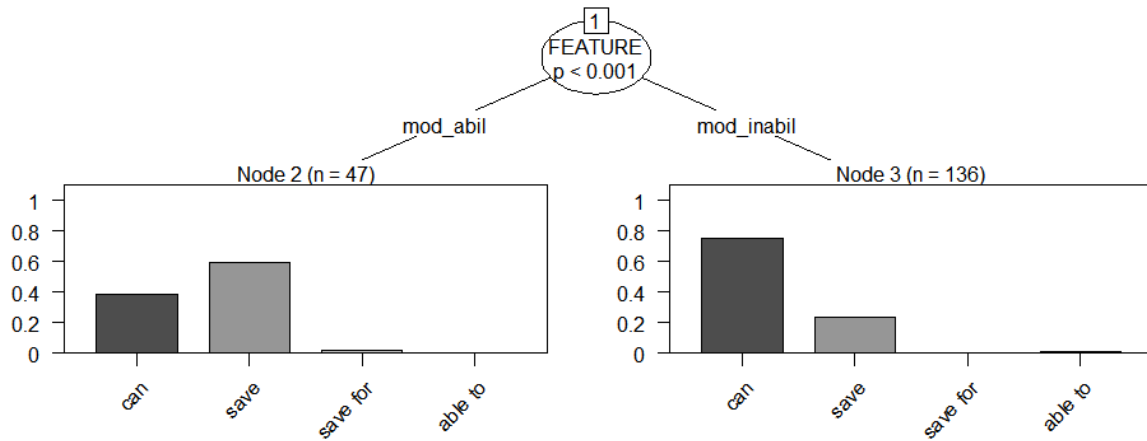


Figure 85: Conditional inference tree for ability in SIP ST_FORM~YEAR_DET+FEATURE

The split remains even if the text type is included as a further possible predictor variable into the analysis (see Figure 86). The tree consisting of three splits shows that in positive abilitative tokens the form *save* is slightly more likely to be attested than the form *can*. The second level split is based on the text type variable ($p=0.005^{**}$) and the third split is based on the year of attestation ($p=0.022^{*}$). According to the decision tree, a difference between pre- and post-1906 attestations can be observed in negated abilitative contexts in speech related attestations. While tokens dating prior to 1907 show an equal likelihood of forms based on *can* and *save*, post-1906 data shows a greater amount of forms based on the morpheme *can*. In written and intermediate attestations that express inability, a preference of forms based on *can* is observed (cf. Node 7).

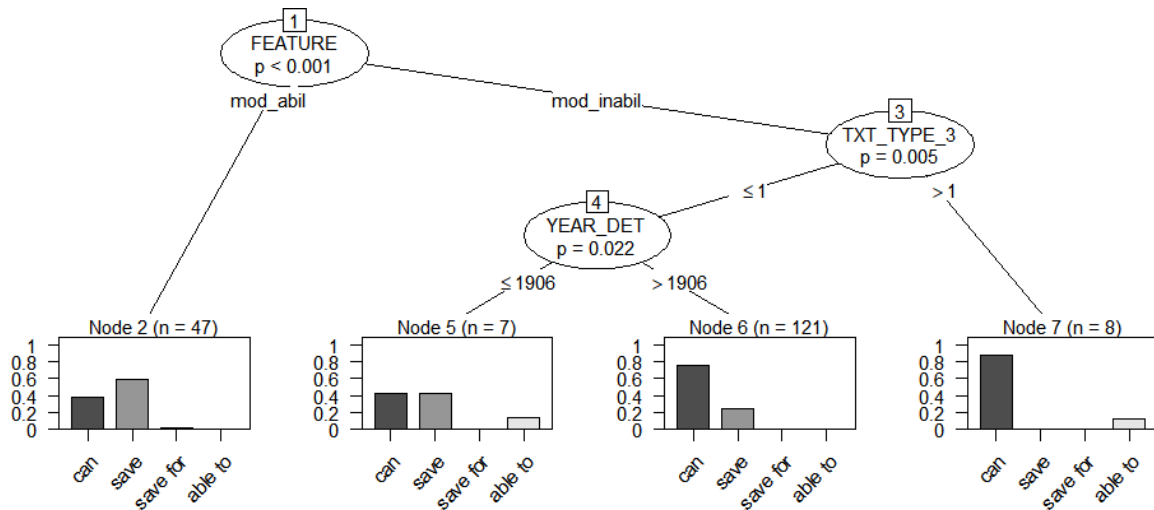


Figure 86: Conditional inference tree for ability in SIP ST_FORM~YEAR_DET+FEATURE+TXT_TYPE_3

If the author is added as an additional possible predictor variable for the choice of form, the predictor variables and their importance for the choice of form change. The resulting tree consists of two splits, the highest-level split being based on the author ($p<0.001^{***}$) and the second level

split being based on the text type ($p < 0.001^{***}$).¹⁵¹ A closer look at the authors reveals that 79.17% of the *can* attestations have their origin in SSEM authored documents. In contrast, only one third of the *save* attestations could be referred to SSEM writers. The high occurrence of *can*-based forms can thus be traced to sources connected with the SSEM.

8.5.2.3 Diachronic analysis of abilitative in BIS

After having looked at ability in SIP, the present section focusses on the forms which were found in the early BIS dataset. The sections have the same structure so that first, an overview of the attested forms will be provided before the focus will be on *when* the features were attested across time. The ctree algorithm will then be utilised to test whether the data at hand provides useful information regarding when a stabilisation or change in the use of abilitative forms took place.

8.5.2.3.1 Attested forms

Only 29 abilitative tokens were attested in the early BIS dataset. The most common forms were based on the morpheme *save*, occurring in 24 tokens (= 82.76%). As in SIP, the form has grammaticalised out of the full verb *save* meaning ‘to know’ (see sentence (144)). As the construction *save* + VERB was also attested in QPPE and SPPE, its development may have taken place on the overseas plantations. As sentences (145) and (146) show, the form occurs in both negated as well as non-negated contexts in early BIS.

- (144) *Me no savy.*
1SG NEG know
‘I do not know.’

(Tanna Island 1871; Forbes 1875: 251)

- (145) *‘Harry ‘e sabby walk about?’*
Harry PM ABIL walk
‘Is Harry able to walk?’

(Lehili 1916; Lynch 1923: 330)

- (146) *Harry ‘e no got. ‘Im ‘e no sabby pay copperah*
Harry PM NEG have 3SG PM NEG ABIL pay copra
‘Harry has none (no money). He is unable to pay/buy copra.’

(Lehili 1916; Lynch 1923: 328)

Five of the 29 attested ability forms are built with the morpheme *can*. Four of these *can*-based tokens are negated (cf., for instance, (147)). However, no clear preference can be observed as the forms *cannot*, *cant* and *no can* are all attested. More data would be required to make claims about

¹⁵¹ See https://www.dropbox.com/s/uzzwikbmzhtmg9x/Ctree_Ability_SIP.png?dl=0 (last access 29 September 2021).

whether *can*-based forms were in use as variants of *save*, to make claims about whether they were used by Europeans only, or whether they occur in the data due to author and editor revisions.

- (147) 'Man 'e **no** **can** savvy. Might you me catch 'im quick time.
 man PM **NEG** **ABIL** know DUB PL.INCL catch TR quickly
 'Humans are unable to know. We may get there quickly.'

(Espiritu Santo 1934; Harrison 1937: 146)

8.5.2.3.2 Timeline and boxplot approach

Figure 87 shows the timeline of ability datapoints in BIS across time. Again, the dashed line was used to separate forms used to express inability from forms used to encode ability. The earliest form attested is *can* in 1871. The negated form *cannot* is attested in 1872. Both forms represent single attestations. Similarly, the form *can't* was only attested once in the BIS dataset in 1913. Although *no can* is attested once pre-1900 and once in the late 1930s, there are gaps in which it is not attested.

Save first occurs in affirmative ability contexts in 1883 and is first attested in negated contexts in 1878. According to the timeline, *save* and its negated counterpart are the forms that spread the most across time.

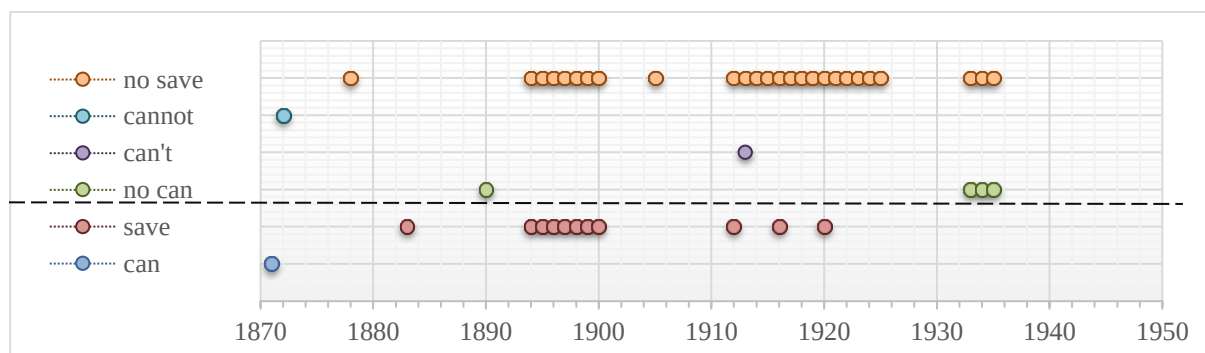


Figure 87: Timeline of ability variants in BIS¹⁵²

If we exclusively distinguish between the main morphemes (M_MORPH), the resulting boxplot in Figure 88 indicates that the median of *save* lies in 1915, whereas the median of *can*-based forms dates to 1890. Although this may be an indicator that *can*-based forms were used initially as borrowed forms and slowly became replaced, the amount of datapoints is too low to make such claims.

¹⁵² A boxplot based on the fine-grained classification of forms (ST_FORM_F) can be found in Appendix I. Focussing on *mod_abil* only, a clear preference of the form *save* is observed. In negated abilitative contexts *no save* and *no can* are attested. The median of *no save* dates to the year 1916 and the median of *no can* to 1912.

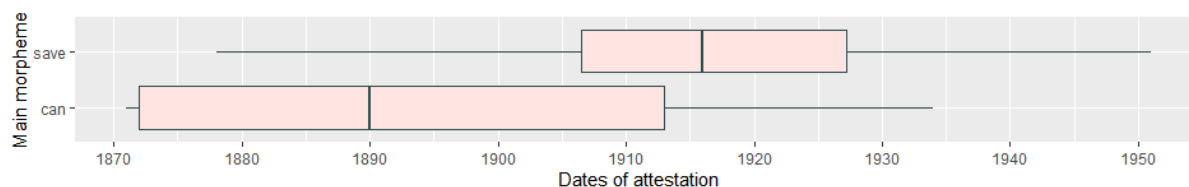


Figure 88: Boxplot of ability variants in BIS based on *M_MORPH*

8.5.2.3.3 Testing for the impact of the years of attestations on the choice of form

Since the distribution across time is not a reliable indicator of whether *time* represents a predictor variable in the forms chosen, the ctree algorithm was applied. This was also necessary to test whether the dominance of a certain form is based on the quality of the dataset, i.e. that the author or text type turn out to be the main predictor variables.

First, the ctree algorithm was used to test whether the year of attestation (*YEAR_DET*) has an impact on the choice of form (*ST_FORM*). The resulting tree structure, as displayed in Figure 89, shows a significant split in the year 1897 ($p=0.046^*$). However, as soon as other possible predictor variables are included into the analysis, such as the text type, the author and whether the focus is on ability or inability (*FEATURE*), a single node with no splits is obtained. Thus, there is no significant and reliable time split observable. One reason for this may be the generally low amount of attestations. Overall, the data shows a preference of the form *save*.

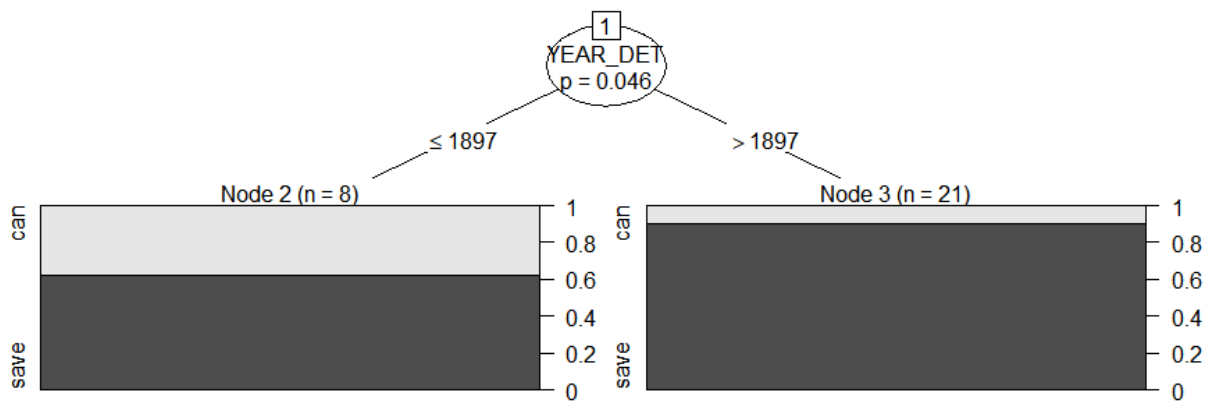


Figure 89: Conditional inference tree for ability in BIS *ST_FORM~YEAR_DET*

8.5.2.4 Diachronic analysis of abilitative in TP

In contrast to BIS, 262 abilitative utterances were attested in TP. The present section focusses on the forms that occurred in the variety and *when* they were attested. The results of the ctree algorithm will be discussed to investigate whether a significant time split can be observed which reveals which external events may have had an impact on changes in the variety.

8.5.2.4.1 Attested forms

The most dominant form attested in TP is based on the morpheme *can*. It occurs with and without the preverbal negative marker *no* in the data (cf. for instance (148)).

- (148) *Iesu nau kan olrait iu.*
 Jesus now ABIL heal 2SG
 ‘Jesus can heal you now.’

(Malakuna ~1935; Methodist Mission 1935: 13)

Save was identified in the early TP data as another form encoding ability. As in the other two observed varieties, the auxiliary has its origin in the lexical verb *save* meaning ‘to know’ and follows the common grammaticalisation path, as exemplified in sentences (149)-(151).

- (149) *Missis me no sawy road*
 misses 1SG NEG know way
 ‘Misses, I do not know the way.’

(Bismark Archipelago 1890; Kunze 1897: 70)

- (150) *White man allsame woman, he no savee fight*
 white man like woman 3SG/PM NEG know.how.to fight
 ‘The European is like a woman who does not know how to fight.’

(Kolombangara 1880; Coote 1882: 206)

- (151) *Man nating i no save mek-im mirakel*
 man ordinary PM NEG ABIL make-TR miracle
 ‘The ordinary man cannot perform miracles.’

(Alexis Harbour ~1942; Wolf 1942: 10)

Another form encoding ability, which is exclusively attested in TP, is the form *enough* (realised, inter alia, as *inap*, *e-nuff*, *nuf*, *i nap*). The form fulfilled various functions in the early TP data. It is likely that it entered TP as a derivative of the StE adjective with the meaning ‘enough, sufficient’. The form is attested as an adjective in sources that date to the 1930s (cf. van Baar 1930: 33) and 1940s:

- (152) *Baloose ‘im ‘e no nuff long you-me threefeller gold!*
 airplane 3SG PM NEG enough PREP 1TRI.INCL gold
 ‘The airplane is not large enough/sufficient for our gold.’

(Lae 1926-28; Idriess 1941: 240)

- (153) *Kaikai bilong yupala i-no inap long salim nambaut long Japan*
 food POSS 2PL PM-NEG enough PREP send-TR around PREP Japanese
 ‘Your food is not sufficient to spread it to the Japanese.’

(Salamaua 1943-45; Kerr 1985; NLA MS9002 PM8)

However, the English adjective seems to have developed into a verbal adjective in TP since the form is attested with both adjectival as well as verbal characteristics. For instance, it can take the *-im* particle to encode transitivity where required, as example (154) illustrates.

- (154) *dil-im* *fish,* ***inap-im*** *olgeder* *man*
 share-TR fish **make.enough-TR** all man
 ‘Share the fish; make it sufficient for all.’

(Mugil Mission ~1930; van Baar 1930: 8)

In addition, there are attestations of *enough* in the data that imply that the form could function as a *conjunction of purpose* as well, as the sentence extracted from a propaganda war pamphlet in (155) shows. Furthermore, *inap long* grammaticalises into a terminative preposition in locative and time-referential contexts (cf. Chapter 9 and sentences (156) and (157)). However, the form is also attested as an abilitative marker, as exemplified in (158).

- (155) *Yupala* *mas* *lus-im* *ologeda* *piles* *bilong* *nabis* *na* *mek-im* *liklik* *piles*
 2PL OBLIG leave-TR all place POSS beach and make-TR little place

long *bus* ***inap long*** *Japan* *i-no* *kan* *paen-im* *yu*
 PREP bush **so.that** Japanese PM-NEG ABIL find-TR 2SG
 ‘You must leave the villages near the beach and build villages in the bush area so that the Japanese can’t find you.’

(New Guinea 1943-45; Kerr 1985; NLA MS9002 PM23)

- (156) *Bus* *bilog* *Morobe* ***enap long*** *Lae* *i* *klia* *long* *ol* *Japan*
 bush PREP Morobe **TERM** Lae PM clear PREP PL Japanese
 ‘The bush from Morobe to Lae was cleared from all Japanese.’

(New Guinea 1943-45; Kerr 1985; NLA MS9002 PM28)

- (157) *Yesus* *i* *stop* *long* *Nasaret* ***inaf long*** *trifelo* *ten* *yar* *belong* *em*
 Jesus PM LOC PREP Nazareth **TERM** thirty year POSS 3SG
 ‘Jesus stayed in Nazareth until he turned 30.’

(Alexis Harbour ~1942; Wolf 1942: 10)

- (158) *Ol* *i-no* ***inap log*** *qo* *log* *Australia*
 3PL PM-NEG **ABIL** go PREP Australia
 ‘They are unable to go to Australia.’

(New Guinea 1943-45; Kerr 1985; NLA MS9002 PM18a)

As Kuteva et al. (2019: 415) argue, it is not uncommon for verbs expressing suitability to grammaticalise into the abilitative marker. A closer look at the dates of attestations is necessary in order to understand the grammaticalisation path of the forms.

8.5.2.4.2 Timeline and boxplot approach

The timeline in Figure 90 shows that the earliest attested form in TP is *can’t*.¹⁵³ However, it is only attested in the 1840s and thus, does not seem to have been a common marker for encoding inability. In both negated as well as non-negated contexts *save* represents the second earliest attested form. *No save* occurs in 1890 for the first time and the form *save* is found in the data from 1908 onwards. By contrast, *no can* is attested for the first time in the written data in 1913 and *can*

¹⁵³ As previously, the dashed line was used to separate forms used to encode inability from those used to encode ability.

in 1926. The occurrence of ability forms based on the marker *enough* do not date before 1940. The negated form *no enough long*, as well as non-negated *enough* and *enough long*, are attested in 1940 for the first time.

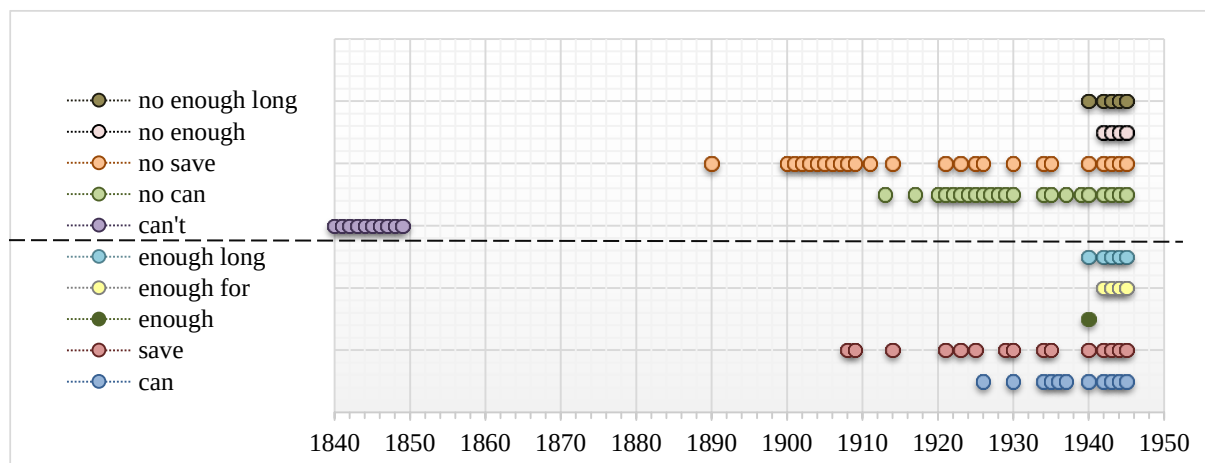


Figure 90: Timeline of ability variants in TP

As pointed out earlier, *enough* represents a multifunctional item in the early data. Independent of the function which the form fulfils, it is attested rather late in the written material. The earliest attestations of the form *enough* are found in van Baar's *Pitshen-Wörterbuch* which was probably written around 1930. A closer look into the dictionary reveals that though many functionalities of *enough* are listed, there is no evidence so far that the form was used to express ability. The first attestations of *enough* as an ability marker can be found in Kutscher's *Wörterbuch deutsch-pidgin-english* which dates roughly to the 1940s. For instance, 'inexplorable' is translated as *i no nap lg painim* (Kutscher 1940: 114). What is interesting is that Kutscher's dictionary lists not only *enough*, but also *can* and *save* as forms to express ability. In entries such as the one for 'inexplorable' two translations are found, namely *ol i no ken painim* and *i no nap lg painim* (Kutscher 1940: 114). This shows that the forms *can* and *enough* were used interchangeably. However, the majority of abilitative utterances in the dictionary are encoded with a form of *can*. Thus, during the late 1930s the form *can* may still have dominated.

As the timeline shows, all forms that were attested in the written data (with the exception of *can't*) are attested until the end of the observed period, especially in the years 1942-1945.¹⁵⁴ Thus, a clear preference for a single form cannot be observed. Variation in the choice of form seems to have prevailed until at least the 1950s.

¹⁵⁴ The high number of 1942-1945 attestations can be inter alia referred to the high amount of war pamphlets that form part of the TP database.

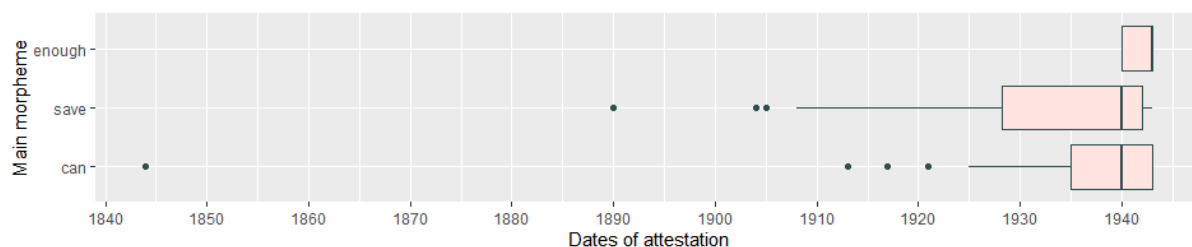


Figure 91: Boxplot of ability variants in TP based on *M_MORPH*

The boxplot which is based on the main morphemes used to encode abilitative modality demonstrates that the medians of *save* and *can* date to 1940. The median of *enough* dates to the year 1943. To understand the developments in regard to the ability marker in TP, more data and data covering the years after 1950 is required.¹⁵⁵

8.5.2.4.3 Testing for the impact of the years of attestations on the choice of form

Although the timeline and boxplot approach seem to show that no clear preference of the form is made before 1950, the ctree, analysing the impact of the year of attestation on the form, shows a significant split in the year 1934. The same split is obtained when the variable *FEATURE*, indicating whether the form encodes ability or inability, or both, the *FEATURE* and text type variables, are included as further predictor variables.¹⁵⁶ Data attested before 1935 shows a 60% chance to being the form *save* and a 40% chance to being the form *can*. In contrast, data that is attested after 1934 is most likely to make use of the form *can*. The form *save* is less frequently used and, in addition, forms based on the morpheme *enough* are attested:

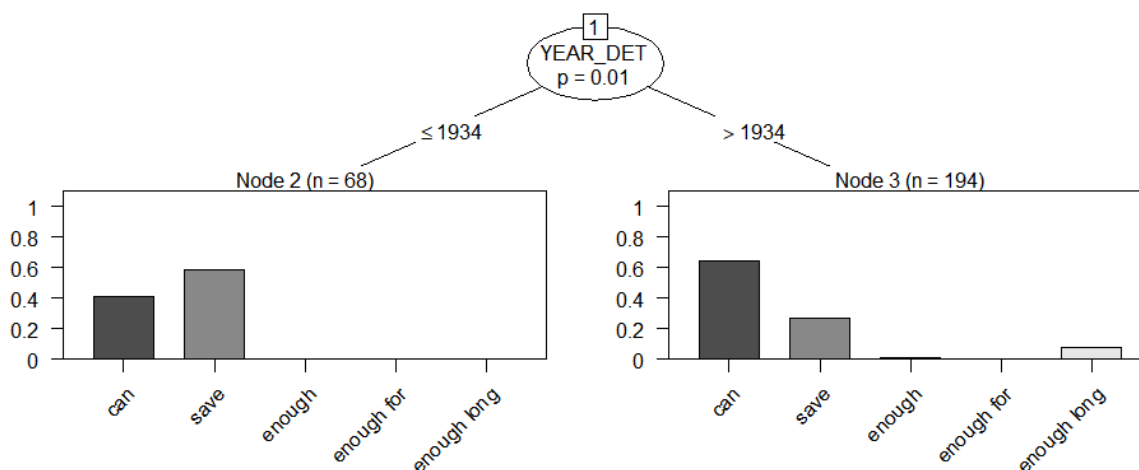


Figure 92: Conditional inference tree for ability in TP *ST_FORM~YEAR_DET*

If the author is considered as a further predictor variable, Figure 93 is obtained. The year 1934 remains the most significant predictor variable ($p=0.039^*$) regarding the choice of the form

¹⁵⁵ A boxplot based on *ST_FORM_F* can be found in Appendix I.

¹⁵⁶ The p-level changes depending on which variables are included into the analysis. *ST_FORM ~ YEAR_DET*: $p=0.01^*$; *ST_FORM ~ YEAR_DET + FEATURE*: $p=0.02^*$; *ST_FORM ~ YEAR_DET + FEATURE + TXT_TYPE_3*: $p=0.029^*$.

(ST_FORM). Data attested before 1934 is more likely to show the form *save* than the form *can*, although both forms are attested. The split on the second level is based on the author. While in post-1934 attestations the majority of authors show a clear preference of a form based on *can* (despite *save* and *enough long* being used as well), there are only two authors, namely Fransciscus Wolf and N. C. A. Helton who before 1943 only use the form *save*. Post 1942 the form *can* represents the dominant form used by the authors, although *save* and *enough* represent variants.

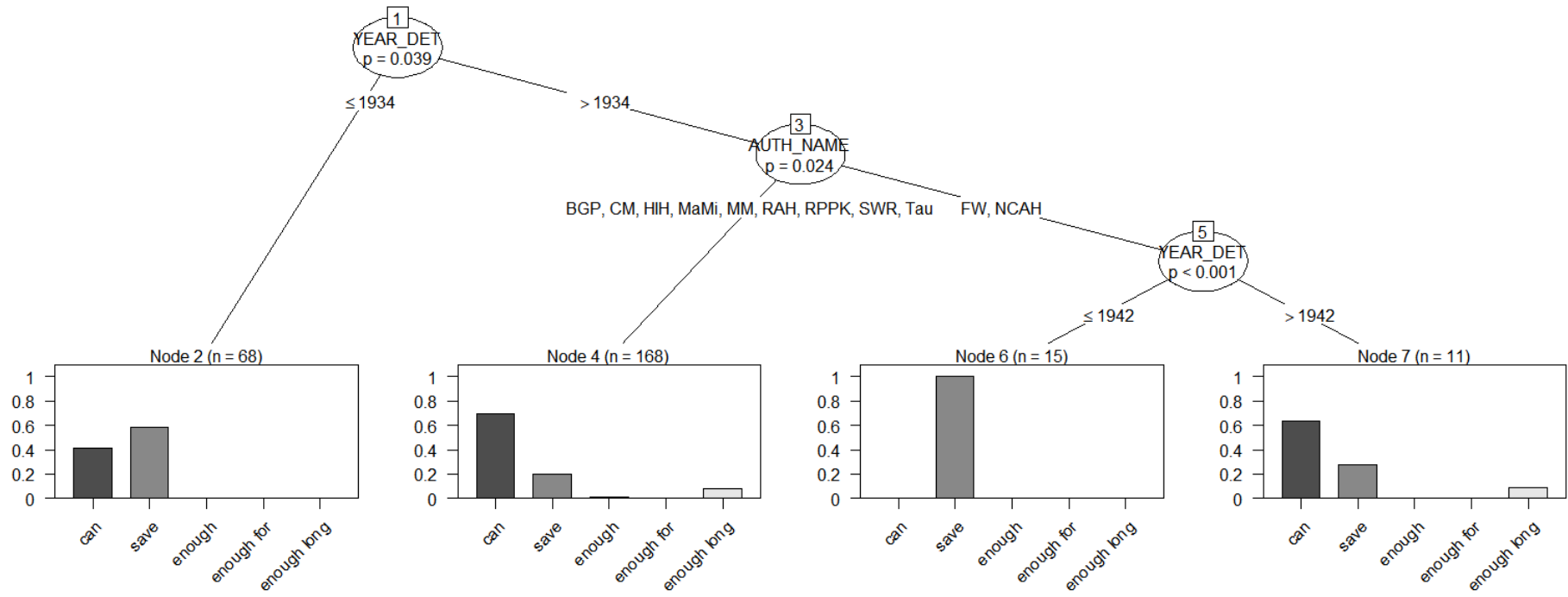


Figure 93: Conditional inference tree for ability in $TP\ ST\_FORM \sim YEAR\_DET + FEATURE + TXT\_TYPE\_3 + AUTH\_NAME$

8.5.2.5 Diachronic comparative analysis of abilitative in MPE

The individual analysis of abilitative markers in SIP, BIS and TP has shown that in the period under investigation variation was the norm. In SIP forms based on *save* and *can* were attested until the end of the observed period. While *save* and *can* were attested to express ability, the forms *no save* and *can't* were the major forms to encode inability. The ctrees analyses showed that the year of attestation has no impact on the choice of form used to encode the notions of ability and inability.

Regarding BIS, the forms *save/no save* and *can/no can/can't/cannot* were attested, whereby the form *save/no save* was the dominant form. Although the year 1897 evoked a split if the focus was on the year of attestation, as soon as other predictor variables were considered, no splits were obtained. A further problem represented the low amount of datapoints collected for BIS.

The analysis of abilitative markers in TP showed that TP made use of ability forms based on *save* and *can* as well. The negated forms seem to have been *no save* and *no can*. Interestingly, a significant split in the year 1934 was observed with data before 1935 showing a slightly higher occurrence of *save* in comparison to *can*, and data after 1934 showing higher occurrences of the form *can* and the attestation of new forms based on the morpheme *enough*. Even if the author was included, 1934 remained a split-evoking year.

The ctrees algorithm was applied to the datapoints of all three varieties. Because the previous analyses showed that the varieties differ in terms of how the form *can* is negated, the ctrees is based on the fine-grained variable ST_FORM_F, which lists not only the main morpheme but also the negative marker and the infinitive marker if applicable. A ctrees testing the impact of the year of attestation, negation vs. non-negation and the text type on the choice of form confirms that major differences exist in the forms used to encode inability.¹⁵⁷ On the highest-level the data is split according to whether the forms encode ability or inability (FEATURE; $p < 0.001^{***}$). With positive ability forms, the variety does not represent a significant predictor variable. However, the second level split indicates that the variety is a highly significant predictor variable when encoding inability ($p < 0.001^{***}$). The algorithm clusters TP and BIS together and shows that SIP seems to differ in encoding inability. Node 4 and Node 5 further split the TP and BIS data based on the year of attestation ($p < 0.001^{***}$). Nevertheless, these results should be taken with caution because only 29 instances of abilitative cases in BIS were taken into consideration. The high amount of occurrences of the form *can't* explain why SIP is separated from BIS and TP. If the

¹⁵⁷ The ctrees is too large to be displayed here, but can be found online at https://www.dropbox.com/s/oc3ks4srwxohgur/Ctrees_Ability_MPE.png?dl=0 (last access 29 September 2021).

author is additionally considered, a ctree with ten splits on eight levels is obtained, whereby five are based on the author. The author seems to have the greatest impact on the choice of form.

What can be learned from the written attestations is that the grammaticalisation of abilitative markers was not completed by the end of the 1940s. Data post 1950 is required to observe how and when individual forms tuned into the main ability markers in the varieties. Nonetheless, the observed differences may give us some general tendencies about the marker's development. Forms based on *save* were used to encode ability in all three varieties. There is evidence that *can* was also used to encode ability in TP and SIP and possibly also in BIS. However, differences can be observed in how *can* is negated when encoding inability. While in TP *can* was negated through the common preverbal marker, SIP used the forms *can't* and *cannot*, which are still attested today. Furthermore, TP differs from the other two varieties in that the form *enough* grammaticalises as an additional marker. The grammaticalisation of *enough* seems to have happened after the end of the labour trade. Abilitative auxiliaries, which are attested in the contemporary versions of all three varieties, were already in use before 1950, albeit still competing with other forms.

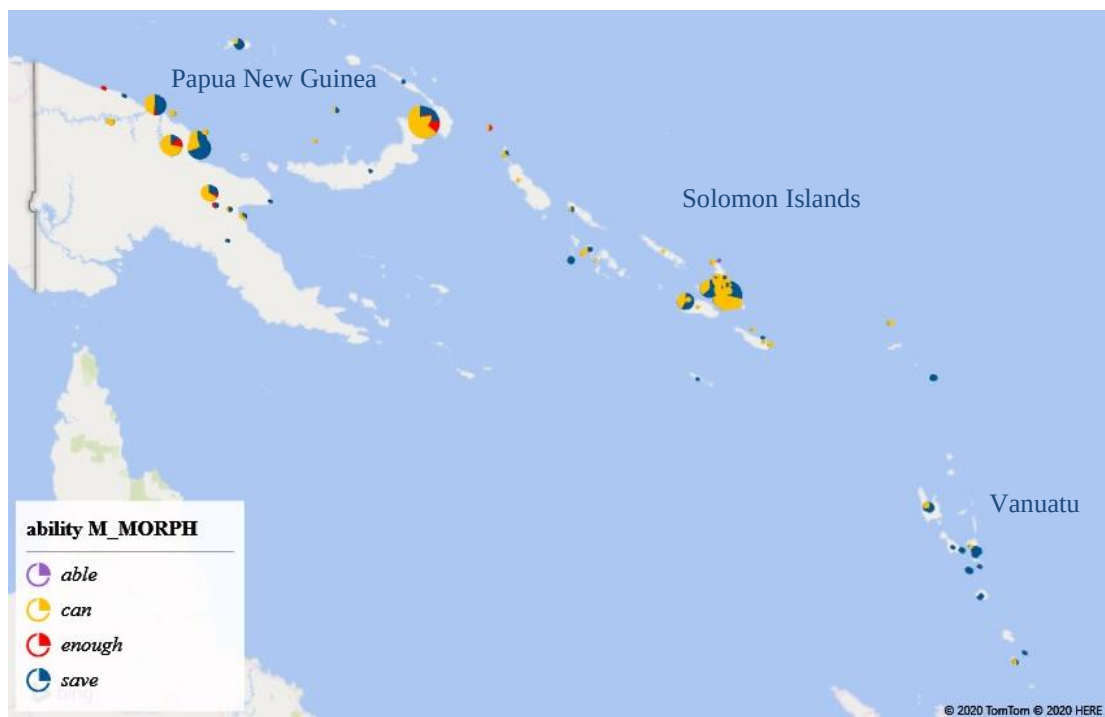


Figure 94: Map of Melanesian Islands, showing attestations of ability M_MORPH variants

8.6 Permissive

The present section will focus on *permission* and *prohibition* and the auxiliary markers used to encode the modalities in the historical data of the three MPE varieties. First, some methodological considerations will be discussed before the findings will be presented.

8.6.1 Methodological considerations

The analysis of permissive and prohibitive modality will focus on overt markers identified in the varieties. The various forms attested in the early data that encode permissive and prohibitive modality are listed in Table 17 and will be further explained with example sentences.

FACTOR	VARIANTS
ST_FORM	<i>can</i> <i>no can</i> <i>can't</i> <i>cannot</i> <i>save</i> <i>no save</i> <i>no more save</i> <i>must not</i>
FEATURE	mod_perm mod_prohib
M_MORPH	<i>can</i> <i>save</i> <i>must</i>
YEAR_DET	1832 ... 1950

Table 17: Linguistic coding for permission

However, for the sake of completeness, it should be noted that zero marking was common as well. For instance, sentence (159) and (160) show that affirmative or negated imperatives without overt markers represented one alternative way of expressing permission and prohibition.

- (159) *You lotu 'long one fella Deo dass all.*
 2SG glorify PREP one MODIF God only
 'Thou shalt have no other gods before me.'

(New Guinea 1927; Matches 1930: 80)

- (160) *Ju no sitil.*
 2SG NEG steal
 'You shall not steal.'

(Vunapope 1934; Baker 1945: 330)

Focussing on the attested preverbal auxiliaries, *can* was identified as one form to put into code permissive modality, as sentence (161) illustrates. In addition, the early data gives rise to the question whether *save* also served as a preverbal permissive marker. For instance, in sentence (162), which was extracted from Hogbin's translation of the rules for native officials in Guadalcanal, the form *savvy make-im Court* is used to encode that the district officer is allowed to put a person on trial if that person has committed a crime.

- (161) *Spose boy he lie long Court, Kiap he can calaboose im*
 if boy PM lie PREP court chief PM **PERM** imprison TR
 'If a person lies in court, the chief can (=has the permission to) imprison him.'

(Kokopo 1927; McCarthy 1926-1952; PMB 616)

- (162) *this fellow* *District Officer* *him* *he* **savvy** *make-im* *Court* *along*
 this MODIF district officer 3SG PM **PERM** make-TR court PREP

 this fellow *man,* *this fellow* *woman.*
 this MODIF man this MODIF woman
 ‘The district officer can (=is allowed to) put this man/this woman on trial.’
(Solomon Islands 1943; Hogbin 1944: 279)

In order to encode prohibition, the forms are negated with the negative markers *no* and *no more*, resulting in *no save*, *no more save* and *no can* (cf. sentences (163)-(165)). Similar as with *can* in abilitative modality, the forms *can't* and *cannot* were also attested (cf., for instance, sentence (166)).

- (163) *Him he no savvy talk, no savvy make noise.*
 3SG PM NEG PERM talk NEG PERM make noise
 ‘One is not allowed to speak (and) not allowed to make noise.’
 (Solomon Islands 1943; Hogbin 1944: 278)

- (164) Suppose we school we **no more** **savey** pight we all same woman
 if 1PL school 1PL **NEG** **PERM** fight 1PL like woman
 ‘If we go to the (missionary) school we are not allowed to fight, we are like women.’
 (Baunani 1912; Deck February 1912: 3; AU PMB MS 1253)

- (165) *Yu* **no** **ken** *kil-im* *man.*
 2SG **NEG** **PERM** kill-TR man
 ‘You shall not murder.’
- (Alexis Harbour 1934; Baker 1945: 330)

- (166) *You me cant hold on long two Master, we must [...] hold fast long Jesu*
 1PL.INCL **PROH** hold on PREP two master 1PL OBLIG [...] hold fast PREP Jesus
 ‘We cannot serve two masters, we must hold onto Jesus.’
 (Inakona 1932; Cowie 01.08.1932; PMB 1150)

A further variant which occurred in the early data was *must not* (cf. sentence (167)). It must be stated, however, that the marker only occurred once in the speech of a European. The complete utterance of the European is “I am going to speak good along altogether. You must not speak. Suppose you no speak I will give you one case of tobacco.” and contains acrolectal as well as pidgin features. Due to the fact that there is only a single attestation, it can be assumed that the use of the marker represents a case of author modification, or that the marker was used by the European but will not have been used by Pacific Island speakers.

- (167) You **must not** speak
 2SG **PERM NEG** speak
 ‘You shall not speak.’
- (Ugi Islands 1908-1926; Dickinson 1927: 120)

As pointed out in Section 8.5, the boundaries between ability and permission are not always clear-cut. In example (166), *cant* was identified as a prohibitive marker due to the fact that the sentence was extracted from a religious conversation which referred to the Ten Commandments. However, if the contextual situation had not been clear, *cant* may have been interpreted as an ability marker

as well. This is because the semantic concepts of ability and permission are closely related. Kuteva et al. have shown that markers that encode ability frequently grammaticalise into markers of permission (Kuteva et al. 2019: 345). It is a common phenomenon that the two notions cannot be clearly distinguished. For instance, in the English sentence *Melissa can come*, the auxiliary may represent a case of ability or permission. Without further contextual information a clear interpretation is impossible. The early data contained several sentences in which it could not be clearly identified whether the auxiliary, independent of its form, was encoding ability/inability or permission/prohibition. Unclear cases were not included in the analysis.

Next to ST_FORM, the data was coded for FEATURE (*mod_perm* and *mod_prohib*) and for M_MORPH, the latter factor solely distinguishing between the major morphemes used.

8.6.2 Findings and discussion

In the present section, the results of the diachronic analysis of permissive/prohibitive markers will be outlined. The section will start with some remarks on the general data distribution before each variety will be considered individually starting with SIP in Section 8.6.2.2, continuing with BIS in Section 8.6.2.3 and focussing on TP in Section 8.6.2.4. Since the application of the ctree algorithm would be pointless for the TP and BIS data, the focus of the analysis was on the forms and the timeline. Section 8.6.2.5 will focus on the results of the study from a comparative perspective.

8.6.2.1 General data distribution

The number of datapoints that contain a permissive/prohibitive auxiliary amounts to 248 sentences only. The tokens are distributed unequally among the varieties. While 192 of the 248 instances are found in the TP dataset (= 77.41%), only 21.37% (= 53/248) are attested in the SIP data and only 1.61% (= 4/248) in the BIS data. In addition, Figure 95 shows that especially in TP the great amount of attestations accumulates in the years 1927 and 1943.

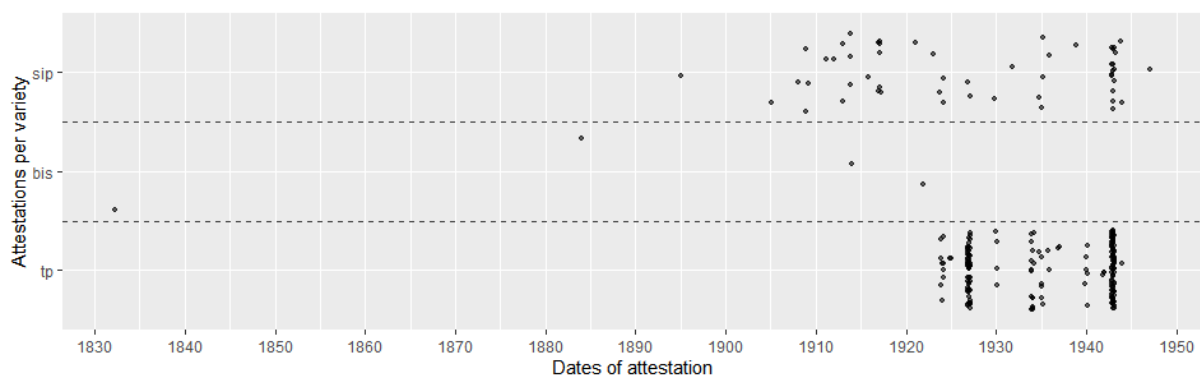


Figure 95: Distribution of permissive datapoints across time per variety

This indicates that the high amount of TP attestations were found in a small amount of sources only and, thus, may not be diagnostically significant. Although the database for the analysis of permissive/prohibitive modality is far from ideal, a look at the attested forms will be used to investigate whether we can nonetheless learn about the development of the markers in the three varieties.

If the time from 1830 to 1950 is treated as a single period and the focus is placed on the main morphemes used to encode permission/prohibition, the three varieties show the following morphemes:

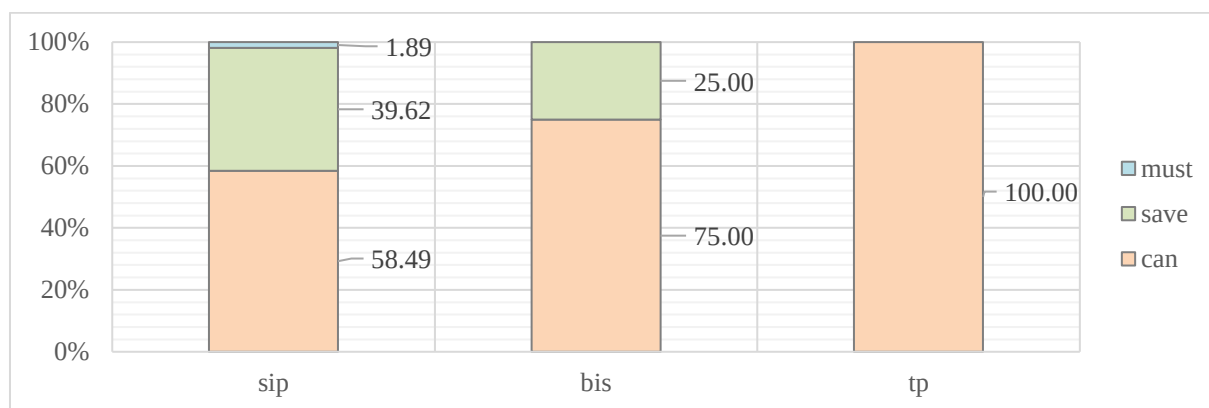


Figure 96: Relative frequencies of permissive *M_MORPH* variants per variety (BIS and SIP *save* ~ *can*: $p=0.544$)

As Figure 96 demonstrates, in TP the main morpheme used in permissive/prohibitive semantic contexts is *can*, which is also the main morpheme attested in 58.49% (= 31/53) of the SIP tokens and in 75% (= 3/4) of the BIS tokens. However, in SIP around 40% (= 21/53) and in BIS 25% (= 1/4) of the forms are based on the morpheme *save*.¹⁵⁸ It should be kept in mind that in BIS only four permissive/prohibitive tokens were attested and the qui-square test comparing the distribution of *save* and *can* in BIS and SIP is not significant ($p=0.544$). Thus, the results are not diagnostically meaningful.

8.6.2.2 Diachronic analysis of permissive in SIP

To encode permission in SIP, the markers *save* and *can* were attested as the primary morphemes in preverbal position. They were used to encode permission as exemplified in (162) and (168). In prohibitive contexts, the morphemes were negated with the common preverbal negative markers *no* and *no more*. Thus, *no save*, *no more save* and *no can* were attested to express prohibition, as shown in sentence (163), (164), (169) and (170). In addition, the from StE-derived contracted

¹⁵⁸ A map showing where the permissive/prohibitive markers were attested is displayed in Figure 104.

form *can't* was used to encode prohibition, as sentences such as (166) confirm. The single attestation of *must not*, which was shown in (167), was assigned to the SIP data as well.

- (168) [...] *no man can talk other way*
 [...] NEG man **PERM** talk other way
 'Nobody can/is allowed to say anything to the contrary.'
 (Solomon Islands 1905; Watkinson 1906: 8; AU PMB DOC 439)

- (169) *H'm b'long tambo – no can sell*
 3SG PREP taboo **NEG** **PERM** sell
 'It is taboo/forbidden – cannot be sold'
 (Roviana Island 1909; Burnett 1911: 120-121)

- (170) *Why we no can go American camp now? Why tambu?*
 why 1PL **NEG** **PERM** go American camp now why forbidden
 'Why are we not allowed to go to the American camp now? Why is it forbidden?'
 (Guadalcanal 1944; Anonymous 01.07.1944: 4)

The timeline in Figure 97 shows when the individual forms were attested across time. It shows that *no save* represents the earliest form being first attested in 1895 and latest attested in the collected data in 1943. *Save* is only attested with the negative marker *no more* in 1912. The negated form *can't* is attested from 1909 until 1947 in the written data. *No can* is only attested in 1910 and 1944.

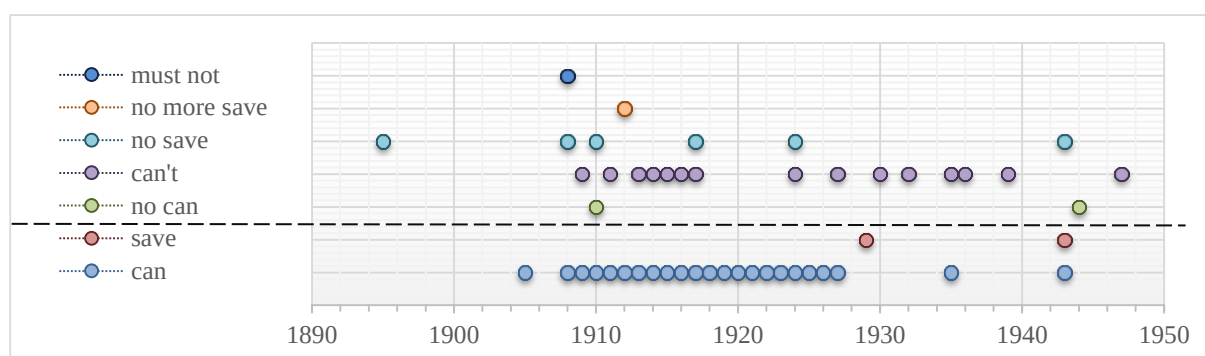


Figure 97: Timeline of permissive/prohibitive variants in SIP

The single attestation of *must not* seems to reflect a direct borrowing from StE which did not enter the variety as a variant form. As previously shown, the marker was attested in a source in which acrolectal features were mixed with Pidgin English features. In permissive contexts, the verbs *save* and *can* are attested in preverbal position. While *can* is attested from 1905 until 1943, there are only two years in which *save* was attested, namely in 1929 and 1943.

Figure 98 shows a boxplot based on the main morphemes (M_MORPH) and the variable YEAR_DET. Thus, periods of attestation were replaced by determined years of attestation. What can be observed is that both *save* and *can* represent variant morphemes to express permission/prohibition in the observed period. If the medians are compared, it becomes visible that the median of the morpheme *can* dates to the year 1923 and the median of *save* to the year

1943, which may already be an indicator that a change in the choice of forms occurs. However, the figure does not provide information about further predictor variables and does not take frequencies and statistical significance into account.



Figure 98: Boxplot of permissive/prohibitive variants in SIP based on *M_MORPH*

A ctree measuring the impact of the year of attestation on the choice of form (*ST_FORM*) evokes a significant time split ($p = 0.049^*$) in the year 1939, as shown in Figure 99.

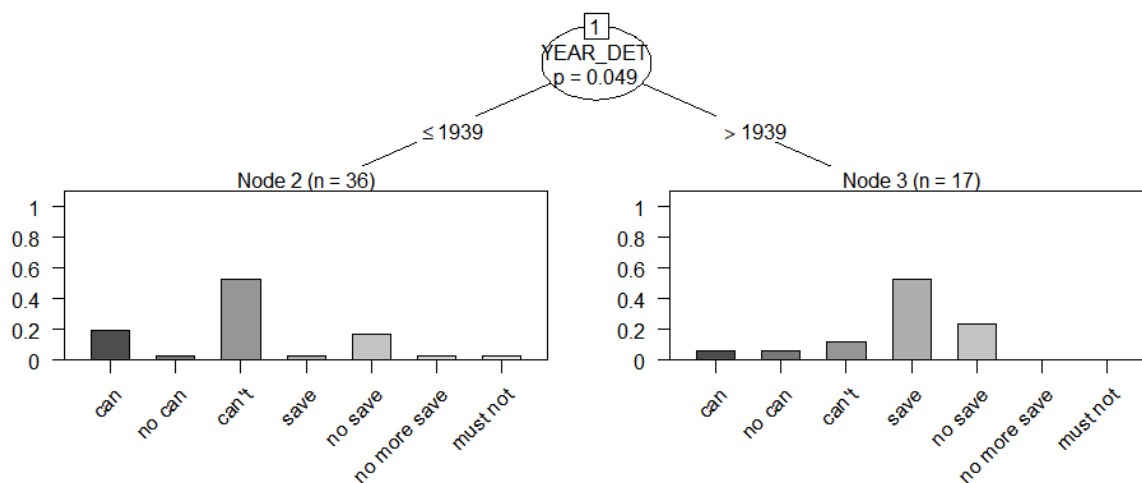


Figure 99: Conditional inference tree for permissive/prohibitive in SIP *ST_FORM*~*YEAR_DET*

While the collected data stemming from before 1940 shows a higher amount of the forms *can't* and *can*, in post-1939 data the forms *save* and *no save* are most likely to occur. This split can however only be observed if the analysis is based on the variable *ST_FORM*. If the study concentrates on the main morphemes (*M_MORPH*), no splits are obtained.

If the text type is included either alone or together with the author as further possible predictor variables to test their impact on the choice of form, the ctree algorithm creates a tree with a single, very significant split based on the text type ($p < 0.005^{**}$ and $p < 0.007^{**}$).

As Figure 100 shows, in speech-like attestations *can* and *can't* are most likely to occur, whereas in written and intermediate attestations the form *save* and *no save* seem to be favoured. The split remains the same if the focus is placed on the main morphemes used to encode permissive/ prohibitive modality ($p < 0.004^{**}$). The time variable does not represent a significant predictor variable. As only 53 tokens were identified in total, the results have to be treated with caution.

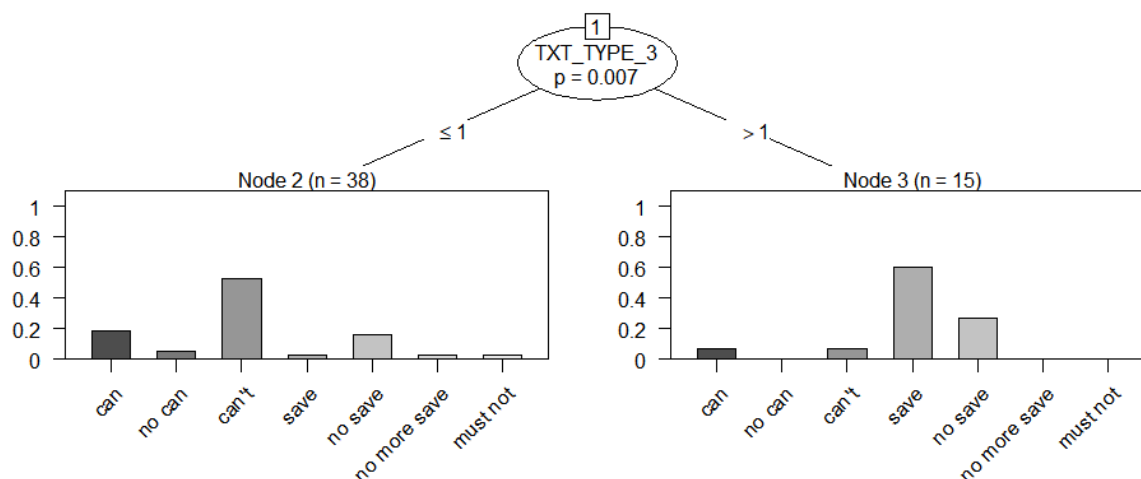


Figure 100: Conditional inference tree for permissive/prohibitive in SIP ST_FORM~YEAR_DET+TXT_TYPE_3+AUTH_NAME

8.6.2.3 Diachronic analysis of permissive in BIS

The BIS data contained only four attestations of prohibitive sentences. The datapoints and their dates of attestations are portrayed in Figure 101. It is striking that only prohibitive sentences were identified. As example (171) shows, the form *no can* was attested very early and outside of the New Hebrides, when Bennett brought a Vanuatu girl to the UK. *Can't* was attested twice, once in 1844 and once in 1922, as example (172) indicates. Sentence (173), which refers to 1914, demonstrates that negated *save* may have represented a further variant to encode prohibition.

- (171) *Now little girl-s no can go in the dark.*
 now little girl-PL NEG PERM go PREP ART dark
 'Now little girls can't/are not allowed to go outside in the dark.'

(UK 1832; Bennett 1883: 7)

- (172) *You can't make trade here, ground here blong company*
 2SG PROH make trade here ground DEM POSS company
 'You are not allowed to trade here, this is the ground of the company'

(Vila; Lengwe 06.12.1922)

- (173) *'Suppose one woman make him all same [...] he no sabby go belong school*
 if ART woman make TR alike [...] PM NEG PERM go PREP school
 'If a woman behaves like this, ..., she is not allowed to go to school.'

(Port Vila 1914; Lynch 1923: 106-107)

Due to the low amount of data available for BIS, no statements about the development of permissive/prohibitive modality across time can be made.

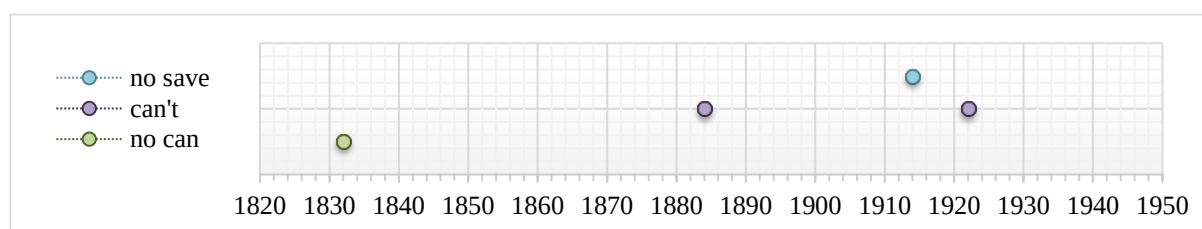


Figure 101: Timeline of permissive/prohibitive variants in BIS

8.6.2.4 Diachronic analysis of permissive in TP

The early TP dataset contains the highest number of permissive/prohibitive tokens. 72.92% of the tokens express prohibition, whereas the remaining 27.08% encode permission. In TP only *can* and negated *no can* were attested; an example of each were displayed in sentence (161) and (165).

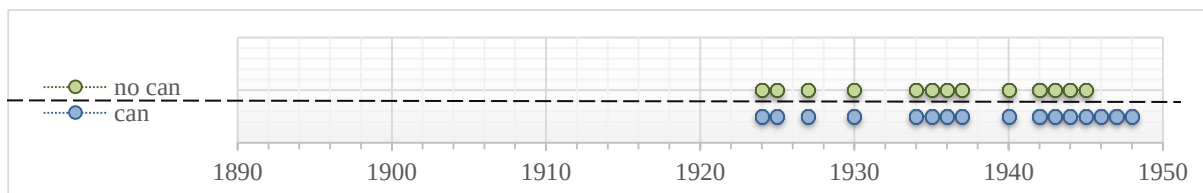


Figure 102: Timeline of permissive/prohibitive variants in TP

Although many instances are counted, their attestations only spread across the years 1924 to 1948, as Figure 102 shows. In addition, the datapoints accumulate in the years 1927 and 1943 (cf. Figure 95). The reason for this is that most of the datapoints encoding prohibitive or permissive modality were found in war pamphlets and the McCarthy Patrol rules.

It is interesting to note that the form *can* also represented TP's dominant form to express ability. As pointed out in the methodology, it is common for permissive markers to grammaticalise out of abilitative markers. From the late attestations of permissive/prohibitive markers, it may be suggested that overt means to encode permission and prohibition developed later and out of the earlier analysed ability markers. It is also conceivable that the late occurrence is based on the surviving written records and the type of documents that survived. However, as prohibitive modality was expressed in earlier data as well through the means of imperatives or phrases such as *he tambu* (cf. sentence (74)), the late attestations of preverbal auxiliaries are likely to indicate a late development of auxiliary markers.¹⁵⁹

8.6.2.5 Diachronic comparative analysis of permissive in MPE

The analysis of permissive/prohibitive tokens showed that in none of the varieties the year of attestation served as a predictor variable for the choice of form. Figure 103 displays the tree which is obtained when the effect of the year of attestation, the variety, the text type, the author and whether the form encodes permission or prohibition is tested by including the data of all three varieties. The tree consists of three highly significant splits. While the first one is based on the factor FEATURE ($p < 0.001^{***}$), indicating that different forms are preferred in prohibitive versus permissive contexts, the second level splits are based on the variety ($p < 0.001^{***}$). If a specific

¹⁵⁹ It may be possible that the grammaticalisation of the abilitative marker *enough long* started when *can* grammaticalised from an ability marker into the permissive marker.

year and concomitant with this a specific historical event had a significant impact on the diverging choice of permissive features, a split based on the year of attestation should have been attested. While TP shows a clear preference of the form *can*, the morphemes *can* and *save* represent competitive forms in SIP and BIS. No significant time split can be observed based on the data at hand.

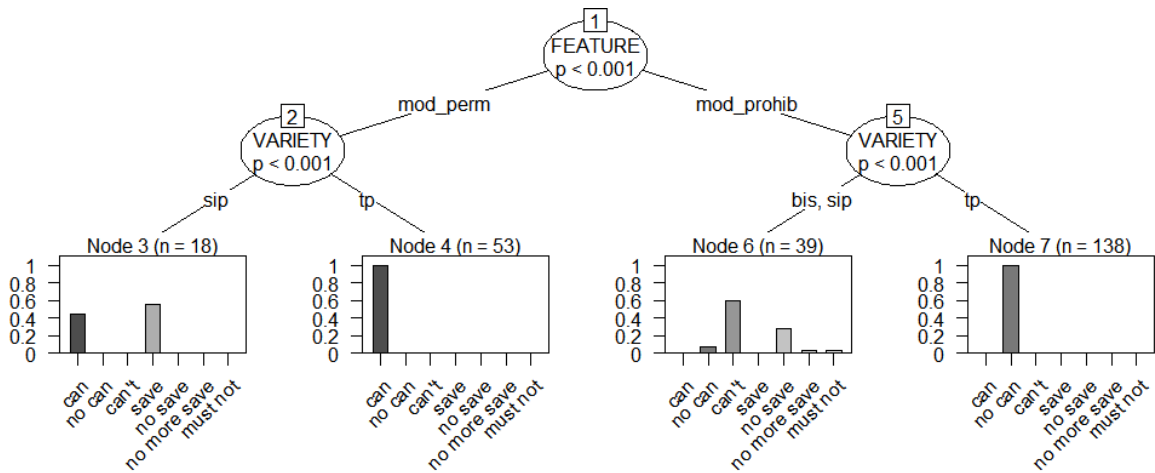


Figure 103: Conditional inference tree for permissive/prohibitive in the three MPE varieties $ST\_FORM \sim YEAR\_DET + VARIETY + TXT\_TYPE + AUTH\_NAME$

The number of tokens is too low to learn about the feature development and the stabilisation of permissive/prohibitive markers in SIP and BIS. Since contemporary SIP and BIS only make use of *save* and the data does not show a preference of this form yet, it can be assumed that the grammaticalisation was not completed by 1950. In TP the form *can* seems to have grammaticalised already by the late 1940s. A clear reason cannot be determined. It seems as if permissive/prohibitive auxiliaries only developed late in the histories of the varieties and that beforehand, imperatives and other means (such as, for instance, phrases with *he tambu*) were made use of. The results should not be overvalued as the frequencies of attestations were too low to make sustainable claims.



Figure 104: Map of Melanesian Islands, showing attestations of permission *M_MORPH* variants

8.7 Speculative

The last section in this chapter will focus on the three forms *ating*, *maet* and *mebi* which, though sometimes being translated as ‘perhaps’ or ‘maybe’, may be classified as markers used to encode speculative modality. Some initial remarks on the methodology will be outlined before the findings of the analysis will be introduced.

8.7.1 Methodological considerations

The morpheme-by-morpheme analysis showed that the early MPE varieties show differences in terms of the markers used to encode speculative modality. The coding is summarised in Table 18 and shows that the forms *I think*, *might*, *maybe* and *might be* were identified as variants.

Factor	Levels
ST_FORM	<i>I think</i> <i>maybe</i> <i>might</i> <i>might be</i>
FEATURE	mod_spec
STRUCTURE	clause-initial preverbal
YEAR_DET	1832 ... 1950

Table 18: Linguistic coding for speculation

Ating has grammaticalised out of the StE first-person epistemic parenthetical *I think*. As Thompson & Mulac (1991: 313) have shown, speakers’ confidence level in the likelihood of an utterance can be expressed with phrases such as *I think that*, *I guess that* and *I believe that*. Although they have their origin in “pure matrix clauses”, they can behave like modal adverbs (de

Haan 2006: 38). The grammaticalisation is said to be accompanied by *that*-deletion so that the forms can develop into full adverbs (cf. Thompson & Mulac 1991: 313). The fact that *I think* expresses a speaker's speculative confidence level may be the reason why it is attested in dubitative contexts in early MPE as well.

The identification of '*I think*'-strings which encode speculative modality was not without its hazards. Although there is evidence that the 1SG pronoun *me* and a form of *save* to express 'know, think' were already in use by the time the varieties were earliest attested, the 1SG pronoun *I* and the verb *think* were occasionally used, or at least attested, in the early years of MPE as well. Thus, occurrences of *I think* had to be analysed carefully in regard to their meaning to identify whether the two morphemes represent a modality marker/adverb or whether they encode the first-person singular pronoun and the verb *think*.

Europeans may have used the string of words with their original meaning and it is also easily conceivable that Pacific Islanders, who were constantly exposed to the lexifier language or that spoke an acrolectal form of MPE, may have used *I think* with its StE meaning. However, it is unlikely that speakers who had the 1SG pronoun *me* in their linguistic repertoire and used it in *all* instances in which a 1SG pronoun occurred, made use of *I think* with its StE meaning.¹⁶⁰ Instead, it is more reasonable to assume that they reanalysed the meaning of the string. *I think* may have been reinterpreted as a monomorphemic form used to express opinions that are less definite. To put it in Keesing's (1988: 102) words, although the surface strings were mutually intelligible, the "speakers of European and indigenous languages [may have] us[ed] different grammatical analyses". This assumption makes an analysis and clear assignment of *I think*-strings difficult.

I searched for further indicators to identify the meaning of the individual strings. The orthographic spelling in the original source may serve as one important indicator. When written as a single word, such as *ajtink*, I assumed that the form was no longer used as in StE. Another indicator was to look at how 1SG pronouns are generally encoded in the source document. If *I* occurs only in the *I think*-string but in all other cases the 1SG pronoun *me* is used, this is a clear indicator that *I think* has grammaticalised into an adverb/epistemic marker. For instance, in sentence (174) it is unlikely that the speaker used the 1SG pronoun *I* since the 1SG pronoun *me* is used not only in the same sentence, but also in the complete utterance of the speaker.¹⁶¹ In sentence (175), however, we can assume that *I think* is used as in StE as we find the phrase *I no*

¹⁶⁰ As intraspeaker variation is possible, all instances in which a 1SG pronoun occurred need to be considered. If a speaker uses both 1SG pronouns *I* and *me* interchangeably, it cannot be necessarily assumed that *I think* functioned as an epistemic marker.

¹⁶¹ There are eleven attestations of 1SG *me*, while *I* only occurs in the *I think* string.

think in the clause which directly precedes the sentence shown here.¹⁶² As the negative marker is inserted between *I* and *think*, the two items are not used as a monomorphemic form. Cases such as this and, for instance (176), had to be identified and excluded from the analysis.

- (174) **I think** *me shoot him master Kolbe.*
SPEC 1SG shoot TR master Kolbe
 ‘I might have shoot master Kolbe.’
 (Vunapope 1912; *Reichskolonialamt Records* 1912: 29 in Tyron & Charpentier 2004: 380)

- (175) **I think** *he plenty cross that schooner no takee-him come him friend!*
1SG think PM very cross that schooner no take-TR come TR friend
 ‘I think they were angry that the schooner did not return their friends (from the plantation).’
 (Malakula 1905; Grimshaw 1907: 273)

- (176) **I tink** *so. You got medicine belong smell?*
1SG think so 2SG have medicine PREP smell
 ‘I think so. Do you have amonia?’
 (Ambrym 1892; Lamb 1905: 165)

A further problem arises in utterances in which a European quotes a Pacific Islander and not himself, which implies that author modification is likely to come into play. The speaker may have used *I think* as a monomorphemic epistemic marker which the European with no linguistic affinity may have reinterpreted as StE *I think*. Thus, it is not always possible to identify whether the *I think*-string functioned as the 1SG pronoun + verb or whether it was already used as an epistemic adverb.

As pointed out above, 1SG pronouns were analysed to decide whether *I think* strings should be considered as *pronoun+verb* or as an *epistemic marker*. Nonetheless, the risk remains that tokens were excluded although in the ‘original’ speech act they were used as epistemic adverbs by the substrate speaker. Therefore, I choose to err on the side of caution.

In addition, *might* and *maybe* were identified as alternative forms to encode dubitative modality as exemplified in sentence (177) and (178). Both forms have their origin in Standard English. *Might*, for instance, is an auxiliary verb which is used in the lexifier language when speakers make a guess about something. In the early MPE data the form occurs almost exclusively in clause-initial position. For the analysis only those sentences were included in which *might* occurred together with a verb phrase (including zero-verbs), which means that sentences such as (179) were excluded. A fourth variant identified was *might be* as shown in sentence (180).

¹⁶² **I no think** *that fellow he make bad for misinari* (Malakula 1905; Grimshaw 1907: 273).

- (177) Suppose I no go ashore, **might** 'he come
 if 1SG NEG go ashore **SPEC** PM come
 'If I do not go ashore, she might come.'

(Malu'u 1904; Young 1925: 187)

- (178) Look out Missus. **Maybe** dey tink you debbil-debbil!
 Watch out missis **SPEC** 3PL think 2SG spirit
 'Watch out, missis. They may think you are a ghost/spirit!'

(Malakula 1928; Cheesman 1949: 111)

- (179) Me learn about Jesus, **might** one yam **might** two yam [...]
 1SG learn PREP Jesus **maybe** one year **maybe** two year [...]
 'I learned about Jesus for maybe one or two years.'

(Baunani 1911; Anonymous 1910-1911: 43; PMB Doc 439)

- (180) [...] me fright mary **might be** hit him me too.
 [...] 1SG afraid woman **SPEC** hit TR 1SG too
 'I was afraid (the) woman might hit me too.'

(Tulagi; Talatova 09.01.1923)

The data was further coded for STRUCTURE indicating whether the modality marker occurs in clause-initial or preverbal position.

8.7.2 Findings and discussion

In this section, the results of the analysis of the speculative markers *might*, *maybe*, *might be* and *I think* and their occurrence in the three MPE varieties will be presented. The section starts by providing an overview of the general data distribution before a closer look at the attestations of the markers and their development across time will be provided, starting with SIP in Section 8.7.2.2, continuing with BIS in Section 8.7.2.3 and focussing on TP in Section 8.7.2.4. The diachronic comparative analysis of speculative markers in Section 8.7.2.5 will bring together the results of the individual analyses.

8.7.2.1 General data distribution

The number of utterances in which speculative modality is overtly marked amounts to 180 tokens and, as such, the feature is rare in the early MPE data. The greatest number of tokens were found for SIP with 132 instances (= 73.33%). In contrast, in the early dataset of BIS only 20 (= 11.11%) and in the TP dataset 28 sentences (= 15.56%) were found. This already indicates that the frequencies are too low to make reliable judgments. Moreover, when focusing on the data distribution, it becomes noticeable that the datapoints do not spread equally across time:

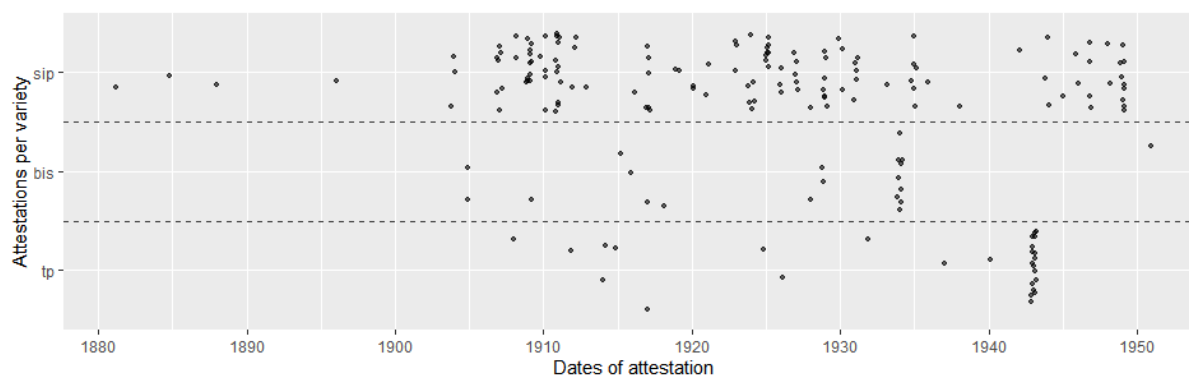


Figure 105: Distribution of speculative datapoints across time per variety

If the complete time period is considered as a single period only (see Figure 106) and the focus is placed on the frequencies of the individual forms, it can be observed that SIP shows a clear preference of *might* which was attested in 66.67% (= 88/132) of the tokens. *I think* occurs in 28.79% (= 38/132) of the datapoints and is the second most common form. *Maybe* is only attested in 3.03% (= 4/132) and *might be* in 1.52% (= 2/132) of the tokens. In the BIS data, three of the four forms occurred, whereby *might* and *I think* were attested as the dominant forms prevalent in 45% (= 9/20) and 40% (= 8/20) of the tokens. The form *maybe* was attested in 15% (= 3/20) of the speculative modality utterances. In contrast to SIP and BIS, the most frequently occurring form in the TP dataset is *I think* (25/28= 89.29%) and the forms *might* and *might be* were not attested at all. The form *maybe* occurred in 10.71% (= 3/28) of the tokens.¹⁶³

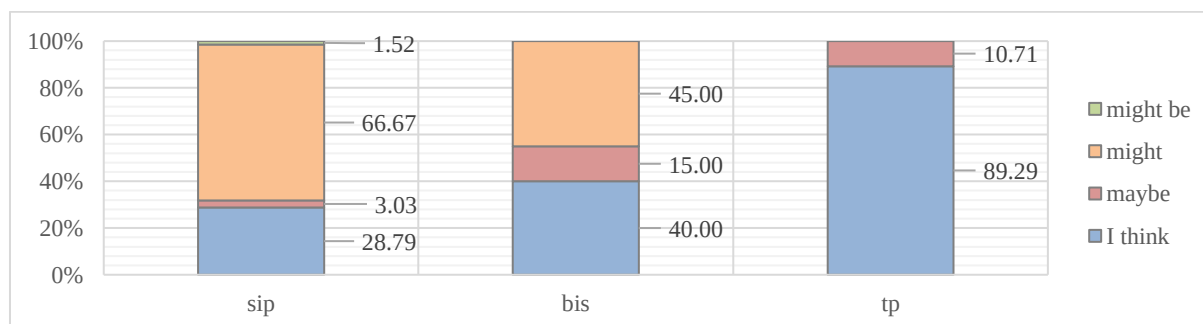


Figure 106: Relative frequencies of speculative variants per variety

8.7.2.2 Diachronic analysis of speculative in SIP

This section will focus on the results of the diachronic analysis of speculative particles in SIP. As in previous chapters, the attested forms will be exemplified before their dates of attestations will be considered. Finally, a look at the decision trees will show whether there is statistical rigor to claim that time is an important indicator for the choice of form.

¹⁶³ A map showing where the speculative markers were attested is displayed in Figure 117.

8.7.2.2.1 Attested forms

As Figure 106 showed, all variants of the variable ST_FORM *might*, *might be*, *maybe* and *I think* were attested in the early collected data of SIP. *Might* was only attested four times in preverbal position. In the remaining 84 cases it was fronted and preceded the subject-referencing pronoun, as exemplified in (181). The marker *I think* only occurred in clause-initial positions as well, as shown in (182). Sentence (180) and (183) show an early attestation of the forms *might be* and *maybe* to encode speculative modality in SIP.

- (181) [...]suppose me wait **might** me no ready when the Lord Jesus comes
 [...] if 1SG wait **SPEC** 1SG NEG ready when ART Lord Jesus come-3SG
 ‘If I wait, I may not be prepared when the Lord Jesus comes’
 (Malaita 1946; Deck 1946: 2; AU PMB DOC 442)

- (182) **I think** tomorrow you die
SPEC tomorrow 2SG die
 ‘You might die tomorrow.’
 (The Stories of the Crew; NIV June 1947: 8; AU PMB DOC 439)

- (183) No- dam-fear! **maybe** he true, but me no ride ‘im!
 NEG damn fear **SPEC** PM true but 1SG NEG ride TR
 ‘I have no fear! It might be real, but I will not ride on it.’
 (Sydney 1920-1926; Collinson 1926: 228)

8.7.2.2.2 Timeline and boxplot approach

The timeline in Figure 107 shows that the earliest attestation of the form *might* goes back to the 1880s. The form is attested until the end of the period under investigation. The form *I think* is already attested in 1881 for the first time and it appears throughout the observed period. In contrast, *might be* and *maybe* were only attested in the 1920s in the written data that provided the basis for the analysis.

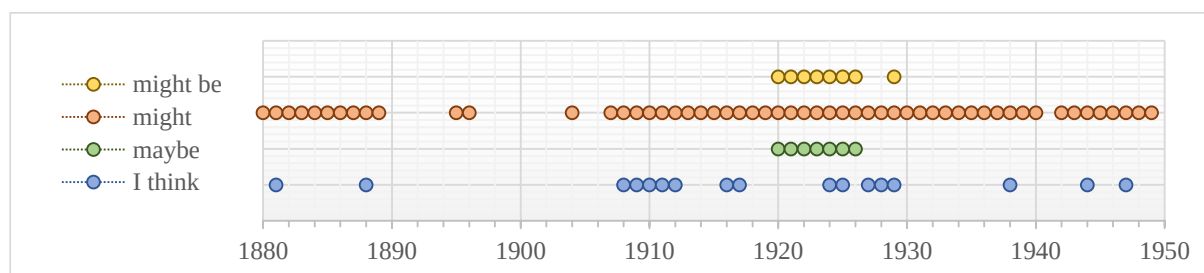


Figure 107: Timeline of speculative modality variants in SIP

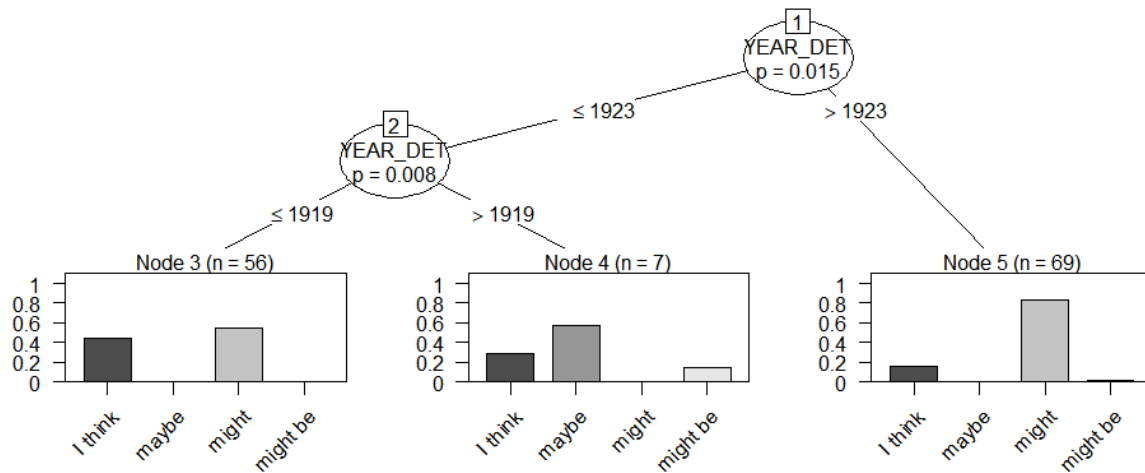
The boxplot results being displayed in Figure 108 reveal a similar result. The median of *I think* lies in the year 1911, whereas the median of *might* dates to the year 1926. Thus, although both forms are attested throughout the complete observed period, the bulk of *I think* forms were attested earlier than *might*.



Figure 108: Boxplot of speculative modality variants in SIP

8.7.2.2.3 Testing for the impact of the years of attestations on the choice of form

A ctree taking the year of attestation as a possible predictor variable for the choice of form in speculative contexts gives us a tree with two splits (see Figure 109). The first split, which is significant with $p=0.015^*$, indicates that the data behaves differently depending on whether it was attested before or after 1923.

Figure 109: Conditional inference tree for speculative in SIP $ST\_FORM \sim YEAR\_DET$

The second split shows that pre-1924 data shows further very significant differences depending on whether the forms were attested before 1920 or in the period 1920 until 1923. Utterances attested before 1920 are almost equiprobable to either encode speculative modality with the form *I think* or *might*. Tokens attested from 1920 until 1923 are, according to the tree, likely to use the forms *maybe*, *I think* and *might be*, with *maybe* being the most likely form (see Node 4). Interestingly, *might* is not attested in this period but the number of tokens is only seven. Especially as the data dating from 1924 onwards shows a clear preference of the form *might*, it seems reasonable to assume that the non-attestation of *might* in the years 1920-1923 is based on the peculiarity of specific texts or the availability of texts for that time.

Growing a ctree which takes the year of attestation ($YEAR\_DET$) and the position in which the speculative marker occurs ($STRUCTURE$) as possible predictor variables, results in the tree structure displayed in Figure 110. The split on the first level differentiates clause-initial

markers from preverbal ones ($p < 0.001^{***}$). In preverbal position the form *might* is most probable, but *I think* does not occur at all. In addition, two significant splits based on the year of attestation can be observed when speculative markers appear in clause-initial-position (cf. Nodes 2 and 3). Utterances attested before 1918 are almost equiprobable to encode dubitative modality in clause-initial position with the forms *I think* or *might* ($p = 0.027^*$). In the period 1918-1923 the dubitative is expressed equally likely with the forms *I think*, *might* and *maybe*. Clause-initial dubitative markers that appear after 1923 are most likely to encode dubitative modality with the marker *might* ($p = 0.014^*$). This result does not change if the text type is added as a further possible predictor variable (see Appendix II). Clause-initial markers dating to the year 1923 and belonging to the text category 1 are even likelier to use the marker *might*, than texts of category 2 and 3.

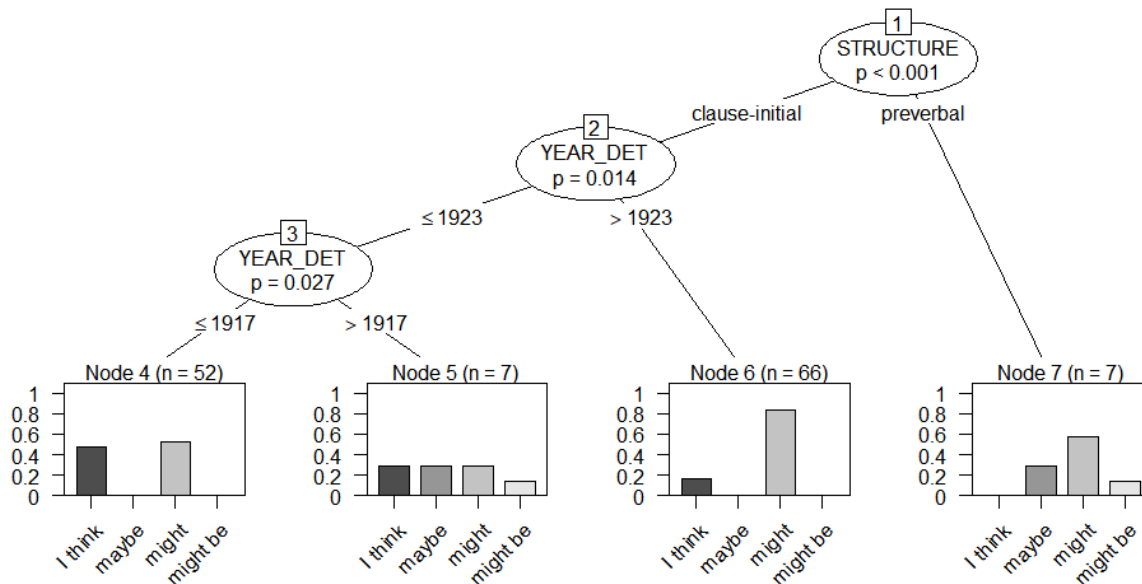


Figure 110: Conditional inference tree for speculative in SIP $ST\_FORM \sim YEAR\_DET + STRUCTURE$

However, if in addition the predictor variable *AUTH_NAME* is added to see whether the author has an impact on the form, the resulting ctree consists of three splits, whereby the highest-level split is based on the author (cf. Figure 111; $p < 0.001^{***}$). The second-level split is based on the year of attestation and evokes a split in 1928 ($p < 0.001^{***}$). It is interesting that only three authors, namely Eric Muspratt, Clifford Collinson and Arthur Morris Hocart are separated from the remaining 26 authors. Comparing the end nodes it becomes clear that the three authors are the only ones in showing a preference of the form *maybe*. As *n* equals eight in Node 7 in Figure 111, it can be assumed that the general tendency of the early SIP data is to preferably use *might* and *I think*, whereby the latter form is less frequently used. Thus, *might* represents the dominant form.

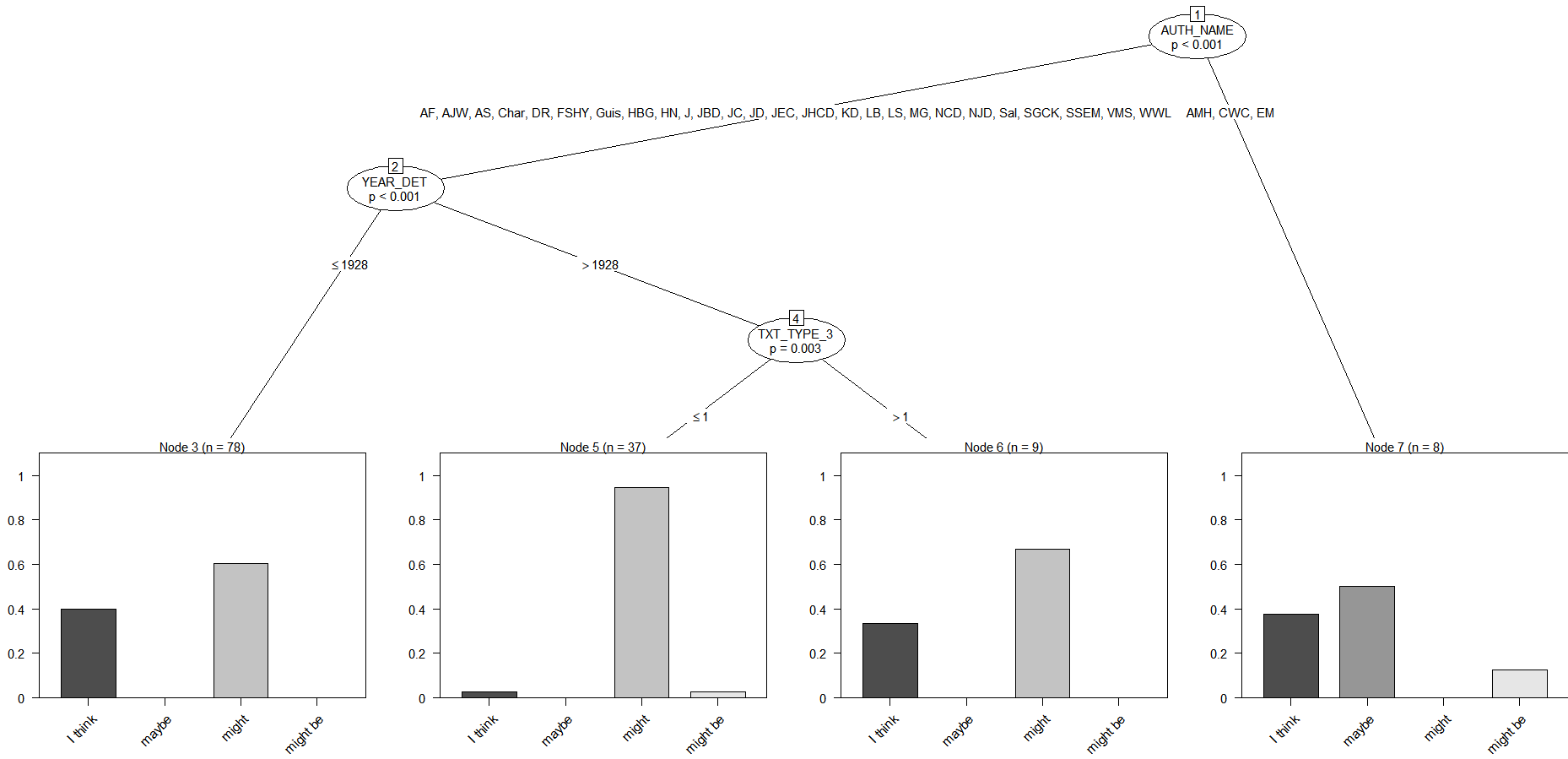


Figure 111: Conditional inference tree for speculative in SIP ST_FORM~YEAR_DET+STRUCTURE+TXT_TYPE_3+AUTH_NAME

8.7.2.3 Diachronic analysis of speculative in BIS

After having looked at SIP, the present section will focus on the development of speculative markers in BIS. First, the forms that were attested in the early data will be presented before their diachronic development will be traced with the help of a timeline, boxplot and ctree analysis.

8.7.2.3.1 Attested forms

In the early BIS data, three of the four forms to encode speculative modality were attested. Sentence (184)-(186) show that the attested forms are *might*, *maybe* and *I think*. All attested forms usually occur in clause-initial position with a single exception which is shown in (187), in which the marker *maybe* is used in preverbal position.

- (184) *Man* 'e no can savvy **Might** you me catch 'im quick time.
 one PM NEG ABIL know **SPEC** 1PL.INCL catch TR quickly
 'One cannot know. Maybe we will reach there quickly.'

(Espiritu Santo 1933-35; Harrison 1937: 146)

- (185) **Maybe** dey tink you debbil-debbil!
SPEC 3PL think 2SG ghost
 'They may think you are a spirit.'

(Malakula 1928; Cheesman 1949: 111)

- (186) **I think** all he lose long saltwater
SPEC 3PL PM lose PREP sea
 'They may have got lost in the sea.'

(Malo Pass 1915; Hawkesby 20.01.1915)

- (187) *Byme by* white marster **maybe** go along place [...]
 FUT white master **SPEC** go PREP place [...]
 'Soon the white master may go to the place ...'

(Malakula 1929; Shurcliff 1930: 167)

8.7.2.3.2 Timeline and boxplot approach

Figure 112 shows the dates of attestations of the three markers in BIS. The earliest form which was attested in 1905 is *I think*. The form seems to persist until the end of the observed period with a datapoint in 1951. The marker *might* is first attested in 1907 in the BIS data. There is also an attestation in the 1930s but no later attestations are found in the collected data.

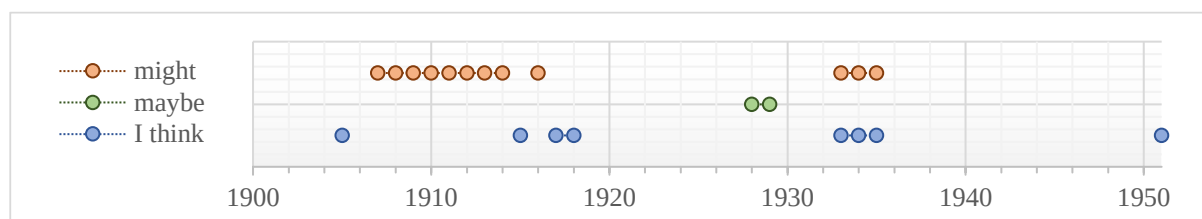


Figure 112: Timeline speculative modality variants in BIS

However, as there is almost no material that covers the time 1940 until 1950, no claims can be made in regard to the sudden absence of the feature. *Maybe* is only attested in 1928 and 1929.

The boxplot displayed in Figure 113, in which periods were replaced by concrete determined years of attestations, shows a similar result. However, the median of *I think* tokens dates to 1918, whereas the form *might* has its median in 1934.

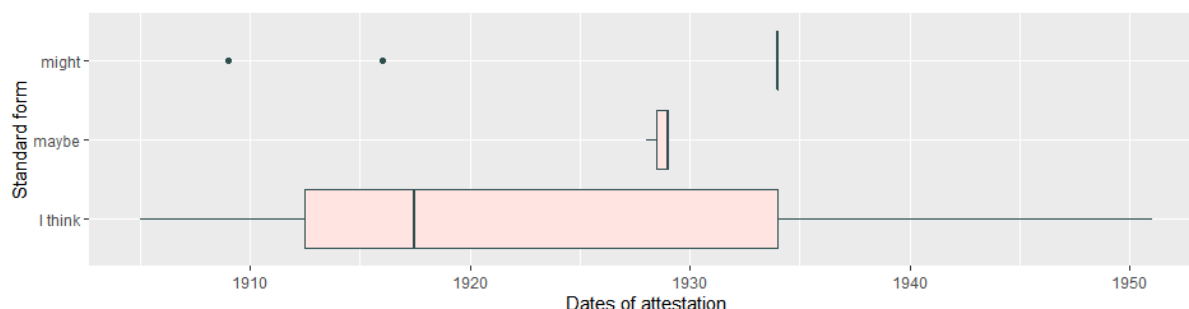


Figure 113: Boxplot of speculative modality variants in BIS

8.7.2.3 Testing for the impact of the years of attestations on the choice of form

Independent of whether only the year of attestation was considered as a possible predictor variable or whether it was combined with the position, the text type or author, no splits were obtained. This shows that in the small number of tokens obtained for BIS, no time splits are observed. Instead, the form *might* represents the dominant attested form.

8.7.2.4 Diachronic analysis of speculative in TP

The present section will demarcate how early TP encoded speculative modality. After introducing the attested forms, their development across time will be in focus.

8.7.2.4.1 Attested forms

In early TP only two of the three forms were attested. As sentence (188) and (189) show, *maybe* and *I think* represent the markers used to encode speculative modality. Both were attested in clause-initial positions only.

- (188) *majbi* *em* *i-daj,* *majbi* *em* *i-stap,*
 SPEC 3SG PM-die SPEC 3SG PM-stay
 ‘Maybe it dies, maybe it stays./ It may die or it may stay.’

(Sepik ~1943; Hall 1943: 48)

- (189) *Masta* *me* *no* *savee* *steer-im good* *I think* *wheel* *he* *no* *strong* [...]
 master 1SG NEG ABIL steer-TR good SPEC wheel PM NEG strong [...]
 ‘Master, I am unable to steer, the wheel might be rotted.’

(Rabaul 1925; Groves 1925: 6; AU PMB MS 612)

8.7.2.4.2 Timeline and boxplot approach

The timeline of speculative modality markers displayed in Figure 114 shows that the form *I think* is attested for the first time in speculative contexts in 1908. It spreads throughout the complete period. The form *maybe* dates to 1917 and 1943. Although it was only attested in two years, the form might have been used by some speakers.

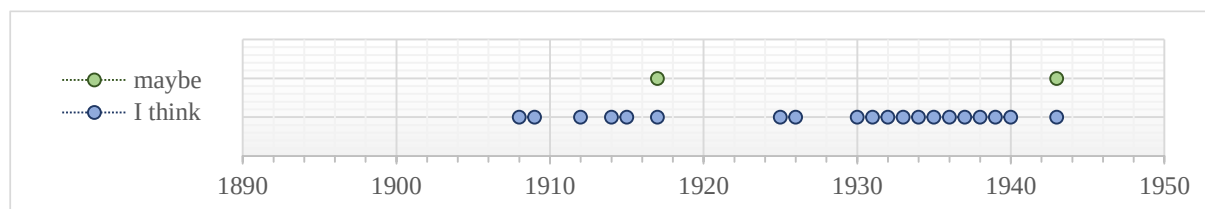


Figure 114: Timeline of speculation modality variants in TP

Looking at Figure 115 which shows the boxplots of the two forms, it can be seen that for both forms the mean is in 1943. The figure is misleading as it appears as if *maybe* occurred throughout the period with high frequencies.

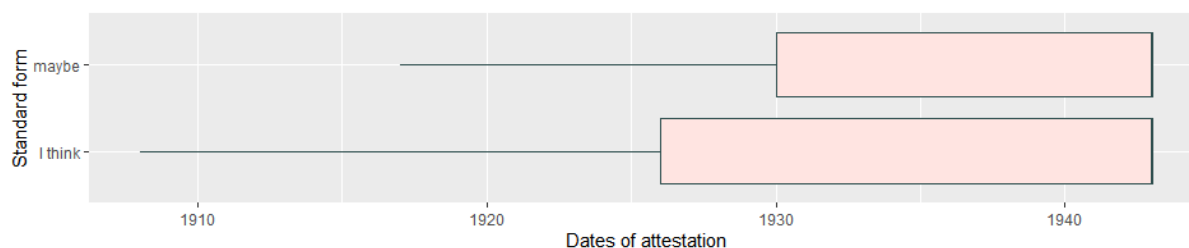


Figure 115: Boxplot of speculation modality variants in TP

8.7.2.4.3 Testing for the impact of the years of attestations on the choice of form

Growing ctrees by including one or several of the predictor variables results in a single node with no splits. Thus, none of the predictor variables has an impact on the choice of form used to encode speculative modality. *I think* represents the dominant form in TP while the form *might*, which is attested in SIP and BIS, does not occur at all.

8.7.2.5 Diachronic comparative analysis of speculative in MPE

If the encoding of speculative forms is analysed by taking together the tokens obtained for all three varieties considering the position in which the marker occurs, the text type and the variety as possible predictor variables, the following tree is obtained:

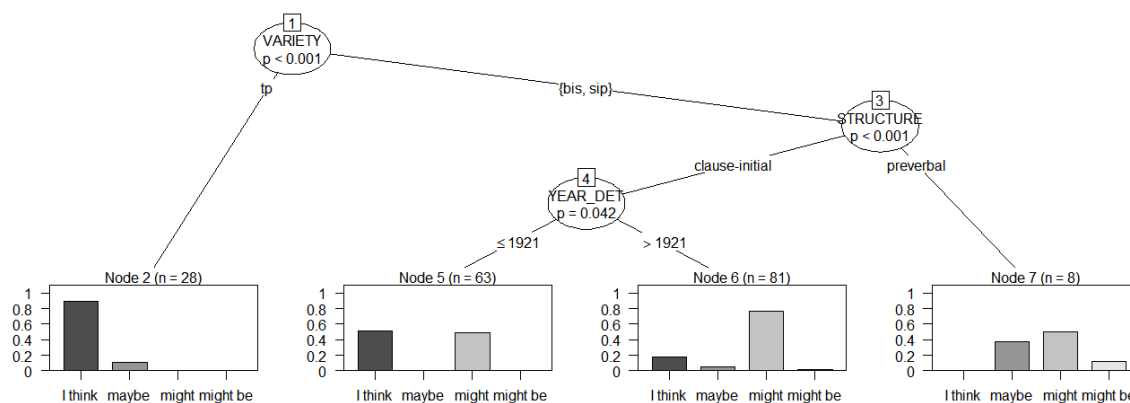


Figure 116: Conditional inference tree for speculative in the three MPE varieties $ST\_FORM \sim YEAR\_DET + VARIETY + STRUCTURE + TXT\_TYPE\_3$

The tree consists of three splits on three levels. The primary highly significant split is based on the variety and indicates that TP differs from the other two varieties in the choice of the speculative marker, independent of the year of attestation. In the small amount of attestations available for TP, a clear preference of *I think* is attested. The second split is based on the position in which the marker occurs (STRUCTURE; $p < 0.001^{***}$) and the third one on the year of attestation (YEAR_DET; $p = 0.005^{**}$). The varieties SIP and BIS seem to be very similar in the form they use in dubitative contexts. According to the ctree, clause-initial dubitative markers dating before 1921 are almost equiprobable in being encoded through *I think* or *might*, whereas post-1921 attestations are more likely to encode dubitative modality with the form *might*. Differences between TP and the two varieties seem to have existed right from the varieties' beginnings, considering that in TP there are no attestations of *might*. If the author is added into the analysis as a further possible predictor variable, the resulting tree looks similar.¹⁶⁴ The first and the third split remain the same; however, the second split is no longer based on the position but on the author (AUTH_NAME; $p < 0.001^{***}$). Four authors are separated by the split as they use the forms *maybe* and *might be*.

As a difference in the use of the speculative marker was observed between TP and the other two varieties independent of the time variable, the differences may be based on events that occurred prior to the end of the labour trade. The fact that the marker *might* is attested in SIP and in BIS but not in TP is of particular interest.

A closer look at early QPPE and SPPE material was taken to investigate whether clause initial *might* was used in the plantation varieties that are said to be the precursors of the MPE varieties. While I found an instance of clause-initial *might* to encode uncertainty in the QLD data (cf. (190)), the marker could not be attested in SPPE. Of course, the absence is no evidence that

¹⁶⁴ See https://www.dropbox.com/s/0xbb8jq7qfp049x/Ctree_Speculative_MPE.png?dl=0 (last access 29 September 2021).

the feature was not used in SPPE due to the fact that the variety is only scarcely documented. There is clear evidence that *might* was used on the plantations in Queensland. However, while Baker (1993) came to the conclusion that *might* ‘perhaps’ has its origin on QLD plantations (first attestation in 1906) and spread to Vanuatu (first attestation in 1914) and the Solomon Islands (first attestation 1937), the results of the present data may suggest a different scenario. In my sample *might* is first attested in the pidgin used by the recruiter John Cromar on board of the Queensland labour recruiting vessel *Fearless* in the Malaita province (cf. sentence (191)) and later in the pidgin used by the Solomon Islander Peter on Malu’u (cf. Young 1925: 150). Peter was one of the first converts of the QKM who had returned to the Solomon Islands in 1894 to spread the word of God. The earliest attestations of *might* in my data can thus be traced to the Solomon Islands. Therefore, there is the possibility that the feature spread from the Solomon Islands to Queensland and Vanuatu. However, Cromar worked on a Queensland recruiting vessel and Peter had first come into contact with Pidgin English while serving on the plantations in Queensland. It is thus still possible that the *might* marker originated in the pidgin as spoken in Queensland.

- (190) *B'mbi* **might** *you* *catch'em!*
 by.and.by *SPEC* 2SG catch-TR
 ‘You might catch it soon.’

(Queensland 1909; Banfield 1913: 144)

- (191) **Might** boat's crew shoot' em.
 SPEC boat's crew shoot 3SG/TR
 ‘Perhaps the boat's crew shoot him.’/The boat's crew may have shot him.’

(Ata'a 1880s; Cromar 1935: 316)

Again, it needs to be pointed out that the analysis in the present section is based on a very small amount of data which makes it difficult to apply statistical methods and to make generalisable claims.

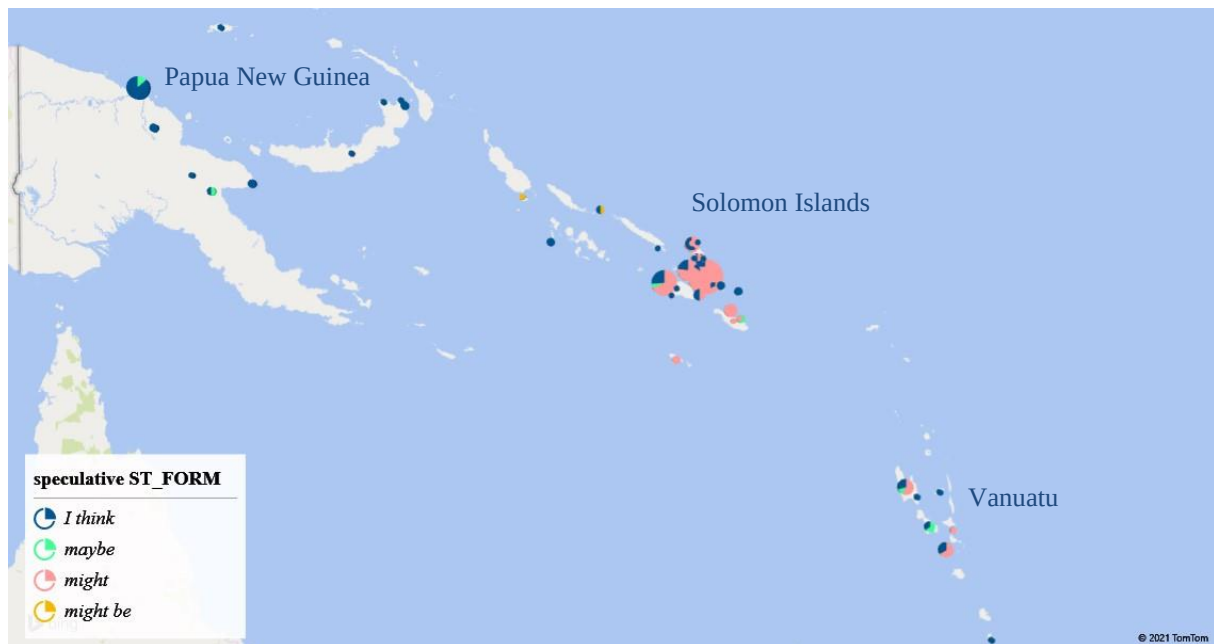


Figure 117: Map of Melanesian Islands, showing attestations of speculative *M_MORPH* variants

8.8 Summary

The present chapter discussed the diachronic development of modality in SIP, BIS and TP. It started with an introduction of the theoretical background (8.1) before it provided insight into how contemporary SIP, BIS and TP encode *volitive*, *abilitative*, *permissive* and *speculative* modality (8.2). Previous diachronic research was reviewed in Section 8.3. The subsequent Sections 8.4 - 8.7 focussed on volition, abilitative, permissive and speculative modality respectively. Each chapter started with some methodological considerations before the analysis of attested forms and their change across time was presented.

The individual results of the diachronic analysis of *volitive*, *abilitative*, *permissive* and *speculative* modality showed that one would wish for more pre-1950 data to have survived in the written historical records. The general scarcity of data availability and the uneven distribution of datapoints across time make it difficult to trace the development of modality markers in the three MPE varieties. Nonetheless, some important observations could be made:

In regard to *volition markers*, differences seem to have their origin in the plantation history rather than in the end of the labour trade. The study provided evidence that TP showed a preference of the form *like* throughout its development. *Want to*, which was predominantly attested in early SIP and BIS, was only scarcely attested in TP. As both forms based on the morphemes *like* and *want* were attested to have been in use on QLD plantations but on Samoan plantations only instances of the former could be found, this may be an indicator that the forms have their origin in the different plantation histories. However, based on the data at hand, it

became visible that the stabilisation of the contemporary forms had not been completed in SIP and BIS by the end of the observed period (= 1950).

For ability markers the amount of available datapoints was too low to make final claims. In SIP, the variety which is best documented in the present dataset, the grammaticalisation process does not seem to have been completed by the end of the 1940s. A great amount of variation was observed. What is interesting, however, is that in non-negated contexts, the occurrence of *save* was attested to be slightly higher than the use of *can*. In negated contexts, forms based on *can* represented the dominant option. The coexistence of forms such as *no save*, *cannot* and *can't* in negated contexts seemed at first glance striking. However, it should not be forgotten that contemporary SIP shows exactly the same degree of variation with the forms *no save*, *kanduit*, *kan* and *kanot* (the first representing the dominant form). In the BIS data, the form *save* dominated and, based on the data at hand, a change in the attestations of *can* and *save* became apparent in 1897. However, the data gap in the years 1935-1950 makes an interpretation of the results almost impossible. The most interesting developments could be observed in the TP data. The ctree algorithm showed a significant split in the year 1934. While in and before 1934 the data shows a higher chance of the form *save* than of *can* to encode ability, in post-1934 data a greater likelihood of *can* is observed. In addition, forms based on the morpheme *enough*, which has grammaticalised into the marker used in contemporary TP, occur for the first time in the written data around 1940.

The data that was available for the analysis of permission markers was not sufficient to make claims about their development in the three varieties. It seems as if permissive markers developed rather late, although it needs to be kept in mind that the non-attestation does not necessarily mean that markers were not in use before the here documented first written attestations. The permissive attestations cannot be used to answer the question whether the end of the labour trade was responsible for the diverging permission features as attestations only start in 1905. A possible reason why SIP and BIS develop in a similar way and different from TP may be that abilitative and permissive modality are closely connected and that permissive forms frequently grammaticalise out of ability markers.

In regard to the speculative marker the most interesting observation was that while all varieties made use of the *I think*-string, the form *might* was only attested in SIP and BIS. The analysis showed that *might* was first attested in the Solomon Islands in the PE used by a Queensland recruiter and later in the PE used by a QKM convert. Since the marker was attested in interactions between Queensland recruiters and Pacific Islanders, it is still likely that it spread to Vanuatu, where the marker was first attested in 1907, via Queensland. This would also explain why it was not attested in the TP dataset.

9 Case Study: Selected prepositions

The possessive or genitive is expressed by means of the word ‘belong’ [...] All other cases are indicated by means of the preposition ‘along’.

(Landtmann 1927: 454)

In early descriptions of the three MPE varieties, prepositions are frequently described as *simple*. While some early writers claim that PE only disposes of a single preposition that has its origin in the English word *belong*, other authors limit the number of prepositions to two, namely *belong* and *along*. Neuhauss (1911: 121) argues, for instance, that words are “juxtaposed without grammatical sentence formation with the help of ‘belong’”,¹⁶⁵ and Imhaus (1890: 39) claims that prepositional forms have derived from the English word *belong*. In contrast, Reed (1943: 281) claims that “[p]repositions are limited to two in number: *along* (or *long*) and *bilong*, which mean ‘to,’ ‘at,’ ‘toward,’ ‘on,’ ‘in,’ ‘as,’ ‘of,’ ‘from,’ and ‘with’ according to the context”. While Reed does not suggest a functional distinction between the use of the two prepositional forms, there are also early writers that assign distinct functional properties to both forms. The morpheme *belong* is described as fulfilling a possessive and genitive function, whereas *along* is described as used in all other cases. This view is *inter alia* found in Lambert (1942: 91) who, referring to 1921, claims that:

[t]he frequent use of “belong” (or “belonga”) is confusing, and “along” is worse. Loosely speaking, “belong” is possessive. “Knife belong me” is “my knife” and about the only way to translate “Dis fellow knifie belong dis fellow mary belong house belong Keop” would be “The knife of the native woman who lives in the Captain’s house” – a pretty clumsy way of making your point. “Along” generally expresses movement or approach: “Ship stop along place”.

A similar functional distinction is identified by Landtmann (1927: 454) who nevertheless claims that other prepositions than *belong* and *long* exist as well, such as *out*, *inside*, *up* and *down* which would often not occur in isolation “but together with ‘along’”.

My initial morpheme-by-morpheme analysis supported Landtmann’s claim that even though *belong* and *along* are the dominant prepositional forms, they do not represent the only prepositions in the early MPE varieties. Neither the generalised functional classification nor the claim of the existence of only two prepositions is justified in the pidgin language varieties under investigation.

In addition, some of today’s available grammars show that the varieties have developed prepositions in addition to *belong* and *long* and that there are more complex rules than a dipartite functional classification. For instance, Bislama makes use of the prepositions *long*, *blong*, *from*, *wetem* and *olsem* and of complex prepositions, which are forms that are built with the preposition

¹⁶⁵ “[...] hier werden tatsächlich die Worte ohne grammatische Satzbildung, meist mit Hilfe von ‘belong’ lose nebeneinander gestellt” (Neuhauss 1911: 121).

long to encode more complex meanings, such as *ananit long* or *aotsaed long*. In addition, verbal prepositions such as *kasem*, *bitim*, *ronem*, *agensem*, *raonem*, and *folem* exist as well. Tok Pisin is described as having four main prepositions *long*, *blong*, *wantaim*, *olsem*, as well as complex prepositions such as *antap long* and *insait long*. Moreover, prepositional meanings can be encoded with the help of serial verbs. Solomon Islands Pijin is said to have the prepositions *long*, *blong*, *fo*, *olsem*, *from* and *widim*, as well as adverbs functioning as prepositions and verbal prepositions. Differences in the prepositional forms that are used in SIP, BIS and TP to encode specific semantic functions have developed over time. Thus, the purpose of the present chapter is to trace back this development.

As the focus of the general study is on *when* the varieties developed diverging features, this chapter will be restricted to prepositions for which differences between the varieties were observed during the morpheme-by-morpheme analysis. The focus will be on selected semantic functions that prepositions can fulfil, which is why the present chapter starts with a theoretical background on prepositions (Section 9.1). Section 9.2 will give an overview over previous diachronic studies on the development of prepositions in the three varieties. Based on the theoretical background, some general remarks regarding the extraction and classification of the prepositional datapoints will be made in Section 9.3. The focus of the subsequent analyses will be on the following semantic prepositional functions: *comitative* and *instrumental* (9.4.1), *terminative* (9.4.2), and *adessive* (9.4.3). Each section will be accompanied by some theoretical background and methodological considerations, before the findings will be presented. Concluding remarks will summarise the results of the study at the end of this chapter.

9.1 Theoretical background

Prepositions and postpositions together with ambipositions and circumpositions form the category of *adpositions*, which can be defined as “grammatical tools which mark the relationship between two parts of a sentence”, whereby one of these parts represents “the element which an adposition governs” (Hagège 2010: 1). While prepositions “precede the word which they ‘govern’” and postpositions follow it, ambipositions may either precede or follow it (Chalker & Weiner 1994: 310). Circumpositions are adpositions which consist of two parts, whereby one part precedes the word governed and the other part follows it (cf. Hagège 2010: 115).¹⁶⁶

¹⁶⁶ Even if Hagège does not explicitly mention circumpositions in his brief initial definition of adpositions (cf. 2010: 8), he does not exclude them but considers them to be “less frequent possibilities” or a “rather uncommon phenomenon” (Hagège 2010: 115).

Even though prepositions do not exist in all languages around the world, they are a very common phenomenon. They represent a “fundamental part of speech” in that they enable to link lexemes that form part of a sentence in order to build “dependency relationships” (Hagège 2010: 5). Their existence in pidgins and creoles, however, has been highly debated for a long time. Some creolists, such as Bickerton, claim that pidgins do not possess “a full range of [...] prepositions”¹⁶⁷ (2016: 109), but that verbs would undertake the functionalities of adpositions instead. With this, Bickerton refers to serial verb constructions (SVC) in which several verbs “act together as a single predicate, without any overt marker of coordination, subordination, or syntactic dependency of any other sort” (Aikhenvald 2006: 1). For example, in a sentence such as (192), the SVC *take knife cut bread* encodes an instrumental meaning so that no overt adpositional marker is required.

- (192) *He take knife cut bread.*
 3SG take knife cut bread
 ‘He cut the bread with a knife.’

Consequently, Bickerton and other creolists supporting this view assume that the fewer adpositions there are, the higher the number of serial verbs in the creole variety (Bickerton 2016: 117). Creolists who hold contrary views have provided counterevidence showing that several creoles use prepositions alongside serial verb constructions (cf., for instance, Muysken 1988: 296).¹⁶⁸

Nonetheless, there is only a small number of studies available that focuses on prepositions in pidgins and creoles, or on prepositions from a typological perspective. The most extensive typological study has been conducted by Hagège (2010) who analysed adpositions from a syntactic as well as semantic point of view. The latter is in focus of the present chapter which aims at identifying *when* specific forms developed in the three varieties to encode distinct semantic functions.

The identification of cross-linguistic semantic functions of prepositions is not an easy task since idiomatic phrases as well as polysemy (implicating the blurring of local, temporal and other semantic categories) exacerbate a clear classification of prepositions according to their semantic function (cf. Hagège 2010: 277). Nonetheless, Hagège tries to distinguish the *major meanings* which adpositions encode cross-linguistically and identified three major domains, namely, the domain of *core-meanings*, the domain of *spatio-temporal* meanings and the domain of *non-*

¹⁶⁷ It should be noted that it remains unclear what *full* means.

¹⁶⁸ Further research needs to be conducted in order to analyse whether there is a correlation between the number of adpositions and the number of serial verb constructions from a typological perspective. Unfortunately, WALS does not have a contribution on SVCs so far.

spatio-temporal meanings (Hagège 2010: 258, 261-262). Each of the three domains consists of further subsets of functions so that at least 51 semantic functions of adpositions can be differentiated, which are reduplicated in Table 19.

The domain of core meanings can also be referred to as the domain of *actancy* meanings because the group “concern[s] the grammatical functions of subject, object, indirect object, and adnominal complement” and thus contains adpositions which mark agentive, patientive, attributive and possessive meaning (cf. Hagège 2010: 273). In contrast, the domain of non-core meanings can be referred to as the domain of *circonstants* or *circumstantials* and equals the English category of adverbials (cf. Hagège 2010: 273).¹⁶⁹

Semantic domains	Names of semantic functions	English Adp as examples
I. CORE MEANINGS	(1) agentive	<i>no Adp in English</i>
	(2) patientive	<i>no Adp in English</i>
	(3) attributive	<i>to</i>
	(4) possessive	<i>'s, of</i>
II. NON-CORE MEANINGS		
A. SPATIO-TEMPORAL		
(a) Static	(5) inessive: - spatial	<i>in, within</i>
	- temporal	<i>in, on, at, as of</i>
	(6) adpudessive	<i>at (X)'s</i>
	(7) adessive	<i>at, by, beside, near</i>
	(8) abessive: - spatial	<i>out of, from, beside</i>
	- temporal	<i>since, as early as</i>
	(9) obessive	<i>in front of, opposite</i>
	(10) suessive	<i>on</i>
	(11) superessive	<i>above, over</i>
	(12) subessive	<i>under, below, down</i>
	(13) preessive: - spatial	<i>before</i>
	- temporal	<i>before, ago, pending</i>
	(14) possessive: - spatial	<i>behind, after, beyond</i>
	- temporal	<i>after, in, within</i>
	(15) circumessive	<i>around</i>
	(16) medioessive	<i>among</i>
	(17) interessive	<i>between</i>
(b) Non-static	(18) illative	<i>to, into</i>
	(19) allative	<i>towards</i>
	(20) terminative: - spatial	<i>down to, as far as</i>
	- temporal	<i>till</i>
	(21) ablative	<i>from</i>
	(22) obversive	<i>in front of</i>
	(23) surlative	<i>onto, above, over</i>
	(24) superversive	<i>above, over</i>
	(25) sublative	<i>under, below, down</i>
	(26) prelative	<i>before</i>
	(27) postlative	<i>behind, after, beyond</i>
	(28) circumlative	<i>around</i>
	(29) mediolative	<i>among</i>
	(30) interlative	<i>between</i>
	(31) perlative: - spatial	<i>across, through, via</i>
	- temporal	<i>for, during, in</i>
	(32) prolative	<i>past</i>
	(33) secutive	<i>with, along, following</i>

¹⁶⁹ According to Hagège (2010: 273), the term *adverbial* is “confusing and less accurate” since it refers to a category and not the function and since it does not inform about the “semantic content of Adp-phrases and case-marked complements”. This is why he prefers the term *circumstantials*.

B. NON-SPATIO-TEMPORAL		
	(34) proprietive	<i>having</i>
	(35) exclusive	<i>without</i>
	(36) exceptive	<i>except, save, bar(ing)</i>
	(37) comitative	<i>with, along with</i>
	(38) instrumentive	<i>with, by, through</i>
	(39) mediative	<i>in the manner of, à la</i>
	(40) motivative	<i>because of, for, thanks to</i>
	(41) concessive	<i>despite</i>
	(42) comparative	
	- of equality	
	- equative	<i>as</i>
	- assimilative	<i>like</i>
	- of inequality	<i>than</i>
	(43) essive, translative, mutative	<i>qua, as, off, out, to</i>
	(44) purposive	<i>for</i>
	(45) adversative	<i>against</i>
	(46) pertentive	<i>about, with respect to</i>
	(47) roborative	<i>according to, depending on, following</i>
	(48) adnumerative	<i>in proportion to</i>
	(49) additive	<i>in addition to, besides</i>
	(50) substitutive	<i>instead of</i>
	(51) hypothetical	<i>in case of</i>

Table 19: The semantic system of adpositions (Hagège 2010: 261-262)

Focussing on non-core meanings, the *spatio-temporal* domain includes adpositions that encode the notions of space and place and may or may not involve some movement. Thus, they can be further classified into static versus non-static (Hagège 2010: 285). While most semantic functions have static and non-static counterparts, *terminative*, *perlative*, *prolative* and *secutive* are only non-static. There are also non-spatio and non-temporal relational and notional meanings that adpositions can encode, which are referred to as *non-spatio temporal* adpositions by Hagège (2010: 273).

Several remarks are necessary regarding the various semantic functions listed in Table 19. It is important to note that languages around the world will not necessarily encode the listed semantic functions with the help of adpositions. Other means, such as affixes or other lexico-grammatical markers may be used instead (cf. Hagège 2010: 259). Cross-linguistically, adpositions are less common than affixes to encode core meanings but they are more likely than affixes to encode the non-core meanings (Hagège 2010: 274).¹⁷⁰

It needs to be further pointed out that the listed semantic functions are not exclusive. Due to the numerous amounts of semantic contents which adpositions can encode, a comprehensive, all-encompassing cross-linguistic classification of all semantic functions of adpositions remains

¹⁷⁰ In WALS, Dryer (2013) surveys the order of adpositions and noun phrases in 1,184 languages. Only 30 languages (2.53%) do not use adpositions, which means that the remaining 1,154 languages use either postpositions, prepositions, inpositions or a combination of the three mentioned adpositional types. Since Dryer (2013) does not differentiate between core and non-core meanings, a direct comparison with Hagège (2010) is not possible.

a “Sisyphean task” (2010: 282) as “there are virtually as many roles as there are possible different participants in a state of affairs” (Luraghi 2003: 17). Moreover, the semantics of prepositions is dependent on the context in which they occur. Therefore, prepositional forms may be able to encode diverging semantic functions depending on the context.

In the present chapter, I will only focus on a selected number of prepositions, or rather, semantic functions, for which the morpheme-by-morpheme analysis provided evidence that the three varieties developed diverging forms. These include the *comitative*, *instrumental*, *terminative* and *adessive* semantic functions.¹⁷¹ While *comitatives* and *instrumentals* belong to the non-spatio-temporal domain, *terminatives* belong to the non-static and *adessives* to the static spatio-temporal domain in Hagège’s model (2010). Before the methodological steps will be further outlined, previous research on adpositions in MPE will be briefly introduced.

9.2 Previous diachronic studies on prepositions in MPE

Since studies on prepositions are generally neglected, it is not surprising that research focussing on prepositions in MPE is rare as well. Usually, prepositions are only addressed in grammars of the varieties. Thus, Beimers’ (2009: 99-105) grammar on SIP contains a detailed section on the variety’s prepositions and their multiple purposes, including simple, complex, adverbial and verbal prepositions. Similarly, Crowley (2004: 127-139) discusses simple, complex and verbal prepositions in his *Bislama Reference Grammar*¹⁷² and in Verhaar’s reference grammar of TP (1995: 235-252) or Mühlhäusler’s (1985c: 366-368) *Syntax of Tok Pisin* the prepositional phrase in Tok Pisin is addressed.

Studies exclusively focussing on prepositions in the MPE varieties are rare but the study by Kurzon (2002) on the preposition *long* in Bislama or the study by Lee (1996) on prepositions in Solomon Islands Pijin are examples. The latter, although being based on contemporary data of SIP, inter alia discusses possible sources of the preposition *fo* in Pijin. However, as “[e]vidence from earlier stages of Melanesian Pidgin is not readily available to” Lee (1996: 388), a comparative analysis of early forms is absent. He assumes that Pijin borrowed *for* as a dative form and extended its use to encode purposive and infinitival constructions “as part of a universal tendency which both Bislama and Tok Pisin have resisted in the retention of *long* for dative uses and *blong* for the purposive use” (Lee 1996: 390).

¹⁷¹ It needs to be noted that diverging forms were also observed in the encoding of other semantic prepositional categories but could not be covered in the scope of this work. I selected the *comitative*, *instrumental*, *terminative* and *adessive* semantic functions since these allowed a ‘relatively’ unambiguous classification. Further diachronic studies on differences in prepositional forms are in progress (cf. Schäfer in progress a and b).

¹⁷² See also Crowley’s (1990b) article *Serial verbs and prepositions in Bislama*.

The most extensive study available for Bislama demonstrates the development of prepositions from a diachronic perspective. Crowley (1990a), tracing the development from early Beach-la-Mar to modern Bislama, shows that early Beach-la-Mar (1840s-1860s) had only a small number of overt prepositions, which were *for* ‘benefactive’, *belong* ‘possessive, origin, oblique’, *with* ‘instrumental’, *about* ‘about’ and *all the same* ‘like, as’ (Crowley 1990a: 194), the latter being a world-wide feature of Pidgin Englishes in the 19th century (cf., for instance, Baker & Huber 2001: 117). Comparing the semantic functions which the prepositions fulfilled in this early period with the prepositions used in modern BIS, he shows that *belong*, *all the same* and *with* have their origin in these early days (Crowley 1990a: 195-197). Moreover, he exemplifies how the semantic functions, which the prepositions are used for, have been redistributed. A summary of his results can be found in Table 20, which is adapted from Crowley (1990a: 266).

	1840s-60s	1870s-90s	1900-30s	1940-
Poss	<i>blong/blonga</i>	<i>blong/blonga</i>	<i>blong</i>	<i>blong/blo</i>
P/wh		<i>blong/blonga</i>	<i>blong</i>	<i>blong/blo</i>
Or	<i>blong/blonga</i>	<i>blong/blonga</i>	<i>blong</i>	<i>blong/blo</i>
Purp		<i>blong/blonga</i>	<i>blong</i>	<i>blong/blo</i>
Ben	<i>fo</i>	<i>blong/fo</i>	<i>blong?/long/longa/fo</i>	<i>blong/blo</i>
Loc	$\emptyset$	<i>long/longa/$\emptyset$</i>	<i>long/longa/$\emptyset$</i>	<i>long/lo</i>
Goal		<i>long/longa/$\emptyset$</i>	<i>long/longa/$\emptyset$</i>	<i>long/lo</i>
Sce		<i>long/longa/from</i>	<i>long/longa</i>	<i>long/lo</i>
Obl	<i>blong/blonga?</i>	<i>long/longa</i>	<i>long/longa</i>	<i>long/lo</i>
Inst	<i>wit</i>	<i>wit?/long/longa</i>	<i>wetem?/long/longa</i>	<i>wetem/long/lo</i>
Com		<i>long/longa</i>	<i>wetem</i>	<i>wetem</i>
Caus		<i>from?</i>	<i>from?/long/longa</i>	<i>from</i>
Sim	<i>olsem</i>	<i>olsem</i>	<i>olsem</i>	<i>olsem</i>
Con. loc	<i>abaot</i>	<i>?</i>	<i>?</i>	<i>long/lo</i>

Table 20: Development of Bislama prepositions (adapted from Crowley 1990a: 266)

The development of verbal prepositions such as *kasem* (< *catch him*, ‘as far as, until’), *bitim* (< *beat him*, ‘past, beyond, than’), *agensem* (< *against him*, ‘against’), *roanem* (< *round him*, ‘around’) and *folem* (< *follow him*, ‘in accordance to, according to’), by contrast, are the result of modern grammaticalisation processes (Crowley 1990a: 321).

To the best of my knowledge, such a detailed description of the development of prepositions is not available for SIP and TP. Mühlhäusler (1985b), who studied TP from a diachronic perspective, refers to the development of prepositions in TP but not to the same extent. Providing early jargon examples he shows that during the jargon stage prepositions were usually omitted (cf. Mühlhäusler 1985b: 85). Moreover, he mentions that three basic prepositions developed. While *long* would have functioned as the general locative marker and represented the earliest preposition, *bilang* would have developed for possessive encoding and *wantaim* to encode the comitative. In addition, he shows that no further prepositions were necessary as serial verb constructions were commonly used to encode more complex semantic functions (Mühlhäusler

1985b: 138). New complex prepositions such as *egens long*, *akros long*, etc. are argued to have developed in the post-pidgin stage, which, according to Mühlhäusler, did not start before the end of World War II.

Keesing & Jourdan (1997: 404) only casually mention prepositions and demonstrate with an early example of Solomon Islands Pijin that by the early 1890s *bilong* had already grammaticalised into the possessive preposition and *long* into an “all-purpose preposition [...] to mark direction, location, and instrument”. In addition, they show that the plantation Solomon Islands Pijin of the 1930s further developed prepositional verbs, which they define as “forms which have morphology of transitive verbs, but carry prepositional meanings”, such as *wetem*, *agensem*, *aboutem* and *roanem* (1997: 408; cf. also Keesing 1988: 181). However, they do not indicate when these forms were attested for the first time.¹⁷³

Diachronic comparisons of the development of prepositions in the three MPE varieties do not exist. Baker (1993) lists the multi-purpose preposition *(a)long* and the genitive preposition *belong* in his list of earliest attestations, but he does not differentiate between the various semantic functions *(a)long* may fulfil. Additionally, his list contains an entry for *close up* ‘near (by)’ (Baker 1993: 20). He does not differentiate between an adverbial and prepositional use of the form but shows nonetheless that *close up* represents a NSWPE feature which entered into QPPE and from there spread to Vanuatu, the Solomon Islands and a bit later to New Guinea.¹⁷⁴ Baker does not list other prepositional forms and does not focus on the further development of prepositions as his main interest was to show the influence of NSWPE and QPPE on the Pacific MPE varieties.

All other available comparisons are based on the contemporary forms of the varieties and restricted to simple prepositions. Crowley (1990a: 12) claims that “[t]here are some significant interdialectal differences in the basic preposition systems”, summarising the differences in a table, which was adapted by Tryon & Charpentier (2004: 397) and is displayed in Figure 118.

According to Crowley (1990a) and Tryon & Charpentier (2004), differences exist in the prepositional forms used to encode the purposive, characteristics, place of origin, cause/motive, instrumental and comitative. It remains unclear, however, why these differences can be observed in the varieties, whether those differences existed right from the varieties’ births or whether the varieties diverged across time.

¹⁷³ Furthermore, because they assume that the Pijin spoken on the plantations did not change a great deal between the years 1930 and 1960, they also include data collected from Pijin speakers in the 1960s that learned the variety in the 1930s.

¹⁷⁴ The first written attestations identified by Baker (1993: 20) are as follows: NSWPE: 1826; QPPE: 1844; BIS: 1886; SIP: 1888; PAP: 1923; DNG: 1923.

	Bislama	Pijin	Tok Pisin
– Possession (man’s dog); Part-whole (roof of house)	<i>blong</i>	<i>blong</i>	<i>bilong</i>
– Characteristic (church-going woman)	<i>blong</i>	<i>fo</i>	<i>bilong</i>
– Purpose (cup for tea)	<i>blong</i>	<i>fo</i>	<i>bilong</i>
– Place of origin	<i>blong</i>	<i>from</i>	<i>bilong</i>
– Cause	<i>from</i>	<i>long</i>	<i>long</i>
– Destination (to, towards); Source (from); Goal (give it to me)	<i>long</i>	<i>long</i>	<i>long</i>
– Instrument (with)	<i>long/wetem</i>	<i>long/weitem</i>	<i>long/wantaim</i>
– Accompaniment (with)	<i>wetem</i>	<i>weitem</i>	<i>wantaim</i>
– Similarity (like, as)	<i>olsem</i>	<i>olsem</i>	<i>olsem</i>

Figure 118: Prepositions in contemporary MPE varieties (Tryon & Charpentier 2004: 397)

9.3 Primary methodological steps

In order to analyse prepositions in the early MPE data, 10,456 prepositional tokens had to be categorised and coded according to the semantic categories introduced by Hagège (2010). This primary attempt already revealed that polysemy and polyfunctional forms hampered an explicit classification of several tokens. Especially in the initial years of the varieties’ development, it seems as if early language observers are right in claiming that the prepositions are restricted to *(a)long* and *belong*. Thus, the forms are characterised by polysemy “where local, temporal, and other semantic categories apparently melt together” (Hagège 2010: 277). Since specific forms used to encode a specific function did not exist, each preposition had to be analysed and interpreted in its contextual environment. For instance, in (193), the preposition *bilong* can fulfil two semantic purposes. It may encode an attributive function which marks the beneficiary, or it may encode the possessive semantic function. Only with the contextual information that the sentence was extracted from a pamphlet written by the Australian government and was directed towards the Wiwiak inhabitants, it could be ruled out that *bilong* encodes a possessive semantic function. This demonstrates once again that the semantic function of prepositional forms is highly dependent on the context in which they occur. Contextual information could not be collected for all early language data, resulting in ambiguous tokens which could not be included into the analysis.

(193) TOK **BILONG** OLOBOI BILONG WIWIAK
 talk ATT PL-boy ABES Wiwiak
 ‘Message for all boys from Wiwiak.’

(Wewak 1943-45; Kerr 1985; NLA MS9002 PM22)

In a second step, the forms used in the MPE varieties to express the various semantic functions were compared in order to identify those semantic functions which were encoded in *different* ways in the three varieties.¹⁷⁵ From these, those semantic functions were chosen which could most

¹⁷⁵ In *different* ways may also mean that differences regarding frequencies in the choice of forms were observed.

clearly and straightforwardly be identified and analysed. As mentioned earlier, the semantic classification of adpositions is highly dependent on the context in which they occur. However, some semantic notions can be more easily identified than others and leave less room for possible false interpretations. Moreover, those semantic functions were selected, for which striking forms were attested to exclusively occur in one of the three varieties. Finally, three semantic notions were chosen for a closer analysis, which are the *comitative*, *instrumental*, *terminative* and *adessive* semantic function. Following the classification by Hagège (2010), the first two belong to the non-spatio-temporal semantic domain, whereas the latter two belong to the spatio-temporal domain. Although the chapter is restricted to four semantic prepositional functions only, it will provide initial insight into how the varieties developed from a comparative perspective across time.

9.4 Selected Prepositions

The following sections will focus on the four semantic prepositional functions and their encoding in SIP, BIS and TP across time. Section 9.4.1 starts with *comitatives* and *instrumentals*, 9.4.2 focuses on the *terminative* semantic function, which is followed by *adessives* in 9.4.3. Some of the sections require further theoretical background, whereas others directly begin by considering the prepositions used to encode the semantic function in the contemporary varieties. Each section provides some remarks regarding the methodological steps before the findings are introduced and discussed.

9.4.1 Comitatives and instrumentals

The morpheme-by-morpheme analysis revealed that the three MPE varieties make use of diverging relator morphemes to encode the comitative and instrumental semantic function.¹⁷⁶ Although comitatives and instrumentals can be generally encoded through various means such as, for instance, case affixes, they occur in form of prepositions in the three early MPE varieties.

9.4.1.1 Theoretical background

Comitative and *instrumental* are the designations which are used to label “grammaticalized semantic relations between participants of an event” (cf. Stolz et al. 2013: 214). They are frequently treated together as they are “considered two sides of the same coin” (Stolz et al. 2006: 23). Comitative is the label used for situations in which one participant (= *accompanee*) performs

¹⁷⁶ A relator morpheme can be defined as a free or bound morpheme which links and encodes a specific semantic-syntactic relation between two constituents.

an action together with another participant (= *companion*) (cf. Stolz et al. 2006: 17). In contrast, we find *instrumentals* in situations in which one participant is used as the tool by another participant (= user) in order to perform an action. The main distinction regarding the two functionalities is that comitatives “allow the reinforcement of the preposition by the adverb *zusammen* ‘together’” (Stolz et al. 2006: 44). A third type identified by Stolz et al. (2006: 2) represent *modals*, referring to those cases in which the comitative describes the “adverbial modification of a predicate”.

Most researchers only distinguish between comitatives and instrumentals even though the relators could be categorised according to “more specific relations” (Stolz et al. 2006: 41). Focussing on the German preposition *mit*, Stolz et al. (2006: 41-42) show that there are more than 14 different subcategories according to which the preposition can be classified. These fine-grained classifications differ regarding the animacy and control of the participants but are not formally distinguished in German, in which *mit* represents the relator used to encode all functions. As Stolz et al. point out, if “direct primary markers are missing”, fine-grained classifications are “hard to come by”, which is why usually no more than “the two ‘classic’ thematic micro-roles” are distinguished (Stolz et al. 2006: 41). Yet, languages have the possibility to realise semantic differences through diverging forms and some varieties will use different forms for different micro-level classifications (cf. Stolz et al. 2006: 41).

Nonetheless, most studies do not differentiate between more than the two classic semantic roles and there are languages around the world which do not even keep these two broad categories formally apart. For instance, though the *World Atlas of Language Structures* identifies the use of distinct markers as the most common strategy (213/321=66.36%), 23.36% of the listed languages use identical markers (Stolz et al. 2013: 214).¹⁷⁷ Contact languages are most likely to use an identical relator to encode comitatives and instrumentals (54/75=72%). Only nine of the 75 listed pidgins and creoles listed in APiCS use different forms for encoding comitative and instrumental contexts. The remaining twelve varieties make use of an overlapping form (12/75=16%). The latter implies that two relator markers exist, of which one marker can only be used to encode one function (either the comitative or instrumental) and the other marker may encode both comitative and instrumental relations (cf. Maurer 2013c: 276).

The three contemporary MPE varieties are characterised by using overlapping relator particles, with a specific form being used in instrumental semantic contexts which cannot be used in comitative contexts and a second form being predominantly used in comitative semantic

¹⁷⁷ 10.28% of the listed languages use overlapping markers. Afrikaans was excluded from the WALS sample to avoid an overlap, as it appears in the list of languages in APiCS.

contexts, but which is increasingly used in instrumental contexts as well. Although the MPE varieties are similar from a typological perspective, they differ in the concrete forms used to encode the comitative and instrumental function. In the following section, the formal realisations of comitatives and instrumentals in contemporary SIP, BIS and TP will be presented.

9.4.1.2 Comitatives and instrumentals in contemporary MPE varieties

In *contemporary SIP* the form *wetim* can encode both comitative and instrumental semantic contexts, as exemplified in sentence (194) and (195). The form has its origin in StE *with him*. The relator *long* is only used in instrumental contexts, as exemplified in (196).

- (194) *Sekson blong mi bae kam wetem mifala* (comitative)
 section POSS 1SG FUT come INST 1PL
 'My family will come with us.'
 (Jourdan 2002: 261)
- (195) *Samfala busar-em man ia wetem naef* (instrumental)
 Someone stab-TR man DEM INST knife
 'Someone stabbed this man with a knife.'
 (Jourdan 2002: 261)
- (196) *Samfala stik-im dogi blong mi long naefi.* (instrumental)
 someone jab-TR dog POSS 1SG INST knife
 'Someone has jabbed my dog with a knife.'
 (Jourdan 2002: 227)

In *modern BIS* *wetem* represents the prominent form in comitative semantic contexts (see (197)) and *long* is the dominant form used to encode instrumentals (see (198)). As in SIP, *wetem* derives from StE *with him* and can also be used in instrumental contexts, as shown in example (199).

- (197) *Mi no-mo slip wetem abu blong mi* (comitative)
 1SG no-more live COM grandmother POSS 1SG
 'I don't live with my grandmother anymore.'
 (Crowley 2004: 26)
- (198) *Mi kat-em bred long naef* (instrumental)
 1SG cut-TR bread INST knife
 'I am cutting the bread with a knife.'
 (Crowley 2004: 129)
- (199) *i had blong yu spoelem wetem finga blong yu* (instrumental)
 PM hard for 2SG spoil INST finger POSS 2SG
 'It's hard to ruin it with your fingers [alone].'
 (Meyerhoff 2013b: online; Example 23-114)

While SIP and BIS show similar forms to encode comitatives and instrumentals, *Tok Pisin* makes use of a distinct relator particle. As example (200) shows, the particle *wantaim*, deriving from StE *one time*, is used to encode the comitative functionality (cf. Smith & Siegel 2013a: 220).¹⁷⁸

¹⁷⁸ The grammaticalisation of *one time* is discussed below.

The form can also be used to encode the instrumental context, as shown in example (201), although *long* represents the preferred instrumental relator particle, as shown in example (202) (Smith & Siegel 2013b: online).

- (200) *Yu laik kam wantaim mi?* (comitative)
 2SG VOL come COM 1SG
 ‘Do you want to come with me?’
 (Volker 2008: 56)
- (201) *Ol i sut-im disla pig ia wantaim disla ol spia [...]* (instrumental)
 3PL PM shoot-TR DEM pig IA INST DEM PL spear [...] (Smith & Siegel 2013b: online; Example 22-123)
 ‘They shot this pig with these spears ...’
- (202) *kaikai long spun* (instrumental)
 eat INST spoon
 ‘eat with a spoon’
 (Volker 2008: 44)

Although the similarity between SIP and BIS and the distinctiveness of TP may imply that the different plantation histories are responsible for the diverging forms, historical data needs to be consulted to prove whether this was in fact the case. Other scenarios are possible as well, such as, for instance, that a form based on *with* existed in Tok Pisin but disappeared after the end of the labour trade. An analysis of the attestations across time is expected to shed light on the developmental path of comitatives and instrumentals in the varieties.

9.4.1.3 Methodological considerations

Before the findings are presented, some methodological considerations need to be outlined. In a first step, the tokens that were glossed as comitatives or instrumentals during the morpheme-by-morpheme analysis were extracted for further analysis. As the focus is on the formal realisation of comitatives and instrumentals, ST_FORM represents the major variable under consideration for which ten variants were identified that are listed in Table 21. The forms *long* and *along* are treated as a single form (= *(a)long*), and it is not differentiated between *one time long* and *one time along* which are summarised under *one time (a)long*.

Only clear comitative and instrumental cases were included into the analysis. Sentences such as (203), which were defined as *modals* by Stolz et al. (2006: 2), were excluded. In addition, relators encoding an *ornative/temporary property*, a *combination*, a *part-whole/permanent property* or *possession* were also disregarded since they do not conform to the classical definition of the instrumental and comitative category.

(203) *Iu kam 'log marimari*
 2SG come **with** mercy
 'You come with mercy'

(Malakuna -1935; Methodist Mission 1935: 11)

FACTOR	VARIANTS
ST_FORM	(a)long along of along with belong by in one time one time (a)long with with him
FEATURE	prep_com prep_inst
TYPE	tool material/substance body part transportation human being abstract instrument monetary tool active comitative/human companion co-operative reciprocal
YEAR_DET	1832 ... 1950

Table 21: Linguistic coding for comitatives and instrumentals

The remaining sentences were coded as *prep_com* or *prep_inst* depending on whether the form was used to encode a comitative or instrumental semantic relation between two participants of an event (= FEATURE). Sentence (204), for instance, represents a sentence that was classified as *prep_com*, as the 1SG pronoun *me* represents the accompanee, the 2SG pronoun *you* represents the companion and the preposition *widim* functions as a relator indicating the relationship that exists between the two participants, since the sentence allows the reinforcement through the word *together*. Sentence (205), by contrast, represents a typical example of a sentence that was encoded as *prep_inst* since the *axe* represents the tool which is used by the 1PL participants (= *Me and Borani*) to perform the action (= *hit Deferi*).

(204) $\left(\begin{array}{cc} \textit{Me} & \textit{me} \\ \text{1SG} & \text{1SG} \end{array} \right)$ $\left(\begin{array}{c} \textit{go} \\ \textit{go} \end{array} \right)$ $\left(\begin{array}{c} \textit{widim} \\ \text{COM} \end{array} \right)$ $\left(\begin{array}{c} \textit{you} \\ \text{2SG} \end{array} \right)$ (comitative)
 ACCOMPANEE ACTION RELATOR COMPANION
 'I go with you.'

(Vanuatu 1907-1914; Jacomb 1914: 102)

- (205) Me and Borani killem Deferi along axe (instrumental)
- 1SG and Borani hit-TR Deferi INST axe
- USER ACTION RELATOR TOOL
- 'I and Borani hit Deferi with an axe.'

(Tulagi Island; Halimae 12.05.1924)

It was previously outlined that more fine-grained categories are usually not differentiated. Nonetheless, it cannot be excluded per se that formal distinctions based on a fine-grained classification system exist. There may be differences regarding whether languages consider certain contexts "closer to the prototype of the Comitative [...] or the Instrumental prototype" (Stolz et al. 2006: 53). Thus, to see whether a distinct marking is used depending on the semantic fine-grained context, a further column called TYPE was added to closer define the type of instrumental and comitative. Figure 119 shows that I differentiated between seven different instrumental subtypes.

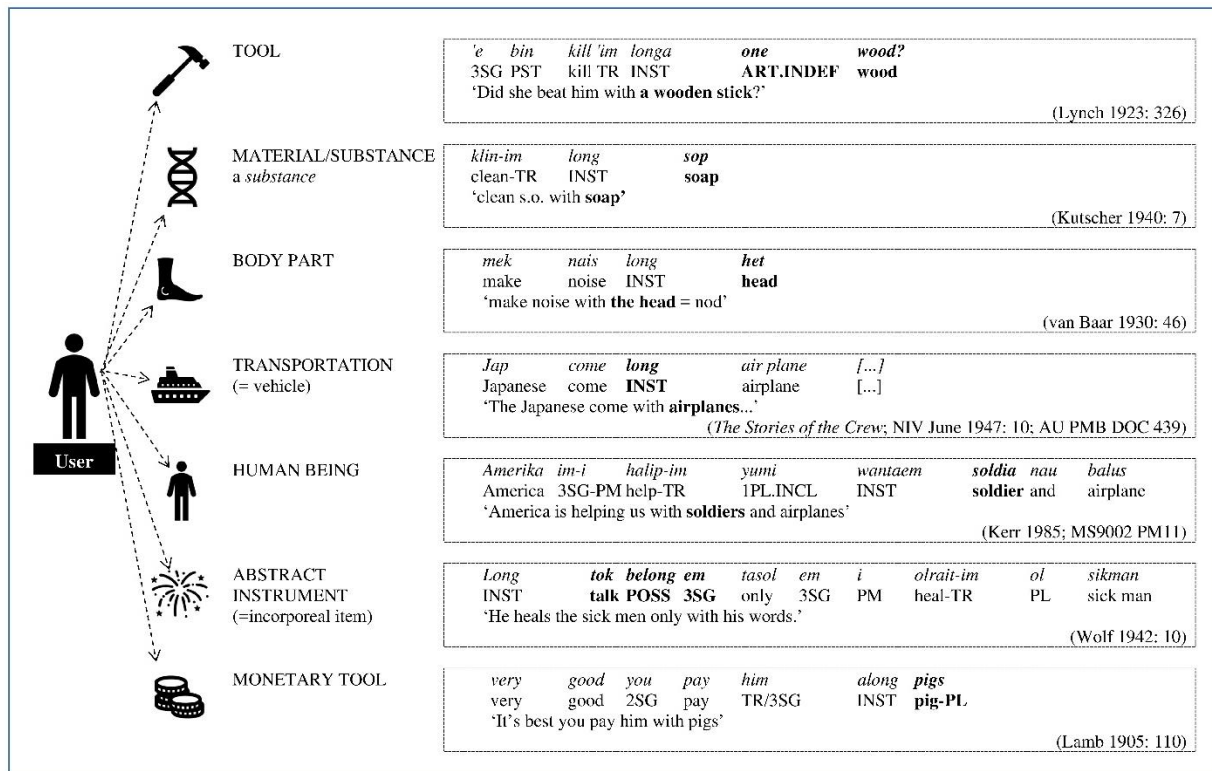


Figure 119: Types of instruments in early MPE

Next to the five categories *body part instrument*, *means of transportation*, *human instrument*, *tool* and *material*, which were introduced by Stolz et al. (2006), I added the instrumental categories *abstract instrument* and *monetary tool*. The category *abstract instrument* is used to refer to tokens in which an incorporeal item is used to carry out an action like encouragement, menace, etc. As there were several sentences in which a monetary tool was used by a user to perform an action, the category *monetary tool* was added.

The more fine-grained classification of comitatives proved difficult. Nonetheless, the remaining clauses were classified according to the three different micro-level categories *co-operative*, *reciprocal*, *active comitative/human companion* which are listed with an example in Table 22. Sentences which did not fit into these categories were not included into the analysis.

One of the reasons for classification difficulties is that languages may differ in terms of whether they regard a situation as reciprocal or not, which as Lehmann & Shin (2005: 12) state, may be “culture-dependent and accordingly be coded in the lexicon”. There are some words which can be considered “inherently reciprocal”, including verbs such as *marry*, *make love*, and *fight*, but the form used may or may not support this assumption (Lehmann & Shin 2005: 11). For instance, in English a person is *married to so*. and not **married with so*. Furthermore, there is the possibility of the development of phrasal verbs in the MPE varieties, which is a factor that may influence the results as well. Since the varieties are analysed from their early beginnings, it is assumed, however, that phrasal verbs had to first grammaticalise.

ACTIVE COMITATIVE/ HUMAN COMPANION	<i>Me</i> 1SG 'I go with you.'	<i>me</i> 1SG	<i>go</i> go	<i>widim</i> COM	<i>you</i> 2SG	(Jacomb 1914: 102)
CO-OPERATIVE	<i>now</i> now 'I took the dinghy with Tingina.'	<i>me</i> 1SG	<i>take him</i> take TR	<i>dinghy</i> dinghy	<i>with</i> COM	<i>Tingina [...]</i> Tingina [...] (Tulagi; Louie 07.01.1925)
RECIPROCAL	<i>Tufelo</i> 1DU 'They talked (together) with Jesus.'	<i>i</i> PM	<i>toktok</i> talk	<i>wantaim</i> (together) COM	<i>long</i> COM	<i>Yesus.</i> Jesus (Wolf 1935: 28)

Table 22: Fine-grained classification categories

A further classification issue arose since comitatives are closely related to coordination. In constructions of the type NP+PREP+NP, the relator particle may encode a comitative or coordinative meaning. Stolz et al. (2006: 47) argue that the difference between comitatives and coordinative participants is that the latter “are assigned the same semantic role and syntactic relation” and can therefore be “inverted without changing the meaning”. Thus, in a sentence such as (206), *wantaim* functions as a coordinator. It is not of importance whether the *kiap* or the *nambawan soldia bilong Ingilis* is mentioned first – the message stays the same.

- (206) *Namabwan* *soldia* *bilong* *Ingilis* **wantaim** *kiap* *bilong* *yu*
first soldier POSS English **and** government.officer POSS 2SG
- bilong* *bifor* *i-sal-im* *disfela* *tok*
PREP before PM-send-TR DEM talk
'The English general and your previous government officer send you this message.'
(New Guinea 1943-45; Kerr 1985; NLA MS9002 PM2)

By contrast, the replacement of comitative participants results in a change in meaning as typically one of the participants is the dominant one (cf. Stolz et al. 2006: 47-48). Example (207) will help to illustrate the difficulty in differentiating between coordination and comitative.

- (207) *Yesus wantaim trifelo i go on-top long Maunten Tabor*
Jesus COM 1TRI PM go on-top of mountain Tabor
 ‘Jesus went with the three on top of mount Tabor.’

(Alexis Harbour -1935; Wolf 1935: 28)

Although the participants of the sentence, namely *Jesus* and *his three disciples*, seem to share their syntactic relation, a dominant participant can be identified. This is, however, only possible if the original wording of the Bible is known or the power relations are clear. Jesus is assigned a higher degree of control. In English Bible translations the sentence is usually rendered as “Jesus took with him Peter, James and John” (Matthew 17, NIV). The use of the verb *take*, which also occurs in other translations, shows that Jesus acts as the dominant participant, which is why *wantaim* was classified as an instrumental particle in the sentence.

Although the above rule of thumb proved helpful for the identification of some comitative sentences, others still allowed for two interpretations and therefore had to be excluded from the analysis. Since comitatives serve as a common source for the grammaticalisation of coordinating *and* (cf. Heine & Kuteva 2004: 327-328), the sentences may nonetheless reveal if and when a comitative developed into a coordinating particle.

In addition, clauses in which the relator fulfilled an *adversative* function (= ‘against’), as exemplified in sentence (208), as well as ambiguous tokens which could not be clearly categorised as fulfilling a comitative or adversative function, were not included in the analysis either.

- (208) *PAIT LONG JAPAN I PINIS NAU [...]*
war PREP Japanese PM over now [...]
 ‘The war against the Japanese has ended.’

(New Guinea 1943-45; Kerr 1985; NLA MS9002 PM99)

A further classification issue came along with the form *one time along*. *One time* does not only occur in prepositional contexts in the data but also functions as an adverb encoding the meaning ‘at once’ or ‘together’.¹⁷⁹ Adverbial attestations of *one time* date earlier than prepositional ones. The temporal adverb *one time* combined with the preposition *long* into *one time long* and thus may be encoded as [together COM]. However, it is also possible that the whole string served to encode the preposition ‘with’. In the latter case *one time long* would have to be glossed as

¹⁷⁹ The adverbial use of *one time* ‘at once’ is a world-wide feature in English-lexified pidgins in the 19th century (cf., for instance, Baker & Huber 2001: 202).

[COM].¹⁸⁰ Since it was not possible to differentiate whether *one time long* functioned as an adverbial plus preposition or whether a speaker used the whole string to encode ‘with’, *one time along* was coded as a preposition, although it should be kept in mind that an adverbial interpretation of *one time* may have been implied. As the analysis will show, the string further transformed as *long* turned into an optional element leading to the grammaticalisation of *one time*.

Before the findings of the analysis are presented, it should be noted that it is not necessarily the case that the forms encoding the comitative and instrumental semantic function grammaticalised at the same time. Therefore, tokens identified as comitatives and those identified as instrumentals were analysed separately before they were examined from a comparative perspective.

9.4.1.4 Findings and discussion

In the following section the findings of the analysis will be presented. The section will first focus on general information regarding comitative and instrumental tokens in the dataset before a closer look at the individual varieties will be taken. First, comitative and instrumental attestations in SIP, then in BIS and finally in TP will be discussed. The section ends with a comparative analysis and a summary of the findings.

9.4.1.4.1 General data distribution

In total, 734 tokens were included into the analysis of comitatives and instrumentals, of which 351 encoded a comitative semantic function and 383 encoded an instrumental semantic function. 138 of the 351 comitatives were identified in SIP (39.32%), 29 in BIS (8.26%) and 184 (52.42%) in TP. The majority of the 383 instrumental tokens were also found in TP, in which 68.15% (= 261/383) of the instrumentals were attested. In contrast, in SIP only 84 (21.93%) and in BIS 38 (9.92%) attestations of the latter were found. This shows that the datapoints, which provide the basis for the analysis, are unequally distributed across the varieties.

In addition, Figure 120 shows that the temporal distribution of comitative and instrumental datapoints differs a great deal in the varieties as well. While attestations of SIP do not start before 1889, comitatives and instrumentals are rather well represented in the period from 1905 to 1950. The BIS dataset contains the earliest attested instrumental and comitative prepositions. While the variety is better represented in the 1910s, comitatives and instrumentals are only rarely attested between 1925 and 1950. Although the majority of datapoints were found in TP for both semantic functions, they accumulate during the period 1925 to 1945.

¹⁸⁰ I used COM as the early attestations of *one time long* are purely comitative.



Figure 120: Distribution of comitative & instrumental datapoints across time per variety

If the period from 1830 until 1950 is considered as a single period only and a look at the frequencies of forms in each variety is made, the following observations can be made regarding *instrumental* prepositional forms:

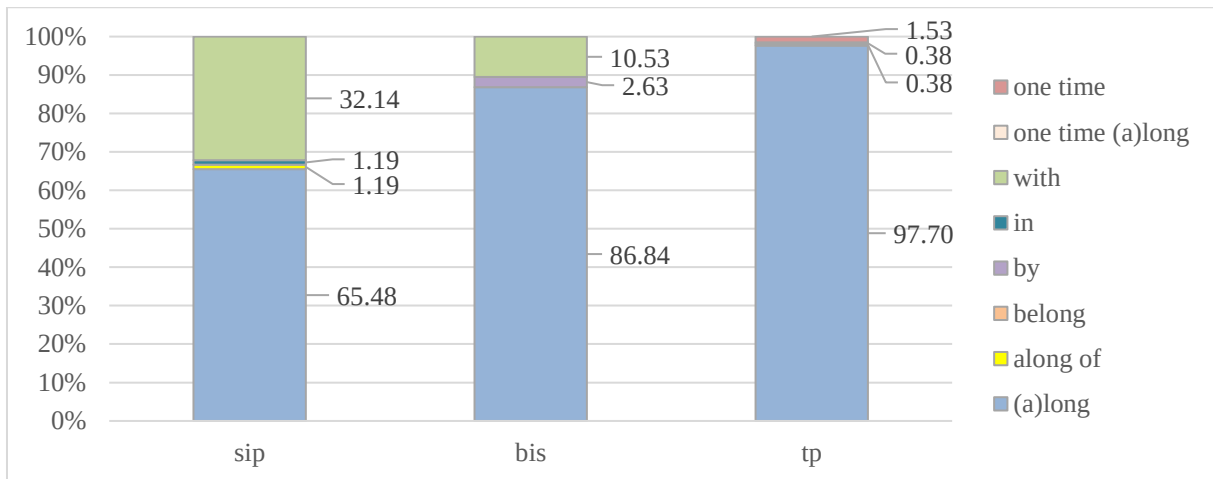


Figure 121: Relative frequencies of instrumental prepositional variants per variety

As Figure 121 shows, *(a)long* represents the form that is attested with highest frequencies in all three datasets to encode the instrumental semantic function. In SIP and BIS the StE form *with* represents the second most dominant form.

If the focus is placed on the encoding of *comitative* prepositional forms only (see Figure 122), a clear dominance of a form is only observable in the BIS data. In BIS, the form occurring with highest frequencies is *(a)long*. Although the form is also the dominant form in the SIP dataset, the StE form *with* was attested in 44.93% of the SIP tokens. In TP, the dominant form represents *one time*, which is not attested at all in SIP and BIS and occurs in 45.65% of the TP data. By contrast, the form *with him*, which is attested in SIP and BIS, does not occur at all in the TP dataset. Thus, differences in the choice of form already seem to have existed before 1950. The following sections will take a closer look at the individual forms attested in each variety and their attestations across time.

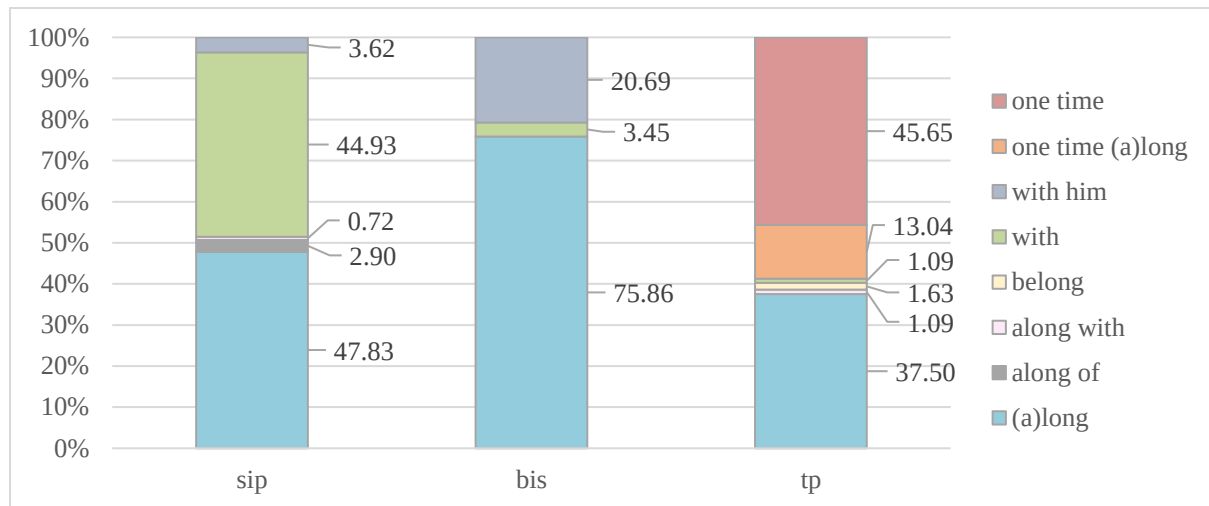


Figure 122: Relative frequencies of comitative prepositional variants per variety

9.4.1.4.2 Comitatives in SIP

This section focusses on comitatives in early SIP. First, the attested forms will be introduced before a closer look at their attestations across time will be taken. In the final section the results of the ctree algorithm are presented in order to test whether *time* represents a reliable factor in predicting the choice of form.

9.4.1.4.2.1 Attested forms

Five different forms to encode comitative semantic contexts were identified in early SIP. The forms which occur with highest frequencies are *(a)long* and *with*, the former being attested in 47.83% (= 66/138) and the latter in 44.93% (= 62/138) of the comitative tokens:

- (209) *me want to go along you [...]*
 1SG want INF go COM 2SG [...]
 ‘I want to go with you.’

(Solomon Islands 1910; Young April 1910; PMB 1150)

- (210) *Time me better me go back long house and talk with Anifelo*
 time 1SG good.COMP 1SG return PREP house and talk COM Anifelo
 ‘When I feel better I return to the house and talk with Anifelo’

(The Stories of the Crew; NIV June 1947: 9; AU PMB DOC 439)

Next to *with*, the relator *with him* was attested as exemplified in (211).

- (211) *Me think more better me go with im you 'long Bina [...]*
 1SG think best 1SG go COM 2SG PREP Bina [...]
 ‘I think it is best if I come with you to Bina.’

(Malaita 1930; Waite 31.03.1930; AU PMB MS 1253)

The form appeared in 3.62% (= 5/138) of comitative semantic contexts and was orthographically realised as *with him*, *with im*, *with em*, *withim* and *widim*. The form is composed of two

morphemes, namely the StE comitative morpheme *with* and the morpheme *him*. Although the latter occurs with different meanings in the early SIP data, for instance, as a 3SG pronoun or as a transitive marker, it functions as a marker of transitivity when combining with *with*. For instance, in example (211) the 2SG form *you* represents the object. Thus, *im* does not encode a 3SG pronoun, but functions as a transitive marker. The transitive inflection *him* usually only occurs with verbs in the early data. However, comitative semantic contexts correspond to transitivity in that an object is required to indicate with whom one is performing an action. This may have been the reason why *him* fused with *with* and grammaticalised into the comitative preposition.

There are also variants of *along* attested in the early SIP data. Sentence (212) provides an example of the form *along of*, which was attested in 2.90% (= 4/138) of the tokens. It needs to be noted that during the encoding of the column ST_FORM, *along*, *along a*, and *alonga* were encoded as *along*. It may be possible that the form *along of* represents an author reinterpretation of the form *along a*. In addition, the form *along with* was attested once in the dataset, the sentence being shown in (213).

- (212) *myself, Jimmy, Jack Jack Tanna, and Oora all go along of him*
 1SG.REFL Jimmy Jack Jack Tanna and Oora 3PL go COM 3SG
 ‘Myself, Jimmy, Jack , Jack Tanna and Oora, they all went with him.’
 (Bundaberg; Charley 24.12.1888 [Australian Station 1888: 22])
- (213) *Me kill him along with Taliniau*
 1SG kill TR COM Taliniau
 ‘I killed him together with Taliniau.’
 (Tulagi; Uiaria, J. 10.04.1924)

In order to learn about the diachronic development of comitative prepositional forms across time, it is necessary to take account of the dates of attestation of the individual forms.

9.4.1.4.2.2 Timeline and boxplot approach

Figure 123 shows when the individual forms were attested across time. The timeline indicates that there are two forms which occur continuously throughout time to encode comitatives in early SIP. Both *along* and *with* are first attested in 1888 and are still in use by the end of the observed period. The form *along of* is only attested pre-1900 and *along with* was only once recorded in 1924. As outlined above, it is likely that the two forms represent editor or author modifications. It is observable that the form *with him* first occurs in 1930 in the written data. Although the form is not attested in many years, it persists until the end of the time period.

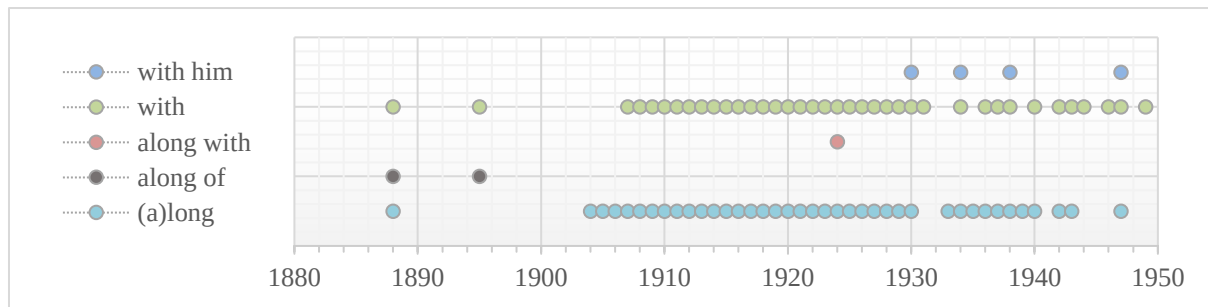


Figure 123: Timeline of comitative variants in SIP

If a boxplot is created based on the variable YEAR_DET so that time periods are replaced by a single determined year, a similar picture is obtained (cf. Figure 124). The medians of the forms *(a)long* and *with* are not very far apart. The timeline gives the impression that a change in the data can be observed around 1930 with the introduction of the form *with him*. At the same time, the box of the form *(a)long* ends and solely the whisker spreads to the late 1940s.

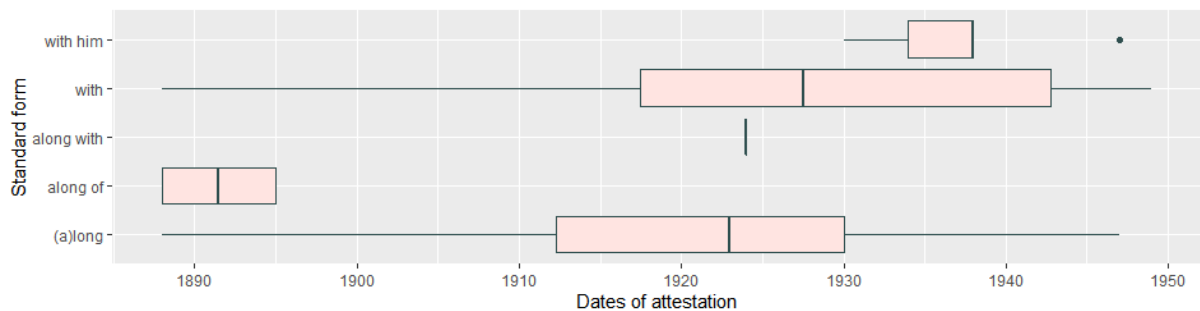


Figure 124: Boxplot of comitative variants in SIP

9.4.1.4.2.3 Testing for the impact of the years of attestations on the choice of form

In order to test whether there is a significant time-based split or whether other possible predictor variables have an impact on the choice of form, a ctree analysis was conducted. If the year of attestation and its impact on the form is analysed, the tree structure in Figure 125 is obtained.

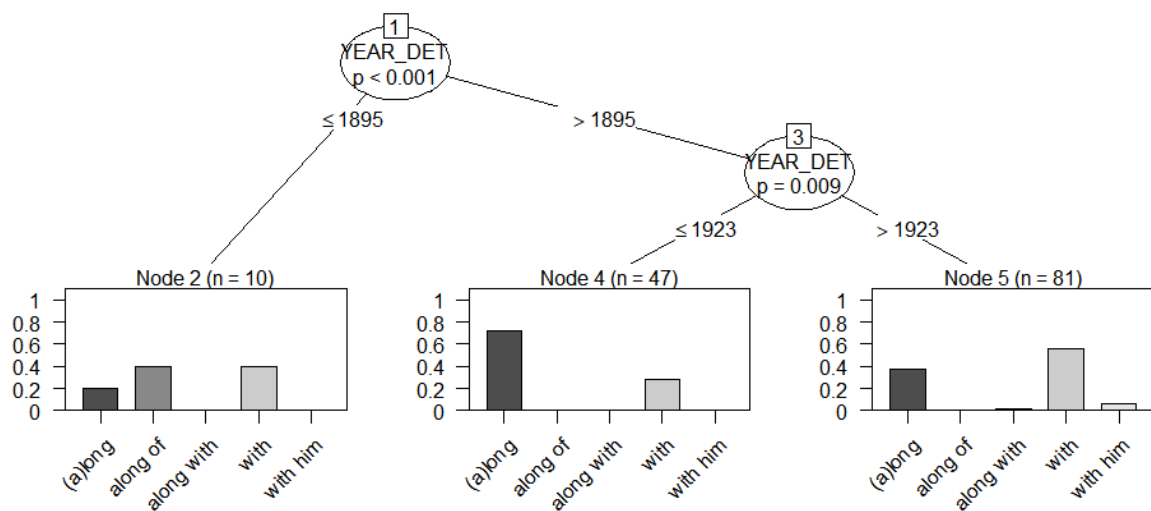


Figure 125: Conditional inference tree for comitative prepositions in SIP $ST\_FORM \sim YEAR\_DET$

The tree consists of two time-based splits. The highest-level split, which is highly significant with $p < 0.001^{***}$, indicates that comitative tokens attested before 1895 are equiprobable to be encoded with *along of* or *with*. There is only a 20% chance that pre-1895 data is encoded with *(a)long*. However, the number of datapoints in the first end node is very low ($n=10$). The second level split, which is very significant with $p=0.009^{**}$, further indicates a difference in comitative encoding dependent on whether the tokens date prior or post 1923. Data attested between 1896 and 1923 is most likely to be encoded with the form *(a)long*. The form *along of* is no longer observed and the probability for *with* is reduced. Attestations dating post 1923, however, are most likely to be encoded with *with*. The likelihood of *(a)long* is reduced and, furthermore, there is a small possibility of the comitative to be encoded with *with him*.

As the text type as well as the author may have an impact on the choice of form as well, the ctree algorithm was applied again, once by adding only the text type variable and once by adding both the text type and author as possible predictor variables. In both cases the resulting tree structure did not differ from the one displayed in Figure 125.¹⁸¹ Thus, the author and text type do not represent significant predictor variables in the choice of comitative forms.

To investigate whether the fine-grained semantic categories (= TYPE) have an impact on the realisation of comitatives, the ctree algorithm was applied analysing the impact of the latter and of the year of attestation on the form. The resulting tree consists of two splits:

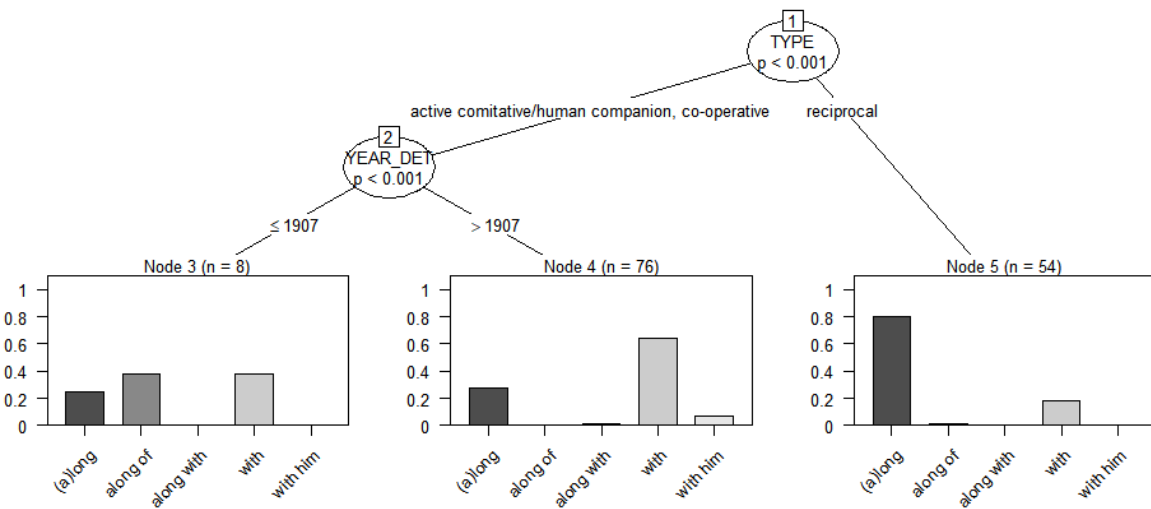


Figure 126: Conditional inference tree for comitative prepositions in SIP ST_FORM~YEAR_DET+TYPE

In contrast to the previous tree structure, the highest-level split is based on the predictor variable TYPE ($p < 0.001^{***}$). The category *reciprocal* seems to behave differently from the remaining comitative categories. As Node 5 indicates, reciprocal comitatives are most likely to be encoded with *(a)long* independent of time. Node 2 splits the remaining comitative categories based on the

¹⁸¹ The p-value of Node 3 changes to 0.019* in the former and to 0.028* in the latter case.

time variable in the year 1907. *Active comitative/human companion* and *co-operative comitatives* attested before 1907 are equally probable to encode the comitative with *along of* or *with*, those attested after 1907 are most likely to be encoded with *with*. Moreover, there is a small chance of the comitative to be encoded with *with him*. Independent of whether only the text type variable or the text type and author variable are added, the tree structure and p-values do not change.

Thus, based on the written attestations of SIP, it can be summarised that independent of the author or text type, significant changes in the choice of form are observable in the years 1895 and 1923, with comitatives attested between 1896 and 1923 being most likely to be encoded with *(a)long* and post-1923 attestations being most likely to be encoded with *with*. These time splits, however, do not turn out as significant factors if a more fine-grained classification is used, which indicates that reciprocal comitatives are most likely encoded with *(a)long*. The remaining comitatives show a preference of the form *with* and the beginning of use of the form *with him* after 1907.

9.4.1.4.3 Instrumentals in SIP

In the following section a closer look at instrumental prepositional forms in SIP will be taken. First, the attested forms will be introduced before their occurrences across time will be considered. Finally, the results of the ctree algorithm will be displayed to see whether time represents a significant predictor variable in the choice of form.

9.4.1.4.3.1 Attested forms

In instrumental semantic contexts four different forms were identified. The most dominant form represents *(a)long*, which was attested in 55 of 84 instrumental prepositional phrases (= 65.48%). In sentence (214) the relator *(a)long* is used to indicate that the knife represents the tool used by the speaker to kill another person. The same tool is also used in example (215), but in this utterance the relator *with* is found.

(214) *He true me want for kill-em along knife*
 PM true 1SG want INF kill-TR INST knife
 ‘It is true, I wanted to stab him with a knife.’

(Lingatu, Santa Isabel; Takwafala 22.03.1938)

(215) *Me kill him die finish with knife*
 1SG kill TR die COMPL INST knife
 ‘I have killed him with a knife.’

(Tulagi; Talatova 09.01.1923)

With is attested in 32.14% (= 27/84) of instrumental semantic contexts. Yet, a closer look at the sentences reveals that *with* occurs predominantly in sentences which in general contain more StE-

like features and are thus more acrolectal. The two remaining attested forms *along of* and *in* are both single attestations. It can thus be assumed that they lack authenticity. The attestations are displayed in (216) and (217).

- (216) [...] *whether man Manoba been kill Government **along of** tomahawk.*
 [...] *whether man Manoba PST kill government **INST** tomahawk*
 ‘... whether the Manoba man killed the government official with a tomahawk.’
 (Bundaberg; Oora 24.12.1888 [Australian Station 1888: 21])

- (217) *me kill him two fellow white man and run away **in** boat*
 1SG kill TR two MODIF white man and escape **INST** boat
 ‘I killed two white men and escaped in/with a boat.’
 (Guadalcanal; Baraghu 18.09.1895 [Australian Station 1895: 23])

9.4.1.4.3.2 Timeline and boxplot approach

The timeline in Figure 127 and the boxplot in Figure 128 show how the four different identified instrumental relator forms spread across time in the early data.

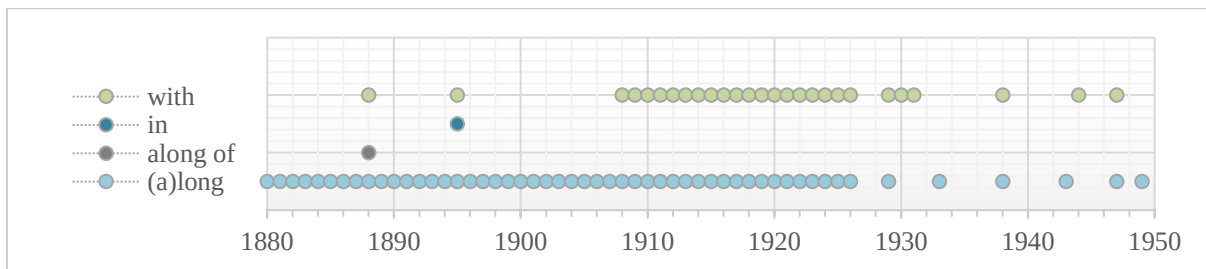


Figure 127: Timeline of instrumental variants in SIP

The earliest attested relator to encode instrumental semantic contexts represents *(a)long* which was attested in the 1880s for the first time. The form prevails until the end of the observed period, and, as mentioned earlier, also represents the most dominant form.

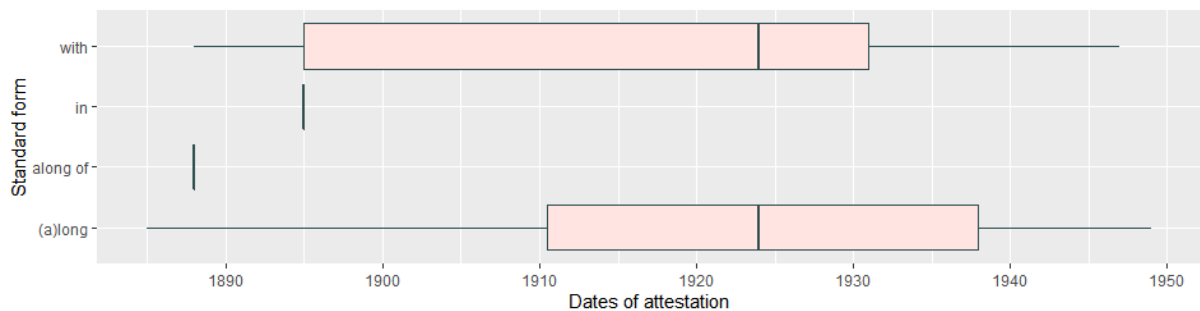


Figure 128: Boxplot of instrumental variants in SIP

With, which represents the second most frequent form, is already attested in 1888 for the first time and is observed until the end of the period under consideration. The two single attestations *in* and *along of* date pre-1900. Based on the timeline and boxplot, no clear preference for one of the forms can be observed.

9.4.1.4.3.3 Testing for the impact of the years of attestations on the choice of form

The ctree algorithm proves that the year of attestation is not a significant predictor variable for the choice of the instrumental preposition. Independent of the possible predictor variables that were considered and on how they were combined, the algorithm resulted in a single stacked boxplot, as Figure 129 shows. Thus, the form *(a)long* was attested as the dominant form independent of time, of the fine-grained category, the author and text type. *With* occurred as the second most dominant form:

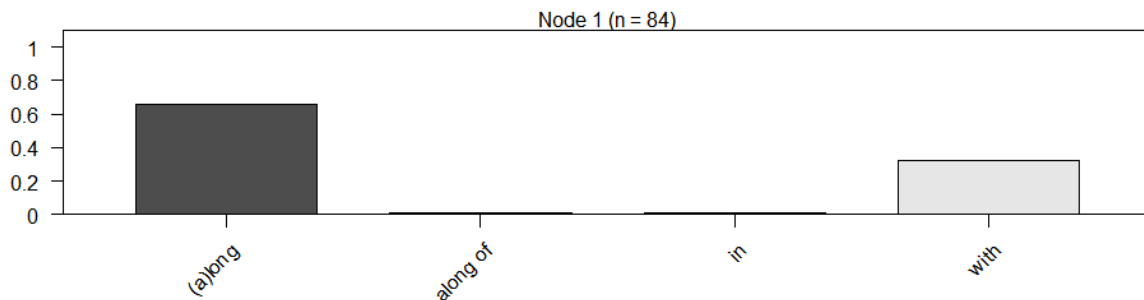


Figure 129: Ctree analysis of instrumentals in SIP resulting in a stacked boxplot

9.4.1.4.4 Comitatives and instrumentals in SIP

So far, comitatives and instrumentals have only been analysed in isolation. Considering all SIP datapoints and investigating the impact of the year of attestation and the variable FEATURE (= comitative vs. instrumental) on the form, the tree in Figure 130 is obtained. The tree consists of three splits on three levels.

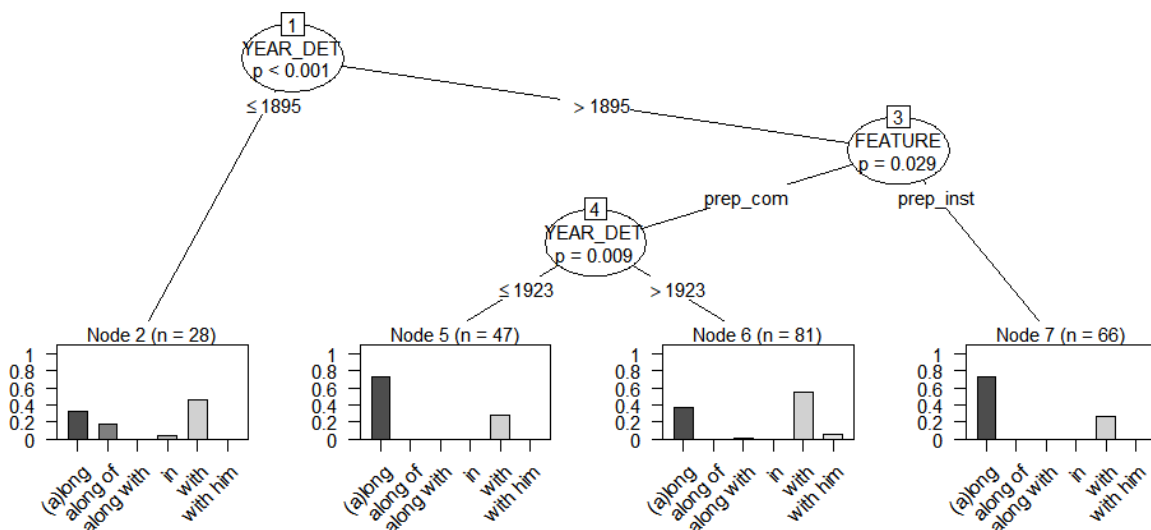


Figure 130: Conditional inference tree for comitative & instrumental prepositions in SIP_ST_FORM~YEAR_DET+FEATURE

The highest-level split is highly significant with $p < 0.001^{***}$ and indicates that data attested before 1896 behaves differently than data attested after 1895. The second level split, which is significant with $p = 0.029^*$ shows that comitatives and instrumental prepositions are encoded

differently in post-1895 data. The third-level split is very significant with $p=0.009^{**}$ and based on the year of attestation. The choice of the comitative form thus seems to show further significant differences depending on whether it occurs prior to 1923 or post 1923. The tree structure stays the same if the text type is added as a further predictor variable, which shows that the text type does not represent a factor that has a significant influence on the choice of form.¹⁸²

Summarising the tree, it can be observed that before 1896, both instrumental and comitative prepositions were predominantly encoded through the forms *with* and *(a)long* with the former being more likely to be used. Only after 1895 differences in the stabilisation of forms can be observed, with instrumental prepositions predominantly being realised with the form *(a)long*. Comitatives, by contrast, are more likely to be encoded with *(a)long* if attested between 1896 and 1923 and show a higher likelihood of being encoded with *with* after 1923. Comparing the Nodes 5, 6 and 7, it thus seems as if the formal differentiation of the two semantic functions started in the second quarter of the 20th century.

If the author variable is added, the resulting tree consists of a single highly significant split which indicates that data attested prior to or in 1895 is most likely to be encoded with the form *with* independent of whether a comitative or instrumental function is expressed. Forms such as *(a)long*, *along of* and *in* may be used in that time as well but are less likely to occur. By contrast, data post 1895 is most likely to be encoded with *(a)long*. Nonetheless, there is still a ca. 40% likelihood of *with* to be used, and the form *with him* represents a further possible form.

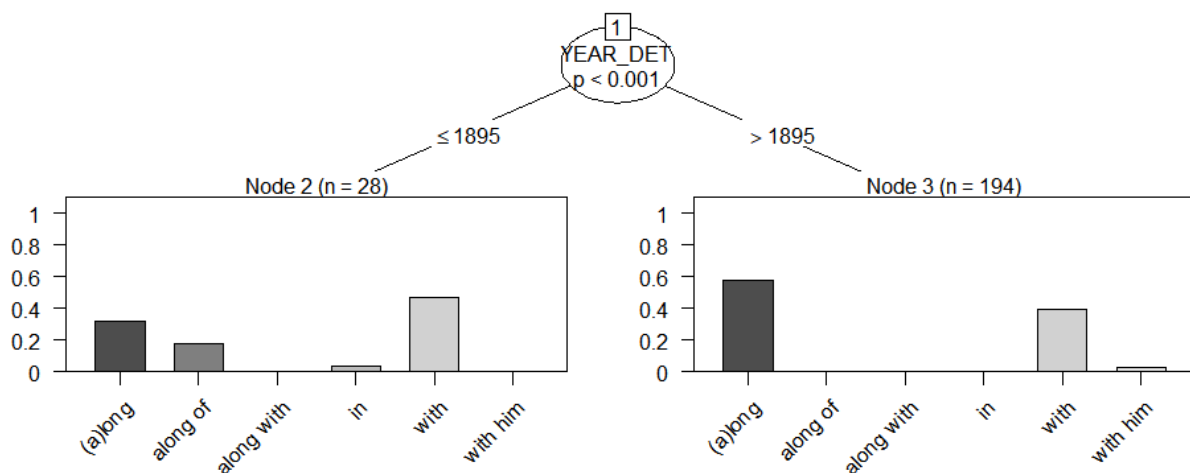


Figure 131: Conditional inference tree for comitative & instrumental prepositions in SIP $ST\_FORM \sim YEAR\_DET + FEATURE + TXT\_TYPE\_3 + AUTH\_NAME$

Interestingly, if a more fine-grained classification is used (TYPE) instead of only distinguishing between comitative and instrumental (FEATURE), a tree consisting of two highly significant

¹⁸² The p-value of Node 3 changes into $p=0.043^*$ and the p-value of Node 4 to $p=0.019^*$.

splits is obtained, independent of whether text type and author serve as further predictor variables.¹⁸³ While the first split is based on the variable TYPE, the second split is based on the year of attestation. The former separates the fine-grained instrumental categories from comitative ones except for the category *reciprocal*, which clusters together with the instrumental categories. While the instrumental categories and the reciprocal one are most likely to be encoded with *(a)long* independent of the time variable, the remaining three comitative categories are equiprobable to make use of *along of* or *with* if attested prior to 1907, with *(a)long* representing a further variant. Post-1907 attestations are most likely to be encoded with *with*, although *with him* seems to have functioned as a further optional form.

9.4.1.4.5 Comitative prepositions in BIS

After having focussed on SIP, the present section focusses on comitatives in early BIS. First, the attested forms to encode comitatives will be introduced before a closer look at their attestations across time will be taken. In the final section the results of the ctree algorithm are presented to test whether *time* represents a reliable factor in predicting the choice of form.

9.4.1.4.5.1 Attested forms

In the early BIS data, only 29 tokens were identified in which a preposition was used as a relator in comitative semantic contexts. The most dominant attested form, as shown in (218), is *(a)long* which occurs in 22 of the 29 tokens (= 75.86%). In addition, there is a single attestation of StE-derived *with* (see (219)), and six attestations of *with him* (= 20.69%). As example (220) shows, the relator *with him* is used in the same way as it is used in SIP (cf. sentence (211)).

- (218) *Lesar he row along me about coconut-s [...]*
 Lesar PM row COM 1SG PREP coconut-PL [...]
 ‘Lesar quarrelled with me about the coconuts.’
 (Ambrym 1914; Dahmansop 14.08.1914)
- (219) *We both come along a boath with Mr. McAlpine.*
 1PL both come PREP boat COM Mr. MacAlpine
 ‘We both came onboard with Mr. McAlpine.’
 (Maryborough; Tara-Haw 06.12.1890 [Australian Station 1891: 67])
- (220) *Me me go widim you*
 1SG 1SG go COM 2SG
 ‘I go with you’
 (Vanuatu 1909; Jacomb 1914: 102)

¹⁸³ See https://www.dropbox.com/s/4t6ardseryn6id3/Ctree_Comins_SIP.png?dl=0 (last access 29 September 2021).

9.4.1.4.5.2 Timeline and boxplot approach

The timeline in Figure 132 (based on the variable YEAR_ATT) and the boxplot in Figure 133 (based on the variable YEAR_DET) show how the identified variants spread across time. It can be observed that the form *(a)long* represents the earliest attested preposition in comitative semantic contexts, being first attested as early as in 1850. In the collected data it is latest attested in 1934, but it needs to be taken into consideration that there is only a small amount of data available for the years 1930 until 1950. The form *with him* is attested in 1907 for the first time and can be observed until the end of the considered time period. Taken into consideration that there is only a small amount of data available for the later years, it is noteworthy that attestations of the form *with him* nonetheless occur.

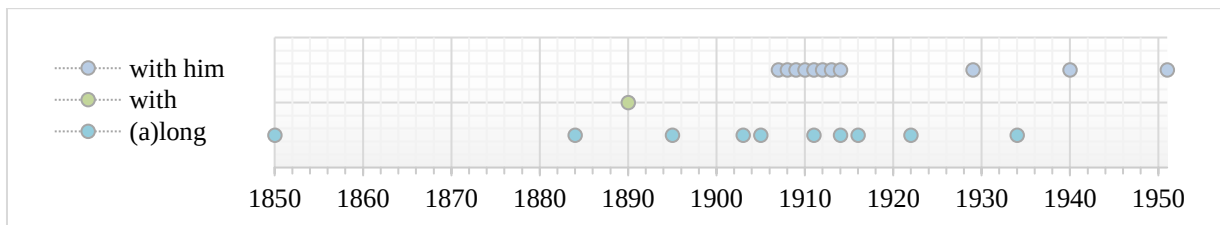


Figure 132: Timeline of comitative variants in BIS

The boxplot confirms these observations and additionally shows that while the median of *(a)long* is in 1908, the median of *with him* is in 1934. The timeline and boxplot give the impression that around 1915 a change in the choice of comitative prepositional forms occurs.

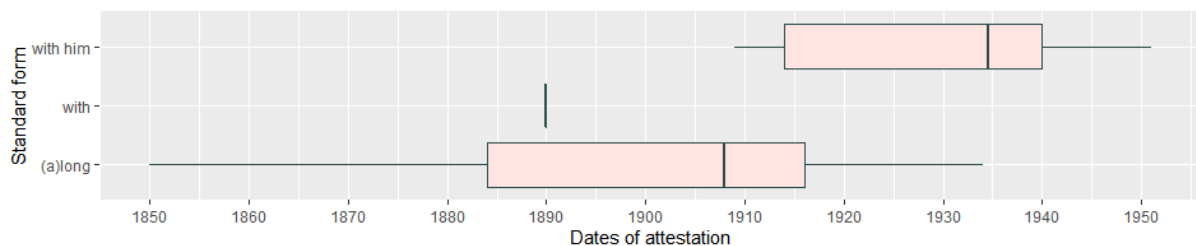


Figure 133: Boxplot comitative variants in BIS

9.4.1.4.5.3 Testing for the impact of the years of attestations on the choice of form

In order to test whether there is some statistical rigor for assuming that around 1915 a change in the preferred form can be observed, the ctree algorithm was applied. As a start, solely the year of attestation and its impact on the form was tested. The resulting tree is displayed in Figure 134 and contains one significant split ($p=0.019^*$). The year 1905 is identified by the algorithm as being significant for the choice of form. Data dating pre-1906 is most likely to show *(a)long* as the comitative relator. Even if it also represents the most likely form in data attested after 1905, there is a 35% likelihood that the form *with him* is used.

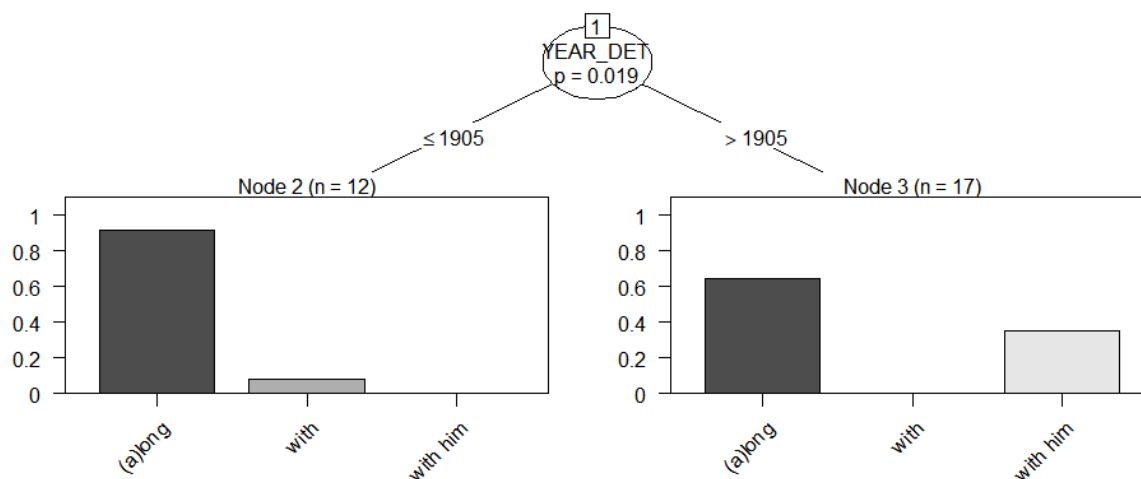


Figure 134: Conditional inference tree for comitative prepositions in BIS ST_FORM~YEAR_DET

Since the text type or the author may have an impact on the results as well, additional trees were created by first adding the text type variable and then adding both, the text type and author as possible predictor variables. If solely the text type and the time variable are integrated into the analysis, the resulting tree is the same as in Figure 134, with $p=0.037^*$. However, if the author variable is integrated, the ctree analysis results in a single stacked boxplot, indicating that neither the year of attestation, nor the text type or author represent reliable predictor variables.

The analysis was repeated adding the variable TYPE (= analysing the impact of the fine-grained semantic categories). Applying the ctree algorithm using the year of attestation and the fine-grained semantic categories as possible predictor variables results in the same tree structure as displayed in Figure 134, with $p=0.037^*$. If the text type or both, the text type and author are additionally considered, the analysis results in a single stacked boxplot. Thus, the fine-grained categories have no impact on the choice of form.

9.4.1.4.6 Instrumental prepositions in BIS

After having looked at comitatives, the present section will focus on instrumental relators that were attested in the early BIS dataset. First, the attested forms are introduced before their dates of attestation will be considered. The ctree algorithm will provide evidence which predictor variables have an impact on the choice of the relator in instrumental semantic contexts.

9.4.1.4.6.1 Attested forms

In total, 38 instrumental prepositional tokens could be extracted from the BIS data collection. Three different forms were identified to encode instrumental prepositional meaning. In 86.84% (= 33/38) of the tokens a form of *(a)long* was attested. For instance, in sentence (221) the relator *long* is used to indicate that the master used a body part instrument (= hand) to hit the boy. The form *with* was only attested in four of the 38 tokens (= 10.53%) and is exemplified in sentence

(222). Moreover, once the form *by* was attested in a letter that was written by Charlie to Mr King in January 1913. The letter is characterised by a mixture of StE and PE features. The sentence in which *by* was attested (see (223)) only shows two non-standard features which are the use of an overt subject in an imperative construction and the use of the lexical item *police* instead of *police officer*. *By* is thus very unlikely to represent a realistic early BIS feature.

- (221) 'Master 'e kill 'im boy **long** 'and all time
 master PM hit TR boy **INST** hand all time
 'The master hit the boy with his hand all the time.'

(Vanuatu 1909; Jacomb 1914: 95)

- (222) Mare Mare shot him **with** Snider.
 Mare.islander shoot 3SG/TR **INST** rifle
 'The Mare Islander shot him with a snider rifle.'

(Tonga 1883; Colonial Office CO225/15 1883-84: 218)

- (223) you please send two police down **by** the Steamer Moackcambo
 2SG please send two police.officers down **INST** ART steamer Mackambo
 'Please send two police officers (to us) with the steamer Mackambo'

(Port Vato 1913; Charlie 31.01.1913)

9.4.1.4.6.2 Timeline and boxplot approach

Although *(a)long* represents the dominant attested form in instrumental semantic contexts, a timeline was created to investigate how the attested forms spread across time. Figure 135 and 136 indicate that *with* represents the earliest form first being attested in 1831. It is latest attested in 1920 but due to the low amount of available BIS datapoints that cover the time period 1925 until 1950, this does not mean that the form was no longer in use. The earliest attestation of *(a)long* dates to 1891 and the form seems to develop into the dominant instrumental form in the first half of the 20th century.

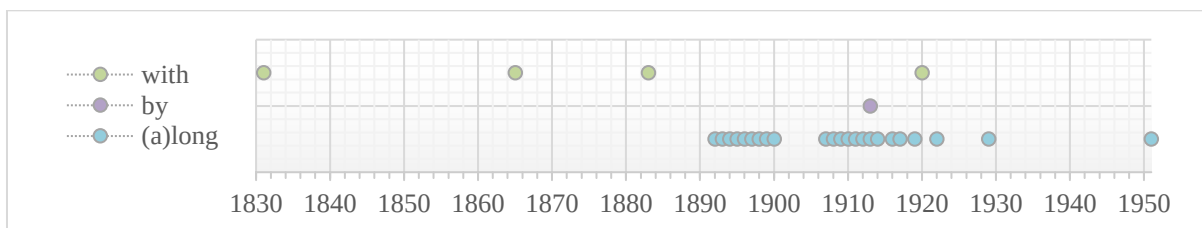


Figure 135: Timeline of instrumental variants in BIS

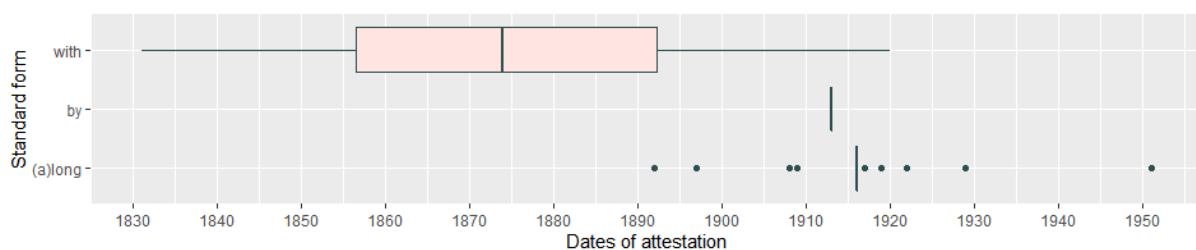


Figure 136: Boxplot of instrumental variants in BIS

9.4.1.4.6.3 Testing for the impact of the years of attestations on the choice of form

In order to test whether there is some statistical rigor for a time-based split, the ctree algorithm was applied to the 38 instrumental datapoints. By considering only the dates of attestation and their impact on the form, the tree in Figure 137 is obtained. The tree consists of a single highly significant split ($p < 0.001^{***}$) which shows that data attested after the year 1908 behaves differently from earlier attested data. Instrumental tokens attested before and in 1908 are slightly more likely be encoded with the relator *(a)long* than with *with*. By contrast, data attested after 1908 shows a dominant use of the form *(a)long*. If the fine-grained semantic categories (TYPE) and/or the text type are added as further predictor variables, the tree structure does not change and the year 1908 still represents the only significant split evoking year ($p < 0.001^{***}$).

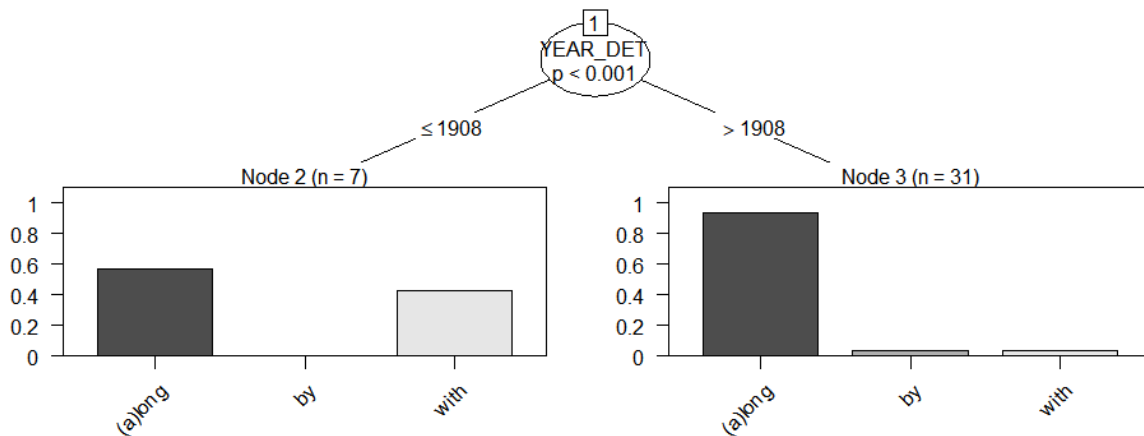


Figure 137: Conditional inference tree for instrumental prepositions in BIS ST_FORM~YEAR_DET

However, if the impact of the year of attestation, the text type and the author is analysed, a tree with two splits is obtained. As Figure 138 visualises, the highest-level split is based on the author ($p < 0.001^{***}$), and the second-level split is based on the year of attestation ($p < 0.001^{***}$).

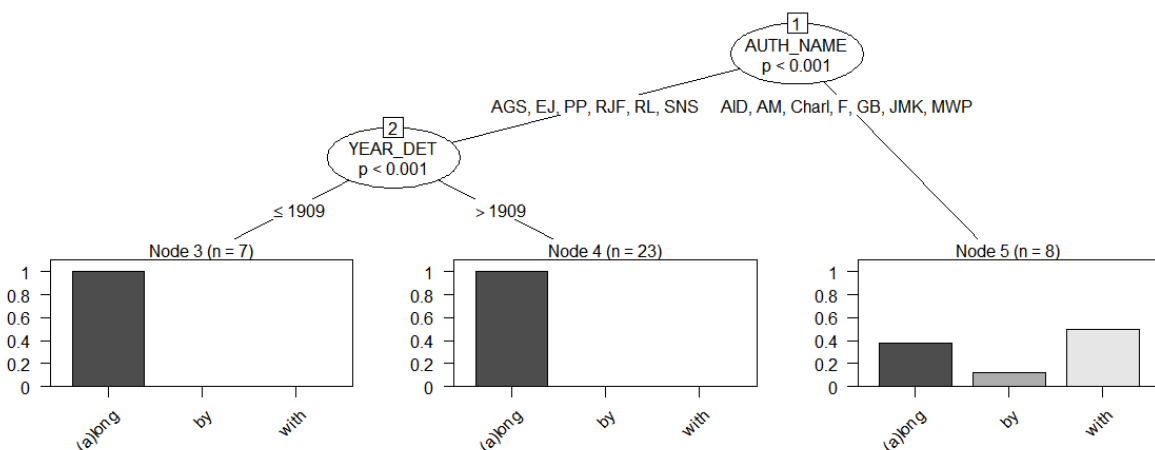


Figure 138: Conditional inference tree for instrumental prepositions in BIS ST_FORM~YEAR_DET+TXT_TYPE_3+AUTH_NAME

Only instrumental tokens occurring in texts produced by the authors Stewart, Jacomb, Fletcher, Pionnier, Lamb and Shurecliff show a time-based split indicating the year 1909 to be of significance. While instrumental relators that are attested prior to 1910 are most likely to be encoded with the form *(a)long* or *with*, those attested after 1909 are only encoded with *(a)long*. It should be noted that *(a)long* already represents the dominant form in pre-1910 tokens.

If the fine-grained semantic categories (TYPE) are considered as a further possible predictor variable, the tree structure does not change. The author represents the most dominant predictor variable ($p < 0.001^{***}$) and, on the second level, a time-based split ($= 1909$) can be observed ($p = 0.003^{**}$). It should be kept in mind that the general number of instrumental tokens in BIS is very low.

9.4.1.4.7 Comitatives and instrumentals in BIS

In a final step the comitative and instrumental tokens were taken together to investigate whether a significant point in time can be determined that demarcates when differences between the comitative and instrumental encoding developed. If only the impact of the year of attestation (YEAR_DET) and of the dual semantic categories (FEATURE) on the form is considered, the resulting tree shows two time-based splits. While the first split indicates the year 1922 to be of importance ($p < 0.001^{***}$), the second split designates the year 1890 to be a significant year ($p = 0.008^{**}$) for the choice of form. It is of interest that these time-based splits are independent of whether comitatives or instrumentals are encoded:

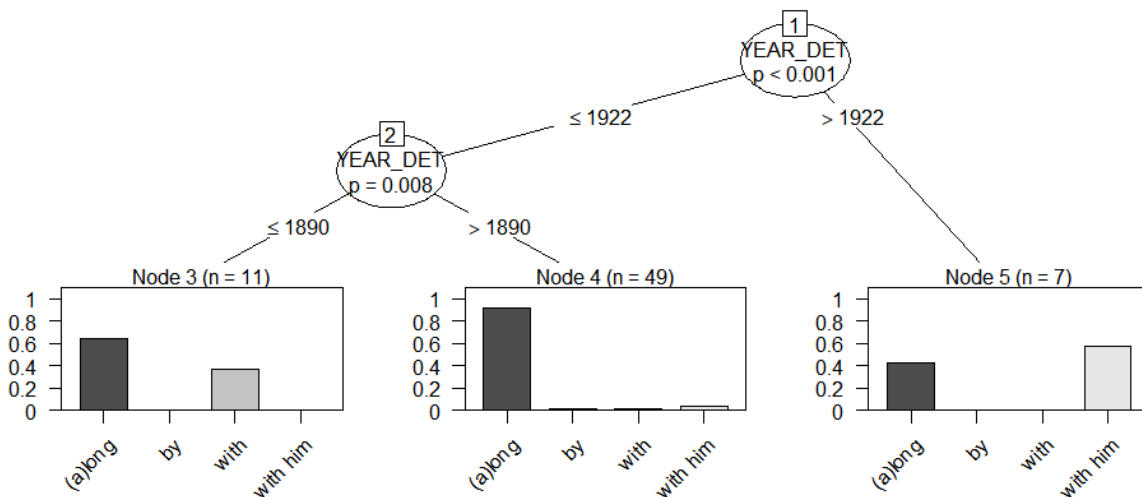


Figure 139: Conditional inference tree for instrumental & comitative prepositions in BIS $ST\_FORM \sim YEAR\_DET + FEATURE$

Adding the text type as a further variable does not change the tree structure, although the p-values differ (Node 1: $p = 0.001^{***}$; Node 2: $p < 0.001^{***}$). Moreover, the tree structure does not change if the dual categories (FEATURE) are replaced by fine-grained semantic categories (TYPE).

However, a change in the significant predictor variable is observed if the author variable is added. Applying the algorithm including the year of attestation, the text type, the dual categories and the author as possible predictor variables results in a tree that consists of a single highly significant split based on the author (see Appendix II). Thus, based on the low amount of data at hand, no statistically significant change can be observed that is dependent on the factor *time* if the author is considered as a possible influencing factor. If the dual categories (FEATURE) are replaced by fine-grained semantic categories (TYPE), the author remains the only significant predictor variable.

9.4.1.4.8 Comitative prepositions in TP

The present section will focus on the 184 comitatives which were extracted from the early collected TP data. First, the attested forms will be introduced before the focus will be placed on their dates of attestations. The results of the ctree algorithm will indicate whether time represents a significant predictor variable in the choice of comitative prepositional forms.

9.4.1.4.8.1 Attested forms

The most dominantly attested form in the early TP data is *one time* as it occurs in almost half of the attested comitative prepositional tokens (84/184= 45.65%). The form is of special interest as it is not found in the other two varieties. It has its origin in StE *at one time*, which in English is used to express that things happen or are done at the same time. It seems as if only the latter part of the English adverb was borrowed but could be used with the same meaning. For instance, in sentence (224) *one time* is used to express that the person carries two children ‘at once’. The meaning of *one time* further expanded, as can be learned from sentences such as (225), in which *one time* is used as an adverb meaning ‘together’. Although it could still be argued that it functions as a temporal adverbial, the form expanded semantically. This becomes clear when looking at comitative tokens such as the early dictionary entry displayed in (226). The example can be translated with ‘to sleep or have sexual intercourse with a woman’. The interesting aspect of this dictionary entry is that the word *vantaim* is parenthesised. By using parenthesis, Kutscher (~1940) shows that this part of the phrase can be omitted without changing the overall meaning. Thus, the dictionary entry may indicate that *vantaim* functioned as an adverbial encoding the meaning ‘together’, but it may also indicate that both *long* and *vantaim long* were in use to encode the comitative. The more *one time (a)long* was used in comitative contexts, the higher the chance that the whole construction was reinterpreted as the comitative preposition. Only in a last step of the

grammaticalisation process was the phrase reduced by taking away the latter part, resulting in *one time* serving as the comitative preposition, as sentence (227) indicates.

- (224) [...] you carry him alltime two fellow **one time**
 [...] 2SG carry TR always two MODIF at.once
 ‘... you always carry two at once/at the same time’
 (Kaiser Wilhelmsland 1911; Dempwolf in Tyron & Charpentier 2002: 379)

- (225) *disfelo singsing i gut-felo sepos olgeder singsing wantaim*
 DEM song PM good-MODIF if all sing together
 ‘This song sounds good if we all sing together.’
 (Mugil Mission ~1930; van Baar 1930: 33)

- (226) *slip (wantaim) long meri*
 sleep (together) COM woman
 ‘have sexual intercourse with a woman’
 (Vunapope ~1940; Kutscher ~1940: 24)

- (227) *NO-GUT YU DAI ONTAIM OL.*
 NEG-good 2SG die COM 3PL
 ‘You should not die with them.’
 (Mandated Territory of New Guinea 1943-45; Kerr 1985; NLA MS9002 PM37)

One time (a)long was attested in 13.04% (= 24/184) of the tokens. Another frequently attested form represents *(a)long* which appeared in 69 of the 184 comitative tokens (= 37.50%). For a direct comparison, again, an example was selected which expresses to have sexual intercourse with someone:

- (228) *i-tok i-lajk pušpuš loŋ mi*
 PM-say PM-VOLIT have.sexual.relations COM 1SG
 ‘(He) said he wants to have sexual relations with me.’
 (New Guinea ~1943; Hall 1943: 83)

The forms *belong*, *with* and *along with* are not attested with high frequencies, the first two forms being attested in three and the remaining in two instances only:

- (229) [...] he puss-puss **belong** this fellow
 [...] PM have.sexual.relations COM DEM fellow
 ‘(she) has sexual intercourse with him’
 (Lamassa 1904; Stephan & Graebner 1907: 123-124)

- (230) *Me bin talk with you now.*
 1SG PST talk COM 2SG now
 ‘I have talked to you now.’
 (Rabaul 1914; Idriess 1941: 32)

- (231) ‘You look out place **alonga with** him [...]
 2SG watch out place COM 3SG [...]
 ‘You watch out the place (together) with him.’
 (Rabaul 1927; Idriess 1941: 32)

Two of the three instances of *with* were traced to the author Ion Llewellyn Idriess. Sentence (230) was extracted from his reproduction of the Proclamation in 1914. Interestingly, authors such as

Reeves (1915: 77-78) and Cameron (1923: 292-293) also reprint the proclamation. While Cameron uses *with* as well, Reeves uses *alonga* (= *Me been talk alonga you now*). Independent of whether *with* was used in the original proclamation or not, it occurs only in one further source. In addition, both attestations of *alonga with* are found in the source produced by Idriess. It is thus rather unlikely that these forms were in use as alternative forms to encode comitative contexts.¹⁸⁴

9.4.1.4.8.2 Timeline and boxplot approach

Figure 140 visualises when the individual comitative markers were attested in the TP data. The preposition *(a)long* is attested in comitative semantic contexts as early as in 1884 and represents the earliest attested form. According to the dataset, the form is used with this functionality until the end of the observed period. The preposition *belong*, which, as shown in the introductory quotes, was usually referred to as the only alternative preposition, is only attested from 1900 to 1910. The attestations of *with* and *along with* date to the year 1914 and are extracted out of the proclamation. As previously mentioned, alternative text versions of the proclamation exist. These show *alonga* in those sentences in which Idriess uses *with*. The form *along with* is replaced with *alonga* in the version by Reeves (1915: 77-78), but Cameron (1929: 292-293) uses once *alonga* and once *alonga with*.

From 1926 onwards, the data indicates that comitatives were additionally rendered by the construction *one time (a)long*. What is interesting is that the first written attestation of *one time* without *along* dates only four years later. Both forms are attested until the end of the period under investigation.

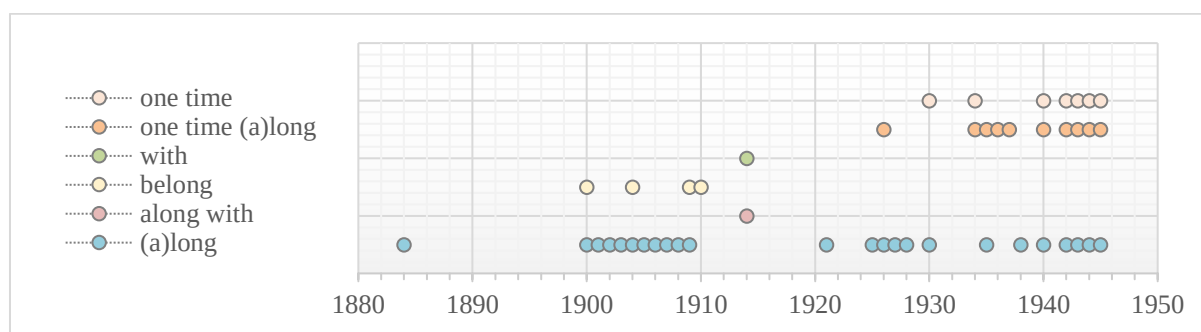


Figure 140: Timeline of comitative variants in TP

¹⁸⁴ As pointed out in Chapter 4, in cases of reprints, I kept and included the earliest dating material. The proclamation forms an exception. As it was clear from all sources that the proclamation text dates to the year 1914 and as all authors introduced the proclamation text in the same way, it was assumed that it is not too much of importance which version to include. The example shows how important it is to always collate the early examples. While Reeves (1915) uses *alonga* in all comitative positions, Cameron (1923) mixes *alonga with* and *alonga* and Idriess (1941) uses *alonga* and *longa with*.

Looking at the boxplot in Figure 141 and comparing the medians of the forms *one time*, *one time along* and *along*, it can be observed that most of the datapoints of each form are attested around 1940. Though *one time* is already used as a comitative preposition, forms such as *one time along* and *along* still seem to coexist by the end of the observed period.

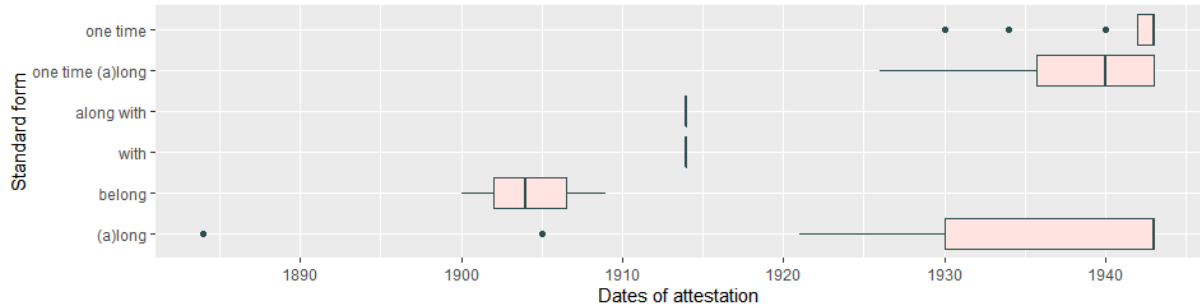


Figure 141: Boxplot of comitative variants in TP

Based on the timeline and boxplot it may be assumed that a change in the data is observable around 1926. To test whether this is true, or whether other possible predictor variables have an impact on the comitative encoding, the ctree algorithm had to be applied.

9.4.1.4.8.3 Testing for the impact of the years of attestations on the choice of form

In a first step, the algorithm was applied to test the impact of the time variable on the form. The resulting tree is displayed in Figure 142 and consists of four time-based splits.

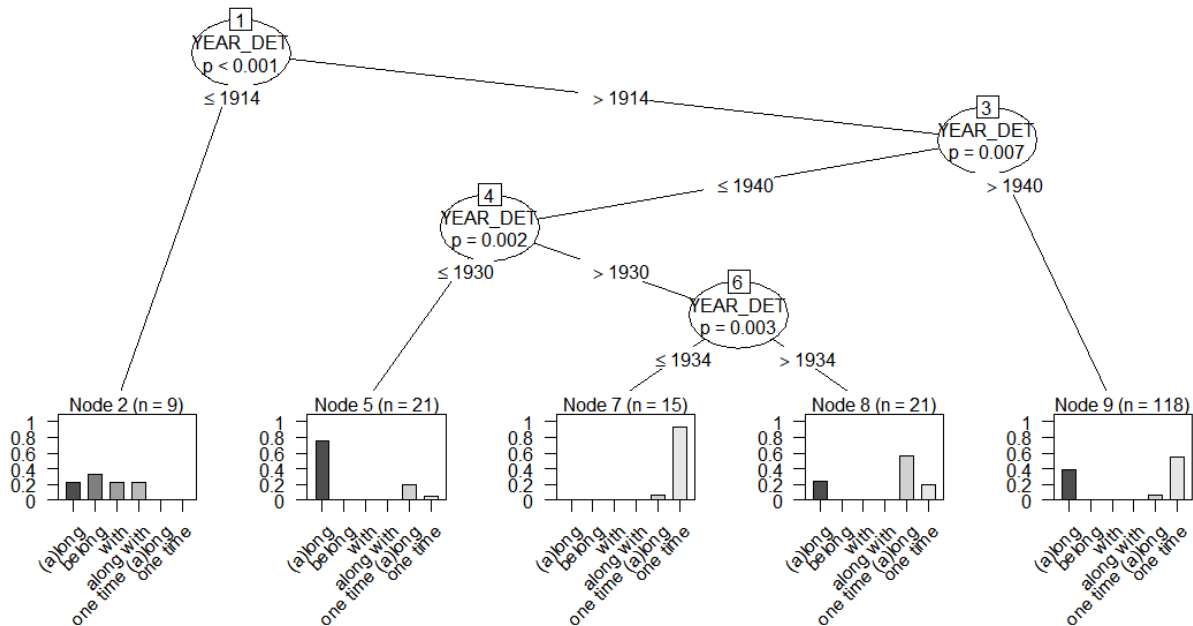


Figure 142: Conditional inference tree for comitative prepositions in TP $ST\_FORM \sim YEAR\_DET$

The highest-level split, which is highly significant with $p < 0.001^{***}$, separates the data attested in and before 1914 from data attested after 1914. In the pre-1915 data, great variation is observed

in the choice of forms with *(a)long*, *belong*, *with* and *along with* being attested (although, as explained above, some of these forms are rather unlikely to reflect realistic variants). Node 3 ($p < 0.007^{**}$) indicates a difference in the data attested before and after 1940. The pre-1941 data is further separated by Node 4 which indicates the year 1930 to be a very significant predicting factor for the choice of form. Data attested between 1914 and 1930 is most likely to show the use of *(a)long* in comitative semantic contexts. Node 5 shows *one time* and *one time (a)long* as possible, albeit less probable, alternative variants. Node 6 further splits 1931-1940 attestations in the year 1934 ($p = 0.003^{**}$). Thus, four different time periods are created for data attested after 1914, which are 1915-1930, 1931-1934, 1935-1940, 1941-1950. When comparing the latter three end nodes, it becomes clear that forms including *one time* turn into the most likely forms. While from 1915-1930 *(a)long* represented the most likely variant, there are no attestations of the latter during the years 1931-1934. Although *(a)long* is attested after 1934 again, *one time (a)long* represents the dominant variant between 1935-1940. After 1940, *one time* turns into the most probable form.

Adding the text type variable to the analysis, the tree in Figure 143 is obtained which consists of four splits on three levels. Two of the splits are evoked by the time variable and the remaining ones are dependent on the text type. The highest-level split still separates the data in the year 1914 ($p < 0.001^{***}$). The second-level split is based on the text type and shows that comitatives in speech-related attestations differ from those attested in written and intermediate attestations. The speech related attestations show differences depending on whether they occur before or after 1930 ($p = 0.012^*$). While comitatives attested prior to 1930 are most likely to be encoded with *(a)long*, post-1930 comitatives are most probable to be encoded with *one time (a)long*. The possibility for *one time* increases as well.

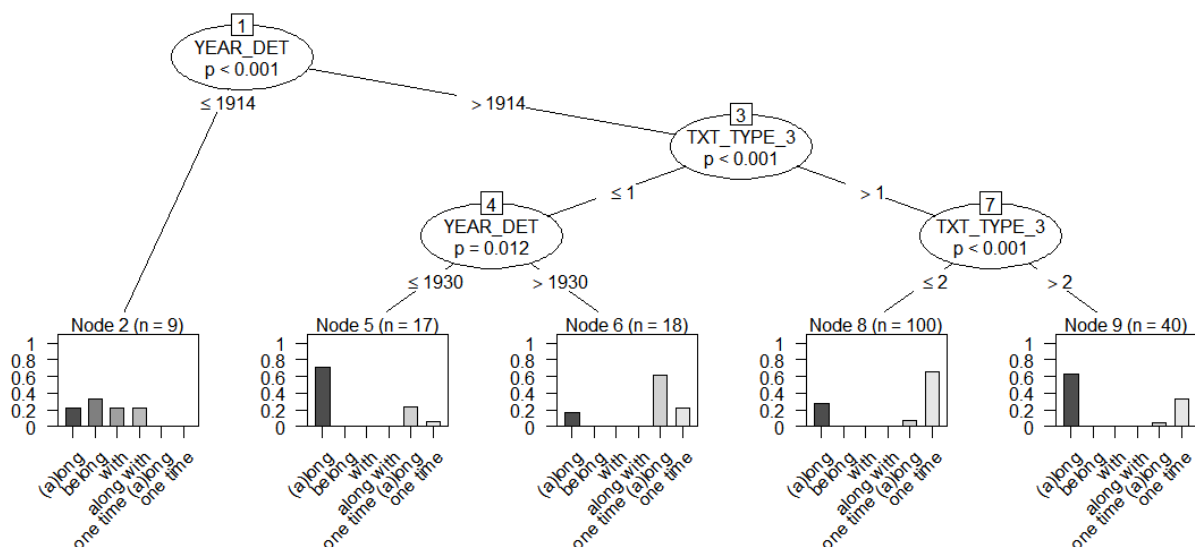


Figure 143: Conditional inference tree for comitative prepositions in TP ST_FORM~YEAR_DET+TXT_TYPE_3

Node 7 shows that there is a significant difference in tokens attested after 1914 depending on whether comitatives were attested in written or intermediate sources. If attested in the former, *one time* is designated as the most likely form. If attested in the latter, *(a)long* is detected as the form with highest probability.

However, if the author is taken into consideration as well, the resulting tree consists of three splits that are solely based on the author. This shows that the author represents the strongest predictor variable for the choice of form.

In a second step it was analysed whether the fine-grained semantic categories (TYPE) have an impact on the encoding of comitatives. If the ctree algorithm is applied to investigate the impact of the year of attestation and the variable TYPE on the form, the tree displayed in Figure 144 is obtained. The tree consists of four splits. The highest-level split is based on the fine-grained categories and splits the reciprocal function from the remaining categories. The latter show a time-based split in the year 1926 (cf. Node 2; $p < 0.001^{***}$), while the reciprocal show a time-based split in the year 1921 (cf. Node 7; $p < 0.001^{***}$). Reciprocal comitatives that were attested before 1921 show a greater degree of variation than those attested after 1921, where a clear preference of the form *(a)long* is observable.

The fourth split is based on the fine-grained categories (TYPE) and splits post-1926 active comitatives from post-1926 cooperative comitatives. While *one time* is the most likely form on both sides of the splits, with cooperative comitatives the form is only slightly more probable than *one time (a)long*.

If the text type is added as a further possible predictor variable, the tree structure in Figure 145 is obtained which shows that the first three splits remain the same as in Figure 144. A change can be observed in the fourth split, which is no longer based on the fine-grained categories but on the text type. A further text type dependent split is also indicated by Node 6 ($p = 0.003^{**}$).

If the author is furthermore considered, a tree consisting of four splits is obtained, whereby three of these splits are based on the author and one is evoked by the fine-grained semantic categories. Again, the results show that in terms of comitative prepositions and their encoding, the author is the most significant predictor variable in the early TP data.

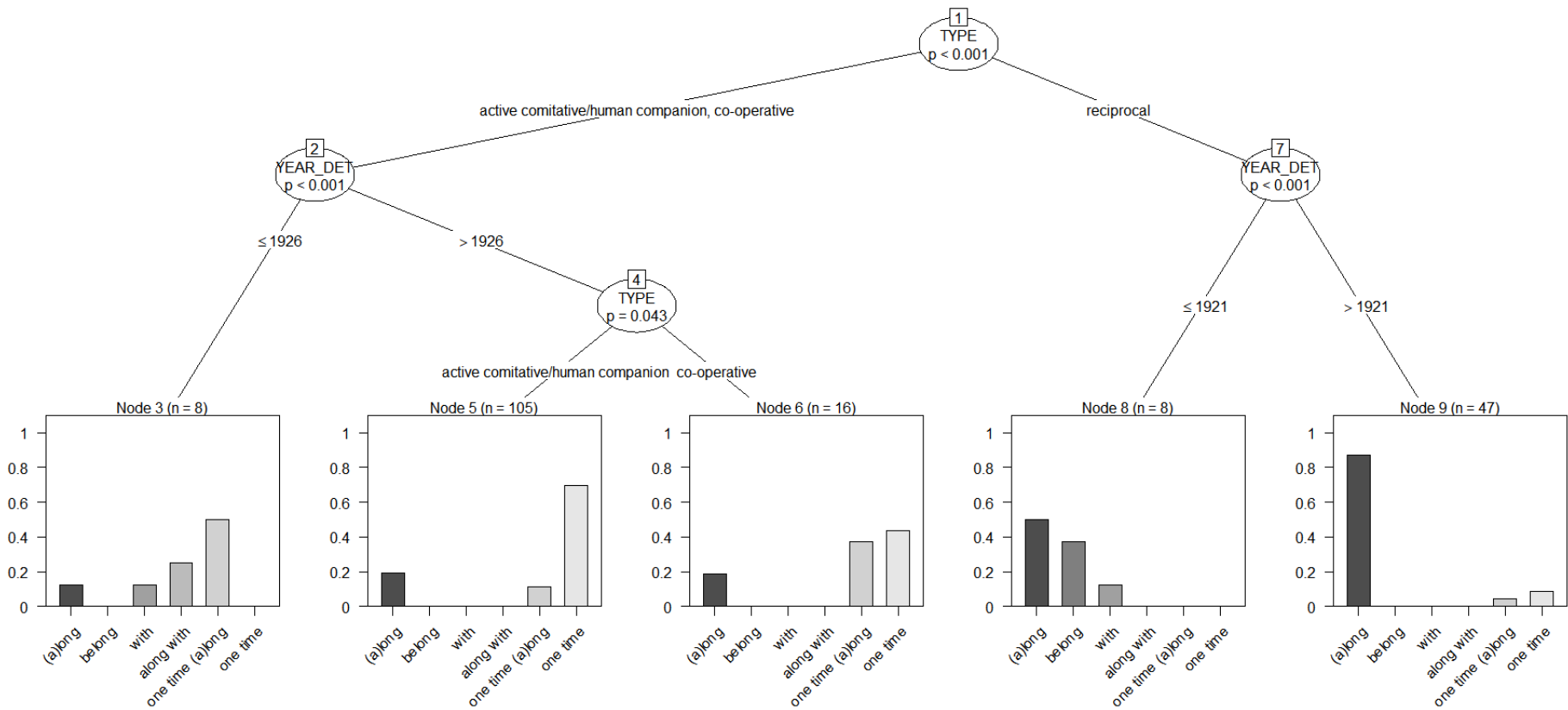


Figure 144: Conditional inference tree for comitative prepositions in TP ST_FORM~YEAR_DET+TYPE

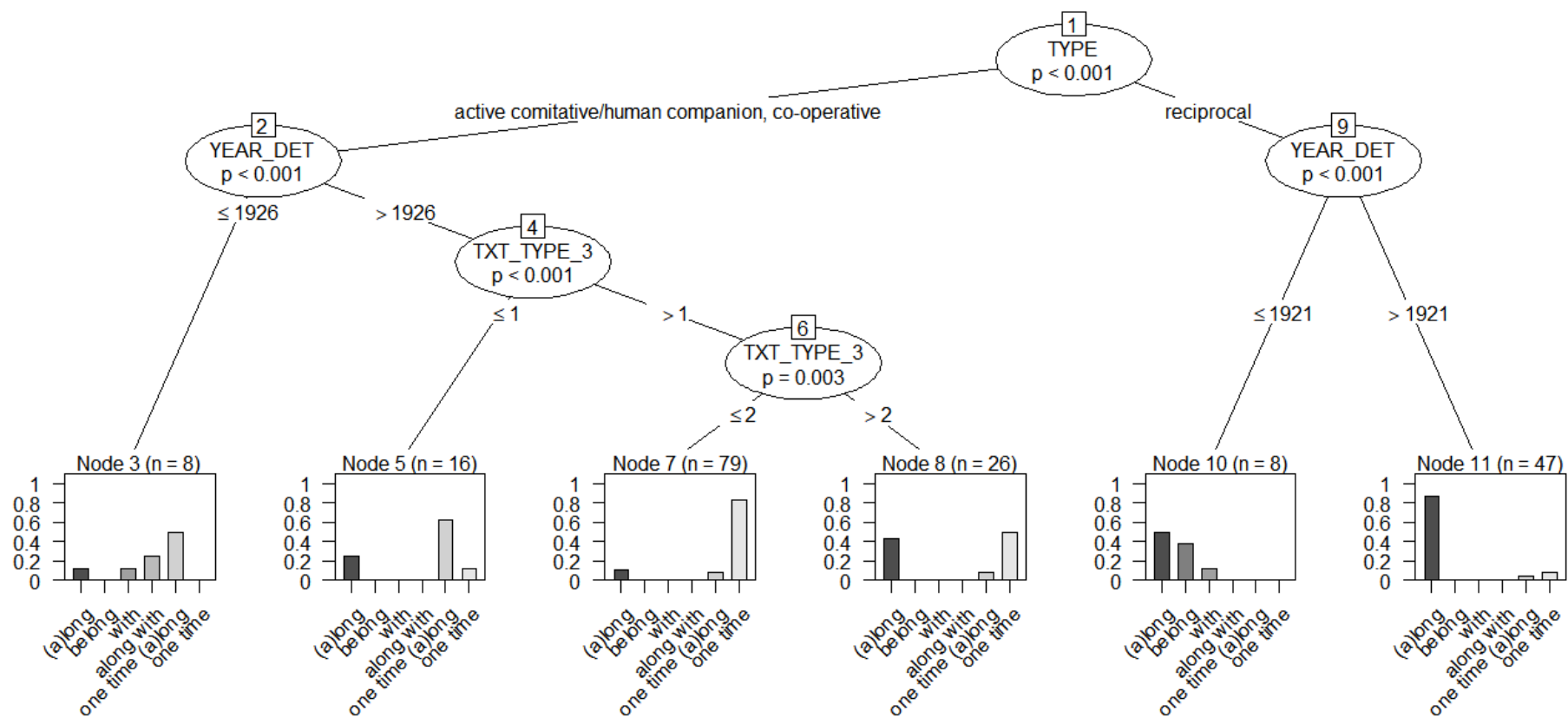


Figure 145: Conditional inference tree for comitative prepositions in $TP\ ST\_FORM \sim YEAR\_DET + TYPE + TXT\_TYPE\_3$

9.4.1.4.9 Instrumental prepositions in TP

In the following the focus will be placed on instrumental tokens in early TP. With 261 tokens, the early TP data collection contains the highest number of instrumental prepositions. This section will start by introducing the attested relators before it will focus on their attestations across time. The ctree algorithm will provide insights on whether time represents an important predictor variable in how instrumental prepositions are encoded in TP.

9.4.1.4.9.1 Attested forms

Four different prepositional forms were attested to have been in use in instrumental semantic contexts. The most common form in the early data represents *(a)long*, which was attested in 97.70% (= 255/261) of the tokens. Example (232) illustrates the use of *(a)long* in an example in which a body part instrument (= teeth) occurs.

- (232) *mek-nais* **long** *tit*
 make-noise **INST** teeth
 ‘crunch with teeth’

(Vunapope ~1940; Kutscher 1940: 133)

One time is attested four times (4/261= 1.53%) and was attested twice with instrumentals that relate to a tool and twice in semantic contexts in which the instrument is a human being, as example sentence (233) shows. The remaining two attested forms *belong* and *one time long* are only attested once. As becomes visible in sentence (234), *belong* is attested with a body part instrument and *one time along* in sentence (235) is followed by an abstract instrument.

- (233) *Amerika* *im-i* *halip-im* *yumi* **wantaem** *soldia* [...]
 American 3SG-PM help-TR 1PL.INCL **INST** soldier [...]
 ‘The Americans help us/provide us with soldiers ...’

(Sepik 1943-45; Kerr 1985; NLA MS9002 PM10)

- (234) *Whitey man* *he* *tight* *him* **belong** *hand* [...]
 white man PM hit TR **INST** hand [...]
 ‘The European hits it with his hand ...’

(Madang 1894; Cayley-Webster 1898: 34)

- (235) **Wan taim log** *qud* *fela* *fasin* [...]
 INST good MODIF fashion [...]
 ‘with good behaviour’

(Malakuna 1935; Methodist Mission 1935: 15)

9.4.1.4.9.2 Timeline and boxplot approach

Although there is a clear dominance of the form *(a)long* in instrumental contexts, a timeline was created to investigate when the individual forms were attested across time. As Figure 146 demonstrates, the relator *(a)long* is attested from 1908 to 1945 and spreads most across time.

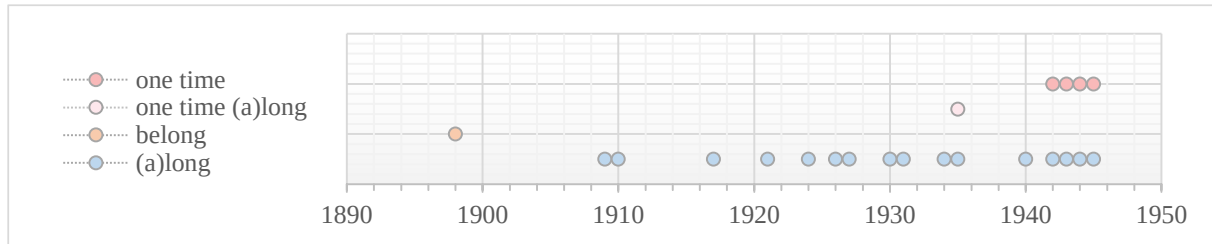


Figure 146: Timeline instrumental variants in TP

The form *one time* is attested in the 1940s and *one time along* in the 1930s for the first time to encode instrumental contexts. *Belong* represents the earliest attested form but was only attested once in 1898.

If time periods are replaced by single years (YEAR_DET), similar results are obtained:

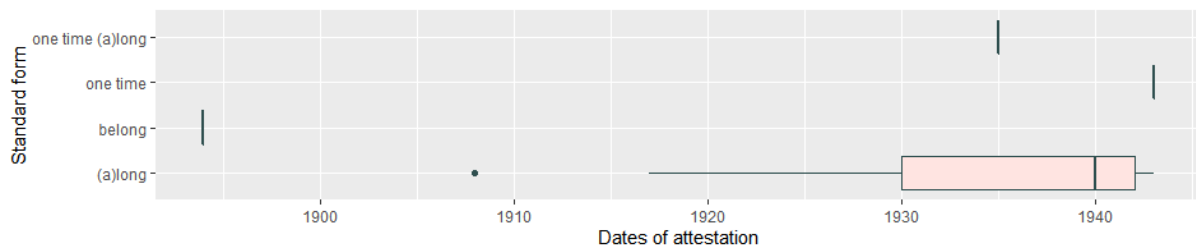


Figure 147: Boxplot of instrumental variants in TP

The relator *(a)long* represents the form that spreads most across time, while other forms seem to represent exceptions. Thus, no time split is assumed. From the timeline it can be learned that the majority of datapoints of *(a)long* date to the 1940s.

9.4.1.4.9.3 Testing for the impact of the years of attestations on the choice of form

Against expectations, the analysis reveals two highly significant splits in the years 1942 (cf. Node 1; $p < 0.001^{***}$) and 1926 (cf. Node 2; $p < 0.001^{***}$) if the ctree algorithm is performed analysing the impact of the year of attestation on the form:

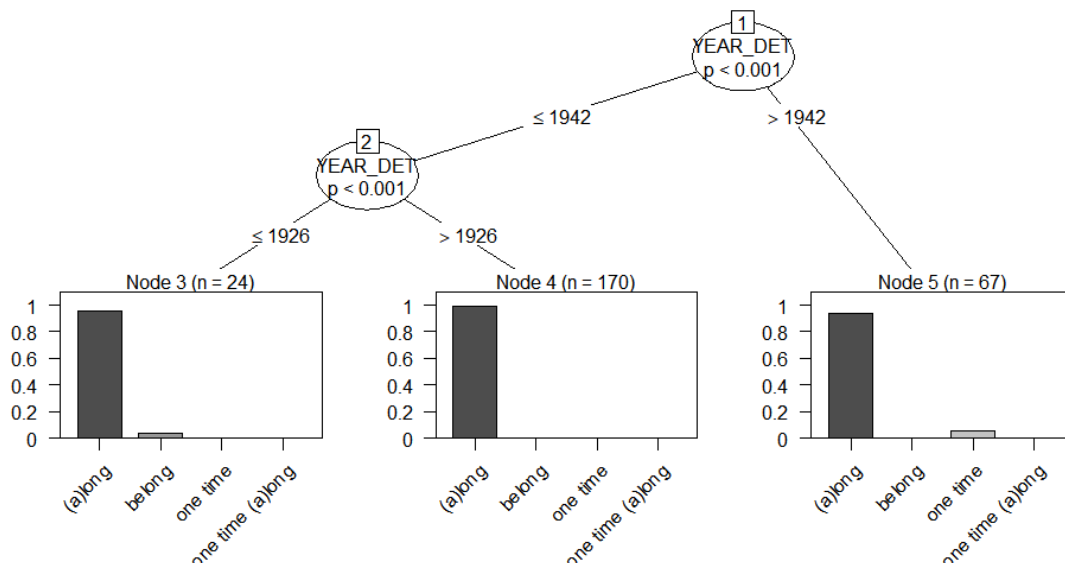


Figure 148: Conditional inference tree for instrumental prepositions in TP ST_FORM~YEAR_DET

However, all three end nodes show a clear preference of the form *(a)long*. The differences between the end nodes are that while in Node 3 *belong* represents a variant of *(a)long*, in Node 4 *(a)long* is the exclusive form and in Node 5, *one time* is displayed as a possible variant.

If the text type is added, the year of attestation remains the most important predictor variable. As shown in Figure 149, the years 1942 and 1926 are still split-evoking years.

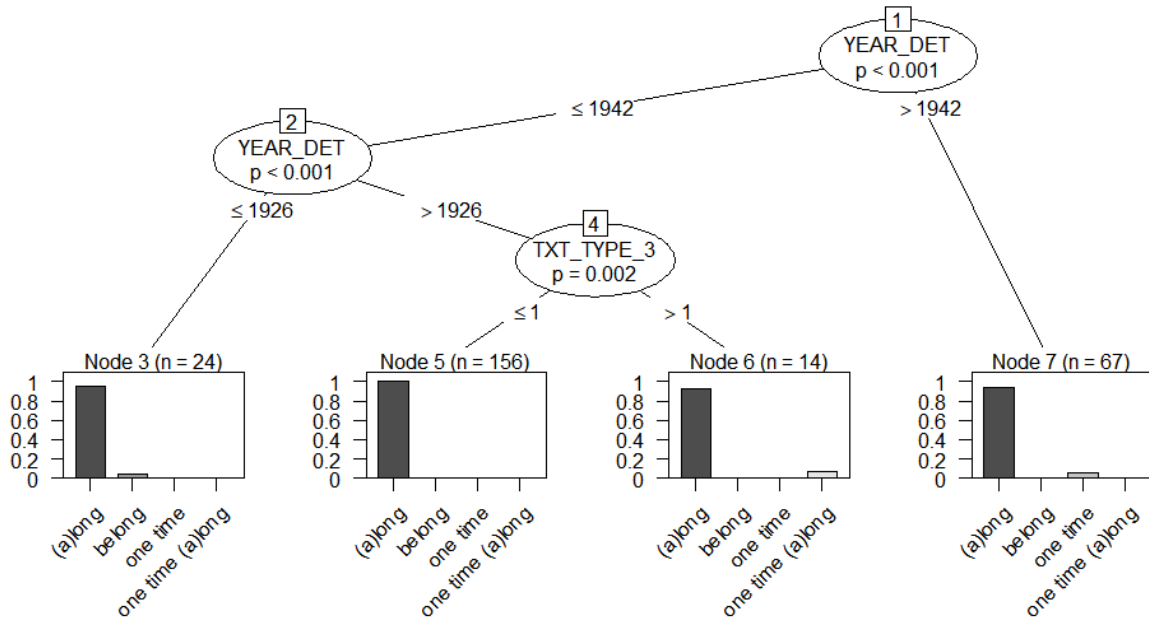


Figure 149: Conditional inference tree for instrumental prepositions in TP ST_FORM~YEAR_DET+TXT_TYPE_3

The only difference that can be observed is that data attested between 1927 and 1942 shows a split based on the text type. While speech-related attestations in that time period are exclusively encoded with *(a)long*, there is a 10% chance in written and intermediate attestations to be encoded with *one time (a)long*.

If the possible predictor variable TYPE and thus the fine-grained classification system is included into the algorithm, a tree consisting of two highly significant splits is obtained. While the highest-level split is based on the fine-grained categories, the second split is based on the year of attestation (cf. Appendix II). While for instrumental relators that refer to body part instruments, a material/substance, a tool, a monetary tool or a transportation tool a time split can still be observed in 1926, relators encoding abstract instruments and human instruments are most likely to be encoded with *(a)long*, or less likely with *one time* and *one time (a)long* independent of time. If the text type variable is subjoined, the tree structure does not change.

If the author variable is added as an additional possible predictor variable, the tree structure changes, and a split based on the year of attestation is no longer observable. Instead, two splits based on the author are yielded. The differences observed thus seem to depend on the authors only and a clear preference of the form *(a)long* to encode instrumental semantic contexts is observed.

9.4.1.4.10 Comitatives and instrumentals in TP

If all TP comitative and instrumental tokens are considered as a whole and the ctree algorithm is applied examining the impact of the year of attestation (YEAR_DET) and the dual categories comitatives versus instrumentals (FEATURE), a tree consisting of seven splits is obtained.¹⁸⁵ The highest-level split indicates that comitatives and instrumentals seem to behave differently (FEATURE; $p < 0.001^{***}$). The individual time-splits which are observable for instrumentals and comitatives are those that were previously discussed. The same applies if the text type is added. The highest-level split still separates comitatives from instrumentals and the remaining splits are those that were discovered when comitatives and instrumentals were analysed separately. What is interesting is that when the dual categories (FEATURE) are replaced by the fine-grained categories (TYPE), the highest-level split is based on the latter.¹⁸⁶ While *active comitative/human companion*, *co-operative* and *human instrument* cluster together, the remaining categories are distinguished from them. While the former show a time-based split in 1926, the latter show a time-based split in 1914. In addition, while the latter show a clear preference of the form *(a)long* after 1914, despite a further fine-grained category split (cf. Node 4), *active comitative/human companion*, *co-operative* and *human instrument* show a clear preference of the form *one time* after 1926. It is possible that the human instrument clusters rather with two comitative categories since it is in reference to a human being. In contrast, the reciprocal category clusters with instrumental tokens showing *(a)long* as the most likely form.

If the impact of the author, the year of attestation, the text type and the dual categories is analysed, no significant time-based splits are observed. Instead, the resulting tree yields five author-based splits and two splits based on whether the preposition has a comitative or instrumental meaning.

9.4.1.4.11 Comparative analysis of comitatives and instrumentals in MPE

After having looked at comitatives and instrumentals in each of the varieties, the results will be compared to learn more about the varieties' divergence.

¹⁸⁵ The tree structure is too large to be shown here but is available at https://www.dropbox.com/s/ev499on52ju8rq9/Ctree_Comins_TP_ST_FORM~YEAR_DET%2BFEATURE.png?dl=0 (last access 29 September 2021).

¹⁸⁶ See https://www.dropbox.com/s/nkeshihs23agpap/Ctree_Comins_TP_ST_FORM~YEAR_DET%2BTYPE.png?dl=0 (last access 29 September 2021).

9.4.1.4.11.1 Comparative analysis of comitatives in MPE

The individual analyses of comitative prepositional forms in the three MPE varieties have shown that SIP, BIS and TP behave differently in how they encode comitatives. It also became clear that depending on the contact variety, different predictor variables were of importance.

For the SIP data, the years 1895 and 1923 were detected as significant split evoking years, independent of the text type and author variable. Data attested before 1895 was characterised by a high degree of variation, whereas *(a)long* was attested as the most likely form to encode comitatives from 1896 to 1923. From 1923 onwards, *with* represented the most likely form to encode comitative contexts, and the form *with him*, which is the form that is used in SIP today, showed a small probability as well. The earliest attestation of *with him* dates to 1930. It is also of interest that as soon as the more fine-grained semantic categories were included (TYPE), the year 1907 turned out to be an important predictor for the choice of form. While comitatives classified as *reciprocals* showed a preference of the form *(a)long* independent of the year of attestation, the remaining comitatives showed a split in 1907, with post-1907 data being most likely encoded with *with*. The development and stabilisation of *with him* does not seem to have been completed by the end of 1950 as variants still coexisted.

In BIS the year 1905 was determined as a significant year. While prior to 1906 the form *(a)long* was most likely to occur in comitative semantic contexts, after 1905 the forms *(a)long* and *with him* turned out to be equiprobable variants. The fine-grained semantic categories, according to which the prepositions can be classified, had no impact on the form, whereas the author turned into the only significant predictor variable if all possible predictor variables were included into the analysis. The earliest attestation of *with him*, the form that is still used nowadays in BIS, dates to 1907.

The form *with him*, which developed into the major form to encode comitatives in contemporary SIP and BIS, was not attested at all in TP. There are attestations of *with* in data attested prior to 1914, which in general was characterised by a greater degree of variation. Afterwards, no further attestations of *with* are found. From 1914 to 1930 *(a)long* represents the dominantly attested form which is outnumbered by *one time (a)long* and *one time* after 1930. However, these time splits were no longer significant if the author variable was considered. Furthermore, the fine-grained classification had an impact on the choice of comitative forms as well. Tokens classified as reciprocal or passive comitatives were most likely to be encoded with *(a)long*. The remaining tokens showed a significant time split in the year 1926, with attestations dating after 1926 being most likely encoded with *one time*. In all three varieties the present-day forms were already attested but variants could still be observed.

If a ctree based on the data of all three varieties is created analysing the impact of the year of attestation and the variety, the resulting tree shows that the variety is the strongest predictor variable.¹⁸⁷ The TP data is shown to behave differently than the BIS and SIP data ($p < 0.001^{***}$). Node 2 indicates that the year 1895 is the significant year for comitative tokens in BIS and SIP. The data prior to and after 1895 further shows a split based on the variety. Only SIP shows a further time-based split in the year 1923 (Node 8), with *with* turning into the most probable form after 1923. In TP *one time* turns into the most likely form around 1934 and the previously observed time splits can be observed as well. If the text type variable is added, the tree changes Nodes 13-19.¹⁸⁸ The year 1930 only represents a significant predictor variable for TP speech-related attestations. If the author variable is added, the latter turns into the dominant predictor variable. A tree with six splits is obtained, whereby four of these splits are author-based and two are text type-based.¹⁸⁹

9.4.1.4.11.2 Comparative analysis of instrumentals in MPE

In contrast to the comitative analysis, the individual analyses of instrumentals in SIP, BIS and TP have shown that the three varieties are similar in using *(a)long* as the most dominant form to encode the instrumental semantic function.

In SIP *(a)long* was attested with highest frequencies independent of the predictor variables and thus independent of the factors time, text type, author, and type. At the same time, a relatively high frequency of the form *with* was attested. These two forms also represented the major instrumental relators in early BIS. The ctree analysis revealed that although *(a)long* occurred as the dominant form throughout time, prior to 1908/1909 the form *with* represented a common variant in BIS, but the use of the latter was reduced after the mentioned years. Although *(a)long* also represented the dominant variant in TP, a split in the data was observed inter alia in 1942 because afterwards the forms *one time* and *one time (a)long* were both attested once in instrumental position as well. The analysis revealed that the fine-grained semantic categories had an impact on the choice of form, with abstract and human instruments being encoded differently from the remaining instrumental tokens.

¹⁸⁷ The tree is too large to be displayed here but can be viewed online at https://www.dropbox.com/s/99rydh3xwban78k/Ctree_Com_MPE_ST_FORM~YEAR_DET%2BVARIETY.png?dl=0 (last access 29 September 2021).

¹⁸⁸ The tree can be found online at https://www.dropbox.com/s/sknp7ou4b6p10ci/Ctree_Com_MPE_ST_FORM~YEAR_DET%2BVARIETY%2BTEXT_TYPE_3.png?dl=0 (last access 29 September 2021).

¹⁸⁹ See https://www.dropbox.com/s/w33atl2y3xv6k62/Ctree_Com_MPE_ST_FORM~YEAR_DET%2BTEXT_TYPE_3%2BAUTH_NAME%2BVARIETY.png?dl=0 (last access 29 September 2021).

If the ctree algorithm is applied to the instrumental tokens of all three varieties, and the impact of the variety and the year of attestation is measured, the tree in Figure 150 is obtained. The tree consists of three splits. The highest-level split is based on the variety ($p < 0.001^{***}$) and separates SIP from BIS and TP. Only the BIS and TP data are characterised by two further splits, namely Node 2 which is highly significant and based on the year of attestation and Node 4 which is very significant and splits the data based on the variety.

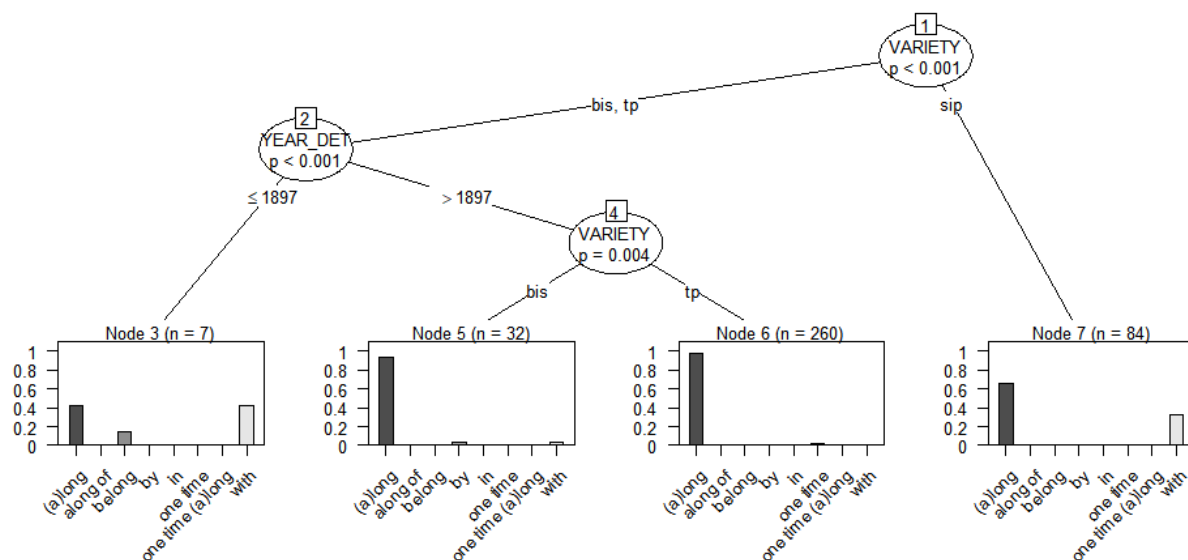


Figure 150: Conditional inference tree for instrumental prepositions in the three MPE varieties $ST\_FORM \sim YEAR\_DET + VARIETY$

Data attested prior to 1897 is equiprobable in encoding instrumentals with *(a)long* or *with*. After 1897, both varieties are most likely to encode instrumentals with *(a)long*, but the attested variants differ. Although *(a)long* is also the most likely occurring form in SIP, there is a more than 35% chance of the instrumental to be encoded with *with*. If the text type variable is added, the tree structure does not change.¹⁹⁰ Only if the author variable is included, the resulting tree consists of four splits, two of which being based on the latter.¹⁹¹

9.4.1.5 Summary

The analysis has shown that all three varieties show a clear dominance of the form *(a)long* throughout the observed period to encode instrumental semantic contexts. In the early years, the data provides evidence that *with* served as a variant in SIP and BIS. It makes sense to assume that Europeans introduced and used *with* in the Pacific in their interactions with Pacific Islanders.

¹⁹⁰ The p-value of Node 4 changes to $p = 0.006^{**}$.

¹⁹¹ See

https://www.dropbox.com/s/z6wfs90z5ajnpyd/Ctree_Ins_MPE_ST_FORM~YEAR_DET%2BVARIETY%2BXT_TYPE_3%2BAUTH_NAME.png?dl=0 (last access 29 September 2021).

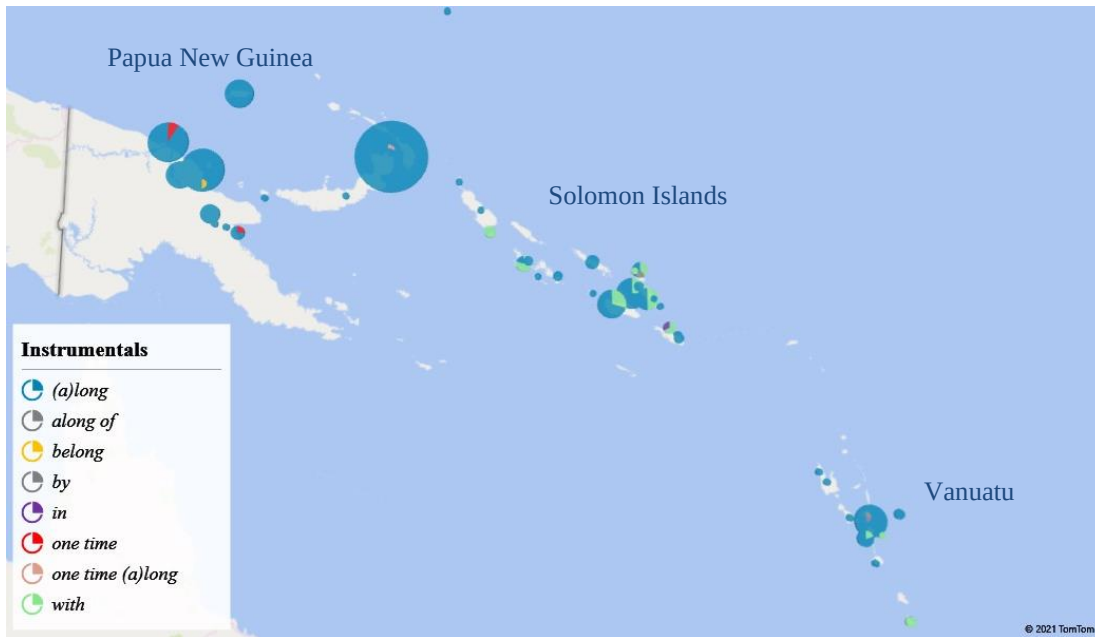


Figure 151: Map of Melanesian Islands, showing attestations of instrumental prepositional variants

Differences between the three varieties' developments were observable when focussing on forms used to encode the comitative function. While the use of *with* increases over time in the SIP and BIS data and first attestations of the verbal preposition *with him* can be attested in 1930 and 1907 respectively, TP only shows three attestations of *with* which all date prior to 1915. After 1915, *one time* develops into the comitative marker in TP. *One time along* was first attested in 1926 and *one time* in 1930. Thus, the forms only seem to have established after the end of the labour trade. In the SIP and BIS data *one time* is only attested as an adverb with the meaning 'at once'.

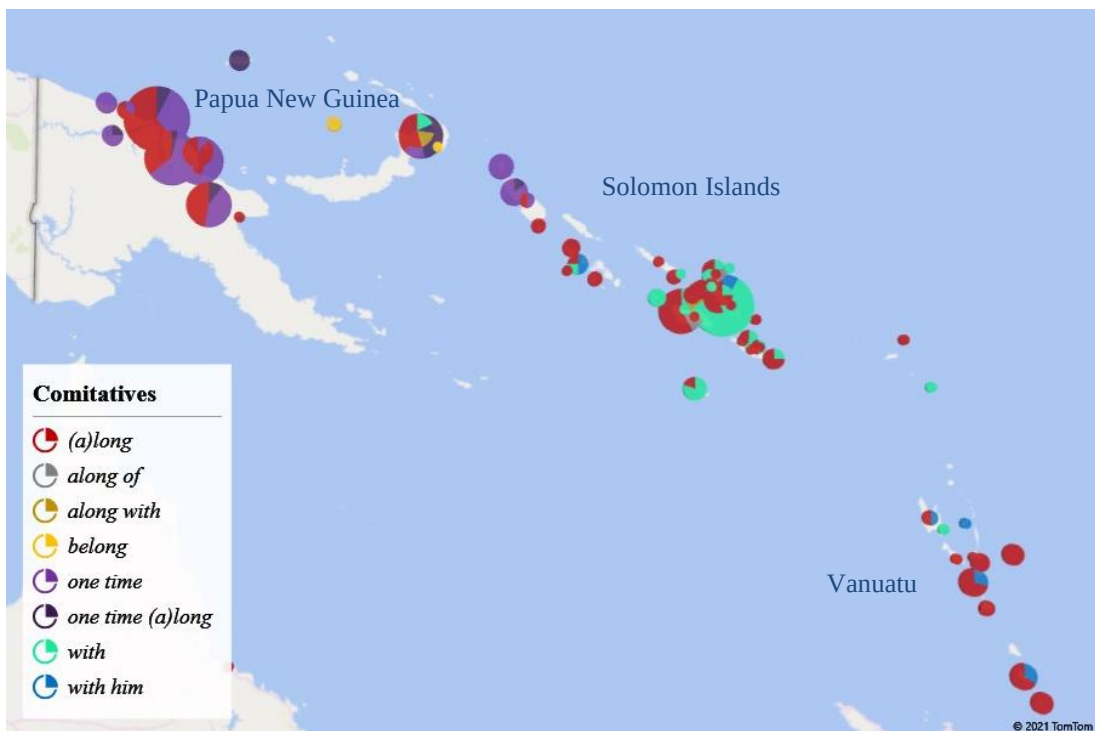


Figure 152: Map of Melanesian Islands, showing attestations of comitative prepositional variants

In the prevalent dataset, only the beginnings of the use of the verbal preposition *with him* in SIP and BIS can be observed. Post-1950 data is required to observe the further development of the forms and to reconstruct the complete history of the forms.

9.4.2 Terminatives

This section will focus on prepositions which express a limitation in space and/or time. Therefore, it focusses on what was classified as the *terminative semantic function* by Hagège (2010: 261, 293). Languages may make use of several markers to encode the terminative function, but they may also make use of a single marker only, independent of whether a limitation in space or time is expressed. English, for instance, belongs to the former as it has several terminative prepositions, involving *down to*, *as far as* to encode a limitation in space and *till*, *until* to encode a limitation in time.

The morpheme-by-morpheme analysis of the early MPE data revealed that distinct markers seem to have developed to encode the terminative in the first half of the 20th century. The aim of the present chapter is thus to investigate whether SIP, BIS and TP developed similar or dissimilar terminative markers and when the varieties began to show differences. The chapter will start by providing an overview over terminative markers used in the contemporary MPE varieties before some methodological considerations will be outlined. In Section 9.4.2.3 the findings will be presented and discussed before a final summary will bring this section to an end.

9.4.2.1 Terminatives in contemporary MPE varieties

The contemporary MPE varieties show differences in the prepositions used to encode the terminative function. In *SIP* the verbal preposition *kasem* is used which has grammaticalised out of the transitive verb *catch him*. While Huebner & Horoi show in 1979 that *kasem* occurs with preceding reduplicated *go* (cf. for instance (236)), the examples listed by Jourdan (2002: 93) and Beimers (2009: 103) provide evidence that the verbal preposition *kasem* can also appear in isolation to encode the terminative function (cf. sentence (237)).

- (236) *Bae mi stap long Solomon go go kasem namba ten long Mei.*
 FUT 1SG stay PREP Solomons TERM number ten PREP May
 ‘I’ll stay in the Solomons until the tenth of May.’

(Huebner & Horoi 1979: 190)

- (237) *Mifala waka kasem belo.*
 1PL work TERM bell
 ‘We worked until the bell rang.’

(Jourdan 2002: 93)

What is interesting is that also in *BIS* the transitive verb *catch him* grammaticalised into a verbal preposition that is used to encode the terminative semantic function. As in *SIP*, the form may express spatial limitation, as in sentence (238), or it may express temporal limitation, as in example (239).

- (238) *Bae yumitu wokbaot kasem en blong taon.*
 FUT 1DU.INCL walk TERM end POSS town
 ‘Let’s walk as far as the end of town.’

- (239) *Olgeta ol-i praktis kasem medelnaet wantaem.*
 3PL 3PL-PM practise TERM midnight at.one.time
 ‘They practised right until midnight.’

(Crowley 2004: 134)

Tok Pisin differs from *SIP* and *BIS* in how it encodes the terminative semantic function. The marker *inap long* has grammaticalised into the terminative preposition which is used to encode spatial as well as temporal limitation (cf., for instance, Verhaar 1995: 247 and Smith 2002: 121). As examples (240) and (241) show, the verb that is preceding the marker can optionally be reduplicated to express duration. Another possibility to express ‘until’ is through verb serialisation with the full verb being followed by *i go i go* and following *na*, as example (242) shows. The full verb may also be followed by *i go* or *i kamap* to encode the semantic function (cf. Verhaar 1995: 103).

- (240) *Wet inap long dram i kol pinis.*
 wait TERM drum PM cold COMPL
 ‘Wait until the drum has cooled off.’

- (241) *Ol i mek-im mek-im inap long taim ol i les.*
 3PL PM make-TR make-TR TERM time 3PL PM tired
 ‘They keep doing this until they are tired.’

(Verhaar 1995: 429)

- (242) *Sik-man i traut i go i go na i no gat kaikai i*
 sick-person PM vomit DUR so PM NEG have food PM
stap long bel bilong en.
 LOC PREP bell POSS 3SG
 ‘The patient keeps vomiting until there is no food in his stomach anymore.’

(Verhaar 1995: 113)

Keeping in mind that *SIP*, *BIS* and *TP* are said to have developed out of *MPE*, the question arises when the varieties started to develop or make use of distinct forms to encode the terminative semantic function. To answer this question and to analyse terminative forms from a diachronic perspective, several methodological steps had to be taken, which are explained in the following.

9.4.2.2 Methodological considerations

During the morpheme-by-morpheme analysis it was detected that the early MPE varieties used eight different ways to encode the terminative semantic function (cf. Table 23).

FACTOR	VARIANTS
ST_FORM	<i>catch him</i> <i>enough long</i> <i>enough</i> <i>long</i> <i>till</i> <i>until</i> <i>V.red+∅</i> <i>V.red+catch him</i>
TYPE	<i>temp</i> <i>spat</i>
FEATURE	<i>prep_term</i>
YEAR_DET	1832 ... 1950

Table 23: Linguistic coding for terminatives

Next to forms such as *until* and *till*, which seem to be direct borrowings from StE, prepositional forms were attested which reminded of the contemporary forms introduced above. Thus, *catch him*, *enough long* and *catch him* following a reduplicated verb (*V.red+catch him*) were identified as variants. Further attested forms were *enough*, *long* and a kind of zero structure (*V.red+∅*). To see whether differences exist depending on whether spatial or temporal limitation is expressed, the data was further coded for temporal (*temp*) versus spatial limitation (*spat*).

A major difficulty in identifying forms used to encode the terminative function represents the polysemy of the prepositions *(a)long* and *belong*. When the preposition occurs with transitional verbs, it cannot always be clearly stated whether there is an end point implied or not. For instance, in a sentence such as *Mary go long ples*, the preposition *long* may encode an illative or a spatial-terminative semantic function. There are even further semantic functions which the preposition *long* may encode in the sentence. Without any further contextual information it is impossible to identify the *true* semantic meaning expressed with the form. This turned out to be a major problem for the analysis of terminatives in the present section. Therefore, I decided that ambiguous datapoints should be excluded from the analysis. Consequently, the number of terminative tokens is very low, which is why the focus of the present chapter will be on the qualitative analysis of the datapoints. The aim during the following analysis is on when forms other than *long* and *belong* became used in terminative semantic contexts.

9.4.2.3 Findings and discussion

In the following section the findings of the analysis will be presented. The section starts by giving a general overview of the data distribution. This will be followed by focussing on terminatives in SIP (9.4.2.3.2), before a look at terminatives in BIS (9.4.2.3.3) and TP (9.4.2.3.4) will be taken. Finally, a comparative analysis and a summary of the findings will be provided.

9.4.2.3.1 General data distribution

In total, 56 tokens were identified in the early MPE data in which a terminal meaning was expressed. Of these, 27 were classified as early SIP (= 47.21%), four as early BIS and 25 as early TP (= 44.64%). Figure 153 shows when the individual datapoints were attested in each variety across time.

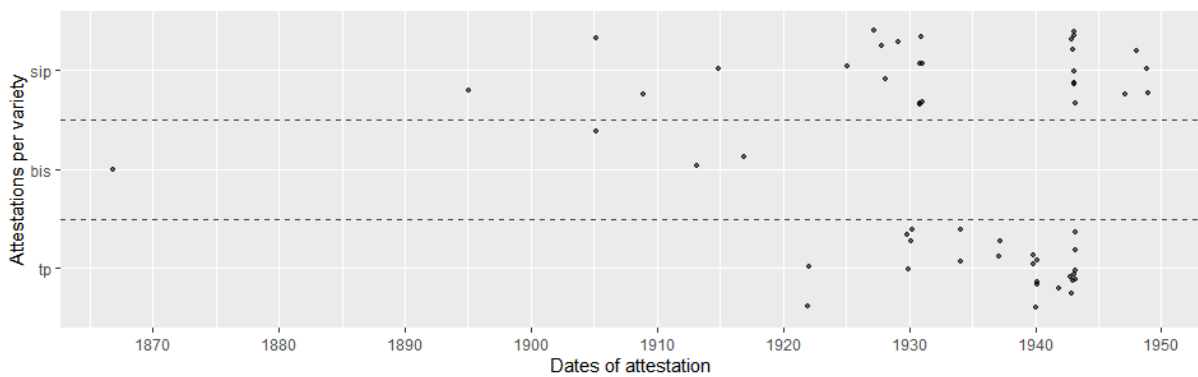


Figure 153: Distribution of terminative prepositional datapoints across time per variety

It can be learned from the figure that most datapoints date after 1900 with the majority of them being attested from 1925 onwards in SIP and TP. The number and temporal distribution of the datapoints indicates that it will be difficult to make generalising claims about the terminative forms' development in the varieties. Nonetheless, the following chapters will focus on the attested forms in each of the three varieties to investigate whether tendencies can be observed.

9.4.2.3.2 Terminatives in SIP

In this section the forms used to encode the terminative function in SIP will be investigated from a diachronic perspective. After introducing the attested forms, I will focus on when the forms occurred across time and what can be learned from this about the development of the variety.

9.4.2.3.2.1 Attested forms

In the early SIP dataset, 27 tokens were identified in which a limitation in space or time was expressed. The analysis of the tokens revealed that English-derived *till* was attested in 33.33% (= 9/27) of these tokens, as exemplified in sentence (243).

- (243) *Suppose the Lord want me to stop long O.P. till I die*
 if ART Lord want 1SG INF stay PREP O.P. **TERM** 1SG die
- me happy to stop.*
 1SG happy INF stay
- ‘If the lord wants me to stay in One Pusu until I die, I am happy to stay.’

(One Pusu 1931; Sullivan February 1931; PMB 1150)

Several remarks about the example sentence are necessary. What is striking is that the sentence contains very common PE features, such as *suppose* ‘if’, *stop* ‘stay’, *long* ‘PREP’ and *me* ‘1SG’. At the same time, several features can be identified which are more common in StE as, for instance, the definite article *the*, the infinitive particle *to*, the 1SG pronoun *I* and the lexical items *Lord* and *happy*. The terminative preposition *till* seems to cluster with the StE features. It is important to note that all nine attestations of *till* appeared in contexts which were characterised by a high amount of StE similar features. In addition, *till* was only attested in temporal contexts.

Next to *till*, there are six (6/27= 22.22%) tokens in the data, in which no overt preposition is used but the sentence nonetheless expresses a temporal or spatial limitation. Such a token is exemplified in (244). Though there is no overt preposition, the use of reduplicated *go* is salient. In the early SIP data, verbs can be reduplicated to modify the aspectual meaning of the verb. Thus, in sentence (245) the reduplication of the main verb *cry* expresses continuative aspect which is reinforced through the adverb *long time* ‘at length’. Although in both sentence (244) and (245) a verb is reduplicated indicating duration, sentence (244) differs in that *go* does not represent the main verb and in that reduplicated *go* appears in a terminative prepositional position.

- (244) *Mary she look out Jesus good fella time He small picannin go go*
 Mary 2SG take.care Jesus good MODIF time PM small child **DUR**
- time he big fella [...]*
 time PM big MODIF [...]

‘Mary took good care of Jesus from when he was a child until he was grown up.’

(One Pusu 1931, Read March 1931; PMB 1150)

- (245) *[...] me too sorry and I cry cry long time*
 [...] 1SG very sorry and 1SG **cry CONT** long time
- ‘...I was very sorry and I was crying for a long time’

(Baunani 1913; Young 1925: 232)

The most frequently attested form represents *catch him*, which was attested 12 times as a verbal preposition (= 44.44%). It needs to be noted that the form served as a multifunctional item in early SIP being attested as a full verb with meanings ranging from ‘take’, ‘catch’ (see example (246)), to ‘reach, arrive at’ as exemplified in sentence (247) and (248). In the latter two examples, *catch him* is attested to denote the end of actions and can be called a terminative verb. It is likely that the verb’s usage was expanded to prepositional environments and developed into a verbal preposition due to this shared terminative meaning.

- (246) *Me catch em one fellow musket*
 1SG **take** **TR** one MODIF musket
 ‘I took one/a musket.’

(Talisi 1902, Bennett 1976: 24)

- (247) *I go away catch-'im Malu along one o'clock on Sunday.*
 PM go away **reach-TR** Malu PREP one o'clock PREP Sunday
 ‘went away and reached Malu at one o'clock on Sunday.’

(Malu'u 1904; Young 1925: 149)

- (248) *Catch em along sundown*
reach **TR** PREP sundown
 ‘arrive (there) at night’

(~1920, Muspratt 1931: 42)

What is interesting is that in 11 of the 12 occurrences, *catch him* is preceded by reduplicated *go*, as example (249) shows. Another interesting fact is that reduplicated *go* might not encode duration in the sentence, because the duration is already expressed through the reduplication of *pray*:

- (249) *[...] me fella everyone we pray pray long early morning go go catch im supper*
 [...] 1PL all 1PL pray.CONT PREP early morning **(DUR) TERM** supper
 ‘...we all prayed from early morning until supper’

(Malaita 1929, Waite 30.08.1929; PMB 1150)

As *go go* almost always precedes *catch him*, it seems as if *go go* and *catch him* need to co-occur in order to express the terminative semantic function. As was shown in Section 9.4.2.1, Huebner & Horoi (1979: 190) identify *gogo kasem* as the terminative prepositional form in 1979. Only sources dating 20 years later identify *kasem* without reduplicated *go* as the terminative prepositional verb (cf. Jourdan 2002: 93 and Beimers 2009: 103). It is thus likely that the development from *gogo kasem* to *kasem* represents a rather late development that took place at the end of the 20th century.

Only a single token was identified in which non-reduplicated *go* was preceding *catch-him*, as shown in (250). Although this may show that reduplicated *go* was not obligatory, it is more likely that it represents an editor modification or that it is an indicator that *catchem* functions as a terminative full verb in the example. This would explain the preceding conjunction *and*:

- (250) *Tomorrow morning I want to go along Hainoria and go catchem Horalen ga*
 next morning 1SG want INF go PREP Hainoria and **TERM** Horalenga
 ‘The next morning I planned to go from Hainoria to Horalenga.’

(Tulagi; Viti, B. 19.09.1927)

9.4.2.3.2.2 Timeline approach

The timeline in Figure 154 illustrates when the individual forms were attested across time. The earliest terminative attestation is encoded with the form *till* and dates to 1895 but the form is only

attested until 1931. As previously mentioned, *till* occurred in sentences which contained several features which are closer to StE. It is possible that the form was used in more acrolectal varieties of the pidgin or that the attestations can be partly explained by editor or author revisions. The forms *go go* (V.red+zero) and *go go catch him* (V.red+catch him) are attested from 1928 and 1929 respectively onwards in the written SIP data. They are still attested by the end of the observed period. The once attestation of *catch him* dates to 1927.

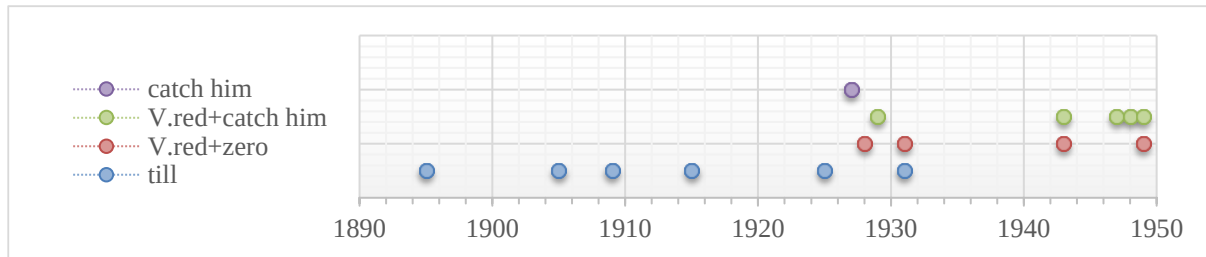


Figure 154: Timeline of terminative variants in SIP

Solely based on the timeline it appears as if a change in the choice of forms is observable around 1928 in the written SIP data. Although there are only 27 instances of terminative tokens, the ctree algorithm was applied to test whether time represents a predictor variable in the choice of form used to encode the terminative semantic function.

9.4.2.3.2.3 Testing for the impact of the years of attestations on the choice of form

Independent of whether only the impact of the year of attestation, the year of attestation and text type or the year of attestation, text type and author on the form are analysed, the resulting tree shows a significant split in the year 1931.¹⁹²

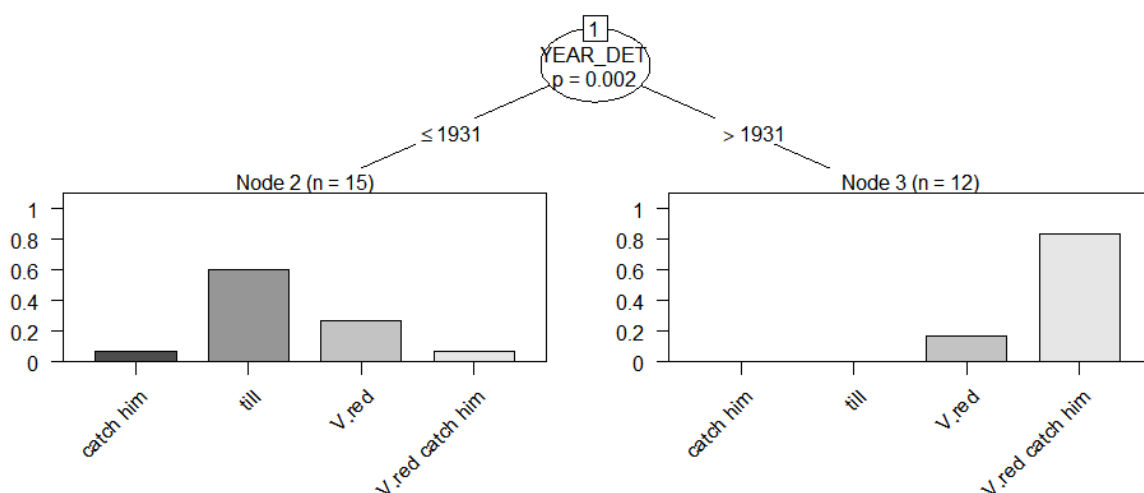


Figure 155: Conditional inference tree for terminatives in SIP ST_FORM~YEAR_DET

¹⁹² If the variables YEAR_DET and TXT_TYPE_3 are included the p-value is very significant with p=0.004** and if the variables YEAR_DET, TXT_TYPE_3 and AUTH_NAME are considered the p-value is very significant with p=0.006**.

As Figure 155 visualises, terminative datapoints attested prior to 1931 are most likely to be encoded with *till*, although the forms *catch him*, *V.red* and *V.red catch him* were already attested in 1927/1928/1929. Data attested after 1931 is most likely to encode the terminative semantic function with *go go catch him* (*V.red catch him*). Although the results are very significant, it should be kept in mind that the number of tokens was only 27.

9.4.2.3.3 Terminatives in BIS

After having focussed on terminatives in SIP, the present section will concentrate on terminatives in BIS. First, the attested forms will be introduced before their dates of attestation will be considered.

9.4.2.3.3.1 Attested forms

Only four sentences were counted in which a terminative prepositional phrase was identified in the BIS data. Once the form *until* and twice shortened *till* were used for encoding termination, the latter being exemplified in (251):

- (251) *I wait, no build-'im school till you come back*
 1SG wait NEG build-TR school TERM 2SG come back
 'I wait, I do not build a school until you return.'

(Port Vila 1905; Young 1925: 144)

As in SIP, the attestations of *until* and *till* occur in very acrolectal contexts in which several StE features can be detected. Thus, their occurrence either is a result of author/editor modification or can be explained by lectal variation.

A single attestation of the form *catch him* to encode termination was attested in the dataset. Sentence (252) shows that *catch him* follows a reduplicated verb but does not require reduplicated *go*. As in SIP, the form *catch him* also occurred with other meanings in the early BIS dataset. It is used as a full verb meaning 'get, receive' and as a full verb meaning 'reach, arrive at', the latter being exemplified in (253).

- (252) *Another day me fellow me work, work, catch him dinner [...]*
 another day 1PL 1PL work.CONT TERM dinner [...]
 'Another day we were working until dinner.'

(Atchin, New Hebrides 1917; Stewart 28.04.1917)

- (253) *[...] four week-s and five day-s me catch him Atchin*
 [...] four week-PL and five day-PL 1SG reach-TR Atchin
 'after four week and five days I/we reached Atchin' / 'four weeks and five days till I/we reached Atchin'

(Atchin, New Hebrides 1917; Stewart 28.04.1917)

The difference between *catch him* 'arrive, reach' and the prepositional use of *catch him* is the position in which the marker occurs in the sentence. When functioning as a terminative verb, as

shown in Chapter 8, *enough* has entered TP via the lexifier language but combined with *long*. It followed the common grammaticalisation path of SUITABILITY (= to be sufficient/enough) < ABILITY but, as the example shows, seems to have grammaticalised simultaneously into a terminative preposition.

- (255) *Long tu-dark i-naf long morning*
 PREP night TERM morning
 ‘from night until morning’

(Alexis Harbour ~ 1940; Catholic Mission October 1940: 8)

Enough was attested twice without following *long* as exemplified below:

- (256) *Trifelo i stop olsem, inaf Yesus i holdim skin bilong ol*
 3PL PM stay like TERM Jesus PM hold-TR skin POSS 3PL
 ‘The three stayed like this until Jesus touched their skin.’

(Alexis Harbour ~1935; Wolf 1935: 30)

9.4.2.3.4.2 Timeline approach

As Figure 157 shows, *long* is the earliest form, which is first attested in 1924. The form is still in use to encode terminative contexts in the 1940s. *Enough* is only attested in 1934 and 1940. The form *enough long* is attested from 1930 onwards. Thus, variation in the choice of forms used to encode terminative prepositions seems to have prevailed until at least 1945. It can be expected that the development towards the exclusive use of *inap long*, which is attested nowadays as the terminative preposition, has only been accomplished at some point in time after 1945.

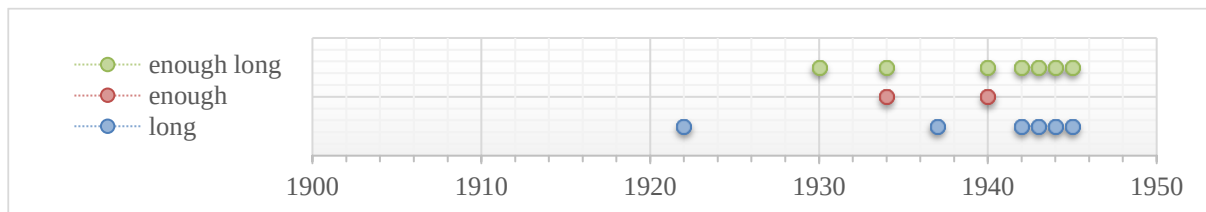


Figure 157: Timeline of terminative variants in TP

9.4.2.3.4.3 Testing for the impact of the years of attestations on the choice of form

As the timeline does not consider frequencies of occurrence and other possible predictor variables, the ctree algorithm was applied. Independent of whether only the year of attestation, the year of attestation and the text type or the year of attestation, the text type and the author are considered as possible predictor variables, the analysis results in a stacked boxplot. This shows that based on the data at hand, neither time, the text type or author have a significant impact on the choice of the terminative encoding, but *enough long* represents the dominant form independent of these factors.

9.4.2.3.5 Diachronic comparative analysis of terminative prepositions in MPE

A comparison of early terminative prepositions in SIP, BIS and TP can only be done with reservations due to the low frequencies of tokens in BIS and the low frequencies of attestations in general.

If the ctree algorithm is applied to the data of all three varieties and the impact of the year of attestation and of the variety on the form is investigated, the tree structure as displayed in Figure 158 is obtained. The strongest predictor variable on terminative prepositions is the variety, which is highly significant with $p < 0.001^{***}$. The first split separates TP from BIS and SIP, showing that if the data can be located to the New Guinea area, the terminative prepositional function is most likely encoded with the form *enough long*. If the data was attested in SIP or BIS, a second significant predictor variable comes into play. While BIS and SIP data attested before 1931 is most likely to be encoded with *till*, post-1931 attestations are most likely to be encoded with *V.red+catch him*. The tree structure stays the same independent of whether the text type is considered as a further predictor variable.¹⁹³

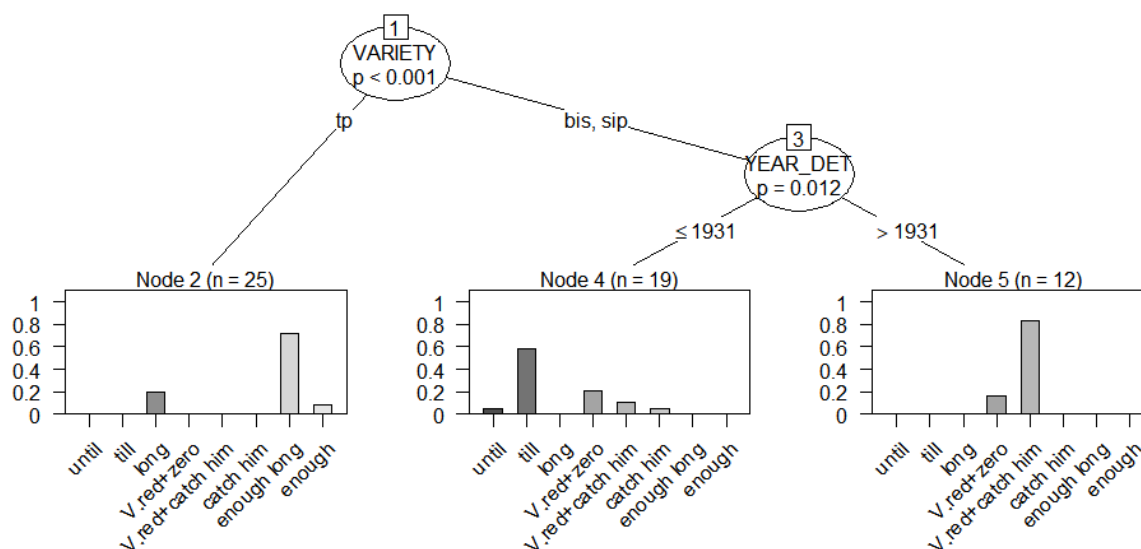


Figure 158: Conditional inference tree for terminatives in the three MPE varieties ST_FORM~YEAR_DET+VARIETY

However, if the author is added as a further possible predictor variable, the tree structure changes. As visualised in Figure 159, the resulting tree consists of three splits, whereby the highest-level split remains being based on the variable VARIETY. However, the second-level split is based on the author. The authors Young, Sullivan, N. Deck, K. Deck and Thurston are separated from the remaining authors. The former are most likely to encode the terminative semantic function with *till*. Interestingly, four of the five authors are of the SSEM which, as mentioned earlier, frequently used StE features or a more acrolectal variety.

¹⁹³ The p-level of Node 2 changes to 0.018*.

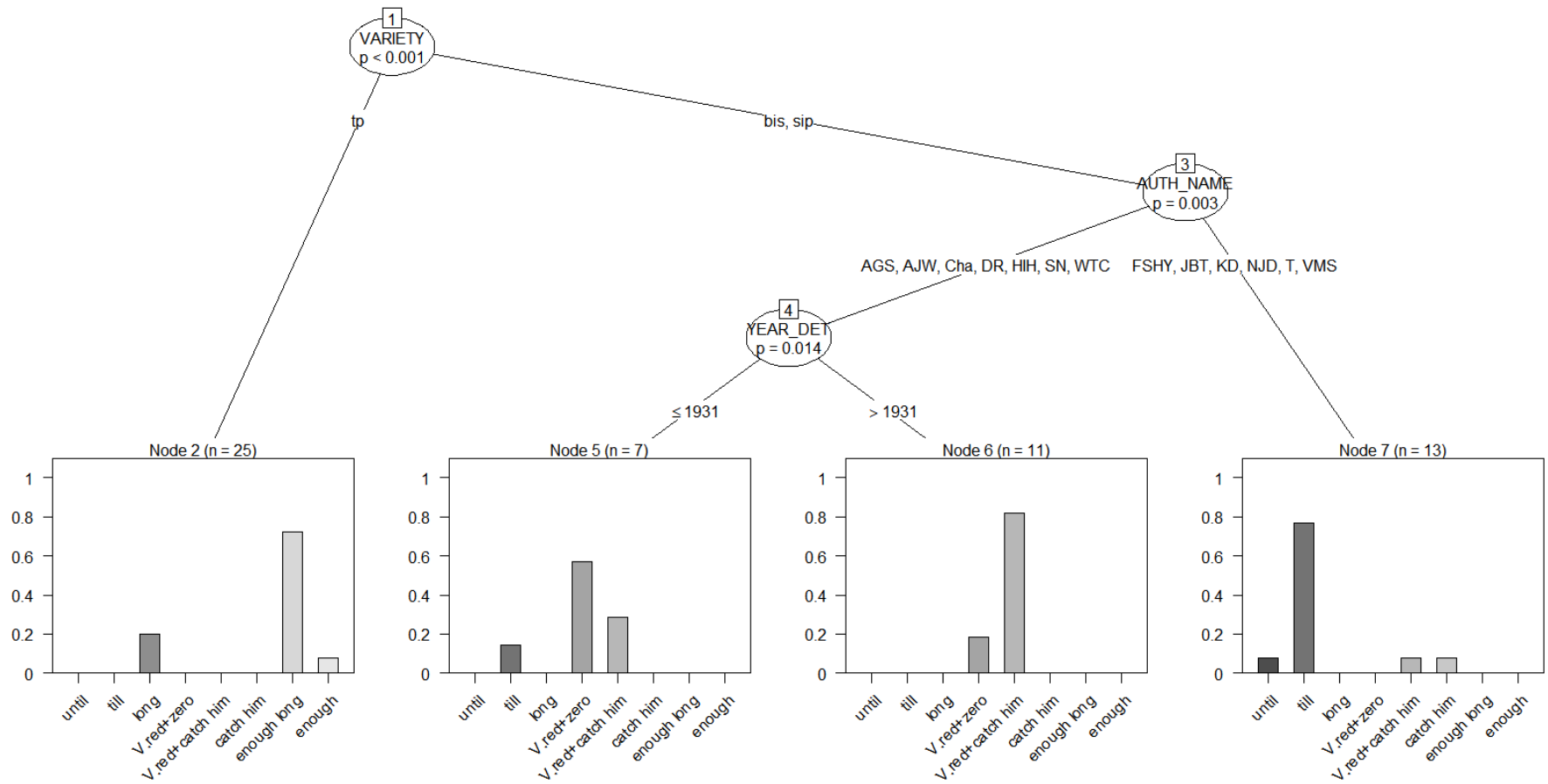


Figure 159: Conditional inference tree for terminatives in the three MPE varieties ST_FORM~YEAR_DET+VARIETY+TXT_TYPE_3+AUTH_NAME

The third split is based on the year of attestation and indicates that BIS and SIP data of the remaining authors which is attested prior to 1931, is equiprobable to be encoded with *catch him* or *V.red+zero*, whereas data attested after 1931 is most likely to be encoded with *V.red+catch him*.

Based on the small number of available SIP tokens, the analysis revealed that a significant difference in the choice of form is observable in 1931, indicating that the form *go go catch him* represents the most probable form for encoding terminative semantic contexts after 1931. Although the BIS data contains a token of the form *catch him* in the year 1917 as well, the form follows a reduplicated full verb (= not *gogo*). Unfortunately, only four tokens encoding the terminative semantic function were identified so that more data is required to learn how the feature has further developed and stabilised. In TP, by contrast, the ctree analysis revealed that neither time nor any other factor can be considered a significant predictor variable for the choice of form but a general dominance of the form *enough long*, which first occurred in 1930 in the written data, is observed. It can be assumed, that the forms were already used earlier in spoken language, especially as there is only a low amount of written terminatives available in total. From the dates of attestations it can be assumed that specific overt marking only developed once the varieties were beginning to stabilise. It is thus likely that the stabilisation took place after the end of the labour trade.

Nonetheless, it remains striking that SIP and BIS share the form *catch him*. It may be possible that the form grammaticalised independently in the two varieties, which would explain why the varieties differ in whether they show reduplicated *go* or not. In both varieties *catch him* may have grammaticalised as the form is likely to have entered the varieties as a terminative verb through QPPE. In TP, which differs most in how the terminative is encoded, the form *enough long* may have grammaticalised instead of *catch him*, because the latter was only attested rather late as a full verb meaning ‘to take, to reach’. By that time the form *enough long* may have already started its grammaticalisation path.

The amount of available datapoints is too low to find evidence for the above claims. Nonetheless, it can be said that differences between the three varieties existed. The contemporary forms were first encoded in 1917 in BIS, 1931 in SIP and 1930 in TP in the written data.

9.4.2.4 Summary

The analysis of terminative prepositional forms showed that although in all three varieties *catch him* was attested as a full verb, it only occurred in terminative prepositional phrases in SIP and BIS. While in BIS the form *catch him* occurred with the reduplicated verb *work*, all but one of

the SIP attestations of the form followed reduplicated *go*. In TP, by contrast, *catch him* was not attested as a verbal preposition at all and instead, the form *enough long* was found to develop out of the StE adjective *enough* into the terminative preposition. What is interesting is that there are no attestations of *enough* as an adjective or verb in the early BIS and SIP datasets.

There are too little attestations to make claims about when the features grammaticalised or first occurred in the *spoken* varieties. Nonetheless, in all three varieties the forms that are prevalent today can be attested in the written data in the late 1920s/early 1930s. The first attestation in the written data does not mean that the forms have not existed in spoken language before that time. However, since the different forms do not occur before 1910, it can be assumed that the forms stabilised only after the end of the labour trade. It is striking that both SIP and BIS make use of a form of *catch him* in terminative prepositional phrases, and that they only differ in whether reduplicated *go* is preceding *catch him* or not.

Early data of QPPE was consulted and analysed regarding its use of the form *catch him*. In QPPE *catch him* is attested as a terminative verb with the meaning ‘reach, arrive at’ but not as a terminative preposition, which supports the assumption that *catch him* grammaticalised only once the labour trade was over.

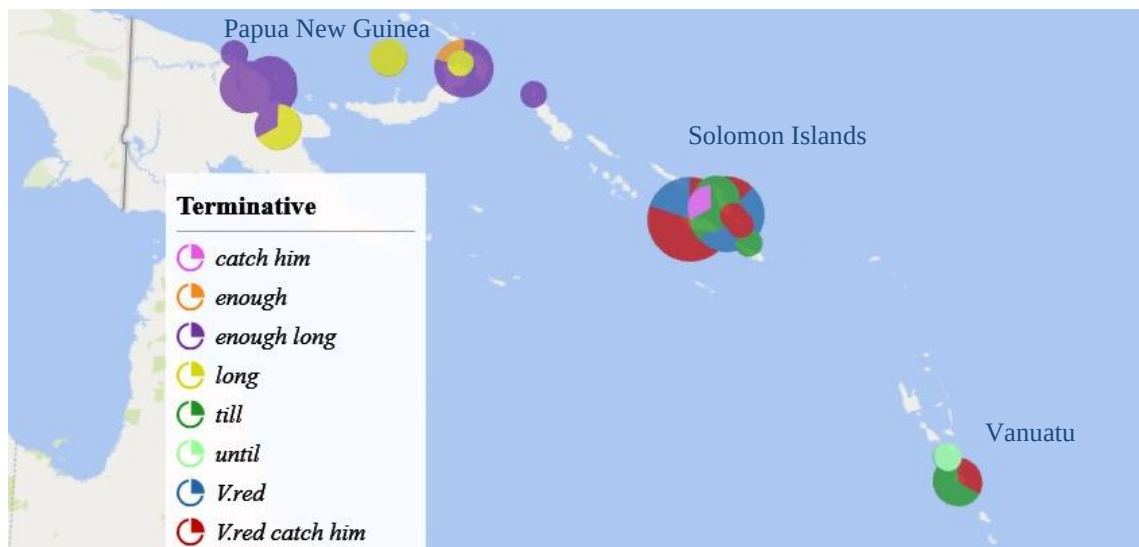


Figure 160: Map of Melanesian Islands, showing attestations of terminative prepositional variants

9.4.3 Adessives

During the glossing process two forms, namely *close to (a)long* and *close up (a)long*, were identified to be used synonymously in the early MPE data but with diverging frequencies depending on the region in which they were attested. Both forms are used to encode the prepositional meaning to “be in contiguity with a place” (Hàgege 2010: 286) and thus refer to what Hàgege (2010) defines as the *adessive* semantic function. The present section will focus on

how the adessive semantic function is encoded in the early MPE varieties across time and whether the use of the forms is regionally restricted.

9.4.3.1 Adessives in contemporary MPE varieties

The present day MPE varieties make use of complex prepositions to encode the adessive semantic function. Complex prepositions are here understood as compounds that consist of at least two words, one of which is the word *long*.

To express the *adessive* semantic function in SIP, the complex preposition *kolsap long* is used, as exemplified in example (257). Similarly, as shown in (258), the complex preposition *klosap long* is used in BIS. Both, *kolsap* and *klosap* derive from StE *close up*. In both varieties, the form can also be used without *long* to “express some kind of meaning relating to place or time without any associated noun phrases” (Crowley 2004: 137). In the latter case, the expressions function as adjectives or adverbs. While the forms used in SIP and BIS have their origin in the same StE morphemes, the form used in TP is based on the StE preposition *close to*. *Close to* has combined with *long* into *klostu long* to encode the adessive semantic function in contemporary TP. An example is shown in (259).

Solomon Islands Pijin:

- (257) *Haos blong sif hem kolsap long stoa ia.*
house POSS chief 3SG **ADES** store DEM
‘The chief’s house is next to the store.’

(Jourdan 2002: 107)

Bislama:

- (258) *Wan skul i stap klosap long ples blong hem*
one school PM LOC **ADES** place POSS 3SG
‘There is a school near his village.’

(Crowley 2004: 33)

Tok Pisin:

- (259) *em i sindaun klostu long yu*
3SG PM sit **ADES** 2SG
‘He is sitting near you.’

(Volker 2008: 37)

The contemporary formal realisations of the adessive semantic function show that the use of the two diverging forms is variety dependent. A diachronic investigation of adessive prepositional attestations is necessary to explore whether a regional distinction in the use of adessive prepositions existed right from the beginnings of the varieties or whether both forms existed as competing variants that were reduced over time in each region. In the latter case, the years of attestations are of importance to discover *when* one of the forms developed into the preferred option.

9.4.3.2 Methodological considerations

To investigate the diachronic development of adessives in SIP, BIS and TP, I extracted the tokens which were previously glossed as *prep_ades*. It needs to be noted that only those sentences were considered for the analysis, in which the forms functioned as prepositions and, thus, as grammatical tools which mark the relationship between two parts of a sentence and precede the word they govern. Moreover, sentences in which the forms functioned as adverbs, as sentence (260) shows, were not included into the analysis.

- (260) *Him he close up*
 3SG PM near
 'It [the village] is near.'

(Vanuatu 1910; Speiser 1913: 78)

The morpheme-by-morpheme analysis revealed that more than the two forms *close to long* and *close up long* were in use in adessive prepositional contexts. As Table 24 shows, eight different forms (ST_FORM) were identified in early MPE. In the following, the individual forms will be introduced with an example sentence.

The majority of the forms are built with either *close to* or *close up*. *Close to* was attested without compounding *long*, as shown in (261), with following *along*, as exemplified in (262), and with phonologically shortened *long*, as example (263) demonstrates. Due to the small number of tokens, the latter two are summarised under *close to (a)long*.

- (261) *Man i-stop clostu Iapan i kis-im bam [...]*
 man PM-LOC ADES Japanese PM receive bomb [...]
 'The man that stay near the Japanese will be bombed.'

(Salamaua 1943-45; Kerr 1985; NLA MS9002 PM27)

- (262) *He tambu for bring-im altogether pigipig close to along house*
 PM forbidden INF bring-TR PL pig ADES house
 'It is forbidden to bring pigs near/close to the house.'

(Solomon Islands 1943; Hogbin 1944: 282)

- (263) *Kru-se ken bring-im mi klos-tu long yu*
 cross ABIL bring-TR 1SG ADES 2SG
 'The cross can bring me close to/near you.'

(Alexis Harbour ~1940; Catholic Mission 1940: 24)

Similarly, *close up* was attested without any additional morpheme, with *along* and with phonologically adapted *long*, as examples (264) – (266) show. In addition, it was attested together with *along of* as exemplified in (267).

- (264) *He stop close up house*
 PM LOC ADES house
 'He was near the house.'

(Bundaberg; Enow 24.12.1888 [Australian Station 1888: 20])

- (265) *Me* **close up along** *door.*
 1SG **ADES** door
 ‘I was near the door.’

(Rendova Islands; Likin, B. 06.11.1908)

- (266) *Iu* *Rod* *i* *qo* **klos ap** **‘Log** *im* *bilog* *ol* *boi*
 2SG road PM go **ADES** 3SG PREP PL boy
 ‘You are the road for all boys to get closer to him.’

(Malakuna ~1935; Methodist Mission 1935: 10)

- (267) *They* *all* *stop* **close up along of** *boat.*
 3PL PL LOC **ADES** boat
 ‘They were all near/close to the boat.’

(Bundaberg; Enow 24.12.1888 [Australian Station 1888: 20])

Close along represents a further variant which was attested in the early data encoding the adessive semantic function:

- (268) *Box* *he* *stop* **close along** *bunk* *bilang* *Jack*
 box PM LOC **ADES** bunk bilang Jack
 ‘The box was near Jack’s bunk.’

(Tulagi Island; Manawa 24.10.1898)

The StE-borrowed form *near* occurred in the early MPE data as well, albeit representing an exception rather than the norm. As example (269) indicates, the feature was attested in sentences which contained many StE features: next to the conjunction *when*, the definite article *the*, the verb *come* in its irregular past tense form and the verb *married* can be identified. Thus, the sentence contains more StE than PE features which either indicates author or editor modification or a very acrolectal form of PE.

- (269) *When* *the* *war* *came* *me* *marri-ed* *one* *girl* **near** *One Pusu [...]*
 when ART war come.PST 1SG marry-PST ART girl **ADES** One Pusu [...]
 ‘When the war came I married a girl near One Pusu’

(*The Stories of the Crew*; NIV June 1947: 10; AU PMB DOC 439)

A final variant that was identified in the data was *long*. The form, however, was not attested in sentences but was only listed as a variant to encode the adessive semantic function in Tok Pisin dictionaries by Kutscher (1940: 23) and van Baar (1930: 53). Based on the data at hand, no further *long*-tokens encoding the adessive semantic function were identified. Nonetheless, *long* may have served as the main preposition to encode the adessive semantic function in all three varieties before more concrete forms developed. As earlier pointed out, (*a*)*long* was one of the earliest prepositions in MPE and was considered an all-purpose preposition (cf. Landtmann 1927: 454). Contextual information was required to understand which semantic function the form encoded in a clause. For instance, in a sentence such as *Mi kam long bot*, the preposition *long* may encode the adessive, illative, allative or instrumental semantic function. Without sufficient contextual information, it is impossible to perform an unambiguous categorisation of *long*. Due to the

polysemy of early prepositional forms and missing contextual information, many prepositional tokens could not be assigned fulfilling a single semantic function only. This may explain the low amount of attestations of *long* to encode the adessive semantic function in my data.

Factor	Levels
ST_FORM	close along close to close to (a)long close up close up (a)long close up along of near long
FEATURE	prep_ades
YEAR_DET	1832 ... 1950

Table 24: Linguistic coding for adessive prepositions

9.4.3.3 Findings and discussion

In the following section the findings of the analysis will be presented. First, some remarks regarding the general data distribution will be made before the focus is placed on the individual varieties. Adessive prepositions in SIP (9.4.3.3.2), BIS (9.4.3.3.3) and TP (9.4.3.3.4) will be considered. Finally, a comparative analysis and summary of the findings will be presented.

9.4.3.3.1 General data distribution

In total, 114 prepositional tokens were identified in the early collected MPE data which encode the adessive semantic function. 39 of these were attested in the SIP data (34.78%), 73 in the TP dataset (63.48%) and only two of them in the BIS dataset (1.74%). In addition to data scarcity, a further problem can be seen in how the datapoints spread across time. Figure 161 shows that even though the majority of datapoints were identified in the TP dataset, the variety is not very well represented as the datapoints spread unequally across time.

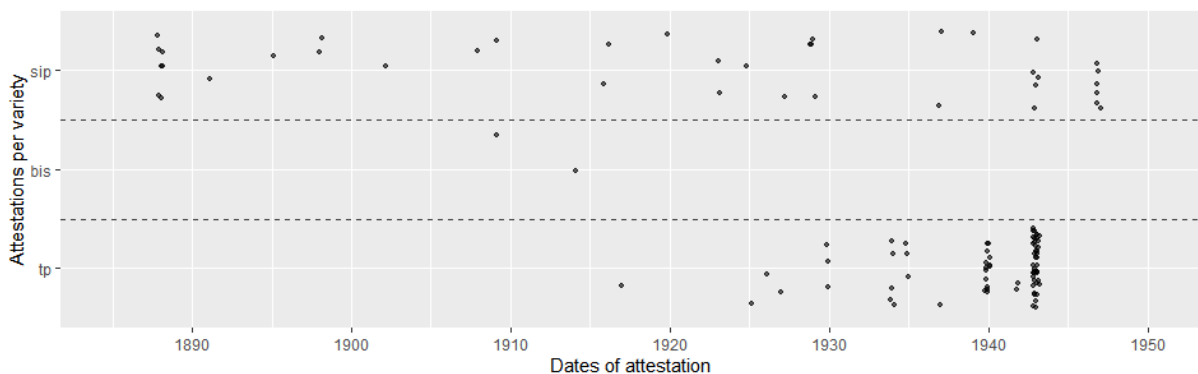


Figure 161: Distribution of adessive prepositional datapoints across time per variety

The earliest adessive attestation in TP dates post 1915 and the majority of datapoints conglomerate in the years 1940 and 1943. Although there were only 39 tokens, it seems as if SIP tokens spread best across time.

If the time period from 1880 to 1950 is treated as a single period, the following general observations regarding the encoding of the adessive semantic function can be made.

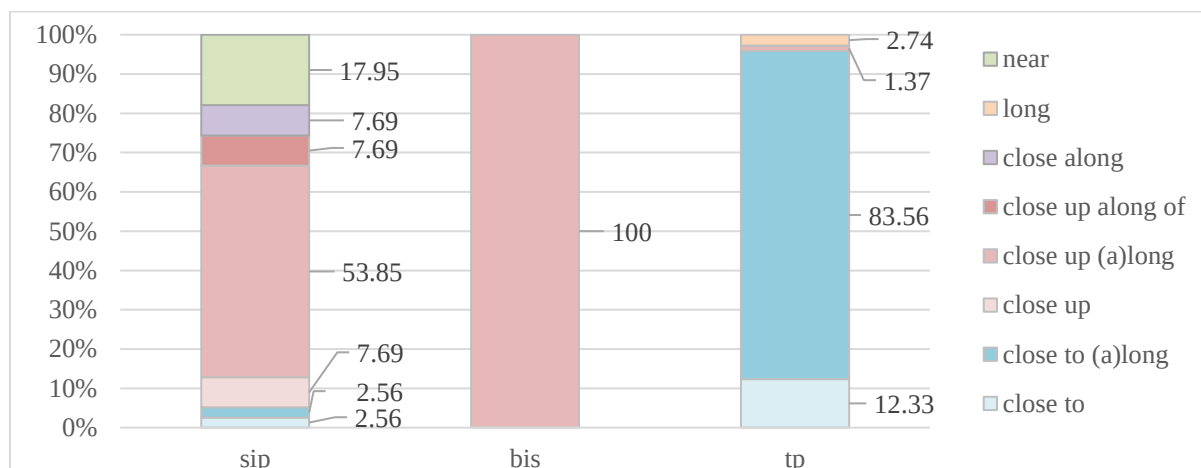


Figure 162: Relative frequencies of adessive prepositional forms in MPE

As Figure 162 shows, in each of the varieties a dominant form is attested. In SIP, more than half of the datapoints are encoded with *close up (a)long*. TP shows a clear dominance of the form *close to (a)long* in adessive semantic contexts. Although the figure indicates a dominance of *close up (a)long* in BIS as well, it needs to be kept in mind that only two adessive token were identified in the dataset of the variety. Furthermore, the figure indicates a great variability in the forms attested in SIP.

9.4.3.3.2 Diachronic analysis of adessive prepositions in SIP

In this section the concrete forms used to encode the adessive semantic function in early SIP will be investigated from a diachronic perspective. First, the attested forms will be introduced before the dates of attestations will be considered. Finally, the results of the ctree analysis will be provided to test whether time can be considered a significant influencing factor for the choice of the adessive prepositional form.

9.4.3.3.2.1 Attested forms

Seven different prepositional forms were attested in the early SIP data. While *close up along* occurred in 21 of the 39 adessive tokens and was exemplified in (265), the variants *close up* and *close up along of* both appeared three times (= 7.69%). An example of the forms in early SIP were displayed in sentence (264) and (267). While the prepositional form *close up* was attested

in three different sources, it is remarkable that the three occurrences of *close up along of* can all be traced to a single source. This may indicate that it was not a reliable form. Furthermore, it seems striking that *near*, as exemplified in (269), represents the second most dominant attested form in the early SIP data with seven occurrences (= 17.95%). A closer look at those seven attestations reveals however that six of the tokens can be traced to a single source, namely *The Stories of the Crew*. The source is characterised by a high degree of mixture between PE and StE features. This begs the question whether the use of *near* in adessive prepositional contexts represents a peculiarity of the text or author, rather than a general common feature in the variety. Nonetheless, it is likely that Europeans may have used the form in their interactions with Pacific Islanders – especially in the initial encounters.

The forms *close to* (cf. sentence (270)) and *close to along* (cf. sentence (262)) represent once attestations in the SIP data. The form *close along* appears three times in three distinct sources and was exemplified in (268).

- (270) *I stop-ped close to Tommy*
 1SG LOC-PST ADES Tommy
 ‘I was near Tommy.’

(Rubiana, Solomon Islands; Billy 23.10.1894 [Australian Station 1894: 20])

9.4.3.3.2 Timeline and boxplot approach

The timeline in Figure 163 demonstrates when the individual forms were attested across time. What can be learned from the figure is that the earliest attested forms in adessive semantic contexts are *near*, *close up along of*, *close up (a)long* and *close up*.

Near is only attested again in 1947, which is the source which consists of a mixture of StE and PE and therefore should be treated carefully. It is unlikely that the form survived in SIP. The form *close up along of*, as previously pointed out, was attested in a single source only and was not attested after 1888. The third mentioned form, *close up along*, is the form that spreads most along the timeline and is attested until the end of the observed period.

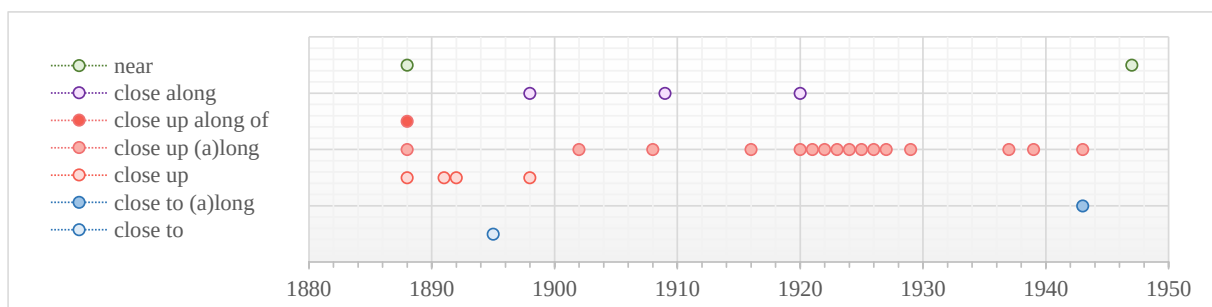


Figure 163: Timeline of adessive prepositional variants in SIP

The form *close up* seems to have existed as a variant of *close up along* until 1898 but does not prevail after 1900. In addition, the single attestation of *close to* dates to 1895 and of *close to along* to 1943. It needs to be noted that the attestation of the latter occurs in a document translated by Herbert Ian Hogbin, an anthropologist, who did several field studies in the BSIP but also in the Mandated Territory of New Guinea. Thus, it is possible that he mixed up forms from the two regions. It is, however, also possible that in some Solomon Island regions the form *close to (a)long* was used.

There are also three attestations of *close along*, which may be the result of European author modification. As many authors assumed *(a)long* to be the major preposition and native English speakers considered *close* as an adjective, they may have modified the original utterance. However, since the form is attested in three different sources, there is a likelihood that the form existed as a variant until the 1920s.

Similar results are obtained when a boxplot is created:

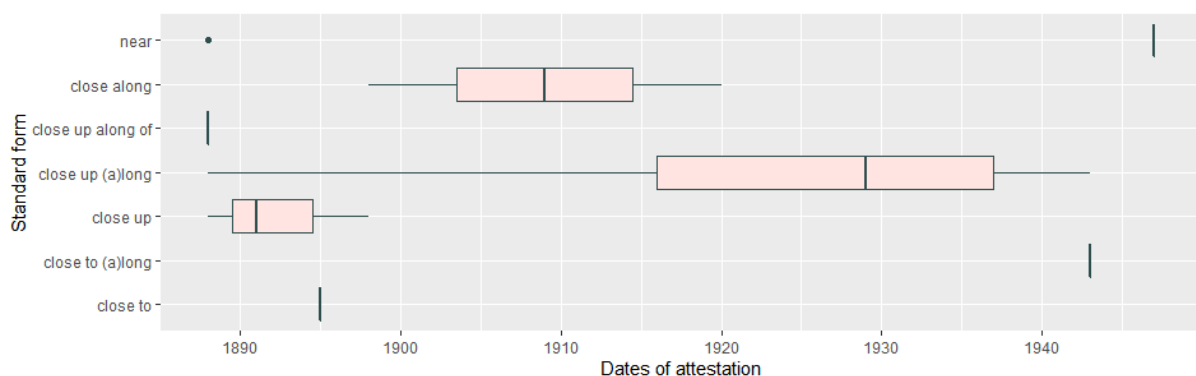


Figure 164: Boxplot of adessive prepositional variants in SIP

As Figure 164 shows, without considering statistical significance, it seems as if around 1898 the amount of variation is reduced and around 1916 the form *close up along* develops into the dominant adessive prepositional form.

9.4.3.3.2.3 Testing for the impact of the years of attestations on the choice of form

In order to test whether the years 1898 and 1916 represent significant factors in the choice of form, the ctree algorithm was applied. If solely the year of attestation and its impact on the form is considered, the tree as displayed in Figure 165 is obtained which consists of two splits.

The highest level split indicates that data attested before and in 1898 differs very significantly from data attested after 1898 in how the adessive semantic function is encoded ($p=0.003^{**}$). The second level split is very significant with $p=0.006^{**}$ and splits the data in the year 1939. As previously observed, before 1898 a high degree of variation in the choice of form existed, whereas from 1899 to 1939, a clear preference of the form *close up along* is observed.

The data attested after 1939 shows *near* as the most likely form. *Close up along* only represents the second most probable form. However, as previously pointed out, *near* was predominantly attested in two sources only. Thus, it is possible that the text type has an impact on the attested forms as well.

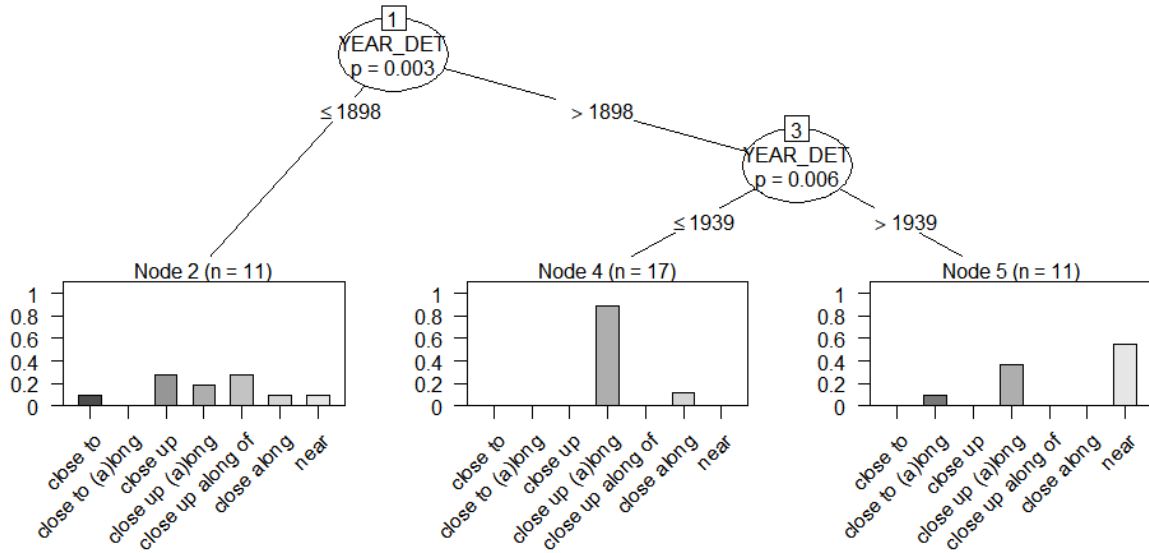


Figure 165: Conditional inference tree for adessive prepositions in SIP ST_FORM~YEAR_DET

If a ctree is created with the year of attestation and the text type as possible predictor variables, the tree in Figure 166 is obtained which consists of two splits. In fact, the year 1939 no longer represents a significant predictor variable. The highest level split of the resulting tree is based on the text type variable ($p < 0.001^{***}$). Only speech-related attestations show a significant time-based split ($p = 0.013^*$) with data pre-1899 being characterised by a great degree of variation and data post 1898 indicating *close up along* as the most likely form. In written and intermediate attestations, however, this time split cannot be observed.

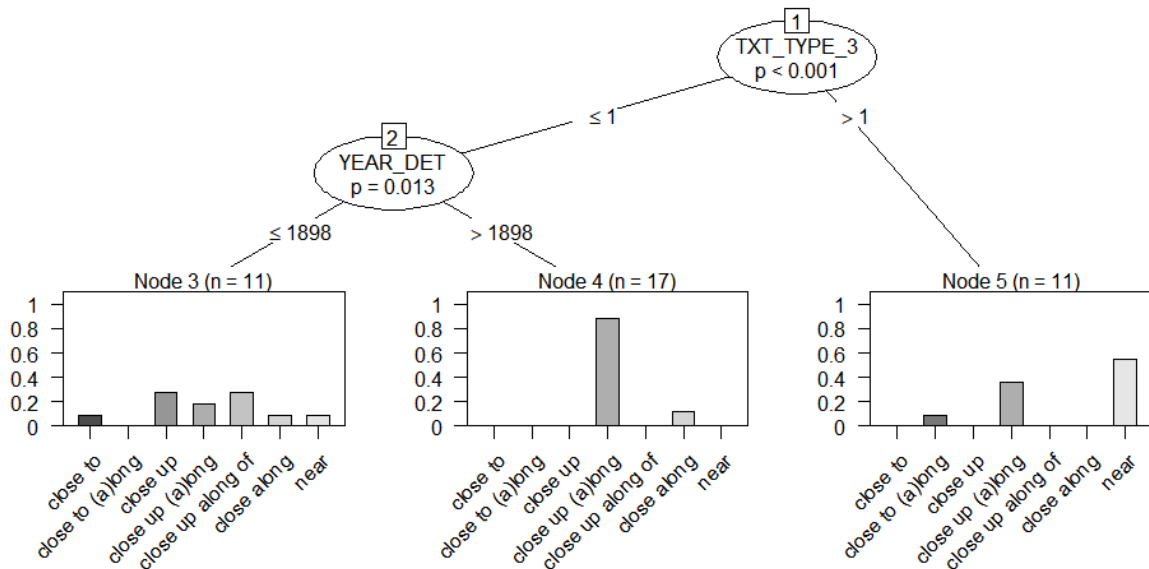


Figure 166: Conditional inference tree for adessive prepositions in SIP ST_FORM~YEAR_DET+ TXT_TYPE_3

The tree structure stays the same if the author is added as a possible predictor variable as well. The text type variable is still highly significant with $p < 0.001^{***}$ and the split in 1898, based on the year of attestation, is significant with $p = 0.025^*$. In general, however, it needs to be kept in mind that the analysis was based on a very small number of tokens.

9.4.3.3.3 Diachronic analysis of adessive prepositions in BIS

In BIS only two adessive prepositions were attested. Both were encoded with the form *close up along* as exemplified in (271). The attestations refer to the years 1907 until 1914. It is possible that prior to these attestations several competing forms existed. It is also probable that variants still existed from 1907 until 1914 or even longer. Unfortunately, no further attestations were found which may help to learn more about the development of the variety.

- (271) [...] one feller something [...] 'e stop 'long big feller bokis **close up long** window
[...] one MODIF thing [...] PM LOC PREP big MODIF box **near** window
'... one thing which is in the big box near the window.'

(Vanuatu ~1914; Jacomb 1914: 99)

9.4.3.3.4 Diachronic analysis of adessive prepositions in TP

In contrast to BIS, the early TP data contained 73 prepositional tokens which encoded an adessive semantic function. In the following, the attested forms will be introduced before the focus will be placed on their attestations across time and on the significance of the results.

9.4.3.3.4.1 Attested forms

In TP four different prepositional forms were identified to encode the adessive semantic function. The dominant form represents *close to long*, which was exemplified in (263) and was attested in 61 of 73 tokens (= 83.56%). The second most common form is *close to*. The form was attested in nine of the tokens (= 12.33%) and a TP example was provided in (261). The form *long*, as pointed out earlier, was only attested twice ($2/73 = 2.74\%$) and appeared in dictionaries only. Although it is likely that *long* represented the common form to encode the adessive semantic functionality before other forms became to be used, a clear identification of *long* as an adessive marker was impossible due to insufficient contextual information. The form *close up (a)long* was attested only once and was displayed in (266).

9.4.3.3.4.2 Timeline and boxplot approach

To learn about the adessive feature development in TP, a timeline and boxplot approach were used. Figure 167 shows that the dominant form *close to (a)long* also represents the earliest form

attested. It first occurs in 1917 in the written data at hand and persists at least until the end of the observed period. The form *close to* is first attested in the early 1940s. As this time period covers the war pamphlets which were written by the Australian government, it is likely that the occurrence of the form in the data is based on an error by the pamphlet producers. It is at least a war time feature, which was not attested earlier. The two attestations of *long* date roughly to the years 1930 and 1940 which shows that *long* could still be used as a variant of the form *close to (a)long* in the 1930s. The single attestation of *close up (a)long* dates to 1935.

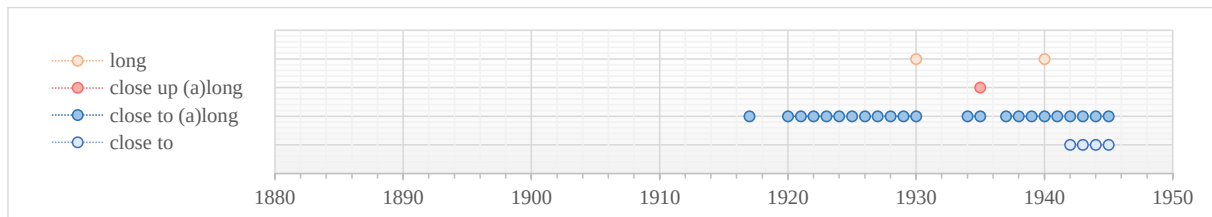


Figure 167: Timeline of adessive prepositional variants in TP

The boxplot in Figure 168 indicates a similar development. Comparing the medians it can be learned that the medians of *close to (a)long* and *close to* date latest.

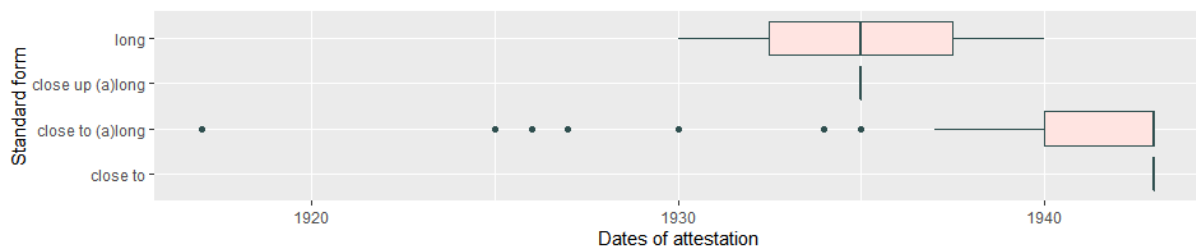


Figure 168: Boxplot of adessive prepositional variants in TP

9.4.3.3.4.3 Testing for the impact of the years of attestations on the choice of form

In order to test whether there is any statistical rigor to claim that the year of attestation has an impact on the choice of the prepositional form used to encode the adessive semantic function, the ctree algorithm was applied. If only the year of attestation and its impact on the form is considered in the ctree algorithm, a stacked boxplot is obtained. This indicates that time does not have a significant impact on the choice of form based on the data sample at hand. If the text type or both, the text type and author variables are added, the results stay the same. Neither the year of attestation nor the text type or author turn out to be significant predictor variables regarding the choice of form in TP.

9.4.3.3.5 Diachronic comparative analysis of adessive prepositions in MPE

Comparing prepositional forms used to encode the adessive semantic function in SIP, BIS and TP, it can be observed that different forms were in use to varying degrees in the three varieties

across time. While in SIP the form *close up (a)long* represented the dominant form in the early data, in TP *close to (a)long* dominated. As in both varieties the other form was attested once as well, it can be assumed that the forms were used in both Melanesian areas at an earlier point in time. However, the written data showed that latest by 1898 *close up along* turned into the dominant form in SIP. The ctree analysis of the TP data revealed that *close to (a)long* represented the dominant strategy to encode the adessive semantic function independent of the possible predictor variables. In BIS only two attestations of the form *close up (a)long* were attested. The small amount of attestations does not allow the making of further claims about when the form grammaticalised.

The results obtained by the individual analyses are supported when applying the ctree algorithm to the complete adessive data tokens of all three varieties at once. Using the year of attestation and the variety as possible predictor variables, the tree in Figure 169 is obtained. The tree consists of three splits. The highest-level split is based on the variety ($p < 0.001^{***}$) and separates TP from BIS and SIP. Node 2 shows that the form *close to (a)long* represents the most likely form in TP. BIS and SIP show two further splits based on the year of attestation. Node 3 indicates that adessive tokens are encoded differently depending on whether they are attested before or after 1898 ($p = 0.005^{**}$), and Node 5 specifies that differences are also dependent on whether the data is attested before or after 1939 ($p = 0.01^{*}$). Since only two BIS tokens were identified which are similar to SIP, the varieties cluster together and are separated from TP.

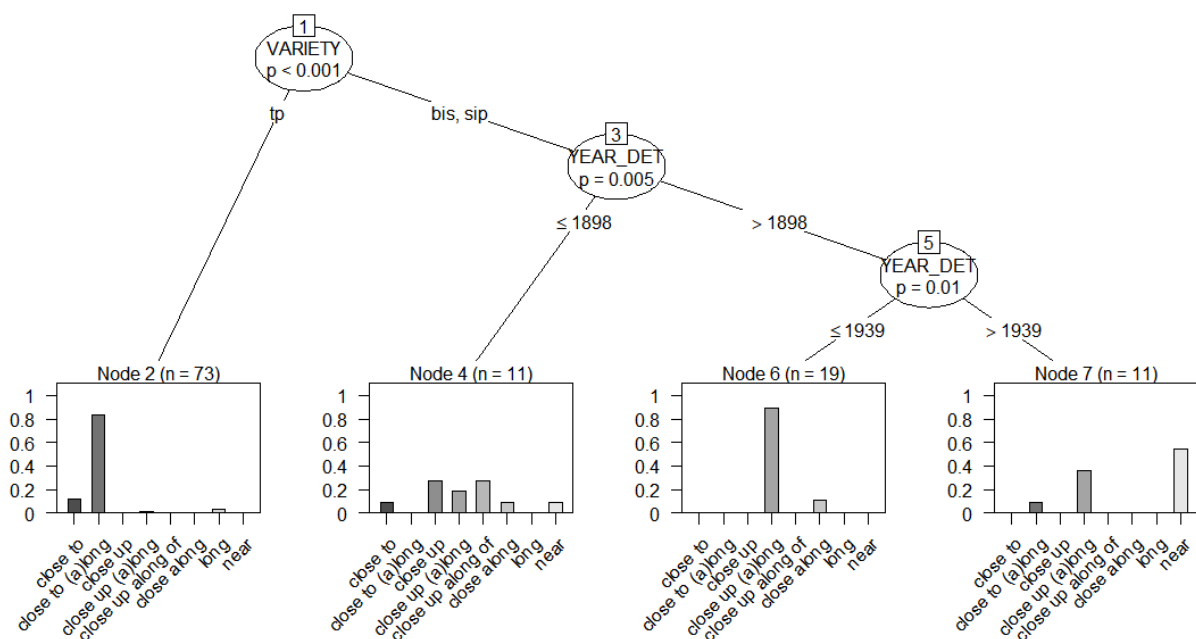


Figure 169: Conditional inference tree for adessive prepositions in the three MPE varieties $ST\_FORM \sim YEAR\_DET + VARIETY$

If the text type variable is added to the algorithm, the highest-level split remains the same, but, as previously observed during the analysis of the SIP data, the second level split is based on the text-type variable and only speech-like attestations show a further time-based split in the year 1898 (cf. Figure 170). Only if the author is added, the author turns into the highest-level split-evoking variable (cf. Appendix II; $p < 0.001^{***}$). While with some authors a further split based on the variety can be observed, splitting TP from SIP data in how they encode the adessive semantic function (Node 2; $p < 0.001^{***}$), with the remaining authors a split based on the text type is observed (Node 5; $p = 0.007^{**}$).

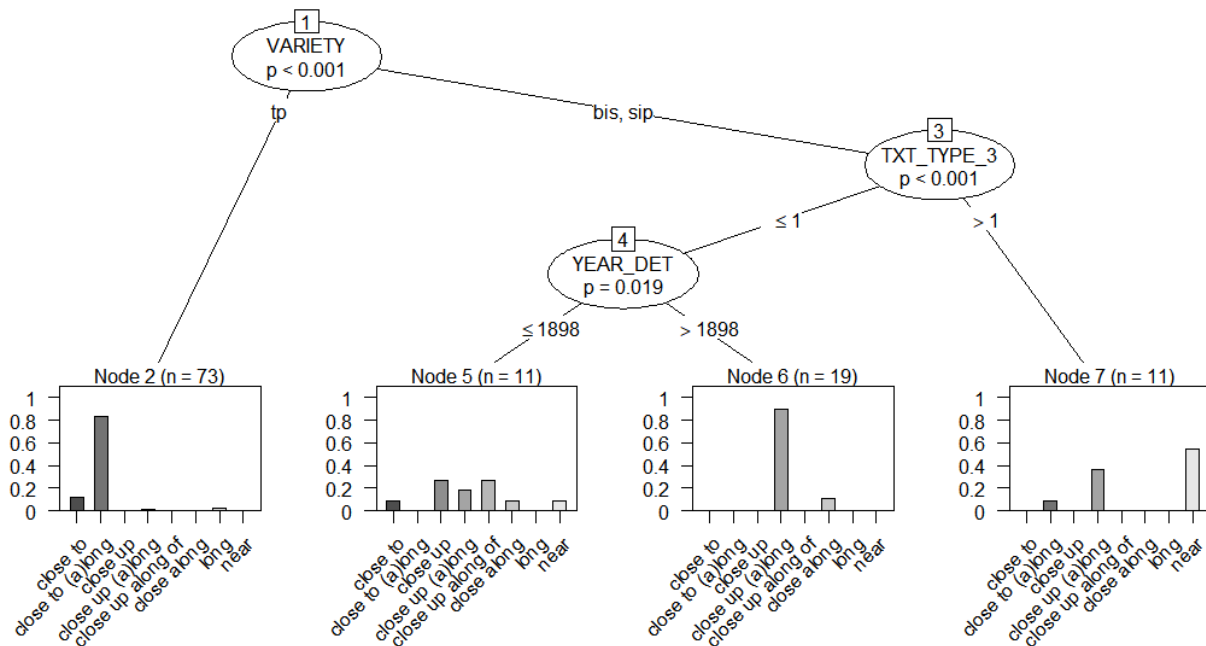


Figure 170: Conditional inference tree for adessive prepositions in the three MPE varieties $ST\_FORM \sim YEAR\_DET + VARIETY + TXT\_TYPE\_3$

As the ctree analysis of SIP identified the year 1898 as the significant year in the stabilisation of the form *close up (a)long* in adessive semantic contexts and thus the process dates prior to the end of the labour trade, it is not very likely that the end of the labour trade was decisive for the stabilisation of the adessive form in SIP. In TP the form *close to (a)long* was identified as the dominant form independent of the time variable. However, the timeline showed that the earliest attestations of adessive markers date to 1917 and, thus, the data at hand cannot be used to answer whether the end of the labour trade led to the dominance of the form. As too little datapoints were identified for BIS, no claims regarding its development can be made at all.

Nonetheless, it remains of interest that the dominant forms, which were attested in the pre-1950 data of the varieties, represent the adessive prepositions that are used in the varieties today. Moreover, it is of interest that while in SIP and BIS a form including *close up* grammaticalises, in TP a form including *close to* develops into the adessive prepositional form. A closer look into early SPPE and QPPE data was taken to analyse whether one specific form

was used in the varieties. The early SPPE data contains neither instances of *close to* nor instances of *close up*. In QPPE, *close up*, *close up (a)long* and *close to* (cf., for instance, Queensland Department of Justice 1906 Part 10) were identified, although the former were attested with higher frequencies. What is of special interest is that most of the *close up along* attestations in QPPE were produced by speakers who originated from the Solomon Islands.

9.4.3.4 Summary

The analysis of the forms showed that while in SIP the form *close up (a)long* dominated, in TP the form *close to (a)long* was attested as the dominant form. However, only the SIP data qualified for answering when the adessive prepositional forms stabilised. As the form to encode the adessive functions seems to have grammaticalised already before 1900, the end of the labour trade cannot be considered the decisive factor. The data of TP dates rather late and showed no significant time splits. The amount of data identified for BIS was too low. Thus, the analysis of adessive semantic prepositions does not provide insights into whether the end of the labour trade resulted in the divergence of the three MPE varieties.

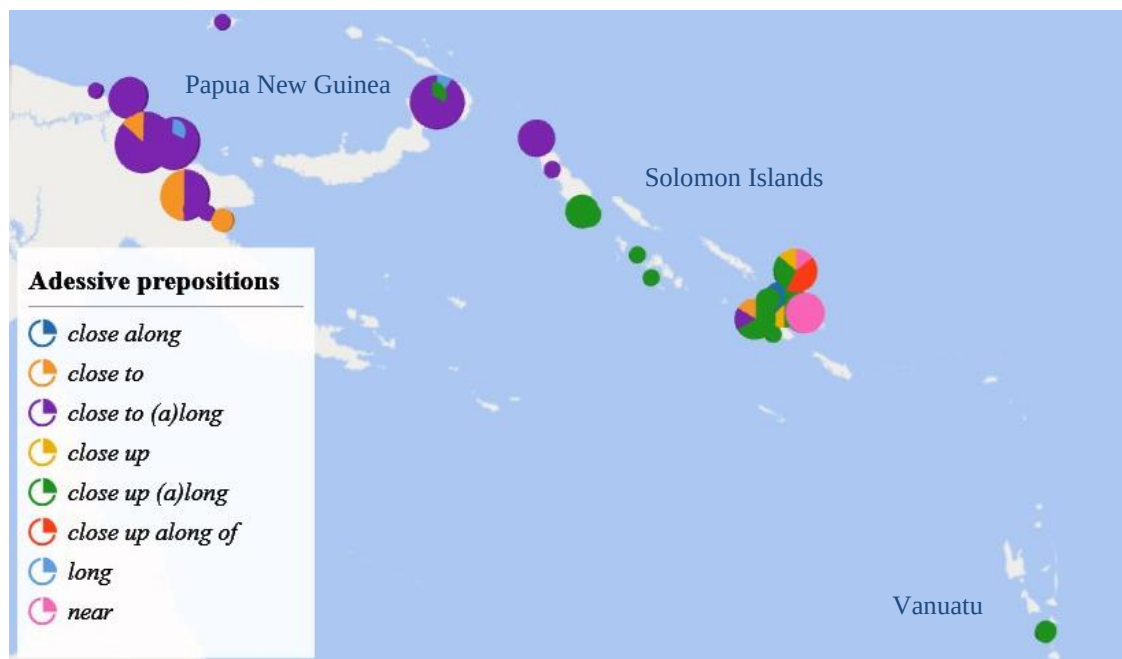


Figure 171: Map of Melanesian Islands, showing attestations of adessive prepositional variants

10 Conclusion

The present study aimed at contributing to our knowledge about the origin and development of the Melanesian Pidgin English varieties Solomon Islands Pijin, Bislama and Tok Pisin. Although several scholars have argued that the end of the labour trade was one of the decisive factors that led to the individualisation of the three MPE varieties, no attempts had previously been made to explore the starting point of their individual development by focussing on an analysis of early language data that goes beyond the comparison of earliest attestations and includes statistical methods to investigate *when* selected features of the varieties began to diverge. This study has examined morphosyntactic variation and change in SIP, BIS and TP from a diachronic perspective by combining qualitative with thorough quantitative-statistical methods.

By analysing four morphosyntactic features, namely demonstratives, relative clauses, modality markers and selected prepositions, I aimed at answering the following research questions, which were introduced in Chapter 1:

- 1) How were the linguistic variables realised in Solomon Islands Pijin, Bislama and Tok Pisin over time?
- 2) Was variation receding, stabilising, or increasing over time within the individual varieties and across the varieties?
- 3) Which external influences may have had an impact on the individual developments of the varieties regarding the features under investigation?
- 4) What do the results tell us about the factors at work in feature selection, as well as grammaticalisation processes in the stabilisation of contact languages in general?

This concluding chapter will attempt to answer these questions and summarises what can be learnt about the development of SIP, BIS and TP based on the early collected data at hand. The chapter begins with a summary of the results of the preceding case study chapters (Section 10.1). In Section 10.2, the years of attestations which turned out to have a significant impact on the formal realisation of the analysed features will be linked to existing theories on the origin and development of the three MPE varieties in order to discuss whether it makes sense to assume that the end of the labour trade was the decisive factor for the individualisation of the varieties. Section 10.3 will then focus on the theoretical implications of the study and will evaluate what the results reveal about form selection and grammaticalisation processes during the development of contact languages in general. Finally, concluding remarks and possibilities for future studies will bring this chapter to an end (10.4).

10.1 Summary of the results

The preceding case studies focused on demonstratives, relative clause markers, modality markers and prepositions in the early MPE varieties. For each of the analysed variables, attested variants

were introduced and their development across time was outlined. Table 25 summarises the variables under consideration, showing the variants attested in the pre-1950 data and the forms which are used in the varieties today. The years in parenthesis indicate when the contemporary form was first attested in the early written data at hand. To identify statistically significant changes based on the factor *time*, the ctree algorithm in R was applied.

Summarising the results of Chapter 6 to 9, Figure 172 displays the previously identified dates which had a significant impact on the choice of form (ST_FORM) if the year of attestation (YEAR_DET) was the only predictor variable considered in the ctree algorithm. As the figure shows, only seven of the 41 time-based splits occur prior to 1906, the date which is considered to mark the end of the labour trade.

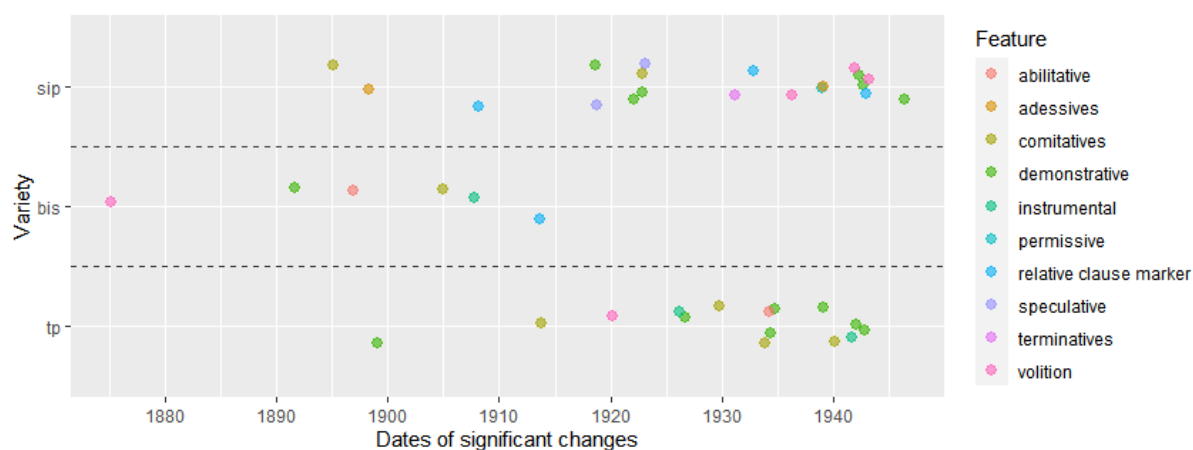


Figure 172: Significant years for the choice of form based on $ST\_FORM \sim YEAR\_DET$

However, the individual case studies showed that changes in the choice of feature forms and in their frequencies were influenced by other predictor variables as well. Figure 173 indicates how the number of significant years is reduced if either the variable FEATURE, TYPE or STRUCTURE was added.

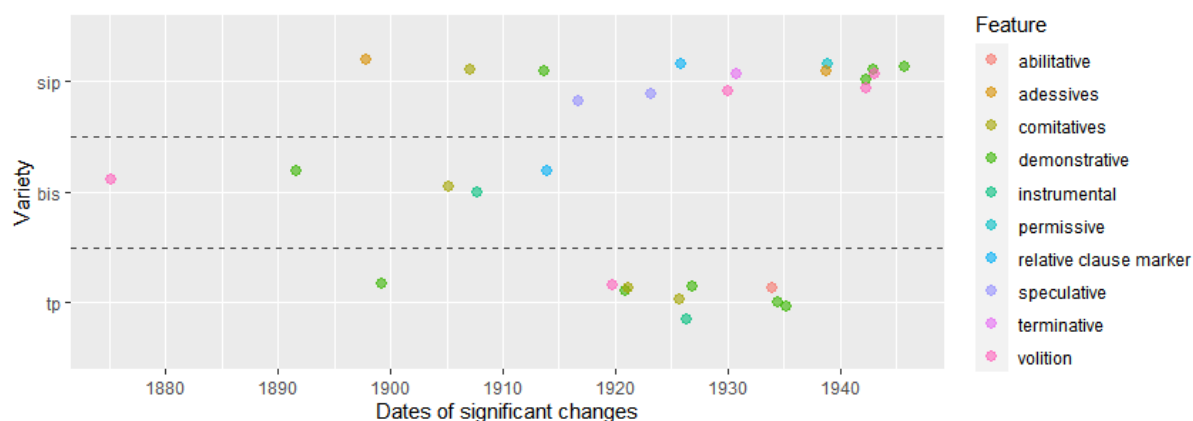


Figure 173: Significant years for the choice of form based on $ST\_FORM \sim YEAR\_DET + TYPE/FEATURE/STRUCTURE$

	SIP		BIS		TP	
	Variants attested before 1950	Contemporary forms	Variants attested before 1950	Contemporary form(s)	Variants attested before 1950	Contemporary form(s)
demonstratives	<i>here, him here, that, that fellow, that one, these, this, this+here, this fellow, this one, those</i>	<i>disfela</i> (1881) <i>desfela</i> (1886) <i>ia</i> (1895) <i>hemia</i> (1913) <i>diswan</i> (1907)	<i>here, him here, that, that fellow, them, them fellow, these, these fellow, this, this+here, this fellow, this one, those</i>	<i>ia</i> (1869) <i>hemia</i> (1894) [<i>disfela</i> (1871)]	<i>here, him here, that, that fellow, that one, these, these fellow, this, this fellow, this fellot+here, this one</i>	<i>dispela</i> (1881) [<i>ia</i> (1925)]
relative clauses	<i>zero, where, that, which, who, who's that</i>	<i>where</i> (1908) <i>zero</i> (1886) <i>hu</i> (1888)	<i>zero, where, that</i>	<i>we</i> (1913)	<i>zero, who</i>	<i>zero</i> (1878) <i>husat</i> (n.a.) <i>we</i> (n.a.)
volition	<i>like, like him, like to, want, want him, want to, want for</i>	<i>laek</i> (1908) <i>laekem</i> (1925) <i>wande</i> (1895 or earlier) <i>wandem</i> (1923) (<i>fo</i>)	<i>like, like him, like to, want, want him, want to, wish to</i>	<i>wantem</i> (1894)	<i>like, like him, like to, want to</i>	<i>laik</i> (1886)
abilitative	<i>can, save, save for</i>	<i>save</i> (1895) <i>fitim fo</i> (n.a.) <i>inaf fo</i> (n.a.)	<i>can, save</i>	<i>save</i> (1883)	<i>can, save, enough, enough for, enough long</i>	<i>inap</i> (1940) [<i>save</i> (1908)]
inabilitative	<i>can't, cannot, no save, not able to</i>	<i>no save</i> (1880) <i>kanduit</i> (n.a.) <i>kan</i> (1880) <i>kanot</i> (1924)	<i>no can, can't, cannot, no save</i>	<i>no save</i> (1878) <i>kanduit</i> (n.a.)	<i>can't, no can, no save, no enough, no enough long</i>	<i>no inap</i> (1941)
prohibition	<i>no can, can't, no save, no more save, must not</i>	<i>no save</i> (1895)	<i>no can, can't, no save</i>	<i>no save</i> (1914)	<i>no can</i>	<i>no ken</i> (1924)
permission	<i>can, save</i>	<i>save</i> (1929)	n.a.	<i>save</i> (n.a.)	<i>can</i>	<i>ken</i> (1924) <i>inap</i> (n.a.)
dubitative	<i>I think, maybe, might, might be</i>	<i>maet</i> (1880) <i>mebi</i> (1920) <i>ating</i> (1881)	<i>I think, maybe, might</i>	<i>might</i> (1907) <i>I think</i> (1905)	<i>I think, maybe</i>	<i>ating</i> (1908)
comitative	<i>(a)long, with, with him, along of, along with</i>	<i>wetem</i> (1930)	<i>(a)long, with him, with</i>	<i>wetem</i> (1907)	<i>(a)long, along with, belong, with, one time, one time (a)long</i>	<i>wantaim</i> (1930)
instrumentals	<i>(a)long, with, along of, in</i>	<i>long</i> (1880) <i>wetem</i> (n.a.)	<i>(a)long, with, by</i>	<i>long</i> (1892) <i>wetem</i> (n.a.)	<i>(a)long, belong, one time, one time (a)long</i>	<i>long</i> (1909) <i>wantaim</i> (1942)
terminative	<i>catch him, gogo catchim, zero, till</i>	<i>kasem</i> (1927)	<i>catch him, till, until</i>	<i>kasem</i> (1917)	<i>long, enough, enough long</i>	<i>inap long</i> (1930)
adessive	<i>close to, close to (a)long, close up, close up (a)long, close up along of, close along, near</i>	<i>kolsap long</i> (1888)	<i>close up (a)long</i>	<i>klosap long</i> (1907-1914)	<i>close to, close to (a)long, close up (a)long, long</i>	<i>klostu long</i> (1917)

Table 25: Pre-1950 variants versus contemporary forms

It needs to be noted that the figure contains all significant year-based splits independent of the level on which they occurred in the ctree. Moreover, the features *terminative* and *adessive* are included even though no predictor variables called FEATURE, TYPE or STRUCTURE existed for the features.

If the text type (TXT_TYPE_3) or the author (AUTH_NAME) were additionally considered, the resulting dates of significant changes can be summarised as in Figure 174 and 175.¹⁹⁴

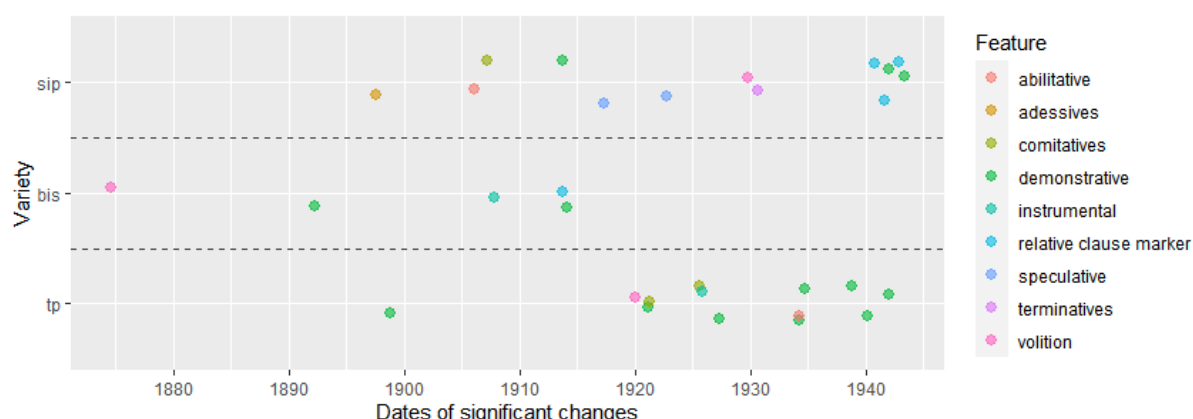


Figure 174: Significant years for the choice of form based on $ST\_FORM \sim YEAR\_DET + TYPE/FEATURE/STRUCTURE + TXT\_TYPE\_3$

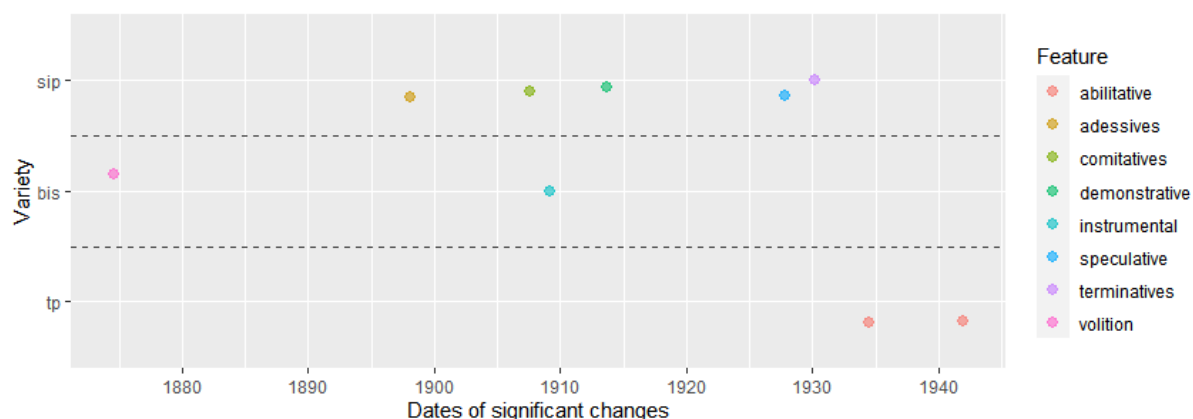


Figure 175: Significant years for the choice of form based on $ST\_FORM \sim YEAR\_DET + TYPE/FEATURE/STRUCTURE + TXT\_TYPE\_3 + AUTH\_NAME$

Several observations can be made when comparing Figures 172-175. The number of significant dates is reduced as soon as other possible predictor variables come into play. Thus, while the dots in Figure 172 indicate 41 significant time-based splits, only nine time-dependent splits remain significant if all predictor variables are considered. The individual chapters showed that the author turned out to be the most important predictor variable in the early data for most of the analysed features.

¹⁹⁴ Again, it should be reiterated that the displayed dots do not indicate that the time-based splits occurred on the highest level.

Another observation that can be made is that in Figures 172-175 most of the significant changes date after 1905, which may indicate that *changes* (although not necessarily a *stabilisation*) in the choice of form began to occur after and as a result of the end of the labour trade. The word *changes* is preferred, as the trees in the previous case study chapters indicated that splits based on the year of attestation do not necessarily indicate a *stabilisation* of a specific form.

Even if the figures suggest that the most significant changes took place after the end of the labour trade, it needs to be noted that a possible time lag has not been considered so far. For SIP a great number of texts was available covering the years 1910 until 1950 but pre-1900 sources are less available and suffer in quality. Thus, it can be expected that there is a time lag and that attestations for 1910 have already existed earlier. Consequently, some of the significant years determined during the ctree analysis may have to be shifted further to the left on the timeline and may reflect changes that date prior to the end of the labour trade. The database of BIS showed a low quality and quantity of sources for the years 1925-1950. Thus, the fact that BIS does not show any significant changes in the formal realisation of the features after the year 1916 can be explained by the latter and does not necessarily mean that no further time-based changes occurred in the variety before 1950. Pre-1900 TP language data is rare and has a lower quality than language data which represents the variety as it was during World War II. Thus, again, a time lag can be expected.

Another observation that was made is that for each of the analysed features different years turned out to be significant. This suggests that the stabilisation of individual grammatical features occurred at different points in time, the same as with ‘natural’ languages (= non-pidgins and non-creoles). While some features may have stabilised prior to the end of the labour trade, others may only have done so after the labour recruitment had come to an end and again others may not have stabilised before 1950. For some features several time splits were observed, indicating that the features changed gradually. Moreover, the years that turned out to be significant for the feature development differed in the three varieties. As individual features and their variants changed at different times in the individual varieties, the localisation of the three varieties was a gradual process. I stress once again that ctrees are restricted to binary choices. Thus, even if ctrees identifies significant splits based on the year of attestation, language change always represents a gradual process.

It is not sufficient to focus on the significant years when aiming to answer the research question *when* the varieties started to diverge. The years displayed in Figures 172-175 only indicate that significant *changes* occurred in the individual varieties but do neither provide

information regarding the forms that were attested before and after the splits, nor about their likelihood of occurrence which are of importance to reconstruct the development of the three varieties.

Focussing on the attested forms that were identified in the case study chapters, it became clear that in order to reconstruct a possible development scenario for the three MPE varieties, we need to differentiate between *when feature forms changed*, *when feature forms grammaticalised*, *when feature forms stabilised* in each variety and *when feature forms began to diverge* among the three varieties.

It became clear that even if significant **changes** in the individual varieties were predominantly observed after the end of the labour trade, some of the **individualising forms** had their origin already prior to the end of the labour trade. For example, the analysis of volition markers revealed that TP showed a clear preference of the form *like* independent of the time variable, whereas in BIS a preference of *want*-based forms was observed after 1875. Thus, differences in the use of the morphemes *want* and *like* were already observed between SIP, BIS and TP prior to the end of the labour trade and may therefore be based in differing labour recruiting histories.

The same applies to the different forms attested to encode *dubitative* modality. While the form *ating* is attested in the early data of all three varieties, only SIP and BIS make use of the additional form *maet*, which was first attested in SIP in 1880 and in BIS in 1907. While clause-initial *might* was attested in QLD as well, the particle could not be attested in SPPE. Although the absence in the latter is no evidence that the feature was not used on Samoan plantations, it is nonetheless likely that the form *maet* is only attested in SIP and BIS due to the varieties' close connection to the QLD plantations.

Moreover, it became apparent that SIP and BIS showed a closer similarity regarding the forms used to encode relative clauses (= particle *where*), comitatives (= preposition *with him*), terminatives (= *catch him*) and adessives (= *close up (along)*). Even if some of these forms were attested rather late (cf. terminatives), it is of interest that there is a closer similarity between SIP and BIS. For example, when focussing on terminatives, the transitive verb *catch him* is first attested in 1927 in SIP as a verbal preposition encoding termination. In BIS, the form seems to have grammaticalised earlier, being attested in 1917 for the first time. *Catch him* might have entered the varieties as a *full verb* via QPPE, as early evidence showed that *catch him* served as a terminative verb in the pidgin as spoken in Queensland. By contrast, in TP *catch him* was only attested as a terminative verb with the meaning 'to reach' in the 1930s. *Catch him* may not have grammaticalised into a terminative marker in TP as, by that time, the form *enough long* may have

already started its grammaticalisation path. Thus, the similarity of forms used in BIS and SIP may be influenced by the fact that the varieties have been in close contact for a long period in time.

Although **divergence** may come along with **stabilisation**, it is not necessarily bound to it. The analysis of the early MPE data indicated that SIP, BIS and TP made use of individual forms even before a stabilisation took place. Even if the varieties were still characterised by a high degree of variability, localised forms were in use. The exact date of a stabilisation of forms is difficult to determine. The case study chapters revealed that although new forms were **grammaticalising**, a great degree of variation was nonetheless common for some of the features. In addition, the moment when a stabilisation took place depends heavily on how the notion of **stabilisation** is defined.

Further findings that could be made when focussing on the concrete forms used to encode the grammatical functions were that the SIP data was characterised by a high amount of forms similar to StE. For example, when analysing demonstratives, the forms *this* and *that* occurred with high frequencies in the data. Even if these forms seemed at first to represent author modifications, the analysis revealed that they had served as true variants which can be learned from fixed deictic grammatical expressions that have developed during this time and are still attested today (cf. *distaem*, *diskaen*, *disaelan*). It can thus be assumed that PE as spoken on the Solomon Islands was closer to the lexifier than the other two varieties. This also became apparent during the analysis of (in)abilitative markers. The analysis revealed that forms with the morphemes *can* and *save* were used in all three varieties. While BIS showed a clear preference of the latter, in TP the former was dominantly attested. In SIP, a difference was observed in whether the particle was used to express ability or inability. While ability was mostly encoded with *save*, inability was dominantly encoded with *can*-based forms. While *save* was negated through the preverbal negative particle *no*, *no can* was not attested at all. Instead, the StE negative forms of the auxiliary, namely *can't* and *cannot*, were attested. This gives rise to suspicion that the forms were directly borrowed from StE. It is of interest that the forms *kan* and *kanot* still serve as variants of *no save* in contemporary SIP.

Returning to demonstratives, it is of interest that postnominal *here* represented the dominant attested form in the BIS data so that BIS behaved most differently from the other two varieties in terms of demonstrative marking. Since the French were present in Vanuatu, and considering in French the locative adverbs *ici* 'here' and *là* 'there' have grammaticalised into *postnominal* elements that form part of the demonstratives (see 6.1), it is possible to imagine that French language structures had an impact on the development and stabilisation of the postnominal marker. Another possibility is that substrate influence led to the use of postnominal *here* since

Bislama's substrate languages make use of postnominal demonstratives as well (Camden 1979: 76; Siegel 2008: 183-184).

New insights about the spread of the relative particle *where* could be gained. While previous studies did not attest the *where* particle in RC constructions in SIP and assumed it to have its origin in Vanuatu (Baker 1993), the present study could show that the particle was attested from 1908 onwards in SIP and from 1913 onwards in BIS. It was shown that the *where* particle might have its origin in the mission pidgin variety as used by the Queensland Kanaka Mission (QKM), or possibly, more generally throughout Queensland, and that it may have further spread to the Pacific Islands through the South Sea Evangelical Mission (SSEM) and their activities in the Solomons. Moreover, the assumption by Crowley (1990a: 330) and Siegel (2008) that the relative particle *where* was a widespread feature in all Melanesian Pidgin English varieties is called into question. In TP zero-marking represented the dominant strategy independent of time and the relative particle *where* was not attested at all. Although non-attestation does not necessarily mean that a feature was not present in a variety, it needs to be considered that the Europeans usually (over-)emphasised peculiarities that they felt deviated from StE use, when documenting the – for them – foreign language. Thus, it is reasonable to assume that they would have pointed out the use of *where* as a relative particle if it had been in use. Therefore, the results support Mühlhäusler's assumption that the *where* particle represents a recent, or at least a post-1950 innovation in TP.

All in all, it became clear that each of the morphosyntactic features showed a different temporal development pattern. In order to obtain a comprehensive picture of the validity of the theories introduced in Section 3.2, all morphosyntactic as well as lexical features and their attestations across time would need to be considered. The following section, which will evaluate the theories on the origin and development of SIP, BIS and TP in the light of the results of the present study, can thus only be understood as a starting point that needs to be extended and re-evaluated in the light of future studies.

10.2 Evaluation of theories on the origin and development of MPE varieties

As outlined in Section 3.2, several hypotheses have been proposed which try to explain how the Melanesian Pidgin English varieties emerged and how differences between SIP, BIS and TP developed. The major debate arose out of the question in how far plantation pidgins, especially those spoken on the plantations in Queensland and in Samoa, influenced the development of the Pacific varieties. Despite these debates, researchers basically agree that the three pidgins further diverged and developed their individual features with the end of the labour trade (cf. Mühlhäusler

1978; Clark 1979; Keesing 1988; Baker 1993; Jourdan & Keesing 1997; Sankoff 2021+). Although this assumption makes sense from a socio-historical perspective, no attempts had been made to do a linguistic comparative analysis to investigate whether the end of the labour trade was in fact the decisive factor for the individualisation of the varieties.

The results of the analysis show that we cannot generalise *per se* when a divergence and/or stabilisation of the three MPE varieties took place. As in non-pidgins and non-creoles, the divergence and stabilisation was gradual rather than abrupt. The forms used to encode the individual morphosyntactic features developed and stabilised at different points in time. Nonetheless, the results allow us to propose a possible scenario for the development of the three MPE varieties.

The assumption that the three MPE varieties began to diverge and individualise with the end of the labour trade is only partly true. It is true that the ctree analyses revealed that most of the statistically significant changes in the three varieties dated to the first half of the 20th century. These results suggest that significant changes in the choice and/or frequency of feature forms might have been propelled by the end of the labour trade. However, as mentioned earlier, the significant dates, in the first instance, show significant changes in the data but do not specify *per se* whether the change involves a reduction of variants and/or the introduction of new, diverging forms.

Again, a differentiation between the notions of *divergence* and *stabilisation* is required. Although most attested forms only began to **stabilise** after the end of the labour trade with overseas plantations, the attestation of forms such as *might* (dubitative modality), *with him* (comitative), *catch him* (terminative), *close up long* (adessive), and *where* (relative clause particle), which were exclusively attested in SIP and BIS and not in TP, made clear that the labour recruitment years (and not the end of the labour trade) seem to have had a major impact on the emergence of **diverging forms**. It is possible that these forms entered SIP and BIS via QPPE. For *close up long*, *might* and *want*, for example, possible predecessors were identified in the QPPE data.

Thus, even before the amount of variants was reduced in the varieties, differences between the variants for encoding features were observed. The varieties SIP and BIS started to *diverge* from TP (or the other way around) before they *stabilised*. Therefore, it can be assumed that the labour recruitment history had a major impact on *diverging* forms attested in the three varieties, although not necessarily on their stabilisation. Many features may have entered all three MPE varieties via QPPE, as suggested and shown by Baker (1993), but the present analysis demarcates that the historical circumstance that the contact between New Guinea labourers and Solomon

Island and Vanuatu labourers had already come to an end in 1885, has left its traces on the varieties' developments as well and contributed to their divergence.

This is also supported by the meta-linguistic evidence of ex-Queensland and ex-Samoan labourers who brought their knowledge of PE to their home countries, as outlined in Section 3.1.3.6. While evidence could be found that both QPPE and SPPE were brought to the Solomon Islands, Vanuatu and German New Guinea by repatriated labourers, the amount of evidence for SPPE brought to the Solomon Islands and Vanuatu and the amount of evidence for QPPE brought to German New Guinea was rather low.

The early separation of New Guinea labourers from Solomon Island and Vanuatu labourers may also explain why TP represents the variety which has stabilised the most by the end of the 1940s. Less variation was observed and competing forms were attested with lower frequencies. This supports Mühlhäusler's (1978) claim that the variety already began to stabilise outside New Guinea, once labourers were no longer recruited to work on Queensland plantations.

It is of interest that although some diverging forms were already attested prior to the end of the labour trade, the analysis of early metalinguistic statements revealed that early travellers and writers did not observe, or at least did not report about *local variation* before 1906, which may have been due to their untrained ears. Moreover, although a growing number of writers observed differences between the varieties from 1926 onwards, they predominantly referred to lexical differences in the geographical areas and not to morphosyntactic variation.¹⁹⁵ One possible explanation for why early travellers and Pacific Island visitors did not comment on morphosyntactic differences of forms may have been that they did not have the linguistic expertise and that several competing forms still coexisted so that the use of diverging forms did not attract their attention.

The analysis further revealed that *mission* Pidgin English varieties had a stronger impact on the development and divergence of the early MPE varieties than previous studies supposed. The relative particle *where*, for instance, was attested in locative relative positions in QPPE but only occurred in an object relative clause in an early letter written by a Pacific Islander of the QKM in Queensland. This attestation allows the assumption that when the QKM turned into the SSEM and moved to the Solomon Islands, this may have contributed to the further spread and establishment of the *where* particle in SIP and BIS. Although Mühlhäusler & Mühlhäusler (2005) claim that the SSEM variety was artificial and constructed based on existing varieties rather than representing PE, its impact on the formation and development of PE in the Solomon Islands (and

¹⁹⁵ Future research should thus include lexical features as well to observe whether a lexical differentiation of the three varieties has started earlier.

to a lesser extent in Vanuatu) should not be underestimated. Even if the mission variety contains a high degree of forms similar to StE and missionaries might have switched between using PE and ‘Simple English’, mission teachers and students will have been exposed to the varieties so that forms closer to StE may have entered PE. Contemporary forms that are similar to StE, such as the SIP inabilitative markers *kan* and *kanot* that were common in the early SSEM data and which are still attested in contemporary SIP, provide evidence for this assumption.

However, not all the *diverging feature forms* can be ascribed to the labour recruitment histories or to an influence of a mission variety. The datasets showed that there were some features of which all variants were attested in all three varieties. For instance, in all varieties the morphemes *save* and *can* were attested as variants to encode abilitative/inabilitative modality. While in SIP and BIS *save* occurred with higher frequencies, which is also the form found in contemporary SIP and BIS, in TP the form *can* was most dominant. Today, *inap* represents the major abilitative/inabilitative marker in TP though *can* may also be used. Even though the end of the labour trade is likely to have propelled the individual development of the varieties, it does not explain so far why the varieties make different choices. The same applies to differences that were identified in the choice of demonstrative forms. Explanations need to be found concerning why BIS encodes demonstratives with postponed *here*, TP dominantly with the particle *dispela*, while in SIP the forms *disfela*, *datfela* and *ia* are in use, although in the early data all three forms were present in all three varieties. One of the possible explanations why the varieties made diverging form selections represents *substrate reinforcement*. Early substrate language data that may have had an impact on the development of the three MPE varieties could not be collected within the scope of this study since the areas were and are characterised by extreme language diversity (see Chapter 1). Nonetheless, substrate influence represents one of the possible factors that may have influenced the form selection, as Siegel’s (2008) study based on contemporary language data of MPE and selected substrate varieties demonstrated.

The present study showed that the degree of exposure to the lexifier language represents a further possible reason for the divergence of the varieties. As SIP was exposed to a greater degree to the English lexifier (or at least to the mission PE which was closer to StE), this may have been one of the reasons why StE similar forms, such as *kan*, *kanot*, *kanduit*, prevailed in the variety.

In sum, it can be stated that several reasons for the divergence of the three MPE varieties and their stabilisation can be observed. Major reasons for *diverging forms* that were identified in the present study are:

- a) the labour recruitment histories and, concomitant with it, the degree of impact of overseas plantation pidgins (QPPE vs. SPPE);
- b) the influence of mission varieties;
- c) substrate reinforcement (which will have increased after the end of the labour trade); and
- d) the degree of exposure to the lexifier language.

By contrast, major reasons for the *stabilisation* of each of the individual varieties, and thus for the reduction of variants, or at least a preference of a variant, represent:

- a) the end of the labour trade (as revealed through the increased number of time-based splits in the first half of the 20th century);
- b) the spread and use of the varieties to further domains in the home areas; and
- c) as shown by Siegel (2008), substrate influence.

For some of the variables the number of variants was still high by the end of the 1940s, and as feature forms continued to change and stabilise after 1950, it is likely that further factors contributed to the stabilisation and/or divergence of the varieties. One of these possible factors, which was not discussed in this study, is the implementation and use of Pidgin English for mass communication. Before the Second World War, written PE was used for religious purposes only. By using *written* PE for mass communication during wartime (i.e. in pamphlets), the varieties were exposed to standardising processes for the first time. Attempts to compile grammars and the fact that PE became increasingly used for mass communication in newspapers in New Guinea and Vanuatu after 1950 are assumed to have had a great impact on the further stabilisation of the MPE varieties.¹⁹⁶

All in all, the results of the analysis show that a combination of the theories proposed by Baker (1993), Mühlhäusler (1978) and *the end of the labour trade* theory seems to be most plausible in explaining the development and divergence of SIP, BIS and TP. This does not exclude a possible influence of early jargons or a preceding Maritime Polynesian Pidgin (Drechsel 2014). Moreover, the varying degrees of exposure to mission Pidgin English varieties contributed to the divergence of the MPE varieties as well.

10.3 Theoretical implications for the stabilisation of contact varieties

The present study provides insights not only into the history of the MPE varieties but also into the processes involved in the development from unstable to stable contact varieties in general. This section will discuss these theoretical implications and thus aims at answering the fourth research question.

¹⁹⁶ For example, in Vanuatu the French issued the newsletter *Bulletin d'information de la Résidence de France*. Although it was initially written in French, great parts of it were translated into Bislama and there are some editions which were completely issued in Bislama (cf. e.g. vol. 8, no. 1/2 (Jan./Feb. 1968), vol. 8, no. 4 (Apr. 1968)-vol. 9, no. 1 (Jan. 1969); vol. 9, no. 3-vol. 9, no. 12 (Mar. 1969-Dec. 1969)).

As outlined in the introduction, researchers do not agree on how exactly pidgin and creole languages came about, which processes were at work during their stabilisation and as such during grammaticalisation and feature selection. Moreover, creolists do not agree on whether these processes are unique to pidgins and creoles, or whether the mechanisms at work during their formation and stabilisation are universal. Further discussions are based on the extent to which pidgins and creoles differ structurally from non-creoles, whether creoles by definition emerge from pidgins and whether language-internal as well as language-external historical data is required to understand a pidgin or creole's history and development.

The data that provided the basis for the present study supports the assumption that the ecology (i.e. the particular setting, the groups involved, levels of contact, etc.) played a decisive role in the development and stabilisation of contact varieties. The data shows that unique linguistic ecologies will have led to the development of unique feature pools in different areas and thus provides evidence in support of the theory of an *evolutionary account of creole formation* (cf. Ansaldo 2009a, 2009b; DeGraff 2014; Yakpo 2021+) and for the assumption that feature pools and feature selection (cf. Croft 2000; Mufwene 2001, 2006; Ansaldo 2009a) played a major role in the formation of the contact varieties. Some of the differences in the three MPE varieties can be explained in that the contact situation and the ecologies differed. For example, German lexical items were only found in the early pidgin data of New Guinea (e.g. *raus* 'throw out' in von Hesse-Wartegg 1902: 53), whereas a French influence could only be observed in the early data attested for the Vanuatu region (e.g. *allez* 'and then' in Lynch 1923: 3309). Furthermore, similarities and differences between the varieties could be explained by the varying degrees to which SPPE and QPPE were brought to the regions by repatriated labourers and by the degree to which the labourers used their pidgin knowledge in renewed contact with Europeans in their home areas.

While the study supports the assumption that the linguistic ecology will have led to individual feature pools in the three areas, it also shows that the selection and stabilisation of features was a *lengthier and more complex process* than suggested models might let us assume. According to Mufwene's *Feature Pool Hypothesis* (2001, 2005, 2006), whether a feature stabilises or not is dependent on its salience and/or on how well a feature is represented in the feature pool. The more frequent and typologically common a feature, the more likely it will be replicated and be taken over into the developing language.¹⁹⁷ While it makes sense to assume that

¹⁹⁷ Again, we come across the problem of defining when a variety can be considered a stable variety. The model by Mufwene shows that the feature pool may result in several output varieties. However, it is not clear how much variation is still allowed to be observed in those varieties to consider them as individual output varieties. In other words, it is not clarified when a *feature pool* turns into an *output variety*.

the relative frequency of a feature was a determining factor in the selection of features, and that, for example, the demographic composition will have contributed to whether a feature was more or less strongly represented in a feature pool, the ctrees indicated that the feature pools seem to have changed their setup over and over across time until a stable contact variety developed. For example, when looking at demonstratives and their development across time in Solomon Islands Pijin, the conditional inference trees showed that while in the first years *this* and *that* represented the dominant demonstratives, and in a subsequent period *this fellow* was attested as the dominant form, in the next following time period *this* returned to be the most frequently attested demonstrative (cf. Figure 22 and 23). Even if these variations may partly be based on the availability and quality of historical data (i.e. underrepresented vs. overrepresented time periods, social imbalance between data derived from European vs. non-European writers), it is likely that feature pools changed their setup across time.

The socio-historical account in Section 3.1 provided evidence that *diffusion* may be one of the reasons why the feature pools may have changed continuously, since diffusion will have led to the introduction of new features from other areas into the individual feature pools. Europeans travelled from one Pacific Island to another, trading vessels encountered various contact varieties and multilingual ship crews will have spoken (individual) forms of the early pidgins and contributed to the spread of features when arriving in new coastal areas or when changing ship crews. The movement of labourers to overseas plantations, as well as to foreign areas in their home islands, will have led to the introduction of new features and/or feature forms into the feature pools of an area. A linguistic example of diffusion in the present study is the relative pronoun *husat*, which was used in a pamphlet produced by the Australian Government. The war pamphlets which were produced in Tok Pisin during World War II were not only dropped over New Guinea and Bougainville and Buka, but also over other parts of the Solomon Islands so that the Tok Pisin feature was found in other data collected for the Solomon Islands. Although *husat* did not establish itself in SIP, other features that were brought to new areas might have done so. For example, in Chapter 7 it was shown that the relative particle *where* is likely to have been brought to the Solomon Islands either by missionaries of the SSEM or by returnees that had been in contact with the mission. That diffusion represented a significant factor in the emergence of creoles has also been shown by Baker (1999) focussing on contact varieties in the Caribbean and by Ansaldo (2009b) analysing Asian contact languages.

Moreover, the present data provided evidence that sociological factors seem to have played an important role in the feature selection as well. For example, it became clear that missionaries had a strong impact on the development of the early contact varieties. Even though

they were lower in number, the social impact of what they said was higher due to the fact that they were teachers and became linguistic role models. Moreover, they used Pidgin English not only in the spoken but also in the written mode and thus may have contributed to the standardisation of the early contact varieties. This shows that socio-cultural factors should not and cannot be ignored when aiming to understand the development of contact languages.

Furthermore, the early data gives rise to the suspicion that the language ecology of the earliest stage shapes the resulting contact variety to a greater degree than later ecology stages. Although it was earlier pointed out that French-originating lexical items were attested in Bislama, the number of features that can be referred to a French origin is rather low. This is remarkable since both the British and French had been present in Vanuatu since 1878 and jointly administered Vanuatu from 1906 until 1980. One possible explanation for this represents the Founder Principle (cf. Chaudenson 1992, 2001; Mufwene 1996, 2001). The contact varieties that were brought to Vanuatu by the first returning labourers may have shaped the resulting contact variety to a great degree which may explain why Bislama is English-lexified despite the fact that it was also French governed.

Moreover, the present study showed that instead of a single linguistic ecology and a single feature pool it is likely that several regionally unique ecologies and feature pools coexisted, which may have influenced each other through diffusion. It became clear that even if a developing contact variety has not stabilised yet and the feature selection has not been completed, a feature pool may have gotten into contact with other *contact and non-contact* varieties. In other words, feature pools may have transformed across time, and it is likely that several feature pools coexisted, influenced each other, and/or combined.

For example, the contact situation in the Solomon Islands had led to a unique linguistic ecology with languages such as QPPE (and, to a lesser degree, SPPE), ship jargons, English, various substrate languages, etc., which all contributed features to a feature pool. The feature selection was not yet completed when the South Sea Evangelical Mission (SSEM) started to select features from this feature pool when creating the PE to be used in its mission. On the mission stations, English nonetheless played an important role and the *not-yet-stabilised* pidgin variety came into renewed strong contact with English. This will of course have influenced the compilation of the feature pool. The feature pool is likely to have differed from other feature pools in other areas in the Solomons where the mission was not present. It is likely that StE forms have been reinforced so that the PE variety showed a higher frequency of StE forms, many of which have stabilised until today.

In this context it is also important to mention that *second and third language acquisition processes* are likely to have had an impact on the formation and development of the pidgin and creole varieties as well. Arends (1989, 1992) claims that L2 acquisitional processes (together with locally-born L1 learners) would represent the driving forces for creole formation (cf. Cardoso 2009). The data in the present study has shown that the Pacific Islanders were usually multilingual. Thus, the influence of second and third language acquisitional processes cannot be ignored. When learning a second or third language, learners frequently transfer elements of the linguistic system of their L1 on the L2. Thus, it is likely that the feature selection may have been influenced by *substrate reinforcement* and features may or may not have been modified or taken over due to L1 *transfer*.¹⁹⁸ Studies such as the one by Siegel (2008) have shown that when comparing present day language material of the MPEs with data of some of their substrates, similarities in their structures can be identified. Therefore, it makes sense to assume that substrate reinforcement also played a role in the past.¹⁹⁹ It is, however, not sufficient to only focus on substrate influence but it is more than likely that adstrate languages influenced the development of the contact varieties as well.²⁰⁰

Indigenous languages continued to be present and spoken alongside the developing pidgincreoles in the Melanesian area until today. Thus, it is likely that the adstrate languages in each of the Melanesian areas will have had an impact on the individual development of SIP, BIS and TP. In addition, most of the transfer and substrate theories that have been put forth only place their focus on the L1 and L2. From the data at hand it becomes clear that the situation must have been more complex in the Pacific Islands. We learned from the accounts of the SSEM, for example, that next to Pidgin English, missionaries tried to teach English to their scholars. Thus, third language acquisitional processes may have played a role in feature selection as well.²⁰¹

The contact varieties discussed in this study developed in multilingual societies. The findings in the present study support the assumption of Muysken (2008, 2021+) that processes such as *borrowing*, *code-switching* and *code-mixing* will have played a role not only in the initial

¹⁹⁸ It should be noted that transfer is not restricted to substrate speakers but may and is likely to have occurred with speakers of the lexifier as well.

¹⁹⁹ In order to compare the early forms of the contact varieties with their substrates, historical data of the substrate languages is required as languages change across time.

²⁰⁰ The difference between substrate influence and adstrate influence is that the former describes the influence of a language on an emerging creole “during original pidginization/creolization processes”, whereas the latter describes the influence of a language “leading to the subsequent development and entrenchment of creole features” (Corum 2015: 15 referring to Huber 2005).

²⁰¹ It cannot be assumed that English necessarily represented the L3 and PE the L2 of those speakers that acquired both languages. For some speakers English may have been the L2 and Pidgin English the L3 and for yet others, English and PE may have been the third, fourth or fifth languages they acquired. A better way to refer to the renewed contact with the English lexifier might therefore be to call it *lexifier reinforcement*.

intercultural encounters, but also in the later development of the varieties (cf. also Johanson 1992, 1993 who prefers the designation *code-copying*). Evidence was found in court proceedings and in some of the SSEM documents, in which code-switching and borrowings were attested.

Other factors, such as the *level of interaction* between the individual groups, *language attitudes* and *government policies* will have further influenced the development of the contact varieties and will have had an impact on whether features originating in PEs that were brought to the areas by returning labourers were reinforced by either substrate languages of the areas, by renewed contact to the lexifier or whether they were replaced by new forms. It is possible that negative attitudes towards a pidgin together with a support of the English language may have led to a higher degree of StE similar features.

It was not the aim of the present study to analyse whether pidgins and creoles represent a special class of languages with specific typological properties, as suggested by McWhorter (2002, 2005), Parkvall (2008) and Bakker et al. (2013). To answer such a question it would be necessary to systematically compare linguistic features of pidgins and creoles with features of non-pidgin and non-creole languages (cf., for instance, Velupillai 2021). Nonetheless, the study has shown that in order to understand the development path of contact languages, it is not sufficient to perform typological studies on present day language data. It became clear that linguistic features may have existed but disappeared or vice versa. To truly understand the linguistic structure of pidgins and creoles and to answer the question whether they form a typological distinct class on their own, we will have to collect and take into consideration historical data of the varieties.

In summary, the findings of the present study suggest that the processes involved in the emergence of pidgins and creoles and in the development of their features are more complex than individual theories might lead us to believe. The analysis of early metalinguistic and linguistic data suggests that it is likely that several of the mechanism which are put forward in the individual theories were involved and interacted during the emergence and stabilisation of the contact varieties.²⁰² None of the existing theories alone is sufficient to explain the complex development of the MPE varieties.

Figure 176 represents an attempt to summarise the findings of the study in a single illustration taking Mufwene's feature pool model as a starting point. In contact situations different languages (represented through the red, yellow and green boxes) come into contact and are likely to contribute features to a feature pool. As the feature pool is based on the unique linguistic ecologies, it is also likely that several regionally unique feature pools coexisted since different

²⁰² Even if I have not referred to creolisation in the study, it is assumed that the first mother tongue speakers of the PEs are likely to have contributed to the stabilisation of language structures as well.

L1s may have been present and/or dominant in an area. The selection of features from the feature pool seems to represent a gradual and lengthy process (at least for the MPE varieties), which is exemplified through the blue feature pool boxes which transform several times over a longer period.

Moreover, as mentioned earlier, it is likely that not only stable output varieties but also unstable varieties (i.e. feature pools) served as input for new contact varieties, which is exemplified in the figure with the second blue feature pool which serves as the input for the emerging dark blue variety.

Even once a stable output variety has developed, common mechanisms of multilingualism such as borrowing, code-switching and code-mixing will have contributed to the further development of the language. In addition, different pidgin varieties may have co-existed, and may have influenced each other through initial and/or renewed contact.

During the process of stabilisation, which is generally understood as the reduction of variation, features might be taken over and may be further modified.²⁰³ It is also possible that features entered the initial feature pool, then disappeared but re-entered into the feature pool, as exemplified with feature ⑤ in the figure. The reason for this is that most of the input languages will continue to be used even after the initial contact situation because the areas are characterised by multilingualism. This is visualised with the help of the red, green and yellow arrows which extend over the complete period. Linguistic (substrate reinforcement, adstrate influence, code-switching, code-mixing, frequency and typological commonness, etc.) and non-linguistic factors (demographic changes in contact situation, level of interaction between input language speakers, exposure to the lexifier, etc.) will have contributed to whether features were retained or not.

As mentioned earlier, it is difficult to define when a variety can be considered a ‘stable’ variety. This *inter alia* is based on that all languages continuously change throughout their lifetime. It may thus be argued that languages can be understood as feature pools which change across time and from which features are selected throughout their history.

Regarding the choice of feature forms that are retained, the study has shown that grammatical forms which resemble StE forms do not necessarily have to be transmitted directly from the lexifier or to be kept right from the start but may have entered the contact language via another contact variety and/or at a later point in time. For instance, the volition verbs *like* and *want* were attested in QPPE. The occurrence of the verbs in SIP and BIS may thus either be directly borrowed from the lexifier or indirectly from QPPE.

²⁰³ It needs to be recalled that stabilisation remains a critical concept in that it is unclear how much variation is still allowed.

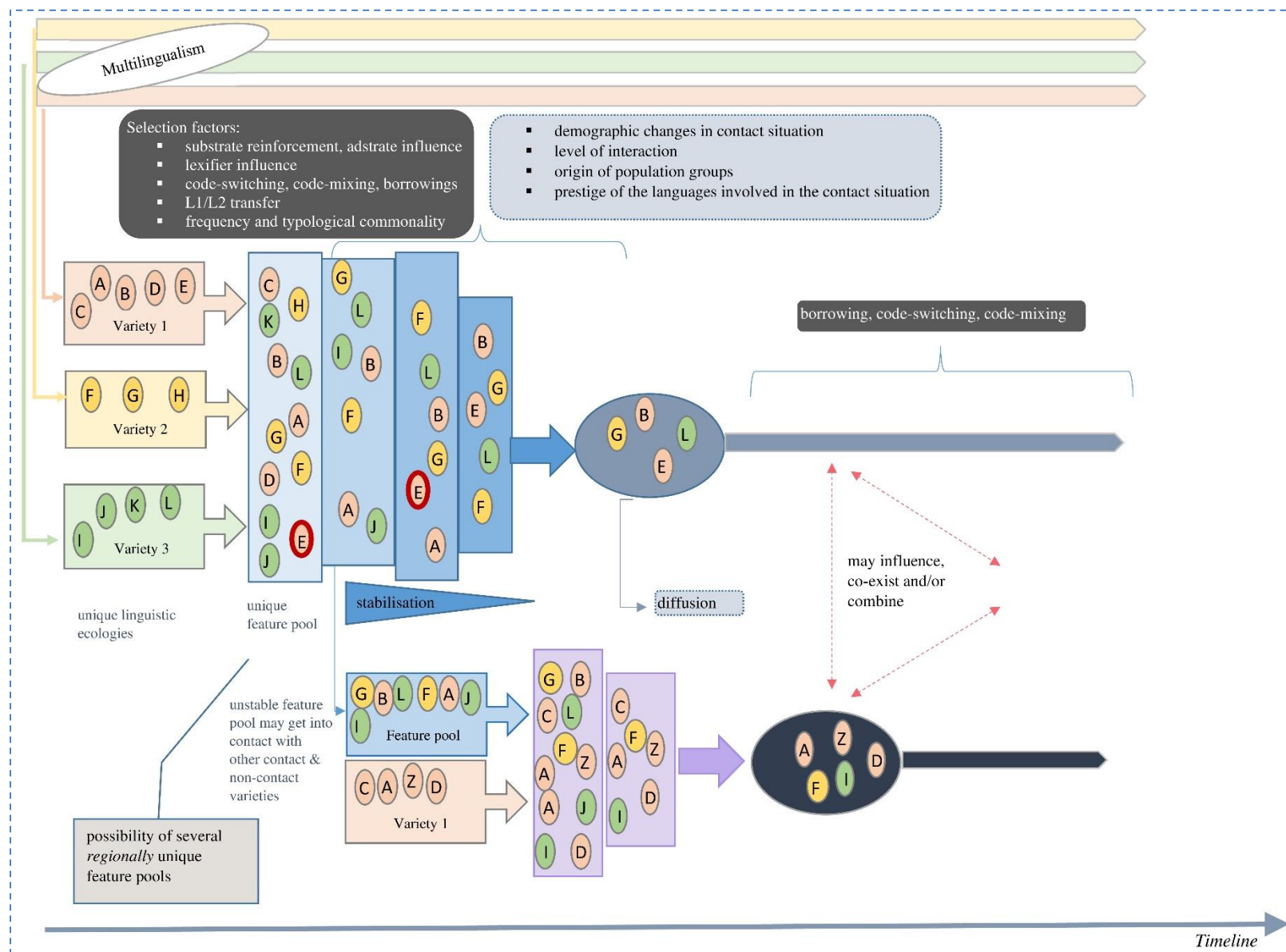


Figure 176: Complexity of contact language feature development (based on evidence of MPE varieties)

Independent of whether lexifier forms were directly transmitted in the early contact situation or via another contact variety, the forms could still go through functional and formal changes. In BIS the form *wandem* indicates that the transitive marker *him* was attached to the morpheme *want* and in SIP the volition verbs are optionally marked with the transitive marker as well.

There are also grammatical functions that were not initially transmitted into the feature pool and, therefore, were not overtly expressed in the initial contact situation. If a contact variety expands and gets used in more domains than it was originally used in, a need might develop to encode grammatical functions that had not been overtly encoded before. Even if grammatical functions were overtly encoded in initial contact situations, it is possible that new forms developed and replaced earlier forms. Three different mechanisms were observed to have played a role during this development in MPE, namely *grammaticalisation*, *reanalysis* and *degrammaticalisation*.

Through grammaticalisation, “lexical items and constructions come in particular linguistic contexts to serve grammatical functions and once grammaticalised, continue to develop new grammatical functions” (Hopper & Traugott 2003: xv).²⁰⁴ Examples of grammaticalisation in the present study represent the forms *save*, *here*, *inap*, *wantaim*, *inap long* and *kitchim*, of which the grammaticalisation paths have been outlined in the previous chapters, which is why no further comments will be made here:

- a) *save* > *save*
'know' ABILITY
- b) *here* > *ia*
'here' DEM
- c) *enough* > *inap*
'be sufficient, enough'; 'be suitable' ABILITY
- d) *enough + long* > *inap long* > *inap long*
'be sufficient, enough'; 'be suitable' PREP.TERM PURP.CONJ
- e) *catch* > *catch+him* > *catch+him*
'reach'; 'arrive' 'reach'; 'arrive' + TR PREP.TERM
- f) *one time* > *one time* > *one time + long* > *one time* > *one time* > *one time*
'one time'; 'together' 'together' + COM COM INST CONJ
'at one time'

The form *ating* showed that the development of grammatical forms does not necessarily have to involve a process of grammaticalisation. The form has its origin in the StE string 'I think'. Although it is possible that in the early contact variety *I think* was used as an acrolectal variant of

²⁰⁴ Cf. Bruyn (2008, 2009) for an overview of processes of grammaticalisation considered to play a role in the development of pidgin and creole languages.

me think which acted as a modal adverb,²⁰⁵ it was shown that there is also a considerable chance that the form is the result of a reanalysis or reinterpretation of the string. While Europeans may have used the string to express 1SG + Verb, Pacific Islanders may have reanalysed the whole string as a single morpheme expressing doubt or speculation.

Of interest is also the development of the form *might*. Although *might* is used in the lexifier language to encode conjunctive modality and, therefore, shares the dubitative/speculative meaning which the form has in SIP and BIS, the form does not seem to be directly transmitted as a modality marker into the MPE varieties. From the historical data it can be learned that *might* was used as an adverb meaning ‘maybe’. During the transmission of the particle from StE into SIP/BIS, the form may thus have been reanalysed or even degrammaticalised.

g)	<i>might</i>	>	<i>might</i>	>	<i>might/maet</i>
	AUX		‘maybe’		SPECULATIVE

Fulfilling a lexical function, the item was flexible in the syntactic position in which it could occur. When it then grammaticalised into the speculative modality marker, its position grammaticalised as well, restricting it to sentence- and clause-initial positions.

The identified grammaticalisation paths turned out to be in accordance with the overall tendencies described in typological literature: from a cross-linguistic perspective relative clause markers frequently originate in demonstratives, adverbial forms (i.e. *here*) and *w*-question words, demonstratives frequently grammaticalize out of locative adverbials such as English *here* and comitatives have been observed to have their origin in verbs encoding the meaning ‘accompany’ (cf., for example, Kuteva et al. 2019). Verbs with the meaning ‘know’ frequently develop into abilitative markers and terminative prepositions have been observed to develop out of terminative verbs.²⁰⁶ Similar observations were made in the present study which indicate that the three MPE varieties align with crosslinguistic tendencies. The grammaticalisation paths seem to be “universal to human experience” (cf. Bybee 2003: 151) and cognitively natural, which might be considered an argument against the *Creole Prototype Theory* and for typological markedness.²⁰⁷

²⁰⁵ The string ‘I think’ can act as a modal adverb in English as well (cf. de Haan 2006: 38).

²⁰⁶ The most interesting form that was covered in this study might be the form *enough* which developed not only into the abilitative marker but also into the terminative preposition in TP. However, the grammaticalisation of the form seems to be cognitively natural. If a person *is enough* or has *enough* knowledge how to do something, the person has an ability. The development of *enough long* as a terminative preposition is of interest since the word *enough* ‘enough’ first functioned as an adverb while *long* was used as the preposition but the string ‘enough long’ grammaticalised into the terminative preposition.

²⁰⁷ It needs to be noted that substrate and adstrate influence might have an impact on the grammaticalization paths as well.

In sum, the present study has shown that the origin and development of contact languages is more complex than individual theories let us assume. The reasons why the theories do not completely depict the complexity of the emergence of pidgins and creoles are that their creators usually based their assumptions on present day language data and focussed on a single (or a restricted) amount of varieties. The present study showed, however, that historical metalinguistic as well as linguistic data is needed to avoid the risk of simplifying the circumstances under which pidgins and creoles emerged. Even if the use of historical linguistic and sociolinguistic data is not without its hazards, it is essential and inevitable to trace and reconstruct the development scenarios of contact varieties.

Moreover, the present study, which was solely based on three related varieties, revealed that it is inappropriate to measure the development of the various pidgins and creoles around the world by the same yardstick. As differences could already be observed in the factors that influenced the individual development of the three MPE varieties, it can be expected that the contact situations that led to the evolution of pidgins and creoles all over the world were unique as well. We need to collect historical data – linguistic as well as metalinguistic – and researchers need to collaborate and share their results to identify similarities and differences in the development of pidgin and creole languages. This is the only way in which we can obtain a comprehensive understanding of the development of pidgin and creole languages around the world.

10.4 Concluding remarks and outlook

The diachronic dimension of pidgins and creoles is essential for a proper understanding of their origin and development. As previously pointed out, studies focussing on differences and similarities between the three MPE varieties were frequently based on present language data (cf. Siegel 2008) or focussed on historical data of just one of the three varieties (cf., for instance, Mühlhäusler 1978). Moreover, usually a qualitative approach was taken. Studies that included a comparison of the three varieties were rather interested in whether a QPPE or SPPE influence can be observed, were based on comparative feature lists and did not focus on the later individualisation of the varieties. Although the end of the labour trade is generally considered as the decisive factor for the divergence of the MPE varieties, no attempts had been made so far to analyse this assumption based on early language data and by using *statistical modelling techniques*.

The reason why statistical methods are usually avoided by historical creolists is that the early documentation of the varieties is scarce and that data that researchers are able to collect is

usually biased. However, the biased nature of historical evidence actually increases the necessity for an application of statistical methods to analyse whether generalising conclusions can be drawn from the data. The surviving historical data of various sizes needs to be compared qualitatively but also quantitatively and significance testing is required to decide whether perceived patterns are significant or simply due to chance.

The selection of appropriate methods to investigate the historical data, however, represented a challenge. The fact that the data is characterised by a diversity of spelling variants and that the amount of data is very scarce made it impossible to apply corpus linguistic analytical tools to the data. Moreover, as decisions regarding the number and length of time periods can lead to different analyses results (cf. Gries & Hilpert 2012: 136), a pre-defined categorisation was avoided. A tool was required which can tell if there is any kind of meaningful grouping regarding the year of attestation. This tool also had to indicate whether a diverging amount of input by different authors has an impact on the representativeness of the data. Due to the complexity of the present study, focussing on four morphosyntactic categories in three different varieties with unequally distributed datapoints across time (and an unequal input of datapoints by the early authors), *conditional inference trees* were considered as the best tool for the analysis in the present study, as they allow us to determine whether there is a dependence between the known attested forms of the individual features and the years of attestation. The ctree algorithm was chosen as it can even cope with “highly correlated predictor variables” (Strobl et al. 2008: 1) and as it allows us to also include other possible factors that might help to “predict best the known outcomes of the dependent variable” (Bernaisch et al. 2014: 14).

The results of the analysis have shown that although the historical records of the three MPE varieties are biased, they can nonetheless be processed statistically. The ctrees provided a useful tool to identify the major predictor variables that are responsible for a change in the data and to detect how various predictor variables interact. Even though the author variable turned out to be the most dominant predictor variable in many cases, this does not invalidate the data or mean that the author is not reliable, but shows that given the nature of the sources, information about the author is important.

Ctrees helped to identify inconsistencies, author idiosyncrasies, and factors that played a statistically significant role in *changes* in the individual datasets.²⁰⁸ However, it also became clear that these changes do not necessarily lead to an individualisation of the varieties and do not

²⁰⁸ Cf. for example, the occurrence of *near* in Section 9.4.3. The ctree as displayed in Figure 166 showed that the occurrence of *near* as an adessive marker can be explained by the factor text type (TXT_TYPE_3) which helped to classify the occurrences of *near* as unreliable.

necessarily explain the origin of diverging forms. In order to evaluate whether the significant dates of changes can be taken as indicators for causes of the individualisation, a qualitative analysis was inevitable. Moreover, researchers need to know their databases well enough to critically evaluate the ctree results. If the input data does not cover a certain timespan and, consequently, ctrees do not identify any significant changes in the data for this timespan, it would be wrong to claim that the variety did not change over time. Researchers need to be aware that the input data has an impact on whether significant changes can be identified or not.

Thus, the present study provided evidence that the application of statistical modelling techniques to early historical pidgin and creole language data can help us to evaluate its reliability. At the same time it showed that the statistical methods available so far can complement but not replace a qualitative analysis. In the future, collaborations between statisticians and creolists may lead to new innovative ideas to analyse the biased language data. Although several strategies were developed to meet the challenges of working with historical written data and to avoid a biased sample, some limitations need to be conceded which might be a further starting point for future research.

Although it became already visible that BIS language data was underrepresented in the collected material, during the case study analyses it turned out that the size of the database was more problematic than previously expected. The BIS data collection did not contain enough attestations of some of the analysed features which made an assertion about the feature development, as well as a comparative analysis with TP and SIP, impossible.

Being predominantly restricted to early written records,²⁰⁹ and since not all early sources were accessible, all three varieties were over- or underrepresented in certain time periods. As individual researchers have collected early attestations of the three MPE varieties for several decades, the general and future objective should be to assemble the individual data collections into a single database.²¹⁰ Although the datasets of the three varieties would probably still be imbalanced due to the surviving records, which is beyond the control of the researchers, it would contribute to make the database as representative as possible of the early MPE varieties.

The gathering of early collected language material of the three varieties may further contribute in making judgments about the reliability of early attested forms. A major challenge of the present study was the decision about the handling of forms resembling StE. StE-derived forms that were attested with no change in their meaning and function (but which no longer exist

²⁰⁹ Rudolf Poech's (1904) early phonogram record of TP represents the one exception.

²¹⁰ An electronic database that aims to assemble early attestations and descriptions of contact languages represents *The Database of Early Pidgin and Creole Texts* (DEPiCT). The database is, however, still under construction www.uni-giessen.de/cms/depict, last access 29 September 2021).

in the contemporary varieties) represent authentic forms that were used for a given time in the varieties. Yet, they may also be the result of author or editor modification or they may represent a rather acrolectal variety of PE.²¹¹ In the present study, StE resembling forms were not excluded per se since an exclusion would be tantamount to manipulating the data. In addition, such an approach requires a more precise definition of what is understood by StE. In the Pacific contact situation, for instance, not only British StE but also Australian StE would have to be considered, as well as colloquial and regional varieties of both British English and Australian English. In addition, it is necessary to keep in mind that standard varieties of English have changed across time as well, and that it is more than likely that the early Europeans spoke non-standard varieties in their early encounters with Pacific Islanders. Nonetheless, it remains an open question whether some of the StE-like forms represent authentic attestations. An inclusion or exclusion of these forms has an impact on the years that the ctree analysis indicates as significant.

Future studies on the origin and development should moreover focus on substrate and adstrate languages to discover the interplay between substrate influence, adstrate influence and the grammaticalisation of specific forms. As languages change over time, the aim should be to find early evidence of substrate and adstrate varieties in order to compare their early structures with the ones in the early MPE varieties. When focussing on prepositions, it became clear that it is also necessary to learn about substrate and adstrate languages from a present-day perspective in order to come to know how spatial and temporal concepts are encoded in the substrate and adstrate varieties. For this it might also be advisable to work closely with native speakers of the MPE varieties. This might help to identify different culture-dependent conceptualisations more easily.

While substrate influence may have had an impact on the choice of specific forms, *language policies* and *language attitudes* may have had an impact on the development of the varieties as well. Analysing language attitudes from a diachronic perspective and comparing these attitudes to the developments of the varieties may show whether language attitudes had an impact on language change as well. Thus, language policies and attitudes should be considered as further possible influencing factors in the stabilisation process in future projects.

It is very important to point out that the results of the comparison of four features cannot be generalised to the remaining morphosyntactic categories. It is possible that different

²¹¹ In order to decide whether language samples represent PE or not, it would be helpful to have more knowledge available about *English as spoken in the Solomon Islands*, *English as spoken on Vanuatu* and *English as spoken on New Guinea*. As to my knowledge, despite research on Fiji English (cf. Zipp 2014; Biewer 2015), studies on *Melanesian English* varieties do not exist and might be promising for future research activities. English as spoken in these areas should be analysed from a synchronic as well as diachronic perspective to not only learn about the contemporary forms, but also about English as spoken in these areas during the formative decades.

morphosyntactic features behave completely different from the here-analysed features. Therefore, I encourage future studies which expand our knowledge on the individual development of the three Melanesian Pidgin English varieties by performing analyses based on a greater range and number of features. Furthermore, the focus should not be restricted to morphosyntactic features but lexical variation should be considered as well. Studies based on further features are also required to explore the localisation of SIP and BIS more thoroughly.²¹²

Moreover, I encourage researchers to study the three varieties with a focus on the years following World War II, since the data gave rise to suspicion that some of the contemporary forms only developed and/or stabilised in the second half of the 20th century. We will only be able to understand the origin and development of differences in the varieties by performing a diachronic comparative analysis of *all* linguistic features of the varieties and by investigating how the latter have developed from their beginning until today. For this it is necessary to compile a corpus of the varieties that covers not only their early years, but their complete timespan of existence.

This would also allow us to investigate how renewed language contact affects the development of the three varieties. For instance, radio programs such as the *Wantok Program* by *abc Radio Australia* are in TP, SIP and BIS, and are broadcasted in the three island groups which means that Pacific Islanders are exposed to all three varieties. The establishment of airports resulted in facilitated mobility and university students from different Pacific areas get into contact through student mobility programmes. Not only the impact of renewed contact between the three Pidgin English varieties, but also with the English lexifier, and/or regional varieties of English, may further be of interest to understand the development of the varieties and might be promising for future research activities.

Another interesting aspect for future research will be to consider how the internet and modern media have influenced and still influence the three varieties. The use of *digitalk* (i.e. informal written register used in text messages, social media messages and social media posts) will increase the degree to which Pidgin English speakers are exposed to the varieties in a written form. This may lead to a greater acceptance of the Pidgin English varieties and may involve further grammaticalisation processes.

²¹² For example, differences can be observed in the way the two varieties encode purpose prepositions. While in TP and BIS the preposition *belong* grammaticalised as the purpose preposition, in SIP the form *for* turned into the purpose marker (cf. Schäfer in progress a).

References

- Abbott, John H. M. (1908). *Peeps at many lands: The South Seas (Melanesia)*. London: Adam & Charles Black.
- Aboh, Enoch (2009). "Competition and selection: That's all!" Eds. Enoch Aboh and Norval Smith, *Complex processes in new languages*. Amsterdam & Philadelphia: John Benjamins. 317–344.
- Aboh, Enoch and Umberto Ansaldo (2007). "The role of typology in language creation. A descriptive take." Eds. Umberto Ansaldo, Stephen Matthews and Lisa Lim, *Deconstructing creole. The rationale*. Amsterdam: John Benjamins. 39–66.
- Aboh, Enoch and Cécile Vigouroux (Eds.) (2021+). *Variation rolls the dice*. Amsterdam: John Benjamins.
- Adams, James N. (2013). *Social variation and the Latin language*. Cambridge: Cambridge University Press.
- Aho, Billy (23.12.1927). [Witness statement]. Billy Aho. Notes of evidence at trial of, on a charge of, murder. Western Pacific Archives. (MSS & Archives-2003/1-WPHC 4/IV-1928, 515/1928). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Aikhenvald, Alexandra (2006). "Serial verb constructions in typological perspective." Eds. Alexandra Aikhenvald and R. Dixon, *Serial verb constructions: A cross-linguistic typology*. Oxford: Oxford University Press. 1–68.
- Aitchison, Jean (1992). "Relative clauses in Tok Pisin: Is there a natural pathway?" Eds. Marinel Gerritsen and Dieter Stein, *Internal and external factors in syntactic change*. Berlin: De Gruyter Mouton. 295–316.
- Aikhenvald, Alexandra Y. (2010). *Imperatives and commands*. Oxford: Oxford University Press.
- Aldrich, Robert (1993). *France and the South Pacific since 1940*. London: Macmillan & Co.
- Alexander, Gilchrist (1927). *From the middle temple to the South Seas*. London: John Murray.
- Alinei, Mario (2004). "The problem of dating in linguistics." *Quaderni di Semantica: Rivista Internazionale di Semantica Teorica e Applicata* 25 (2): 211–232.
- Anonymous (19.07.1897). "Mining notes." *Kalgoorlie Miner*: 2.
- Anonymous (1910-1911). "In the terrors of the shadow of death." *Not in Vain. What God Hath Wrought Amongst South Sea Islanders, 25th Annual Report of the South Sea Evangelical Mission*: 42–43. Pacific Manuscripts Bureau. PMB Doc 439.
- Anonymous (26.07.1913). "Pidgin-English." *Samoanische Zeitung* 12.39.
- Anonymous (31.03.1938). "Prayers in Pidgin-English." *Advocate*: 32.
- Anonymous (01.07.1944). "Guadalcanal natives: Home front natives. Palaver with 'Big Fella Masta'." *Geraldton Guardian and Express*: 4.
- Anonymous (1945). "What is Pidgin English?" *Solomon Soldiers' News* 1. Pacific Manuscripts Bureau. AU PMB DOC 442.
- Ansaldo, Umberto (2009a). "Contact language formation in evolutionary terms." Eds. Enoch Aboh and Norval Smith, *Complex processes in new languages*. Amsterdam & Philadelphia: John Benjamins. 265–291.
- Ansaldo, Umberto (2009b). *Contact languages. Ecology and evolution in Asia*. Cambridge: Cambridge University Press.
- Ansaldo, Umberto (2010). "Identity alignment and language creation in multilingual communities." *Language Sciences* 32: 615–623.
- Arends, Jacques (1989). *Syntactic developments in Sranan: Creolization as a gradual process*. University of Nijmegen.
- Arends, Jacques (1992). "Towards a gradualist model of creolization." Eds. Francis Byrne and John Holm, *Atlantic meets Pacific: A global view of pidginization and creolization*. Amsterdam: John Benjamins. 371–380.
- Arends, Jacques (Ed.) (1995). *The early stages of creolization*. Amsterdam: John Benjamins.
- Army Education Branch (1944). *Melanesian Pidgin English language guide: First level*. Washington, D.C.: Government Printing Office.
- Baker, Paul (2010). *Sociolinguistics and corpus linguistics*. Edinburgh: Edinburgh University Press.
- Baker, Philip (1987). "Historical developments in Chinese Pidgin English and the nature of the relationships between the various Pidgin Englishes of the Pacific region." *Journal of Pidgin and Creole Studies* 2: 163–207.
- Baker, Philip (1999). "Investigating the origin and diffusion of shared features among the Atlantic English Creoles." Eds. Philip Baker and Adrienne Bruyn, *St. Kitts and the Atlantic Creoles: The texts of Samuel Augustus Mathews in perspective*. London: University of Westminster Press. 315–364.
- Baker, Philip (1993). "Australian influence on Melanesian Pidgin English." *Te Reo* 36: 3–67.
- Baker, Philip (1995). "Some developmental inferences from historical studies of pidgins and creoles." Ed. Jacques Arends, *The early stages of creolization*. Amsterdam: John Benjamins. 1–24.
- Baker, Philip (1996). "Productive fellow." Eds. Darrell Tryon, Peter Mühlhäusler and Stephen Wurm, *Atlas of languages of intercultural communication in the Pacific, Asia and the Americas*. Berlin & Boston: De Gruyter Mouton. 533–536.
- Baker, Philip and Magnus Huber (2001). "Atlantic, Pacific, and world-wide features." *English World-Wide* 22 (2): 157–208.
- Baker, Philip and Lise Winer (1999). "Separating the wheat from the chaff: How far can we rely on old pidgin and creole Texts?" Eds. Philip Baker and Adrienne Bruyn, *St. Kitts and the Atlantic Creoles: The texts of Samuel Augustus Mathews in perspective*. London: University of Westminster Press. 103–122.

- Baker, Sidney J. (1945). *The Australian language: Pidgin English in New Guinea*. Sydney: Angus & Robertson.
- Bakker, Peter (2008). "Pidgins versus creoles and pidgincreoles." Eds. Silvia Kouwenberg and John Singler, *The handbook of pidgin and creole studies*. Oxford: Wiley-Blackwell. 130–157.
- Bakker, Peter, Aymeric Daval-Markussen, Mikael Parkvall and Ingo Plag (2011). "Creoles are typologically distinct from non-creoles." *Journal of Pidgin and Creole Languages* 26 (1): 5–42.
- Bakker, Peter, Aymeric Daval-Markussen, Mikael Parkvall and Ingo Plag (2013). "Creoles are typologically distinct from non-creoles." Eds. Parth Bhatt and Tonjes Veenstra, *Creole languages and linguistic typology*. Amsterdam: John Benjamins. 9–45.
- Bakker, Peter and Aymeric Daval-Markussen (2017). "Typology of creole languages." Eds. Alexandra Aikhenvald and Robert Dixon, *The Cambridge handbook of linguistic typology*. Cambridge: Cambridge University Press. 254–286.
- Ballarè, Silvia and Guglielmo Inglese (2021). "The development of locative relative markers: From typology to sociolinguistics (and back)." *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"*. Online First Article. Amsterdam: John Benjamins.
- Banfield, Edmund J. (1913). *My tropic isle*. London & Leipzig: T. Fisher Unwin.
- Baraghu (18.09.1895). [Witness statement]. Australian Station (Solomon Islands, 1895). "Correspondence respecting outrages by natives on British subjects, and other matters which have been under inquiry during the year 1895 by her majesty's ship, Solomon Islands, 1925." In *Solomon Islands: Improper recruiting by Queensland labour vessel "Sybil"*. From: Bridge, Rear Admiral C. A., at H.M.S. "Orlando" at Sydney. Western Pacific Archives. (MSS & Archives-2003/1-WPHC 4/IV-1896,140/1896). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Barrett, Charles (1954). *Isles of the sun*. London: William Heinemann Ltd.
- Barros, João de (1777). "Do que aconteceu a D. Jorge de Menezes na jornada de Maluco, e de como descobrio as Ilhas dos Papuas: e da armada que partio de Castella pera aquellas Ilhas de Maluco, e da derrota que levou até chegar a ellas." Ed. João de Barros, *Da Asia de Diogo de Couto: Dos feitos, que os portugueses fizeram na conquista, e decubrimto das terras, e mares do oriente*. Lisabon: Regia Officina Typografica.
- Barrow, Lenox (1942-1947). "Outlying interlude" - an account of life in the British Solomon Islands Protectorate 1942-47. Pacific Manuscripts Bureau. AU PMB MS 517.
- Behrmann, Walter (1922). *Im Stromgebiet des Sepik. Eine deutsche Forschungsreise in Neuguinea*. Berlin: A. Scherl.
- Beimers, Garry (2009). *Pijin: A grammar of Solomon Islands Pidgin. A thesis submitted for the degree of Doctor of Philosophy of the University of New England*. University of New England.
- Bennett, George (1883). *An account of Elau, a Malayan Papuan child, and a native of the island of Erromanga, one of the New Hebrides group, and Southern Pacific Ocean*. Sydney: F. Cunninghame & Co.
- Bennett, Judith A. (1976). "Immigration, 'blackbirding,' labour recruiting?: The Hawaiian experience, 1877-1887." *Journal of Pacific History* 11 (1): 3–27.
- Bennett, Judith A. (1987). *Wealth of the Solomons: A history of a Pacific archipelago, 1800-1978*. Hawaii: University of Hawaii Press.
- Bennett, Judith A. (1993). "'We do not come here to be beaten': Resistance and the plantation system in the Solomon Islands to World War II." Eds. Brij Lal, Doug Munro and Edward Beechert, *Plantation workers: Resistance and accommodation*. Honolulu: University of Hawai'i Press. 129–186.
- Bernaish, Tobias, Stefan Gries and Joybrato Mukherjee (2014). "The dative alternation in South Asian English(es): Modelling predictors and predicting prototypes." *English World-Wide* 35 (1): 7–31.
- Biber, Douglas (2009). "Multi-dimensional approaches." Eds. Anke Lüdeling and Merja Kytö, *Corpus linguistics: An international handbook*. Berlin: De Gruyter Mouton. 822–855.
- Bickerton, Derek (1977). *Change and variation in Hawaiian English: Creole syntax*. Honolulu: Social Sciences and Linguistic Institute, University of Hawaii.
- Bickerton, Derek (1981). *Roots of language*. Ann Arbor: Karoma.
- Bickerton, Derek (1984). "The language bioprogram hypothesis." *The Behavioral and Brain Sciences* 7: 173–221.
- Bickerton, Derek (1988). "Creole languages and the bioprogram." Ed. Frederick Newmeyer, *Linguistics: The Cambridge survey, Vol. II*. Cambridge: Cambridge University Press. 267–284.
- Bickerton, Derek (1991). "On the supposed 'gradualness' of creole development." *Journal of Pidgin and Creole Languages* 6 (1): 25–58.
- Bickerton, Derek (2016 [1981]). *Roots of language*. Berlin: Language Science Press.
- Biewer, Carolin (2015). *South Pacific Englishes a sociolinguistic and morphosyntactic profile of Fiji English, Samoan English and Cook Islands English*. Amsterdam: John Benjamins.
- Billy (23.10.1894). [Witness statement]. Australian Station (Solomon Islands, 1894). "Correspondence respecting outrages by natives on British subjects, and other matters which have been under inquiry during the year 1894 by her majesty's ship." In *Proceedings in New Hebrides & Solomon Islands during 1894*. From: Bridge, Rear Admiral C. A., at H.M.S. "Orlando" at Port Sandwich. Western Pacific Archives. (MSS & Archives.2003/1.WPHC 4/IV.1895, 293/1895). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Borchardt, Karl (1926). *Kleines Woerterbuch Deutsch-Tokboi*. Unpublished manuscript.

- Braumann, Christoph (2018). *Alvar de Mendana - Auf der Suche nach der Terra Australis Incognita: Nach den Reiseberichten von 1567 und 1595*. Norderstedt: Book on Demand.
- Brewster, Adolf B. (1937). *King of the cannibal isles*. London: Robert Hale.
- Bridge, Cyprian (1886). "Cruises in Melanesia, Micronesia and Western Polynesia, in 1882, 1883 and 1884; and visits to New Guinea and the Louisiades in 1884 and 1885." *Proceedings of the Royal Geographical Society and Monthly Record of Geography* IX: 545–567.
- Brown, George (1908). *Pioneer missionary and explorer: An autobiography*. London: Hodder & Stoughton.
- Bruyn, Adrienne (1995a). *Grammaticalization in creoles: the development of determiners and relative clauses in Sranan*. Amsterdam: Institute for Functional Research into Language and Language Use (IFOTT).
- Bruyn, Adrienne (1995b). "Relative clauses in early Sranan." Ed. Jacques Arends, *The early stages of creolization*. Amsterdam: John Benjamins. 149–202.
- Bruyn, Adrienne (2008). "Grammaticalization in pidgins and creoles." Eds. Silvia Kouwenberg and John Singler, *The handbook of pidgin and creole studies*. Oxford: Wiley-Blackwell. 383–410.
- Bruyn, Adrienne (2009). "Grammaticalization in creoles: Ordinary and not-so-ordinary cases." *Studies in Language* 33 (2): 312–337.
- Bühler, Karl (1934). *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Jena: Fischer.
- Burnett, Frank (1911). *Through Polynesia and Papua: Wanderings with a camera in Southern Seas*. London: Francis Griffiths.
- Burns Philp and Company Limited (1883-1883). *Burns Philp and Company Head Office deposit*. Deposit N145 and N115. Noel Butlin Archives Centre, Australian National University.
- Burns Philp and Company Limited (1931). *Merchants, traders, shipping, travel & plantation owners. Company records - Minutes, financial records, staff records, photographs, maps and plans. Records relate to Burns Philp's involvement in the Pacific region, particularly Fiji, Kiribati, Norfolk Island, Papua New Guinea, Samoa, the Solomon Islands, Tonga and Vanuatu. 1884-1990. 'Statement of case of James Joseph Larkin'*. AU NBAC Z385-361 Burns Philp and Company holding deposit. Noel Butlin Archives Centre, Australian National University.
- Burt, Ben (1994). *Tradition and christianity: The colonial transformation of a Solomon Islands society*. Amsterdam: Harwood.
- Bybee, Joan (2003). "Cognitive processes in grammaticalization." Ed. Michael Tomasello, *The new psychology of language: Cognitive and functional approaches to language structure*. London & New Jersey: Lawrence Erlbaum Associates. 145-167.
- Bybee, Joan, Revere Perkins and William Pagliuca (1994). *The evolution of grammar: Tense, aspect and modality in the languages of the world*. Chicago: University of Chicago Press.
- Camden, William G. (1979). "Parallels in structure and lexicon and syntax between New Hebrides Bislama and the South Santo language spoken at Tangoa." *Papers in Pidgin and Creole Linguistics* No. 2: 51–118.
- Cameron, Charlotte (1923). *Two Years in Southern Seas*. Boston: Small Maynard & Company.
- Cardoso, Hugo (2009). "Jacques Arends' model of gradual creolization." Eds. Rachel Selbach, Hugo Cardoso and Margot van den Berg, *Gradual creolization. Studies celebrating Jacques Arends*. Amsterdam: John Benjamins. 13–23.
- Catholic Mission (October 1940). *Holahi-haho! Buk bilong singsing: Song book for natives of New Guinea*. Alexis Harbour: Catholic Mission Press.
- Cayley-Webster, Herbert (1898). *Through New Guinea and the cannibal countries*. London: T. Fisher Unwin.
- Chalker, Sylvia and Edmund Weiner (1994). *The Oxford dictionary of English grammar*. Oxford: Clarendon Press.
- Charley (24.12.1888). *[Witness statement]*. Australian Station (New Guinea & Solomon Islands, 1888). "Correspondence respecting outrages by natives on British subjects, and other matters, which have been under inquiry during the year 1888, being continuation of reports of cases dealt with in former years, together with other cases which have since arisen." In *Copies of reports of proceedings in Solomon Islands & New Hebrides*. From: Fairfax. Rear Admiral H., at H.M.S. "Orlando" at Adelaide. Western Pacific Archives. (MSS & Archives.2003/1.WPHC 4/IV.1889, 93/1889). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Charlie (31.01.1913). *[Letter to Mr King]*. Outrages by natives, 1913-1914. Western Pacific Archives. (MSS & Archives-2003/1-NHBS 1/I Vol 1, 12/1913). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Charpentier, Jean-Michel (1979). *Le pidgin Bislama(n) et le multilinguisme aux Nouvelles-Hébrides*. Paris: Selaf.
- Chaudenson, Robert (1992). *Des îles, des hommes, des langues*. Paris: L'Harmattan.
- Chaudenson, Robert (2001). *Creolization of language and culture*. London: Routledge.
- Chaurain (1846). Extract from a letter of Father Chaurain, missionary of the Society of Mary, to the Reverend Father Colin, Superior-General. Port Saint Mary, Mach 2nd, 1846. Anonymous, *Annals of the propagation of the faith: A periodical collection of letters from the bishops and missionaries employed in the missions of the old and the new world and of all the documents relating to those missions, and the institution for the propagation of the faith*. London: 398–407.
- Cheeseman, Evelyn (1933). *Backwaters of the savage South Seas*. London: Jarrolds.
- Cheeseman, Evelyn (1949). *Camping adventures on cannibal islands*. London: Harrap.

- Cheyne, Andrew (1852). *A description of the islands in the western Pacific Ocean*. London: J. D. Potter.
- Churchill, William (1911). *Beach-la-mar. The jargon or trade speech of the Western Pacific*. Washington: Carnegie Institutions.
- Clark, Ross (1977). *In search of Beach-la-Mar: Historical relations among Pacific pidgins and creoles*. Auckland: Department of Anthropology, University of Auckland, Working Papers in Anthropology, Archaeology, Linguistics, Maori Studies 48.
- Clark, Ross (1979). "In search of Beach-la-mar: Towards a history of Pacific Pidgin English." *Te Reo* 22: 3–64.
- Clarker (02.03.1914). [Letter to Mr. King, Grievences, which the Atchin natives asked Mr. Clarker to write down, town by town, and hand to Mr. King, H. B. M. Resident Commissioner]. Native Outrages, 1914 - 1924 (3 parts in 8 folders). Western Pacific Archives. (MSS & Archives-2003/1-NHBS 1/I Vol 1, 17/1914. Folder 8). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Collinson, Clifford W. (06.10.1923). "The Solomon Islands: No. 1." *The Advertiser*: 21.
- Collinson, Clifford W. (1926). *Life and laughter midst the cannibals*. London: Hurst & Blackett.
- Collinson, Clifford W. (1929/1932). *Cannibals and coconuts*. London: George Philip & Son Ltd.
- Colonial Office (1882-1883). *Western Pacific original correspondence. Offices (except Miscellaneous)* CO225/13. Kew, London: The National Archives.
- Colonial Office (1883-1884). *Western Pacific original correspondence. Despatches* CO225/15. Kew, London: The National Archives.
- Comrie, Bernard (1989). *Language universals and linguistic typology: Syntax and morphology*. Chicago: University of Chicago Press.
- Comrie, Bernard (1998). "Rethinking the typology of relative clauses." *Language Design* 1: 59–86.
- Comrie, Bernard and Tania Kuteva (2005). "The typology of relative clause formation in African languages." Ed. Erhard Voeltz, *Studies in African linguistic typology*. Amsterdam: John Benjamins. 209–228.
- Comrie, Bernard and Tania Kuteva (2013a). "Relativization strategies." Eds. Matthew Dryer, Martin Haspelmath, David Gil and Bernard Comrie, *The world atlas of language structures*. Leipzig: Max Planck Institute for Evolutionary Anthropology. 494–497.
- Comrie, Bernard and Tania Kuteva (2013b). "Relativization on Subjects." Eds. Matthew Dryer and Martin Haspelmath, *The world atlas of language structures*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <<http://wals.info/chapter/122>>. Accessed 16 January 2020.
- Comrie, Bernard and Tania Kuteva (2013c). "Relativization on Obliques." Eds. Matthew Dryer and Martin Haspelmath, *The world atlas of language structures*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <<http://wals.info/chapter/123>>. Accessed 16 January 2020.
- Cook, James (1842). "Second voyage (July 1774)." *The voyages of Captain James Cook*. London: William Smith.
- Coote, Walter (1882). *Wandering, south and east*. London: Sampson Low, Marston, Searle & Rivington.
- Coote, Walter (1883). *The Western Pacific, being a description of the group of islands to the north and east of the Australian continent*. London: Sampson Low, Marston, Searle & Rivington.
- Cormack, James E. (1944). *Isles of Solomon*. Washington, D.C.: Review and Herald.
- Corris, Peter (1968). "'Blackbirding' in New Guinea waters. 1883-84: An episode in the Queensland labour trade." *The Journal of Pacific History* 3: 85–105.
- Corum, Micah (2015). *Substrate and adstrate: The origins of spatial semantics in West African Pidgincreoles*. Berlin & Boston: De Gruyter Mouton.
- Cowie, Andrew (01.08.1932). *Letter to Florence Young, Inakona*. South Sea Evangelical Mission, formerly Queensland Kanaka Mission (1890-1942). Correspondence. Pacific Manuscript Bureau. PMB 1150. Reel 5.
- Croft, William (2000). *Explaining language change. An evolutionary approach*. Harlow: Longman.
- Cromar, John (1935). *Jock of the islands*. London: Faber and Faber.
- Crowley, Terry (1990a). *Beach-la-Mar to Bislama: The emergence of a national language in Vanuatu*. Oxford: Clarendon Press.
- Crowley, Terry (1990b). "Serial verbs and prepositions in Bislama." Ed. John Verhaar, *Melanesian Pidgin and Tok Pisin*. Amsterdam: John Benjamins. 57–89.
- Crowley, Terry (1991). "Varieties of Melanesian Pidgin: Separate identities vs. European stereotypes." *Language and Linguistics in Melanesia* (22): 51–65.
- Crowley, Terry (2000). "'Predicate marking' in Bislama." Ed. Jeff Siegel, *Processes of language contact: Studies from Australia and the South Pacific*. Saint-Laurent, Canada: Fides. 47–74.
- Crowley, Terry (2004). *Bislama reference grammar*. Honolulu: University of Hawai'i Press.
- Crowley, Terry (2008). "Bislama: morphology and syntax." Eds. Kate Burridge and Bernd Kortmann, *Varieties of English: The Pacific and Australasia*. Berlin: De Gruyter Mouton. 444–466.
- Cumpston, John S. (1963). *Arrivals and departures Sydney, 1788-1825*. Canberra: Roebuck Society Publication.
- Dahmansop (14.08.1914). [Witness statement. Alleged murder of a native named Lesar, Port Vato, Ambrym, By a native named Dahmansop, Port Vato, Ambrym]. Joint Naval Commission, 1914-1915. Western Pacific Archives. (MSS & Archives-2003/1-NHBS 1/I Vol 1, 12/1914). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Daiber, Albert (1902). *Eine Australien- und Südseefahrt*. Leipzig: G. Teubner.

- Dar, Albert (13.08.1916). [*Letter to Mr. King. Grievences, which the Atchin natives asked Mr Clarker to write down, town by town, and hand to Mr. King, H. B. M. Resident Commissioner*]. Native Outrages, 1914 - 1924 (3 parts in 8 folders). Western Pacific Archives. (MSS & Archives-2003/1-NHBS 1/I Vol 1, 17/1914. Folder 7). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Dar, Albert (11.01.1919). [*Letter to Mr. King. Grievences, which the Atchin natives asked Mr. Clarker to write down, town by town, and hand to Mr. King, H. B. M. Resident Commissioner*]. Native Outrages, 1914 - 1924 (3 parts in 8 folders). Western Pacific Archives. (MSS & Archives-2003/1-NHBS 1/I Vol 1, 17/1914. Folder 6). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Daval-Markussen, Aymeric (2018). *Reconstructing creole*. PhD dissertation. Aarhus: Aarhus University.
- Davies, Gwen (1944). "Fosui at One Pusu." *Solomon Soldiers News* 7: 4. Pacific Manuscripts Bureau. AU PMB DOC 442.
- de Barros, João (Ed.) (1777). *Da Asia de Diogo de Couto: Dos feitos, que os Portuguezes fizeram na conquista, e decubrimento das terras, e mares do oriente. Parte primeira*. Lisabon: Regia Officina Typografica.
- DeGraff, Michel (2003). "Discussion Note: Against Creole Exceptionalism." *Language* 79 (2): 391–409.
- DeGraff, Michel (2014). "The ecology of language evolution in Latin America: A Haitian postscript toward a postcolonial sequel." Ed. Salikoko Mufwene, *Iberian imperialism and language evolution in Latin America*. Chicago & London: The University of Chicago Press. 274–328.
- de Haan, Ferdinand (2006). "Typological approaches to modality." Ed. William Frawley, *The expression of modality*. Berlin: De Gruyter Mouton. 27–70.
- de Haan, Ferdinand (2013). "Semantic distinctions of evidentiality." Eds. Matthew Dryer, Martin Haspelmath, David Gil and Bernard Comrie, *The world atlas of language structures*. Leipzig: Max Planck Institute for Evolutionary Anthropology. 314–317.
- Deck, Joan B. (1944). "Island letters: Letter by Crystal Ufaria to Gwaeta 1941." *Solomon Soldiers News* 7: 2–3. Pacific Manuscripts Bureau. AU PMB DOC 442.
- Deck, Joan B. (1946). "The Story of Rorobo (Part 3)." *Solomon Soldiers' News* 18: 2–3. Pacific Manuscripts Bureau. AU PMB DOC 442.
- Deck, Joan B. (December 1946). "Too many mothers-in-law!" *Not in Vain* 104: 5. Sydney and Brisbane. Pacific Manuscripts Bureau. AU PMB DOC 439. Reel 2.
- Deck, Joan B. (1947). "Visiting mountain villages (continued)." *Solomon Soldiers' News* 27: 2–4. Pacific Manuscripts Bureau. AU PMB DOC 442.
- Deck, Joan B. (1948). "Visitors." *Solomon Soldiers' News* 34: 2–3. Pacific Manuscripts Bureau. AU PMB DOC 442.
- Deck, K. (October-November 1937). *Circular No. 38 October-November 1937*. Pacific Manuscripts Bureau. AU PMB DOC 440. Reel 1.
- Deck, Kathleen (July 1933). "News from One Pusu." *Not in Vain* 51: 3–5. ("Wilt thou be made whole"). Ed. Northcote Deck. Sydney and Brisbane. Pacific Manuscripts Bureau. AU PMB DOC 439. Reel 2.
- Deck, Miss (October 1937). "News from One Pusu." *Not in Vain* 68: 3–4. ("Three Appearings"). Ed. Northcote Deck. Sydney and Brisbane. Pacific Manuscripts Bureau. AU PMB DOC 439. Reel 2.
- Deck, Norman C. (27.09.1932). *Letter to friends, One Pusu*. South Sea Evangelical Mission, formerly Queensland Kanaka Mission (1890-1942). Correspondence. Pacific Manuscripts Bureau. PMB 1150. Reel 3.
- Deck, Norman C. (13.05.1928). *Letter to Emmie, Ngongosila*. South Sea Evangelical Mission, formerly Queensland Kanaka Mission (1890-1942). Correspondence. Pacific Manuscripts Bureau. PMB 1150. Reel 2.
- Deck, Norman C. (June 1934). *Letter to prayers, One Pusu*. South Sea Evangelical Mission, formerly Queensland Kanaka Mission (1890-1942). Correspondence. Pacific Manuscripts Bureau. PMB 1150. Reel 2.
- Deck, Norman C. (October 1925). "Account of a visit to Okwala." *Island Letters: circulars published by the South Sea Evangelical Mission*: 6–8. Pacific Manuscripts Bureau. PMB 1253. Reel 1.
- Deck, Norman C. (January 1942). *South Sea Evangelical Mission Despatches No. 65*. Pacific Manuscripts Bureau. AU PMB DOC 440.
- Deck, Norman C. (December 1943). "The story of To'ouna." *Not in Vain* 92: 4. Sydney and Brisbane. Pacific Manuscripts Bureau. AU PMB DOC 439. Reel 2.
- Deck, Northcote J. (March 1910). "Across Malaita." *Island letters: circulars published by the South Sea Evangelical Mission*: 8. Pacific Manuscripts Bureau. AU PMB DOC 440.
- Deck, Northcote J. (February 1912). "Manuscript: Sunday evening at Baunani." Pacific Manuscripts Bureau. AU PMB MS 1253. Reel 1.
- Deck, Northcote J. (July 1919). "Holding fast the faithful word." *Island letters: circulars published by the South Sea Evangelical Mission*: 2–6. Pacific Manuscripts Bureau. AU PMB MS 1253. Reel 1.
- Deck, Northcote J. (July 1928). *Island letters: circulars published by the South Sea Evangelical Mission* 31. Pacific Manuscripts Bureau. AU PMB MS 1253. Reel 1.
- Department of Geological Survey (1931). *British Solomon Islands Protectorate: Annual Report No. 1531*. London: Her Majesty's Stationery Office (H.M.S.O.).
- Detzner, Hermann (1920). *Vier Jahre unter Kannibalen: Von 1914 bis zum Waffenstillstand unter deutscher Flagge im unerforschten Innern von Neuguinea*. Berlin: August Scherl.

- Deumert, Ana (2004). *Language standardization and language change: The dynamics of Cape Dutch*. Amsterdam & Philadelphia: John Benjamins.
- Díaz-Campos, Manuel and Stephanie Dickinson (2019). "Using statistics as a tool in the analysis of sociolinguistic variation: A comparison of current and traditional methods." Eds. Gabriel Rei-Doval and Fernando Tejedo-Herrero, *Lusophone, Galician, and Hispanic linguistics: Bridging frames and traditions*. Oxon & New York: Routledge. 205–226.
- Dickinson, Joseph H. C. (1927). *A trader in the savage Solomons*. London: H. F. & C. Witherby.
- Diessel, Holger (1999a). *Demonstratives: Form, function and grammaticalization*. Amsterdam & Philadelphia: John Benjamins.
- Diessel, Holger (1999b). "The morphosyntax of demonstratives in synchrony and diachrony." *Linguistic Typology* 3: 1–49.
- Diessel, Holger (2006). "Demonstratives, joint attention, and the emergence of grammar." *Cognitive Linguistics* 17 (4): 463–489.
- Diessel, Holger (2013a). "Pronominal and adnominal demonstratives." Eds. Matthew Dryer, Martin Haspelmath, David Gil and Bernard Comrie, *The world atlas of language structures*. Leipzig: Max Planck Institute for Evolutionary Anthropology. 174–177.
- Diessel, Holger (2013b). "Distance contrasts in demonstratives." Eds. Matthew Dryer, Martin Haspelmath, David Gil and Bernard Comrie, *The world atlas of language structures*. Leipzig: Max Planck Institute for Evolutionary Anthropology. 170–173.
- Drechsel, Emanuel J. (2014). *Language contact in the early colonial Pacific: Maritime Polynesian Pidgin before Pidgin English*. Cambridge: Cambridge University Press.
- Dryer, Matthew (2013). "Order of adposition and noun phrase." Eds. Matthew Dryer and Martin Haspelmath. *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <<http://wals.info/chapter/85>>. Accessed 03 February 2020.
- Dryer, Matthew, Martin Haspelmath, David Gil and Bernard Comrie (Eds.) (2013). *The world atlas of language structures*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Dutton, Tom (1980). *Queensland Canefields English of the late nineteenth century: (A record of interview with two of the last surviving kanakas in North Queensland, 1964) (Pacific Linguistics Series D-29)*. Canberra: Research School of Pacific Studies, The Australian National University.
- Dutton, Tom (1996a). "Other pidgins in Papua New Guinea." Eds. Darrell Tryon, Peter Mühlhäusler and Stephen Wurm, *Atlas of languages of intercultural communication in the Pacific, Asia and the Americas*. Berlin & Boston: De Gruyter Mouton. 215–217.
- Dutton, Tom (1996b). "Hiri Motu." Eds. Darrell Tryon, Peter Mühlhäusler and Stephen Wurm, *Atlas of languages of intercultural communication in the Pacific, Asia and the Americas*. Berlin & Boston: De Gruyter Mouton. 225–231.
- Dutton, Tom (2006). "Language contact and change in Melanesia." Eds. Peter Bellewood, James Fox and Darrell Tryon, *The Austronesians: Historical and comparative perspectives*. Canberra: Australian National University Press. 207–228.
- Dutton, Tom and Peter Mühlhäusler (1984). "Queensland Kanaka English." *English World Wide* 4: 231–263.
- Eden, Charles H. (1872). *My wife and I in Queensland: An eight years' experience in the above colony with some account of Polynesian labour*. London: Longmans, Green, and Co.
- Edwards, John (2009). *Language and identity: An introduction*. Cambridge: Cambridge University Press.
- Elspass, Stephan (2012). "The use of private letters and diaries in sociolinguistic investigation." Eds. Juan Hernández Campoy and Juan Conde-Silvestre, *The handbook of historical sociolinguistics*. Malden, Massachusetts: Wiley-Blackwell. 156–169.
- Enow (24.12.1888). [Witness statement]. Australian Station (New Guinea & Solomon Islands, 1888). "Correspondence respecting outrages by natives on British subjects, and other matters, which have been under inquiry during the year 1888, being continuation of reports of cases dealt with in former years, together with other cases which have since arisen." In *Copies of reports of proceedings in Solomon Islands & New Hebrides*. From: Fairfax. Rear Admiral H., at H.M.S. "Orlando" at Adelaide. Western Pacific Archives. (MSS & Archives.2003/1.WPHC 4/IV.1889, 93/1889). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Erskine, John E. (1853). *Journal of a cruise among the islands of the Western Pacific*. London: Murray.
- Faraclas, Nicholas (2013). "Nigerian Pidgin structure dataset." Eds. Susanne Michaelis, Philippe Maurer, Martin Haspelmath and Magnus Huber, *Atlas of pidgin and creole language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <<http://apics-online.info/contributions/17>>. Accessed 10 March 2021.
- Finney, Malcolm (2013). "Krio structure dataset." Eds. Susanne Michaelis, Philippe Maurer, Martin Haspelmath and Magnus Huber, *Atlas of pidgin and creole language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <<http://apics-online.info/contributions/15>>. Accessed 10 March 2021.
- Finsch, Otto (1888). *Samoafahrten: Reisen in Kaiser-Wilhelms-Land und Englisch Neu-Guinea in den Jahren 1884 und 1885 an Bord des Deutschen Dampfers "Samoa"*. Leipzig: Hirt.
- Forbes, Litton (1875). *Two years in Fiji*. London: Longmans Green & Co.

- Forsyth, Mr. (1942). "A day on a coconut plantation in New Guinea." Ed. Neville Threlfall, *Papers relating to the Methodist Church in Papua New Guinea and Rev Threlfall's research on Papua New Guinea 1875 - 1985*. Pacific Research Archives. AU ANUA 305-66.
- Friederici, Georg (1911). "Pidgin-Englisch in Deutsch-Neuguinea." *Koloniale Rundschau*: 92–106.
- Frommund, Bernhard (1926). *Deutsch-Neuguinea, eine Perle der Südsee: Erlebnisse und Eindrücke eines Deutschen auf Deutsch-Neuguinea 1905 - 1908*. Hamburg: Volkmar.
- Fussell, James (1903). *A kanaka slave: A story of the early Queensland labour traffic*. London: Arthur H. Stockwell.
- Garrie (09.01.1929). [Witness statement]. Rex v. Talatova: Minutes of evidence adduced at trial for murder. Western Pacific Archives. (MSS & Archives-2003/1-WPHC 4/IV-1929, 611/1929). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Giles, William E. (1968). *A cruize in a Queensland labour vessel to the south seas, edited by Deryck Scair*. Canberra & London: Australian National University Press & Hurst.
- Glasser, Cyril (1999). "In foreign parts - reflections on British lawyers abroad." Eds. Gavin Drewry and Charles Blake, *Law and the spirit of inquiry: Essays in honour of Sir Louis Blom-Cooper*. Hague, London & Boston: Kluwer Law International. 179–192.
- Gooskens, Charlotte (2018). "Dialect intelligibility." Eds. Charles Boberg, John Nerbonne and Dominic Watt, *Handbook of dialectology*. New Jersey: John Wiley & Sons. 204–218.
- Gordon, James D. (1862). *The sandalwood trade and traders of Polynesia*. Halifax.
- Gordon, Thomas E. L. (1943). *Notes written by Gordon Thomas as a prisoner of war in Rabaul, 1942 on the subject of Pidgin-English*. Canberra, Australian National University: Australian National University Archives. AU ANUA 584-1.
- Goulden, Rick J. (1990). *The Melanesian content in Tok Pisin*. Canberra: Pacific Linguistics Series B.
- Government Gazette (01.11.1914). "Proclamation." *Government Gazette* 1.2.
- Great Britain (1869). *Parliamentary Papers* 43. London: Her Majesty's Stationery Office (H.M.S.O.).
- Gries, Stefan and Martin Hilpert (2008). "The identification of stages in diachronic data: variability-based neighbor clustering." *Corpora* 3 (1): 59–81.
- Gries, Stefan and Martin Hilpert (2012). "Variability-based neighbor clustering: a bottom-up approach to periodization in historical linguistics." Eds. Terttu Nevalainen and Elizabeth Traugott, *The Oxford handbook on the history of English*. Oxford: Oxford University Press. 134–144.
- Griffin, James (2005). "Origins of Bougainville's boundaries." Ed. Douglas Oliver, *Bougainville before the conflict*. Canberra: Pandanus Books. 72–76.
- Grimshaw, Beatrice (1907). *From Fiji to the cannibal islands*. London: Eveleigh Nash.
- Groves, William C. (1925). *Manuscript of Peter the Island King*. Pacific Manuscripts Bureau. AU PMB MS 612.
- Guppy, Henry B. (1887). *The Solomon Islands and their natives*. London: Swan Sonnenschein, Lowrey & Co.
- Hackney, Amherst and Basil Thomson (Eds.) (1901). *The discovery of the Solomon Islands by Alvaro de Mendaña in 1568: Translated from the original Spanish manuscripts* 1. London: Hakluyt Society.
- Hagège, Claude (2010). *Adpositions*. Oxford: Oxford University Press.
- Hagen, Bernhard (1899). *Unter den Papua's. Beobachtungen und Studien über Land und Leute, Thier- und Pflanzenwelt in Kaiser-Wilhelmsland*. Wiesbaden: Kreidel.
- Halimae (12.05.1924). [Witness statement]. Rex v. Borani & Meako - (Murder): Forwards copies of notes of evidence adduced at the trial, 1924. Western Pacific Archives. (MSS & Archives-2003/1-WPHC 4/IV-1924, 1495/1924). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Hall, Robert A. (1943). *Melanesian Pidgin English: Grammar, texts, vocabulary*. Baltimore: Linguistic Society of America.
- Hall, Robert A. (1966). *Pidgin and creole languages*. Ithaca: Cornell University Press.
- Hammarström, Harald (2008). "Counting languages in dialect continua using the criterion of mutual intelligibility." *Journal of Quantitative Linguistics* 15 (1): 34–45.
- Harrison, Tom (1937). *Savage civilisation*. London: Victor Gollancz.
- Hawkesby (20.01.1915). [Letter to Mr. Gire]. Illegal recruiting and general native complaints, 1915–24 (3 Parts in 9 Folders). Western Pacific Archives. (MSS & Archives-2003/1-NHBS 1/I Vol 2, 9/1915. Folder 8). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Heath, Thomas (1941). "Missionary voyage of the Camden." *The Evangelical Magazine and Missionary Chronicle for the year 1941* XIX: 91–99.
- Heine, Bernd and Tania Kuteva (2004). *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.
- Helton, E. C. N. (1943). *Booklet on Pidgin English as used in the mandated territory of New Guinea with dictionary of nouns and phrases this language is used in conversation with natives, Asiatics and German white mission*. Brisbane: Adams.
- Hernández Campoy, Juan and Natalie Schilling (2012). "The application of the quantitative paradigm to historical sociolinguistics: Problems with the generalizability principle." Eds. Juan Hernández Campoy and Juan Conde-Silvestre, *The handbook of historical sociolinguistics*. Malden, Massachusetts: Wiley-Blackwell. 63–79.

- Hilliard, David (1970). "Bishop G. A. Selwyn and the Melanesian Mission." *The New Zealand Journal of History* 4 (2): 120–137.
- Hinz, Earl (1995). *Pacific island battlegrounds of World War II: Then and now*. Hawai'i: Bess Press.
- Hiromitsu, Iwamoto and Australian National University (2000). "Remembering the war in New Guinea." <<http://ajrp.awm.gov.au/AJRP/remember.nsf/Web-Frames/HomeFrame?OpenDocument>>. Accessed 25 April 2017.
- Hocart, Arthur (1922). "The cult of the dead in Eddystone of the Solomons." *Journal of the Royal Anthropological Institute* (52): 71–112.
- Hogbin, Herbert Ian (1944). "Native councils and native courts in the Solomon Islands." *Oceania* XIV (4): 257–283.
- Hogbin, Herbert Ian (1939). *Experiments in civilization; the effects of European culture on a native community of the Solomon Island*. London: G. Routledge & Sons.
- Holm, John (1988). *Pidgins and creoles: Volume II: Reference survey*. Cambridge: Cambridge University Press.
- Hopper, Paul. J. and Elizabeth Traugott (2003). *Grammaticalization*. Cambridge: Cambridge University Press.
- Horton, Dick C. (1965). *The happy Isles: A diary of the Solomons*. London: Heinemann.
- Hothorn, Torsten, Kurt Hornik and Achim Zeileis (2006). "Unbiased recursive partitioning: A conditional inference framework." *Journal of Computational and Graphical Statistics* 15 (3): 651–674.
- Hothorn, Torsten, Kurt Hornik, Achim Zeileis (2010). *party: A laboratory for recursive partitioning*. <<https://cran.r-project.org/web/packages/party/vignettes/party.pdf>>. Accessed 18 January 2020.
- Hothorn, Torsten, Kurt Hornik and Achim Zeileis (2015). *ctree: Conditional inference trees: The comprehensive R archive network*. <<https://cran.r-project.org/web/packages/partykit/vignettes/ctree.pdf>>. Accessed 18 January 2020.
- Huber, Magnus (1999). *Ghanaian Pidgin English in its West African context. A sociohistorical and structural analysis*. Amsterdam & Philadelphia: John Benjamins.
- Huber, Magnus (2005). "The role of adstrates in the development of Atlantic English Creoles." *Paper presented at the conference Creole Language Structure between Substrates and Superstrates*, Leipzig, 3–5 June.
- Huber, Magnus (2013). "Ghanaian Pidgin English structure dataset." Eds. Susanne Michaelis, Philippe Maurer, Martin Haspelmath and Magnus Huber, *Atlas of pidgin and creole language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <<http://apics-online.info/contributions/16>>. Accessed 20 March 2021.
- Huber, Magnus and Viveka Velupillai (2016). "Sprachkontakt in kolonialen Kontexten II. Das Pidgin-Englische in den ehemaligen deutschen Kolonien." Eds. Thomas Stolz, Ingo Warnke and Daniel Schmidt-Brücken, *Sprache und Kolonialismus: Eine interdisziplinäre Einführung zu Sprache und Kommunikation in kolonialen Kontexten*. Berlin: De Gruyter Mouton. 121–146.
- Huber, Magnus and Viveka Velupillai (2018). "Die Database of Early Pidgin and Creole Texts: Sprachplanung, Sprachideologien und Sprachattitüden gegenüber dem Pidginenglisch in Deutsch-Neuguinea." Eds. Birte Kellermeier-Rehbein, Matthias Schulz and Doris Stolberg, *Sprache und (Post)Kolonialismus: Linguistische und interdisziplinäre Aspekte*. Berlin & Boston: De Gruyter Mouton. 127–160.
- Huddleston, Rodney and Geoffrey Pullum (2016). *The Cambridge grammar of the English Language*. Cambridge: Cambridge University Press.
- Huebner, Thom and Stephen Horoi (1979). *Solomon Islands Pijin: Grammar handbook*. Washington, D.C.: School for International Training.
- Idriess, Ion L. (1941). *Gold-dust and ashes*. Sydney and London: Angus & Robertson.
- Imhaus, E. N. (1890). *Les Nouvelles-Hébrides*. Paris & Nancy: Berget-Levrault et Cie.
- Inglis, John (1887). *In the New Hebrides. Reminiscences of missionary life and work, especially on the island of Aneityum, from 1850 till 1877*. London: T. Nelson & Sons.
- Ivens, Walter G. (1918). *Dictionary and grammar of the language of Sa'a and Ulawa, Solomon Islands: Transcribed by the Right Reverend Dr. Terry Brown*. Washington: Carnegie Institution of Washington.
- Jacobs, Thomas J. (1844). *Scenes, incidents, and adventures in the Pacific Ocean: Or the islands of the Australian seas, during the cruise of the clipper Margaret Oakley under Captain Benjamin Morrell*. New York: Harper & Brothers.
- Jacomb, Edward (1914). *France and England in the New Hebrides*. Melbourne, Sydney, Adelaide & Brisbane: George Robertson & Co.
- Jacques, Norbert (1922). *Südsee: Ein Reisetagebuch*. München: Drei Masken.
- Jamond, J. B. (1906–1912). *The mission at Olal, Ambrym, and Sorcery (Hableou) at North Ambrym by Father J.B. Jamond, S.M., with comments by Father Casimir Bancarel, S.M. 'Quelques Souvenirs sur la Mission de Craig-cove et de P.S.'*. Pacific Manuscript Bureau. AU PMB MS 53-11.
- Johanson, Lars (1992). *Strukturelle Faktoren in Türkischen Sprachkontakten*. Stuttgart: Franz Steiner.
- Johanson, Lars (1993). "Code-copying in immigrant Turkish." Eds. Guus Extra and Ludo Verhoeven, *Immigrant languages in Europe*. Clevedon, Philadelphia & Adelaide: Multilingual Matters. 197–221.
- Johns, Dwight F. (1942). *A pocket vocabulary of Malay, Pidgin English and Japanese phrases for use of U. S. troops in the Southwest Pacific Area*. Australia: United States Army Services of Supply.

- Jourdan, Christine (1985). *Sapos yumi mitim yumi: Urbanization and creolization in the Solomon Islands*. Canberra: Australian National University.
- Jourdan, Christine (1989). "Nativization and anglicization in Solomon Islands Pijin." *World Englishes* 8 (1): 25–35.
- Jourdan, Christine (1990). "Solomons Pijin: An unrecognised national language." Eds. Richard Baldauf and Allan Luke, *Language planning and education in Australasia and the South Pacific*. Clevedon: Multilingual Matters. 166–181.
- Jourdan, Christine (1995). "Stepping-stones to national consciousness." Ed. Robert J. Foster, *Nation making: Emergent identities in postcolonial Melanesia*. Ann Arbor: University of Michigan Press. 127–149.
- Jourdan, Christine (1996). "Where have all the cultures gone?: Sociocultural creolisation in the Solomon Islands." Eds. Jonathan Friedman and James Carrier, *Melanesian modernities*. Sweden: Lund University Press. 34–52.
- Jourdan, Christine (2002). *Pijin: a trilingual cultural dictionary*. Canberra: Pacific Linguistics.
- Jourdan, Christine (2008). "Solomon Islands Pijin: morphology and syntax." Eds. Kate Burridge and Bernd Kortmann, *Varieties of English: The Pacific and Australasia*. Berlin: De Gruyter Mouton. 467–487.
- Jourdan, Christine (2015). *Personal email communication*.
- Jourdan, Christine (2018). "The shades of legitimacy of Solomon Islands Pijin." Eds. Jacqueline Knörr and Trajano Filho, *Creolization and pidginization in contexts of postcolonial diversity: Language, culture, identity*. Leiden & Boston: Brill. 78–95.
- Jourdan, Christine and Roger Keesing (1997). "From Fisin to Pijin: Creolization in process in the Solomon Islands." *Language in Society* 26 (3): 401–420.
- Jourdan, Christine and Rachel Selbach (2008). "Phonology and phonetics of Solomon Islands Pijin." Eds. Kate Burridge and Bernd Kortmann, *Varieties of English: The Pacific and Australasia*. Berlin: De Gruyter Mouton. 164–187.
- Jung, Emil (1885). "Die Arbeiterverhältnisse in der Südsee mit Bezug auf die Entwicklung unserer dortigen Erwerbungen." *Globus* 48: 282–285.
- Kay, John (1872). *The slave trade in the New Hebrides: being papers read at the annual meeting of the New Hebrides Mission, held at Aniwa, July 1871, and published by the authority of the meeting*. Edinburgh: Edmonston & Douglas.
- Keen, Kevin J. (2010). *Graphics for statistics and data analysis with R*. Boca Raton: CRC Press.
- Keenan, Edward and Bernard Comrie (1977). "Noun phrase accessibility and universal grammar." *Linguistic Inquiry* 8: 63–99.
- Keesing, Roger (1988). *Melanesian Pidgin and the Oceanic substrate*. California: Stanford University Press.
- Keesing, Roger (1991). "Substrates, calquing and grammaticalization in Melanesian Pidgin." Eds. Elizabeth Traugott and Bernd Heine, *Approaches to grammaticalization, Vol. 1*. Amsterdam & Philadelphia: John Benjamins. 315–341.
- Kerr, Clark G. (1985). *Japanese and allied propaganda relating to the Japanese invasion of New Guinea in World War II, 1943-1945*. Canberra: National Library of Australia. NLA MS 9002.
- Knibbs, Stanley G. C. (1929). *The savage Solomons as they were and are*. London: Seeley, Service & Co.
- Knoolhuizen, Remco (2006). "Language shift, ethnolinguistic vitality and historical sociolinguistics: Testing the models." LEL Postgraduate Conference.
- Koch, Peter and Wulf Oesterreicher (1985). "Sprache der Nähe - Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte." *Romanistisches Jahrbuch* (36): 15–43.
- Koch, Peter and Wulf Oesterreicher (1994). "Schriftlichkeit und Sprache." Eds. Hartmut Günther and Otto Ludwig, *Writing and its use. An interdisciplinary handbook of international research*. Berlin: De Gruyter Mouton. 587–604.
- Krämer, Augustin (1913). "Redet Deutsch in unseren Kolonien." *Deutsche Kolonialzeitung* 24: 406–407.
- Krämer-Bannow, Elisabeth (1916). *Bei kunstsinnigen Kannibalen der Südsee: Wanderungen auf Neu-Mecklenburg 1908-1909*. Berlin: Reimer.
- Krieger, Herbert W. (1943). *Island peoples of the western Pacific, Micronesia and Melanesia*. Washington: The Smithsonian Institution.
- Kunze, Georg (1897). *Ein schönes Tagewerk in einem Lande der Thränen und Trübsale: Lust, Leid und Arbeit der Anfangszeit auf einsamer Südsee-Insel*. Barmen: Verlag des Missionshauses.
- Kurzon, Dennis (2002). "'Preposition' as functor: The case of long in Bislama." Eds. Susanne Feigenbaum and Dennis Kurzon, *Prepositions in their syntactic, semantic and pragmatic context*. Amsterdam & Philadelphia: John Benjamins. 231–248.
- Kuteva, Tania, Bernd Heine, Bo Hong, Haiping Long, Heiko Narrog and Seongha Rhee (2019). *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.
- Kutscher, P. (~ 1940). *Wörterbuch Deutsch-Pidgin-Englisch*. Vunapope (PNG).
- Labov, William (1982). "Building on empirical foundations." Eds. Winfried P. Lehmann and Yakov Malkiel, *Perspectives in historical linguistics*. Amsterdam: John Benjamins. 17–92.
- Labov, William (1999). *Principles of linguistic change: Internal factors*. Oxford: Blackwell.

- Lamb, Robert (1905). *Saints and savages: The story of five years in the New Hebrides*. Edinburgh & London: William Blackwood & Sons.
- Lambert, Sylvester Maxwell (1942). *A Yankee doctor in paradise*. London: J. Dent & Sons.
- Landtman, Gunnar (1927). *The Kiwai Papuans of British New Guinea: A nature-born instance of Rousseau's ideal community*. London: Macmillan & Co.
- Lawrence, David R. (2014). *The naturalist and his 'beautiful Islands': Charles Morris Woodford in the Western Pacific*. Canberra: Australian National University Press.
- Lawrence, Peter (1964). *Road belong cargo: A study of the cargo movement in the Southern Madang District, New Guinea*. Manchester: Manchester University Press.
- Laycock, Donald (1970). *Materials in New Guinea Pidgin (Coastal and Lowlands): Pacific Linguistics Series D. No. 5*. Canberra: Australian National University Press.
- Le Chartier, Henri (1885). *La Nouvelle-Calédonie et les Nouvelles-Hébrides*. Paris: Jouvet et Cie.
- Lee, Ernest (1996). "On prepositions in Solomon Islands Pijin." Eds. John Lynch and Fa'afu Pat, *Oceanic studies: Proceedings of the first International Conference on Oceanic Linguistics*. Canberra: Australian National University Press. 383–401.
- Lefebvre, Claire (1998). *Creole genesis and the acquisition of grammar: The case of Haitian Creole*. Cambridge: Cambridge University Press.
- Lefebvre, Claire (2004). *Issues in the study of pidgin and creole languages*. Amsterdam: John Benjamins.
- Lehmann, Christian (1984). *Der Relativsatz*. Tübingen: Gunter Narr.
- Lehmann, Christian (1986). "On the typology of relative clauses." *Linguistics* 24 (4): 663–680.
- Lehmann, Christian and Yong-Min Shin (2005). "The functional domain of concomitance. A typological study of instrumental and comitative relations." Ed. Christian Lehmann, *Typological studies in participation*. Erfurt: Akademie Verlag. 9–102.
- Lengwe (06.12.1922). [Witness statement attached to a letter by Peapoe to Sir, 09.12.1922]. Executions of judgements of joint court under protocol of 1914, 1923–1924. Western Pacific Archives. (MSS & Archives-2003/1-NHBS 1/1 Vol 2, 14/1923). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Levey, Stephen (2001). "Relative clauses and register expansion in Tok Pisin." *World Englishes* 20 (3): 251–267.
- Levinson, Stephen (1983). *Pragmatics*. Cambridge: Cambridge University Press.
- Levinson, Stephen (2006). "Deixis." Eds. Laurence Horn and Gregory Ward, *The handbook of pragmatics*. Malden: Blackwell Publishing Ltd. 97–121.
- Levinson, Stephen (2018). "Demonstratives: Patterns in diversity." Ed. Stephen Levinson, *Demonstratives in cross-linguistic perspective*. Cambridge: Cambridge University Press. 1–42.
- Levshina, Natalia (2015). *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins.
- Likin, Bridgit (06.11.1908). [Witness statement]. A. E. Hermes: shooting of native by at Rendova. Mr Mahaffy: appointment as Deputy Commissioner and Assistant High Commissioner Western Pacific Archives. (MSS & Archives-2003/1-WPHC 4/IV-1908, 444/1908). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- London, Jack (1911). *The cruise of the Snark*. New York: Macmillan & Co.
- Louie (07.01.1925). [Witness statement]. Proceedings in regard to charge of misconduct preferred against Dr. C.R.M. Pattison, 1925. Western Pacific Archives. (MSS & Archives-2003/1-WPHC 4/IV-1925, 253/1925). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Luke, Harry (1945). *From a South Seas diary*. London: Nicholson & Watson.
- Luraghi, Silvia (2003). *On the meaning of prepositions and cases: The expression of semantic roles in Ancient Greek*. Amsterdam & Philadelphia: John Benjamins.
- Lynch, Bohun (Ed.) (1923). *Isles of illusion. Letters from the South Seas*. London, Bombay & Sydney: Constable.
- Lynch, John (1998). *Pacific languages: An introduction*. Honolulu: University of Hawai'i Press.
- Lynch, John (2010). "Unmarked transitive verbs in Melanesian Pidgin." Eds. Kenneth McElhanon and Ger Reesink, *A mosaic of languages and cultures: Studies celebrating the career of Karl J. Franklin*. Dallas, Texas: SIL International. 227–245.
- Lyons, John (1968). *Introduction to theoretical linguistics*. Cambridge: Cambridge University Press.
- MacClancy, Jeremy (1981). *To kill a bird with two stones: A short history of Vanuatu*. Vanuatu: Vanuatu Cultural Centre.
- MacFarlane, Samuel (1873). *The story of the Lifu mission*. London: James Nisbet & Co.
- Mackellar, C. D. (1912). *Scenes isles and coral gardens: Torres Straits, German New Guinea and the Dutch East Indies*. London: John Murray.
- MacQuarrie, Hector (1948). *Vouza and the Solomon Islands*. New York: Macmillan & Co.
- Maitz, Péter, Siegwalt Lindenfelser and Craig Volker (forthcoming). "Unserdeutsch (Rabaul Creole German), Papua New Guinea." Eds. Hans Boas, Ana Deumert, Mark Loudon and Péter Maitz, *Varieties of German worldwide*. Oxford: Oxford University Press.

- Manawa (24.10.1898). [Witness statement]. Death of John Fleming, a white trader; circumstances concerning, 1899. Western Pacific Archives. (MSS & Archives-2003/1-WPHC 4/IV-1899, 9/1899). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Markham, Clements (Ed.) (1904). *The voyages of Pedro Fernandez de Quiros, 1595-1606* 1. London: Hakluyt Society.
- Marks, Robert M. (2012). *China: Its environment and history*. Maryland: Rowman & Littlefield Publishers.
- Maski Mike (1937). *Pidgin English dictionary of common nouns and phrases used in conversation with natives in the Territory of New Guinea, with advice to new arrivals*. Townsville: T. Willmetts & Sons.
- Matches, Margaret (1930). *Savage paradise*. New York & London: The Century Co.
- Maurer, Philippe (1995). *L'angolar. Un créole afro-portugais parlé à São Tomé. Notes de grammaire, textes, vocabulaire*. Hamburg: Buske.
- Maurer, Philippe (2013a). "Pronominal and adnominal demonstratives." Eds. Susanne Michaelis, Philippe Maurer, Martin Haspelmath and Magnus Huber, *The atlas of pidgin and creole language structures*. Oxford: Oxford University Press. 122–125.
- Maurer, Philippe (2013b). "Distance contrasts in demonstratives." Eds. Susanne Michaelis, Philippe Maurer, Martin Haspelmath and Magnus Huber, *The atlas of pidgin and creole language structures*. Oxford: Oxford University Press. 126–129.
- Maurer, Philippe (2013c). "Comitatives and instrumentals." Eds. Susanne Michaelis, Philippe Maurer, Martin Haspelmath and Magnus Huber, *The atlas of pidgin and creole language structures*. Oxford: Oxford University Press. 276–279.
- Max Planck Institute (February 2008). "Leipzig Glossing Rules: Conventions for interlinear morpheme-by-morpheme glosses. Leipzig: Max Planck Institute for Evolutionary Anthropology. <<https://www.eva.mpg.de/lingua/resources/glossing-rules.php>>. Accessed 23 October 2016.
- McCarthy, John K. (1926 - 1952). "Translations of various sections of the native ordinance into Pidgin English." *Patrol reports and other papers*. Pacific Manuscripts Bureau. PMB 616.
- McMahon, April M. S. (1994). *Understanding language change*. Cambridge: Cambridge University Press.
- McWhorter, John (1997). "Review of the early stages of creolization." *New West Indian Guide/Nieuwe West Indische Gids* 71 (1 & 2): 174–178.
- McWhorter, John (1998). "Identifying the creole prototype: Vindicating a typological class." *Language* 74 (4): 788–818.
- McWhorter, John (2000). *The missing Spanish creoles: Recovering the birth of plantation contact languages*. California: University of California Press.
- McWhorter, John (2002). "The rest of the story. Restoring pidginization to creole genesis theory." *Journal of Pidgin and Creole Languages* 17 (1): 1–48.
- McWhorter, John (2005). *Defining creole*. Oxford: Oxford University Press.
- McWhorter, John (2011). *Linguistic simplicity and complexity. Why do languages undress?* Berlin: De Gruyter Mouton.
- Melanesian Mission (1877). 'The island voyage 1877'. *Letters of the Bishop of Melanesia, Revs. Alfred Penny, John Palmer and John Still*. In: *The island voyage. Printed accounts of the annual voyages of the Melanesian Mission vessel Southern Cross to Norfolk Island, the Banks Group and Solomon Islands*. Pacific Manuscripts Bureau. AU PMB DOC 216.
- Meltiknumba, Meltiwoom, Melterest, Melterus (30.03.1914). [Letter to Mr. King, Atchen. Grievances, which the Atchin natives asked Mr. Clarker to write down, town by town, and hand to Mr. King, H. B. M. Resident Commissioner]. Native Outrages, 1914 - 1924 (3 parts in 8 folders). Western Pacific Archives. (MSS & Archives-2003/1-NHBS 1/I Vol 1, 17/1914. Folder 8). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Mendaña, Álvaro (1901). "A consecutive account of all that happened in the discovery of the islands which the illustrious Sr ALVARO DAVENDANA [sic] went to discover from the year 1567 to the year 1568." Eds. Amherst Hackney and Basil Thomson, *The Discovery of the Solomon Islands by Alvaro de Mendaña in 1568: Translated from the original Spanish manuscripts*. London: Hakluyt Society. 97–158.
- Metcalfe, John R. (24.05.1937). *Letter to friends*. In: Metcalfe, Rev. John R. (15.09.1920-January 1950). Descriptive newsletters from the Solomon Islands. Pacific Manuscripts Bureau. PMB 68.
- Metcalfe, John R. (1947). *Choiseul Circuit Report* In: Metcalfe, Rev. John R. (1918-1952). Solomon Islands - Methodist Mission Reports. Pacific Manuscripts Bureau. PMB 80. 67–71.
- Methodist Mission (1935). *A Methodist hymn book in Pidgin English: For use in mission services held for indentured labourers*. Malakuna, Rabaul: Methodist Mission.
- Meyerhoff, Miriam (2013a). "Bislama." Eds. Susanne Michaelis, Philippe Maurer, Martin Haspelmath and Magnus Huber, *The survey of pidgin and creole languages: Volume 1: English-based and Dutch-based languages*. Oxford: Oxford University Press. 223–231.
- Meyerhoff, Miriam (2013b). "Bislama structure dataset." Eds. Susanne Michaelis, Philippe Maurer, Martin Haspelmath and Magnus Huber. *Atlas of pidgin and creole language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <<http://apics-online.info/contributions/23>>. Accessed 18 July 2019.

- Michaelis, Susanne and Martin Haspelmath (2013a). "Subject relative clauses." Eds. Susanne Michaelis, Philippe Maurer, Martin Haspelmath and Magnus Huber, *The atlas of pidgin and creole language structures*. Oxford: Oxford University Press. 366–369.
- Michaelis, Susanne and Martin Haspelmath (2013b). "Object relative clauses." Eds. Susanne Michaelis, Philippe Maurer, Martin Haspelmath and Magnus Huber, *The atlas of pidgin and creole language structures*. Oxford: Oxford University Press. 370–373.
- Michaelis, Susanne and Martin Haspelmath (2013c). "'Want' complement subjects." Eds. Susanne Michaelis, Philippe Maurer, Martin Haspelmath and Magnus Huber, *The atlas of pidgin and creole language structures*. Oxford: Oxford University Press. 386–389.
- Michaelis, Susanne, Philippe Maurer, Martin Haspelmath and Magnus Huber (Eds.) (2013). *The atlas of pidgin and creole language structures*. Oxford: Oxford University Press.
- Michelsen, Oscar (1893). *Cannibals won for Christ*. London: Morgan & Scott.
- Montgomery, Henry H. (1893). "Melanesia and the Labour Traffic." *The New Review* 8: 552.
- Moore, Clive (2017). *Making Mala: Malaita in Solomon Islands, 1870s–1930s*. Canberra: Australian National University Press.
- Mosel, Ulrike (1979). "Early language contact between Tolai, Pidgin and English in the light of its sociolinguistic background (1875–1914)." Ed. Stephen Wurm, *Papers in pidgin and creole linguistics 2: Pacific Linguistics Series*. Canberra: Australian National University Press. 163–181.
- Mosel, Ulrike and Peter Mühlhäusler (1982). "New evidence of a Samoan origin of New Guinea Tok Pisin (New Guinea Pidgin English)." *The Journal of Pacific History* 17.03: 166–175.
- Moses, John (1973). "The coolie labour question and German colonial policy in Samoa, 1900–1914." *The Journal of Pacific History* 8: 101–124.
- Mufwene, Salikoko S. (1996). "The founder principle in creole genesis." *Diachronica* 13: 83–143.
- Mufwene, Salikoko S. (2001). *The ecology of language evolution*. Cambridge: Cambridge University Press.
- Mufwene, Salikoko S. (2005). "Language evolution: The population genetics way." Ed. Günter Hauska, *Gene, Sprachen und ihre Evolution*. Regensburg: Universitätsverlag. 30–52.
- Mufwene, Salikoko S. (2006). "Multilingualism in linguistic history: Creolization and indigenization." Eds. Tej K. Bhatia and William C. Ritchie, *The handbook of bilingualism*. Massachusetts: Blackwell. 460–488.
- Mühlhäusler, Peter (1978). "Samoan Plantation Pidgin English and the Origin of New Guinea Pidgin." *Papers in Pidgin and Creole Linguistics. Papers in Pidgin and Creole Linguistics* 54: 67–119.
- Mühlhäusler, Peter (1979). *Growth and structure of the lexicon of New Guinea Pidgin*. Canberra: Pacific Linguistics Series C, Australian National University.
- Mühlhäusler, Peter (1985a). "External history of Tok Pisin." Eds. Stephen Wurm and Peter Mühlhäusler, *Handbook of Tok Pisin*. Canberra: Research School of Pacific Studies, Australian National University. 35–64.
- Mühlhäusler, Peter (1985b). "Internal development of Tok Pisin." Eds. Stephen Wurm and Peter Mühlhäusler, *Handbook of Tok Pisin*. Canberra: Research School of Pacific Studies, Australian National University. 75–166.
- Mühlhäusler, Peter (1985c). "Syntax of Tok Pisin." Eds. Stephen Wurm and Peter Mühlhäusler, *Handbook of Tok Pisin*. Canberra: Research School of Pacific Studies, Australian National University. 341–421.
- Mühlhäusler, Peter (1996). "Pidgins and creoles of Queensland." Eds. Darrell Tryon, Peter Mühlhäusler and Stephen Wurm, *Atlas of languages of intercultural communication in the Pacific, Asia and the Americas*. Berlin & Boston: De Gruyter Mouton. 69–81.
- Mühlhäusler, Peter (1997). *Pidgin and creole linguistics: expanded and revised edition*. Westminister: University of Westminister Press.
- Mühlhäusler, Peter and Philip Baker (1996). "English-derived contact languages in the Pacific in the 20th century (excluding Australia)." Eds. Darrell Tryon, Peter Mühlhäusler and Stephen Wurm, *Atlas of languages of intercultural communication in the Pacific, Asia and the Americas*. Berlin & Boston: De Gruyter Mouton. 497–536.
- Mühlhäusler, Peter, Tom Dutton, Even Hovdhaugen, Jeff Williams and Stephen Wurm (1996). "Precolonial patterns of intercultural communication in the Pacific Islands." Eds. Darrell Tryon, Peter Mühlhäusler and Stephen Wurm, *Atlas of languages of intercultural communication in the Pacific, Asia and the Americas*. Berlin & Boston: De Gruyter Mouton. 401–430.
- Mühlhäusler, Peter and Beverly Mühlhäusler (2005). "Simple English in the South Seas Evangelical Mission: Social context and linguistic attributes." *Language Problems & Language Planning* 29 (1): 1–30.
- Mühlhäusler, Peter, Thomas E. Dutton and Suzanne Romaine (2003). *Tok Pisin texts: From the beginning to the present*. Amsterdam & Philadelphia: John Benjamins.
- Mundhenk, Norm (1990). "Linguistic decisions in the 1987 Tok Pisin Bible." Ed. John Verhaar, *Melanesian Pidgin and Tok Pisin*. Amsterdam: John Benjamins. 345–373.
- Munro, Doug (1995). "The labor trade in Melanesians to Queensland: An historiographic essay." *Journal of Social History* 28 (3): 609–627.
- Munro, Doug and Stewart Firth (1990). "German labour policy and the partition of the Western Pacific: The view from Samoa." *The Journal of Pacific History* 25 (1): 85–102.
- Munro, James K. (1867). *The Mary Ira*. London: T. Cautley Newby.

- Muspratt, Eric (1931). *My South Sea Island*. New York: William Morrow & Co.
- Muysken, Pieter (1981). "Halfway between Quechua and Spanish. The case for relexification." Eds. Albert Valdman and Arnold Highfield, *Historicity and variation in creole studies*. Ann Arbor: Karoma. 52–78.
- Muysken, Pieter (1988). "Are creoles a special type of language?" Ed. Frederick Newmeyer, *Linguistics: The Cambridge survey, Vol. II*. Cambridge: Cambridge University Press. 285–301.
- Muysken, Pieter (2008). "Creole studies and multilingualism." Eds. Silvia Kouwenberg and John Singler, *The handbook of pidgin and creole studies*. Oxford: Wiley-Blackwell. 287–308.
- Muysken, Pieter (2021+). "Substrate influence in Northern Quechua languages." Eds. Enoch Aboh and Cécile Vigouroux, *Variation rolls the dice*. Amsterdam: John Benjamins. 133–160.
- Narrog, Heiko (2012). *Modality, subjectivity, and semantic change: A cross-linguistic perspective*. New York: Oxford University Press.
- Neuhauss, Richard (1911). *Deutsch-Neu-Guinea: in 3 Bnd.* Berlin: Reimer.
- Nevalainen, Terttu and Helena Raumolin-Brunberg (2012). "Historical sociolinguistics: Origins, motivations, and paradigms." Eds. Juan Hernández Campoy and Juan Conde-Silvestre, *The handbook of historical sociolinguistics*. Malden, Massachusetts: Wiley-Blackwell. 22–40.
- Nevalainen, Terttu and Helena Raumolin-Brunberg (2017). *Historical sociolinguistics: Language change in Tudor and Stuart England*. London & New York: Routledge.
- New Hebrides Service de l'Information and New Hebrides Résidence de France. (1868-1896). *Bulletin d'information de la Résidence de France*.
- Nicholson, Ian H. (1990). *Log of Logs: A catalogue of logs, journals, shipboard diaries, letters, and all forms of voyage narratives, 1788 to 1988, for Australia and New Zealand and surrounding oceans*. Queensland: Sunstrip Printers.
- Norden, Hermann (1926). *Byways of the tropic seas: Wanderings among the Solomons and in the Malay Archipelago*. London: H. F. & G. Witherby.
- Nuyts, Jan (2001). *Epistemic modality, language, and conceptualization: A cognitive-pragmatic perspective*. Amsterdam & Philadelphia: John Benjamins.
- Nuyts, Jan (2016a). "Surveying modality and mood." Eds. Jan Nuyts and Johan van der Auwera, *Oxford handbook of modality and mood*. Oxford: Oxford University Press. 1–8.
- Nuyts, Jan (2016b). "Analyses of the modal meanings." Eds. Jan Nuyts and Johan van der Auwera, *Oxford handbook of modality and mood*. Oxford: Oxford University Press. 31–49.
- Okea, Harry (29.12.1911). *[Witness statement]*. Double murder in the Solomon Islands by Harry Okea, 1912. Western Pacific Archives. (MSS & Archives-2003/1-WPHC 4/IV-1912, 283/1912). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Oora (24.12.1888). *[Witness statement]*. Australian Station (New Guinea & Solomon Islands, 1888). "Correspondence respecting outrages by natives on British subjects, and other matters, which have been under inquiry during the year 1888, being continuation of reports of cases dealt with in former years, together with other cases which have since arisen." In *Copies of reports of proceedings in Solomon Islands & New Hebrides*. From: Fairfax. Rear Admiral H., at H.M.S. "Orlando" at Adelaide. Western Pacific Archives. (MSS & Archives.2003/1.WPHC 4/IV.1889, 93/1889). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Oroar, Peter (19.05.1906). *[Letter to Mr. Woodford per J. A. Watkinson]*. Forwards letter from Mr Watkinson, a missionary, reporting instances of returning kanaka bushmen being attacked on their arrival by saltwater men and having certain property stolen, 1906. Western Pacific Archives. (MSS & Archives-2003/1-WPHC 4/IV-1906, 221/1906). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Palmer, Frank R. (1990). *Modality and the English modals*. New York & London: Routledge.
- Palmer, Frank R. (2001). *Mood and modality*. Cambridge: Cambridge University Press.
- Parkinson, Richard (1887). *Im Bismarck-Archipel*. Leipzig: Brockhaus.
- Parkvall, Mikael (2008). "The simplicity of creoles in a cross-linguistic perspective." Eds. Matti Miestamo, Kaius Sinnemäki and Fred Karlsson, *Language complexity. Typology, contact, change*. Amsterdam: John Benjamins. 265–285.
- Parkvall, Mikael and Peter Bakker (2013). "Pidgins." Eds. Yaron Matras and Peter Bakker, *Contact languages. A comprehensive guide*. Berlin & Boston: De Gruyter Mouton. 15–64.
- Paton, Frederick J. (23.04.1893 -31.12.1908). *Journal of the Rev Frederick J. Paton, Presbyterian Missionary, New Hebrides*. Pacific Manuscripts Bureau. PMB MS 32.
- Paton, Maggie W. (1894). *Letters and sketches from the New Hebrides*. London: Hodder & Stoughton.
- Philip, J. E. (1978). *A Solomons sojourn: J. E. Philp's log*. Hobart: Tasmanian Historical Research Association.
- Pionnier, Jean-Nestor (1913). "Pigeon English ou Bichelamar." *Revue de Linguistique et de Philologie Comparée* 46: 9–17, 184–198.
- Poech, Rudolf (1904). *Pidgin English: Recording No. 381*. Phonogrammarchiv der Oesterreichischen Akademie der Wissenschaften.
- Price, Charles A. and Elizabeth Baker (1976). "Origins of Pacific Island labourers in Queensland, 1863-1904: A research note." *The Journal of Pacific History* 11.2 (Labour Trade [Part Two]): 106–121.

- Puanimarau (24.01.1927). [Witness statement]. Murder Charge: Rex v. Apuilavo, Barbar & Lauia, 1927. Western Pacific Archives. (MSS & Archives-2003/1-WPHC 4/IV-1927, 720/1927). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Queensland Department of Justice (1892). *Manuscript*.
- Queensland Department of Justice (1906). *Inquests* JUS/N 346-365.
- Queensland Kanaka Mission and South Sea Evangelical Mission (1887-1995). *Queensland Kanaka Mission and the South Sea Evangelical Mission, Sydney and Brisbane*. Pacific Manuscripts Bureau. AU PMB DOC 439.
- Queensland Supreme Court (1884). *Regina vs. Millman Spall etc.* SCT/CC 49.
- Quinn, Heidi (2009). "Pronoun forms." Eds. Pam Peters, Peter Collins and Adam Smith, *Comparative studies in Australian and New Zealand English: Grammar and beyond*. Amsterdam & Philadelphia: John Benjamins. 31–48.
- R Core Team (2020). *R: A language and environment for statistical computing*. Statistical Computing, Vienna, Austria. <<https://www.R-project.org/>>. Accessed 18 January 2020.
- Ramichi (14.09.1927). [Witness statement]. Guadalcanal Murders: Notes of evidence of trial of murderers viz: Billy Viti, Karagori, Tuva, Tuatakombu, Votungi, Vatsuna, Hinji, Aria, and Sevo, 1927. Western Pacific Archives. (MSS & Archives-2003/1-WPHC 4/IV-1927, 3434/1927). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Rannie, Douglas (1912). *My adventures among South Sea cannibals*. London: Seeley, Service & Co.
- Raumolin-Brunberg, Helena (1996). "Historical sociolinguistics." Eds. Terttu Nevalainen and Helena Raumolin-Brunberg, *Sociolinguistics and language history: Studies based on the Corpus of Early English Correspondence*. Amsterdam: Rodopi. 11–37.
- Read, Dorothy (29.01.1930). *Letter to friends, One Pusu*. Pacific Manuscripts Bureau. PMB 1150. Reel 4.
- Read, Dorothy (March 1931). *Letter to friends, One Pusu*. Pacific Manuscript Bureau. PMB 1150. Reel 4.
- Reed, Stephen W. (1943). *The making of modern New Guinea, with special reference to culture contact in the mandated territory*. Philadelphia: The American Philosophical Society.
- Reeves, L. C. (1915). *Australians in action in New Guinea*. Sydney: W. C. Penfold & Co. Ltd.
- Richards, Rhys (2010). "Sperm whaling on the Solanders Grounds and in Fiordland: A maritime historian's perspective." *NIWA Information Series* 76: 1–56.
- Rivers, William H. R. (1914). *The history of Melanesian society*. Cambridge: Cambridge University Press.
- Roberts, Sarah (1998). "The role of diffusion in the genesis of Hawaiian Creole." *Language* 74: 1–39.
- Roberts, Sarah (2005). *The emergence of Hawai'i Creole English in the early 20th century: The sociohistorical context of creole genesis*. Doctoral dissertation. Stanford: Stanford University.
- Romaine, Suzanne (1985). "Relative clauses in child language, pidgins and creoles." *Papers in Pidgin and Creole Linguistics* A-72 No. 4: 1–23.
- Romaine, Suzanne (1992). "The evolution of complexity in a creole language: Relative clauses in Tok Pisin." *Studies in Language* 16: 139–182.
- Romaine, Suzanne (1995). "The grammaticalization of irrealis in Tok Pisin." Eds. Joan Bybee and Suzanne Fleischman, *Modality in grammar and discourse*. Amsterdam: John Benjamins. 389–428.
- Romaine, Suzanne (1999). "The Grammaticalization of the proximative in Tok Pisin." *Language* 75: 322–346.
- Romaine, Suzanne (2005). "Lexical structure in Pidgins and Creoles." Eds. Alan Cruse, Franz Hundsnurscher, Michael Job and Peter Lutzeier, *Lexicology. An international handbook on the nature and structure of words and vocabularies*, Vol. 2. Berlin & New York: De Gruyter Mouton. 1092–1095.
- Roman Catholic Mission (1883-1933). *Papers by the Roman Catholic Mission*. Pacific Manuscript Bureau. AU PMB MS 53.
- Roman Catholic Mission (1939a). *Bigfelo katolik baibel*. Alexis Harbour: Catholic Mission Press.
- Roman Catholic Mission (1939b). *Skul bilong evangelio*. Alexis Harbour: Catholic Mission Press.
- Romilly, Hugh H. (1886). *The western Pacific and New Guinea: Notes on the natives, Christian and cannibal, with some account of the old labour trade*. London: John Murray.
- Ross, Malcolm (1996). "Mission and church languages in Papua New Guinea." Eds. Darrell Tryon, Peter Mühlhäusler and Stephen Wurm, *Atlas of languages of intercultural communication in the Pacific, Asia and the Americas*. Berlin & Boston: De Gruyter Mouton. 595–617.
- RStudio Team (2019). *RStudio: Integrated development environment for R*. 1.2.5033. Boston, MA: RStudio, Inc. <<http://www.rstudio.com/>>. Accessed 18 January 2020.
- Salisbury, R. F. (03.07.1931). [Letter to R.C., Hog Harbour]. Murder of Sumbalum of Malo by Malitakelco (Molitagaleo), 1931 - 1939. Western Pacific Archives. (MSS & Archives-2003/1-NHBS 1/I Vol 3, 143/1931). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Salisbury, R. (1967). "Pidgin's respectable past." *New Guinea* 2(2): 44–48.
- Sankoff, Gillian (1993). "Focus in Tok Pisin." Eds. Francis Byrne and Donald Winford, *Focus and grammatical relations in creole languages: Papers from the University of Chicago Conference on Focus and Grammatical Relations in Creole Languages*. Amsterdam & Philadelphia: John Benjamins. 117–140.
- Sankoff, Gillian (2021+). "A local history of Tok Pisin: Language contact in Papua New Guinea." Eds. Enoch Aboh and Cécile Vigouroux, *Variation rolls the dice*. Amsterdam: John Benjamins. 57–80.

- Sankoff, Gillian and Penelope Brown (1976). "The origins of syntax in discourse: A case study of Tok Pisin relatives." *Language* 52 (3): 631–666.
- Sankoff, Gillian and Penelope Brown (1980). "The origins of syntax in discourse: A case study of Tok Pisin relatives: (Language article reprinted with minor revisions)." Ed. Gillian Sankoff, *The social life of language*. Pittsburgh: University of Pennsylvania Press. 211–256.
- Sarmiento, Pedro (1901). "A short account collected from the papers which they found in the city of La Plata, concerning the voyage and discovery of the Western Islands, commonly called the Isles of Solomon: The narrative of Pedro Sarmiento." Eds. Amherst Hackney and Basil Thomson, *The discovery of the Solomon Islands by Alvaro de Mendaña in 1568: Translated from the original Spanish manuscripts*. London: Hakluyt Society. 83–95.
- Sayer, Edgar S. (1943). *Pidgin English: A text book, history, and vocabulary of Pidgin English, for writers, travellers, students of the English language*. Toronto.
- Schäfer, Damaris (in progress a). "Purpose prepositions and infinitive-like constructions in MPE varieties from a diachronic perspective."
- Schäfer, Damaris (in progress b). "Pertentive prepositions in MPE varieties from a diachronic perspective."
- Schäfer, Damaris (2017a). "From an unstable to a stable pidgin lexicon – Circumlocutions and their replacements in early Tok Pisin, Solomons Pijin and Bislama." *Society for Pidgin and Creole Linguistics*. Tampere, Finland.
- Schäfer, Damaris (2017b). "The impact of Second World War propaganda leaflets on Tok Pisin grammar and lexicon." *Eleventh Creolistics Workshop*, Gießen.
- Schafroth, Marie (1916). *Südsee-Welten vor dem großen Krieg*. Bern: Franke.
- Schellong, Otto (1934). *Alte Dokumente aus der Südsee: Zur Geschichte der Gründung einer Kolonie. Erlebtes und Eingeborenenstudien*. Königsberg: Gräfe & Unzer.
- Schnee, Heinrich (1904). *Bilder aus der Südsee. Unter kannibalischen Stämmen des Bismarck-Archipels*. Berlin: D. Reimer.
- Schneider, Edgar (2013). "Investigating historical variation and change in written documents. New perspectives." Eds. Jack K. Chambers and Natalie Schilling, *The handbook of language variation and change*. Hoboken: Wiley. 57–81.
- Schuchardt, Hugo (1883). "Kreolische Studien V 'Ueber das Melaneso-englische.': Reproduced and translated by Gilbert, Glenn (1980). *Pidgin and Creole Languages: Selected Essays by Hugo Schuchardt*. Cambridge: Cambridge University Press. 15–23." *Sitzungsberichte der philosophisch-historischen Classe der kaiserlichen Akademie der Wissenschaften in Wien* 105: 151–161.
- Schuchardt, Hugo (1889). "Beiträge zur Kenntniss des englischen kreolisch II." *Melaneso-englisches, Englische Studien* (13): 158–162.
- Sebba, Mark (1997). *Contact languages. Pidgins and creoles*. London: Macmillan & Co.
- Shelton-Smith, W. (22.04.1929). "Pidgin pie: English as she is spoke in New Guinea." *The Brisbane Courier*: 10.
- Shineberg, Dorothy (1967). *They came for Sandalwood: A study of the sandalwood trade in the South-West Pacific, 1830-1865*. Melbourne: Melbourne University Press.
- Shipping Gazette and Sydney General Trade List (Saturday 01.03.1845). "India." *Shipping Gazette and Sydney General Trade List*: 57–60.
- Shurcliff, Sidney N. (1930). *Jungle islands*. New York & London: G. P. Putnam's Sons.
- Siegel, Jeff (1981). "Developments in written Tok Pisin." *Anthropological Linguistics* 23: 20–35.
- Siegel, Jeff (1985). "Current use and expansion of Tok Pisin: Tok Pisin in the mass media." Eds. Stephen Wurm and Peter Mühlhäusler, *Handbook of Tok Pisin*. Canberra: Research School of Pacific Studies, Australian National University. 517–533.
- Siegel, Jeff (1987). *Language contact in a plantation environment: A sociolinguistic history of Fiji*. Cambridge: Cambridge University Press.
- Siegel, Jeff (2008). *The emergence of pidgin and creole languages*. Oxford: Oxford University Press.
- Siegel, Jeff (2013). "Pidgin Hindustani structure dataset." Eds. Susanne Michaelis, Philippe Maurer, Martin Haspelmath and Magnus Huber, *Atlas of pidgin and creole language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <<http://apics-online.info/contributions/70>>. Accessed 03 March 2021.
- SIL International (2017). *Fieldworks Language Explorer*. 8.2.8.42510. Dallas, Texas. <<https://software.sil.org/fieldworks/download/>>. Accessed 11 May 2017.
- Simon, Linda and Hugh Young (1978). *Pijin blong yumi: A guide to Solomon Islands Pijin*. Honiara: Solomon Islands Christian Association.
- Simons, Linda (1983). "A comparison of the pidgins of Solomon Islands and Papua New Guinea." Ed. Stephen Wurm, *Papers in Pidgin and Creole Linguistics* 3. Australia: Australian National University Press. 121–137.
- Simons, Linda (1985). "Malaitan influence on two grammatical particles in Solomon Islands Pijin." *Pacific Linguistics* A72: 53–65.
- Sippola, Eeva (2013). "Ternate Chabacano structure dataset." Eds. Susanne Michaelis, Philippe Maurer, Martin Haspelmath and Magnus Huber, *Atlas of pidgin and creole language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <<http://apics-online.info/contributions/44>>. Accessed 03 March 2021.

- Slomanson, Peter (2013). "Sri Lankan Malay structure dataset. Eds. Susanne Michaelis, Philippe Maurer, Martin Haspelmath and Magnus Huber, *Atlas of pidgin and creole language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <<https://apics-online.info/contributions/66>>. Accessed 03 March 2021.
- Smith, Geoff P. (2002). *Growing up with Tok Pisin: Contact, creolization and change in Papua New Guinea's national language*. London: Battlebridge.
- Smith, Geoff and Jeff Siegel (2013a). "Tok Pisin." Eds. Susanne Michaelis, Philippe Maurer, Martin Haspelmath and Magnus Huber, *The survey of pidgin and creole languages: Volume 1: English-based and Dutch-based languages*. Oxford: Oxford University Press. 214–222.
- Smith, Geoff and Jeff Siegel (2013b). "Tok Pisin structure dataset." Eds. Susanne Michaelis, Philippe Maurer, Martin Haspelmath and Magnus Huber, *Atlas of pidgin and creole language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <<http://apics-online.info/contributions/22>>. Accessed 12 May 2019.
- Smith, Ian (2013). "Sri Lanka Portuguese structure dataset." Eds. Susanne Michaelis, Philippe Maurer, Martin Haspelmath and Magnus Huber, *Atlas of pidgin and creole language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <<https://apics-online.info/contributions/41>>. Accessed 03 March 2021.
- Sociedad Geográfica (1885). "Conflicto Hispano-Alemán: Discurso pronunciado por Don Francisco Coello, en la sesión ordinaria del 27 de agosto de 1885." *Boletín de la Real sociedad geográfica* 19: 220–335.
- Soukup, Barbara (2017). "Historical sociolinguistic philology - a new hybrid discipline, its interests, and its scope." *Open Linguistics* (3): 673–678.
- South Sea Evangelical Mission (1909). "Letters from our native mail bag: Letter by David Overi 17.05.1909." *Not in Vain. What God Hath Wrought Amongst South Sea Islanders, 23rd Annual Report of the South Sea Evangelical Mission*. 37–38. Pacific Manuscripts Bureau. AU PMB DOC 439. Reel 1.
- South Sea Evangelical Mission (March 1932 - July 1956). *Despatches from the South Sea Evangelical Mission*. Pacific Manuscripts Bureau. AU PMB DOC 440.
- South Sea Evangelical Mission (1945–1966). *Solomon Soldiers' News*. Pacific Manuscripts Bureau. AU PMB DOC 442.
- South Sea Evangelical Mission, formerly Queensland Kanaka Mission (1890–1942). *Correspondence*. Pacific Manuscripts Bureau. PMB 1150.
- Speiser, Felix (1913). *Two years with the natives in the Western Pacific*. London: Mills & Boon.
- Steel, Robert (1880). *The New Hebrides and Christian missions: With a sketch of the labour traffic and notes of a cruise through the group in the mission vessel*. London: James Nisbet.
- Stephan, Emil and Fritz Graebner (1907). *Neu-Mecklenburg (Bismarck Archipel): Die Küste von Umuddu bis Kap St. Georg: Forschungsergebnisse bei den Vermessungsfahrten von S.M.S. Möwe im Jahre 1904*. Berlin: Dietrich Reimer Verlag.
- Stewart, A. G. (28.04.1917). *[Letter to Mr King]*. Claim of David of Ambrym against Salvin, 1917 - 1917. Western Pacific Archives. (MSS & Archives-2003/1-NHBS 1/I Vol 2, 9/1917). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Stober, Will E. *Isles of disenchantment: The Fletcher/Jacomb Correspondence, 1913–1921*. Pacific Manuscripts Bureau. PMB 1243. Reel 1.
- Stolz, Thomas, Cornelia Stroh and Aina Urdze (2013). "Comitatives and instrumentals." Eds. Matthew Dryer, Martin Haspelmath, David Gil and Bernard Comrie, *The world atlas of language structures*. Leipzig: Max Planck Institute for Evolutionary Anthropology. 214–217.
- Stolz, Thomas, Cornelia Stroh and Aina Urzde (2006). *On comitatives and related categories: A typological study with special focus on the languages of Europe*. Berlin: De Gruyter Mouton.
- Strauch, H. (1876). "Verzeichnis von 477 Wörtern, gesammelt während des Aufenthaltes S.M.S. „Gazelle“ in Neu Guinea, Neu-Hannover, Neu-Irland, Neu-Britannien und Brisbane." *Zeitschrift für Ethnologie* 8: 405–419.
- Strobl, Carolin, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin and Achim Zeileis (2008). "Conditional variable importance for random forests." *BMC Bioinformatics* 9 (307): 1–11.
- Sullivan, Violet M. (August 1926). *Circular No. 40, One Pusu* Pacific Manuscripts Bureau. PMB 1150. Reel 2.
- Sullivan, Violet M. (February 1931). *Circular No. 68, One Pusu*. Pacific Manuscripts Bureau. PMB 1150. Reel 2.
- Supabo (04.06.1923). *[Letter to Mr. King. Grievences, which the Atchin natives asked Mr. Clarker to write down, town by town, and hand to Mr. King, H. B. M. Resident Commissioner]*. Native Outrages, 1914 - 1924 (3 parts in 8 folders). Western Pacific Archives. (MSS & Archives-2003/1-NHBS 1/I Vol 1, 17/1914. Folder 8). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Sydney Morning Herald (08.01.1943). "Pidgin for soldiers: Training for war in islands." *Sydney Morning Herald* 9.
- Takwafala (22.03.1938). *[Witness statement]*. Sentence passed on Takwafala, native of Malaita:- Submission of case to court of appeal for review, 1938 - 1938. Western Pacific Archives. (MSS & Archives-2003/1-WPHC 4/IV-1938-1939. Series 1, 3159/1938). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Talatova (09.01.1929). *[Witness statement]*. Rex v. Talatova: Minutes of evidence adduced at trial for murder. Western Pacific Archives. (MSS & Archives-2003/1-WPHC 4/IV-1929, 611/1929). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.

- Tara-Haw (06.12.1890). [Witness statement]. Australian Station (New Hebrides, 1890). "Correspondence regarding outrages by natives, disturbances, &c. which have been the subject of inquiry by her majesty's ships in the New Hebrides." In *Forwards printed copies of proceedings of H.M. Ships in New Hebrides, New Guinea, & Solomon Islands*. From: Scott, Rear Admiral Lord Charles, at H.M.S. "Orlando" at Sydney. Western Pacific Archives. (MSS & Archives-2003/1-WPHC 4/IV-1891, 96/1891). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Tedder, James L. O. (16.11.1943). *Police songs from the British Solomon Islands with translation by Belshazzar Gina*. Photocopied typescript. Canberra, Australian National University Archives AU ANUA 445-140.
- "That bell again" (June 1947). *Not in Vain* 106: 12-13. Pacific Manuscripts Bureau. AU PMB DOC 439. Reel 2.
- The Brisbane Courier (02.06.1933). "Sing for hours: Wild, men of Sinarago. Converting witch doctors." *The Brisbane Courier*: 4.
- The Daily Telegraph (09.03.1944). "Pacific pidgin gets jive." *The Daily Telegraph*: 8.
- The London Quarterly Review (July 1854). "Christianity in Melanesia and New Zealand." *The London Quarterly Review*: 85-106.
- The Nautical Magazine (1939). "The New Hebrides." *The Nautical Magazine and Naval Chronicle for 1839. A Journal of Papers on Subjects connected with Maritime Affairs*. London: Simkin, Marshall, and Co.: 602-607.
- The Nautical Magazine (1950). "The New Hebrides, New Caledonia, and the Loyalty Islands." *The Nautical Magazine and Naval Chronicle for 1850. A Journal of Papers on Subjects connected with Maritime Affairs*. London: Simkin, Marshall, and Co.: 425-431.
- "The Stories of the Crew" (June 1947). *Not in Vain Issue* 106: 8-12. Pacific Manuscripts Bureau. AU PMB DOC 439. Reel 2.
- Thomas, Julian (1886). *Cannibals and convicts*. London, Paris, New York & Melbourne: Cassell.
- Thompson, Sandra and Anthony Mulac (1991). "A quantitative perspective on the grammaticalization of epistemic parentheticals in English." Eds. Elizabeth Traugott and Bernd Heine, *Approaches to grammaticalization*. Amsterdam & Philadelphia: John Benjamins. 313-329.
- Thurston, John B. (1871). *Journal of a voyage from Ovalau, Fiji, to the New Hebrides in search of voluntary immigrants*. Typescript.
- Tor Tor (09.01.1929). [Witness statement]. Rex v. Talatova: Minutes of evidence adduced at trial for murder. Western Pacific Archives. (MSS & Archives-2003/1-WPHC 4/IV-1929, 611/1929). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Tropical (24.09.1921). "Pidgin English." *The Sydney Morning Herald*: 7.
- Tryon, Darrell T. (1987). *Bislama. An introduction to the national language of Vanuatu*. (Pacific Linguistics Series D-72.) Canberra: Australian National University Press.
- Tryon, Darrell T. (1991). "Wanem Bislama?" Ed. Robert Blust, *Currents in Pacific linguistics: Papers on Austronesian languages and ethnolinguistics in honour of George W. Grace*. Canberra: Australian National University Press. 509-519.
- Tryon, Darrell T. (1996). "Mission and church languages in Island Melanesia." Eds. Darrell Tryon, Peter Mühlhäusler and Stephen Wurm, *Atlas of languages of intercultural communication in the Pacific, Asia and the Americas*. Berlin & Boston: De Gruyter Mouton. 619-623.
- Tryon, Darrell T. and Jean-Michel Charpentier (2004). *Pacific pidgins and creoles: Origins, growth and development*. Berlin: De Gruyter Mouton.
- Tryon, Darrell T., Peter Mühlhäusler and Philip Baker (1996). "English-derived contact languages in the Pacific in the 19th century (excluding Australia)." Eds. Darrell Tryon, Peter Mühlhäusler and Stephen Wurm, *Atlas of languages of intercultural communication in the Pacific, Asia and the Americas*. Berlin & Boston: De Gruyter Mouton. 471-491.
- Ufaria, Crystal (September 1941). "Crystal's letter to her little sister." *Not in Vain* 83: 3. Pacific Manuscripts Bureau. AU PMB DOC 439. Reel 2.
- Uiaria, Joe (10.04.1924). [Witness statement]. Forwards notes of evidence adduced at the trial of Taliniau, Tahioa & Waira, charged with murder and found guilty of manslaughter, 1924-1924. Western Pacific Archives. (MSS & Archives-2003/1-WPHC 4/IV-1924, 1643/1924). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Uthua, Lizzie (15.07.1920). [Witness statement]. Forwards copies of court record of the trial of Noeli Dili, convicted for murder, 1920. Western Pacific Archives. (MSS & Archives.2003/1.WPHC 4/IV.1920, 2500/1920). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- van Baar, William (~1930). *Pitshen-Wörterbuch*. St. Augustin: Anthropos Institute.
- van der Auwera, Johan and Vladimir Plungian (1998). "Modality's semantic map." *Linguistic Typology* 2: 79-124.
- van der Auwera, Johan and Alfonso Zamorano Aguilar (2016). "The history of modality and mood." Eds. John Nuyts and Johan van der Auwera, *Oxford handbook of modality and mood*. Oxford: Oxford University Press. 9-27.
- van Klaarwater, B. (1934). *Buk-rarin na singsing*. Vunapope: Catholic Mission.
- Vance, Robert C. (December 1946). "A heathen feast." *Not in Vain* 104: 10-12. Pacific Manuscripts Bureau. AU PMB DOC 439. Reel 2.
- Velupillai, Viveka (2012). *An introduction to linguistic typology*. Amsterdam: John Benjamins.

- Velupillai, Viveka (2015). *Pidgins, creoles and mixed languages: An introduction*. Amsterdam: John Benjamins.
- Viveka Velupillai (2021). "The typology of Pidgin and Creole languages." Eds. Umberto Ansaldi and Miriam Meyerhoff, *The Routledge handbook of pidgin and creole languages*. Abingdon, Oxon & New York, NY: Routledge. 384–403.
- Verhaar, John W. M. (1995). *Toward a reference grammar of Tok Pisin: An experiment in corpus linguistics*. Hawai'i: University of Hawai'i Press.
- Verplaetse, Heidi (2003). "What you and I want: A functional approach to verb complementation of modal want to." Eds. Roberta Facchinetti, Frank Palmer and Manfred Krug, *Modality in contemporary English*. Berlin & New York: De Gruyter Mouton. 151–190.
- Viti, Billy (19.09.1927). *[Unsworn witness statement]*. Guadalcanal murders: Notes of evidence of trial of murderers viz: Billy Viti, Karagogori, Tuva, Tuatakombu, Votungi, Vatsuna, Hinji, Aria, and Sevo, 1927. Western Pacific Archives. (MSS & Archives-2003/1-WPHC 4/IV-1927, 3434/1927). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Volker, Craig A. (2008). *Papua New Guinea Tok Pisin English dictionary*. Australia & New Zealand: Oxford University Press.
- von Beck, Carl (1903). "Neu-Guinea." Ed. Franz Hutter, *Das überseeische Deutschland: Die Deutschen Kolonien in Wort und Bild*. Stuttgart: Union. 485–566.
- von Hesse-Wartegg, Ernst (1902). *Samoa, Bismarckarchipel und Neuguinea: drei deutsche Kolonien in der Südsee*. Leipzig: Weber.
- Waiko, John D. (1993). *A short history of Papua New Guinea*. Oxford: Oxford University Press.
- Waite, A. J. (30.08.1929). *Letter to Florence Young, Malaita*. Pacific Manuscripts Bureau. PMB 1150. Reel 3.
- Waite, A. J. (31.03.1930). *Letter to Fellow Helpers, One Pusu*. Pacific Manuscripts Bureau. AU PMB MS 1253. Reel 2.
- Wari (between 17.04.1925–25.06.1925). *[Witness statement]*. Rex v. Niola & Mobe, (Murder): Forwards depositions in re., 1925. Western Pacific Archives. (MSS & Archives-2003/1-WPHC 4/IV-1925, 2325/1925 [not filmed, refer to hard copy]). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Warteioch (13.11.1912). *[Witness statement]*. Forwards copies of notes of evidence and judgement in Rex vs. G. Young. Western Pacific Archives. (MSS & Archives-2003/1-WPHC 4/IV-1913, 416/1913). Special Collections, University of Auckland Libraries and Learning Services, Auckland, NZ.
- Watkinson, Joseph (1906). "News from Mr. Watkinson, November-December 1905." *Not in Vain - What God Hath Wrought Amongst the Kanakas. Report of the Solomon Islands Branch and the 20th Annual Report of the Queensland Kanaka Mission 1905-1906*: 6-8. Pacific Manuscripts Bureau. AU PMB DOC 439. Reel 1.
- Watt, Agnes, C. P. (1896). *Twenty five years' mission life on Tanna*. London: Houlston & Sons.
- Wawn, William T. (1893). *The South Sea islanders and the Queensland labour trade*. Canberra: Australian National University Press.
- Webster, John (1863). *The last cruise of "The Wanderer"*. Sydney: F. Cunninghame & Co.
- Wendeland, Wilhelm (1939). *Im Wunderland der Papuas: ein deutscher Kolonialarzt erlebt die Südsee*. Berlin: Verlag für Volkstum, Wehr und Wirtschaft.
- Werner, Eugen (1911). *Kaiser-Wilhelms-Land: Beobachtungen und Erlebnisse in den Urwäldern Neuguineas*. Freiburg i. Br.: Herder.
- Wickware, Francis (June 1943). "Pidgin English: It may sound like baby talk but soldiers in the South Seas have to learn words and grammar." *Life* 14 (23): 67–70.
- Wilson, Ellen (1911). *The isles that wait: By a lady member of the Melanesian Mission*. London: Society for promoting Christian knowledge.
- Winford, Donald (2018). "Creole tense–mood–aspect systems." *Annual Reviews* 4: 193–212.
- Wolf, Franciscus (1935). *Yesus em i foreman bilong yumi*. Alexishafen: Catholic Mission.
- Wolf, Franciscus (May 1942). *Katekismo Katolik: A Catholic catechism in New Guinea Pidgin for the natives of New Guinea*. Alexishafen: Catholic Mission.
- Woodford, Charles M. (April 1897). "Report on the British Solomon Islands." *Colonial Reports. – Miscellaneous. Western Pacific* 8.
- Woodford, Charles M. (1884). *Revised journal of a voyage from Suva Fiji to the Gilbert group from 19 March 1884 to 22 June 1884*. Canberra: Australian National University Archives AU ANUA 481-267.
- Woodford, Charles M. (1886). *Diary 6 July 1886 - 3 August 1886*. Canberra: Australian National University Archives AU ANUA 481-10.
- Woodford, Charles M. (1890). *A naturalist among the head-hunters. Being an account of three visits to the Solomon islands in the years 1886, 1887 and 1888*. London: George Philip & Co.
- Wright, Roger (2012). "Romance languages as a source for spoken Latin." Ed. James Clackson, *A companion to the Latin language*. Oxford: Wiley-Blackwell. 59–79.
- Wurm, S. (1966). "Pidgin - A national language." *New Guinea* 7: 49–54.
- Wurm, Stephen (2012). "Pacific Pidgin Englishes." Ed. Michael Clyne, *Pluricentric languages: Differing norms in different nations*. Berlin: De Gruyter Mouton. 421–436.

- Yakpo, Kofi (1996). *Grammatikalisierung von Tempus, Aspekt und Modalität in Bislama*. Thesis. Cologne: University of Cologne.
- Yakpo, Kofi (2021+). "Social entrenchment influences the amount of areal borrowing in contact languages." *International Journal of Bilingualism*. OnlineFirst.
- Yakpo, Kofi and Pieter Muysken (Eds.) (2017). *Boundaries and bridges: Language contact in multilingual ecologies*. Berlin & Boston: De Gruyter Mouton.
- Young, Earnest (1899). "Queensland Kanaka Mission." *Not in Vain. What God Hath Wrought Amongst the Kanakas in Queensland being [13th] Report of the Queensland Kanaka Mission. A Message from Malaita in the Solomon Islands*: 5–11. Pacific Manuscripts Bureau. AU PMB DOC 439. Reel 1.
- Young, Florence S. H. (April 1910). *Letter to friends, Solomon Islands*. Pacific Manuscripts Bureau. PMB 1150. Reel 1.
- Young, Florence S. H. (1925). *Pearls from the Pacific*. London & Edinburgh: Marshall Brothers.
- Young, Florence S. H. (July 1929). "Anganiwae Conference." *Not in Vain Issue* 35: 4-5. ("Tend My Sheep"). Ed. Northcote Deck. Sydney and Brisbane. Pacific Manuscripts Bureau. PMB Doc 439. Reel 2.
- Zieschank, Frieda (1918). *Ein Jahrzehnt in Samoa (1906–1916)*. Leipzig: Haberland.
- Zipp, Lena (2014). *Educated Fiji English lexico-grammar and variety status*. Amsterdam: John Benjamins.

Appendices

Appendix I: Additional boxplots

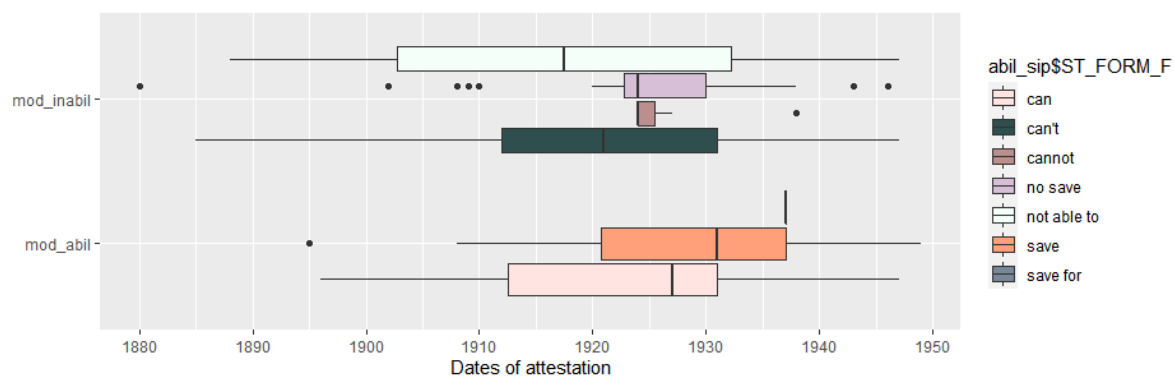


Figure 177: Boxplot of ability variants in SIP based on ST_FORM_F

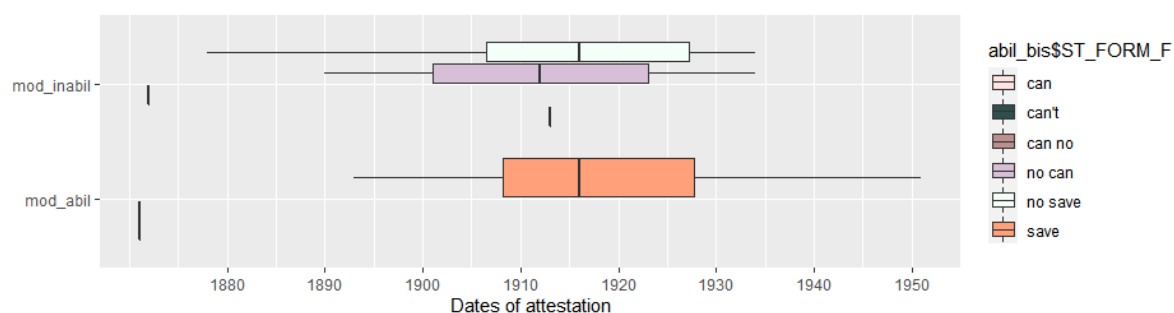


Figure 178: Boxplot of ability variants in BIS based on ST_FORM_F

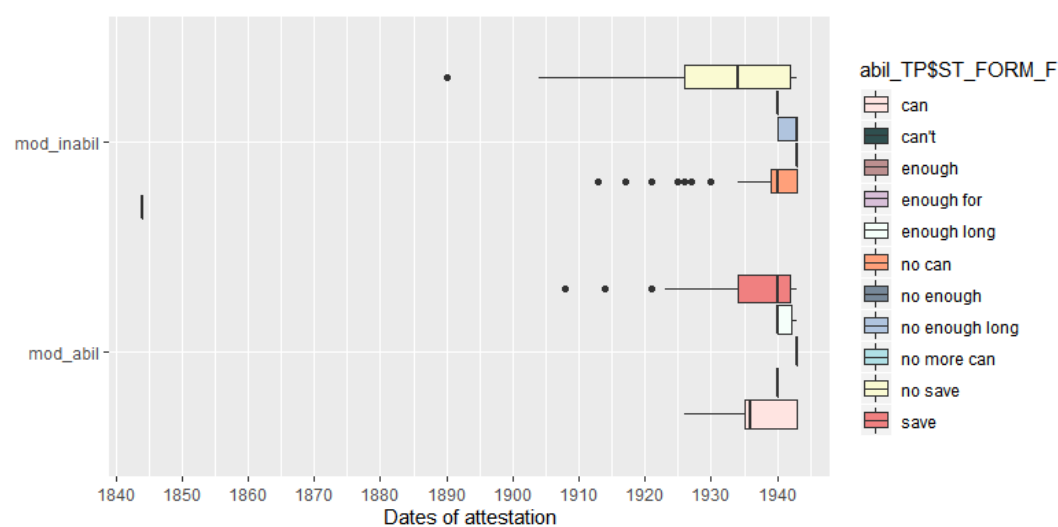


Figure 179: Boxplot of ability variants in TP based on ST_FORM_F

Appendix II: Additional ctrees

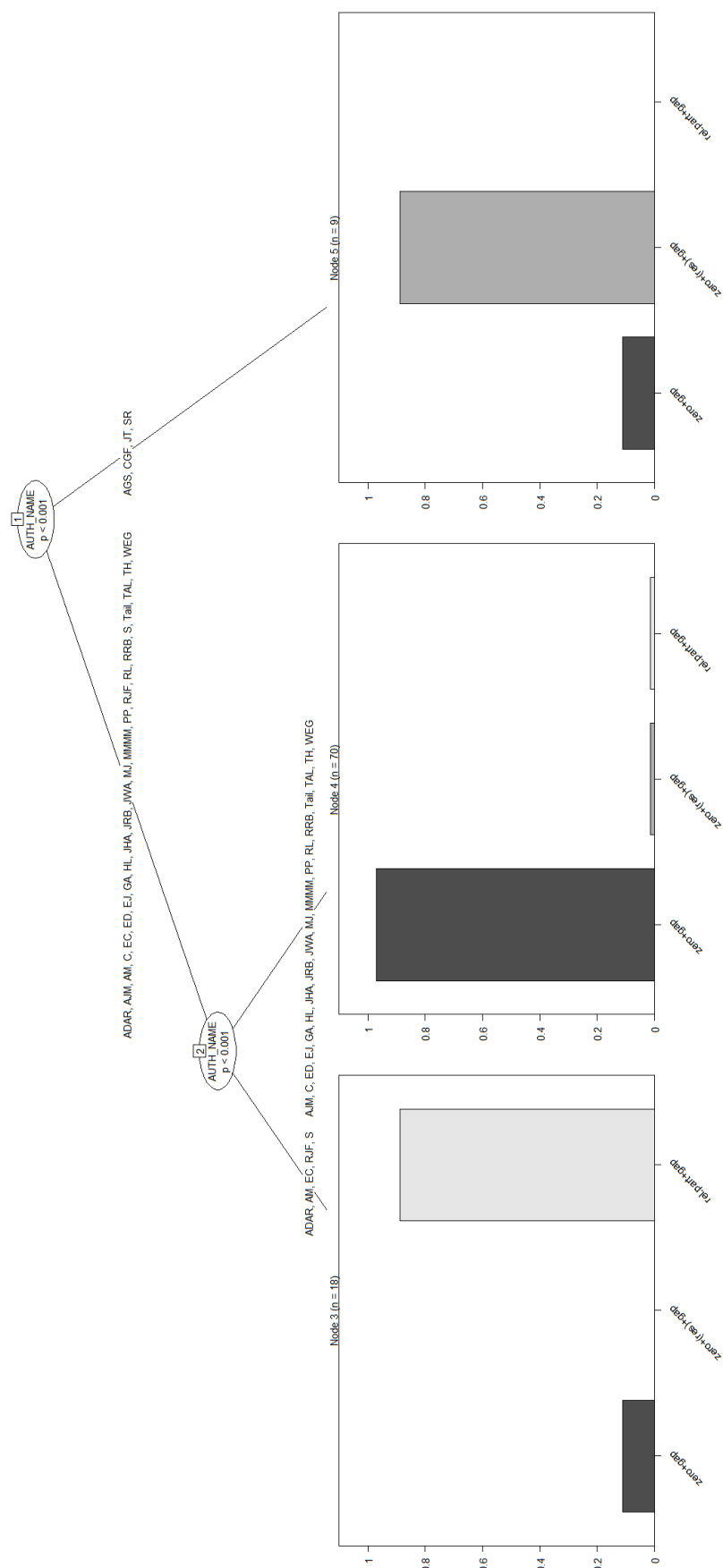


Figure 180: Conditional inference tree for relative clauses in BIS STRATEGY~YEAR_DET+FEATURE+TXT_TYPE_3+AUTH_NAME

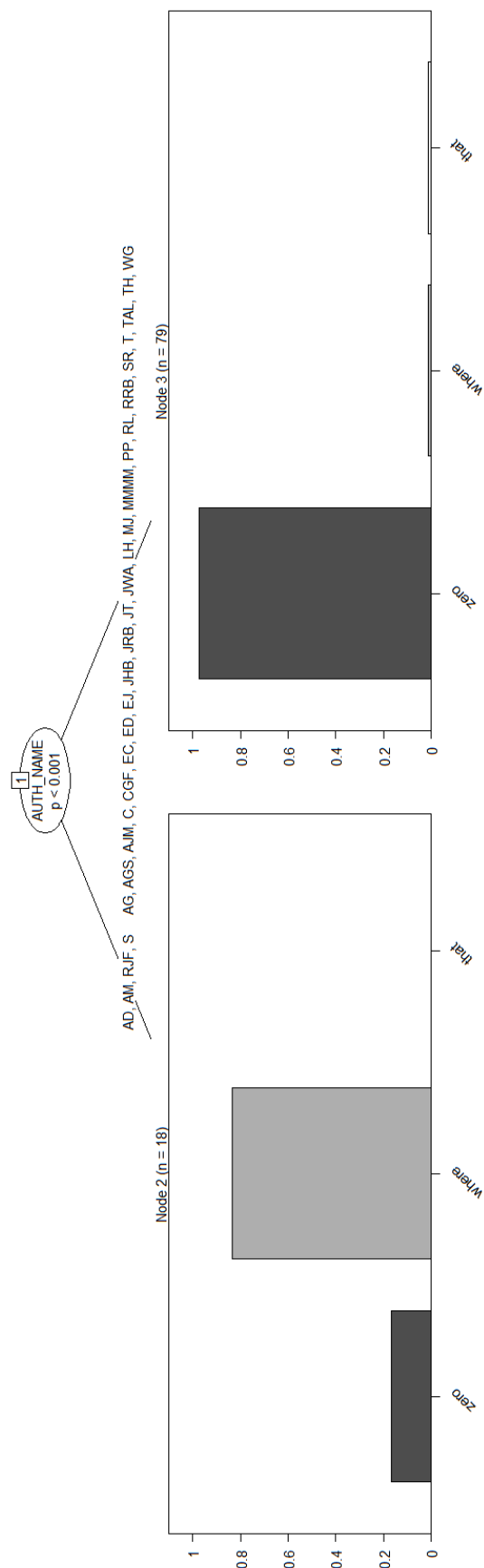


Figure 181: Conditional inference tree for relative clauses in BIS ST_FORM~YEAR_DET+FEATURE+TXT_TYPE_3+AUTH_NAME

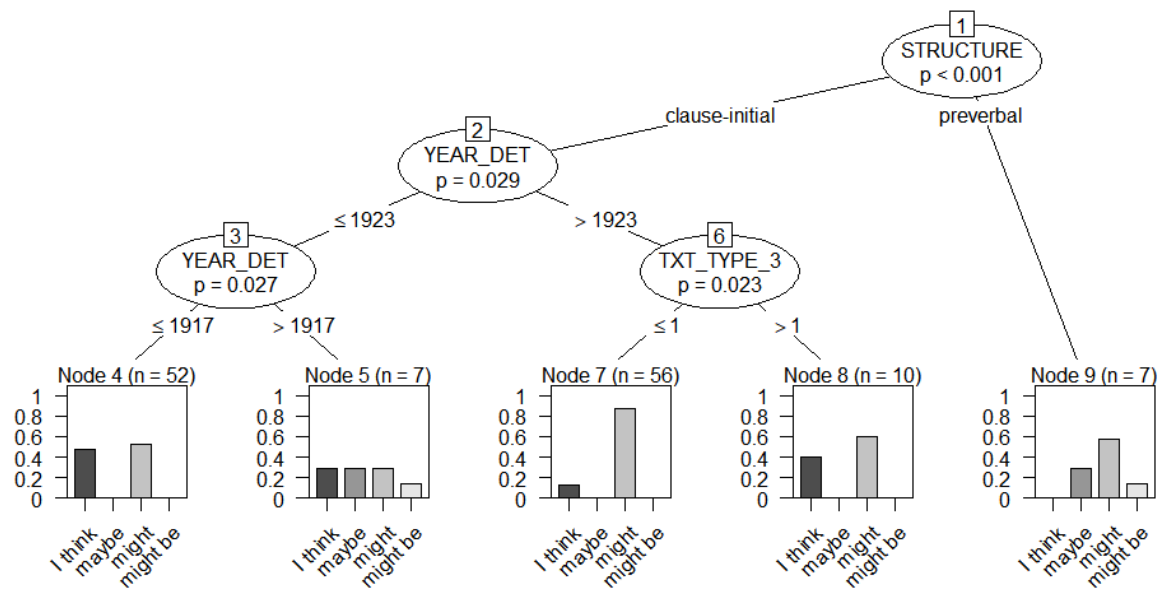


Figure 182: Conditional inference tree for speculative in SIP $ST\_FORM \sim YEAR\_DET + STRUCTURE + TXT\_TYPE\_3$

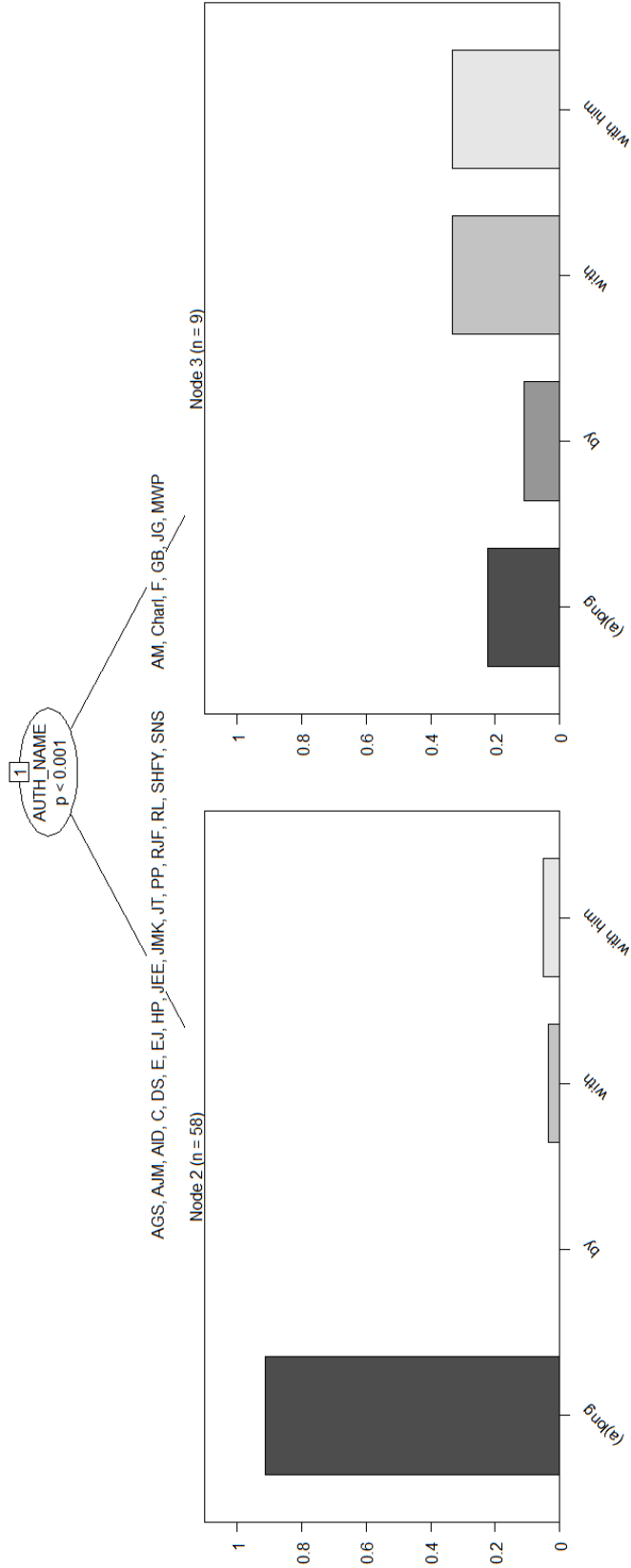


Figure 183: Conditional inference tree for comitative & instrumental prepositions in $BIS_{ST\_FORM} \sim YEAR\_DET + FEATURE + TXT\_TYPE\_3 + AUTH\_NAME$

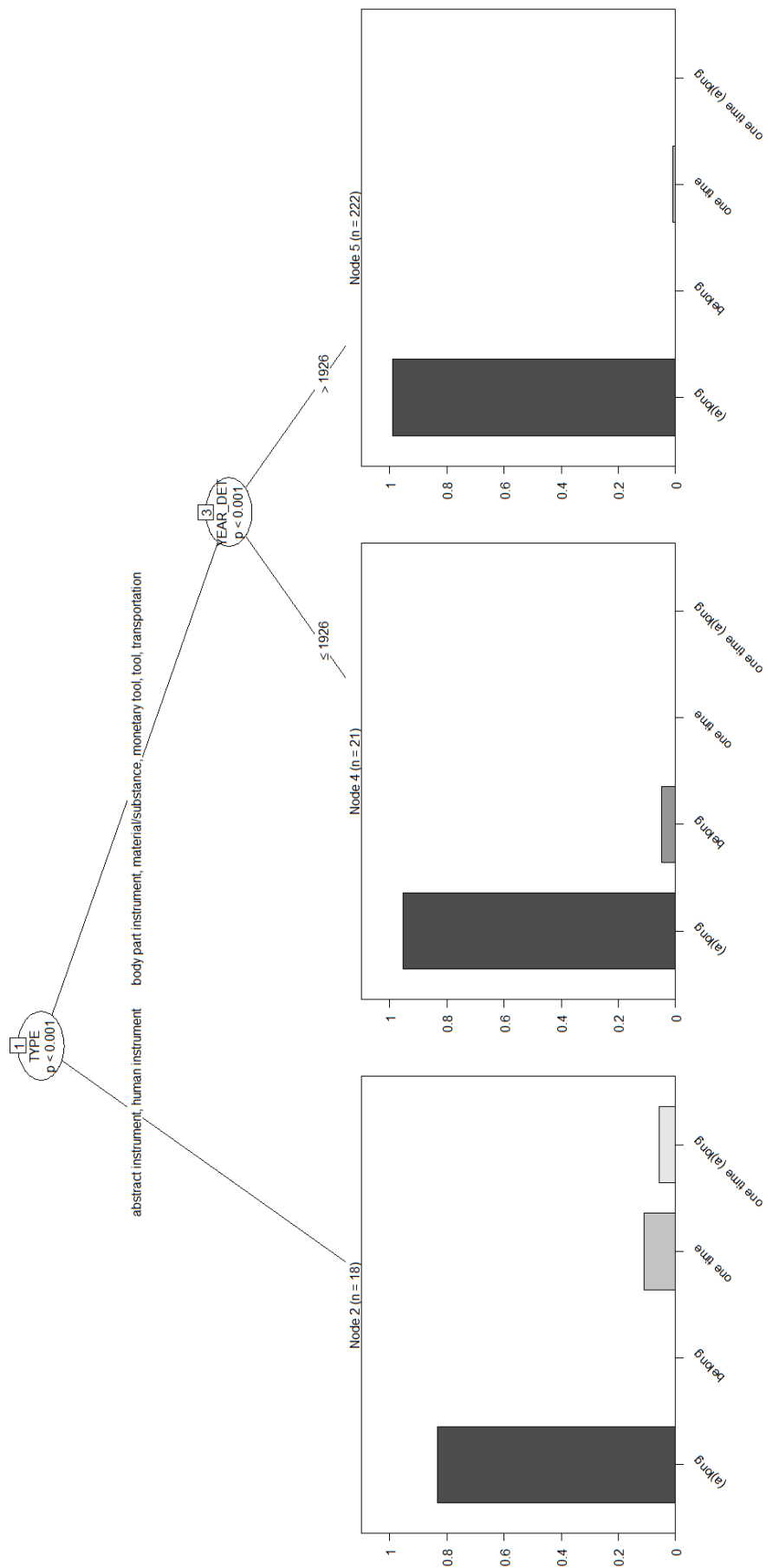


Figure 184: Conditional inference tree for instrumental prepositions in TP ST_FORM~YEAR_DET+TYPE

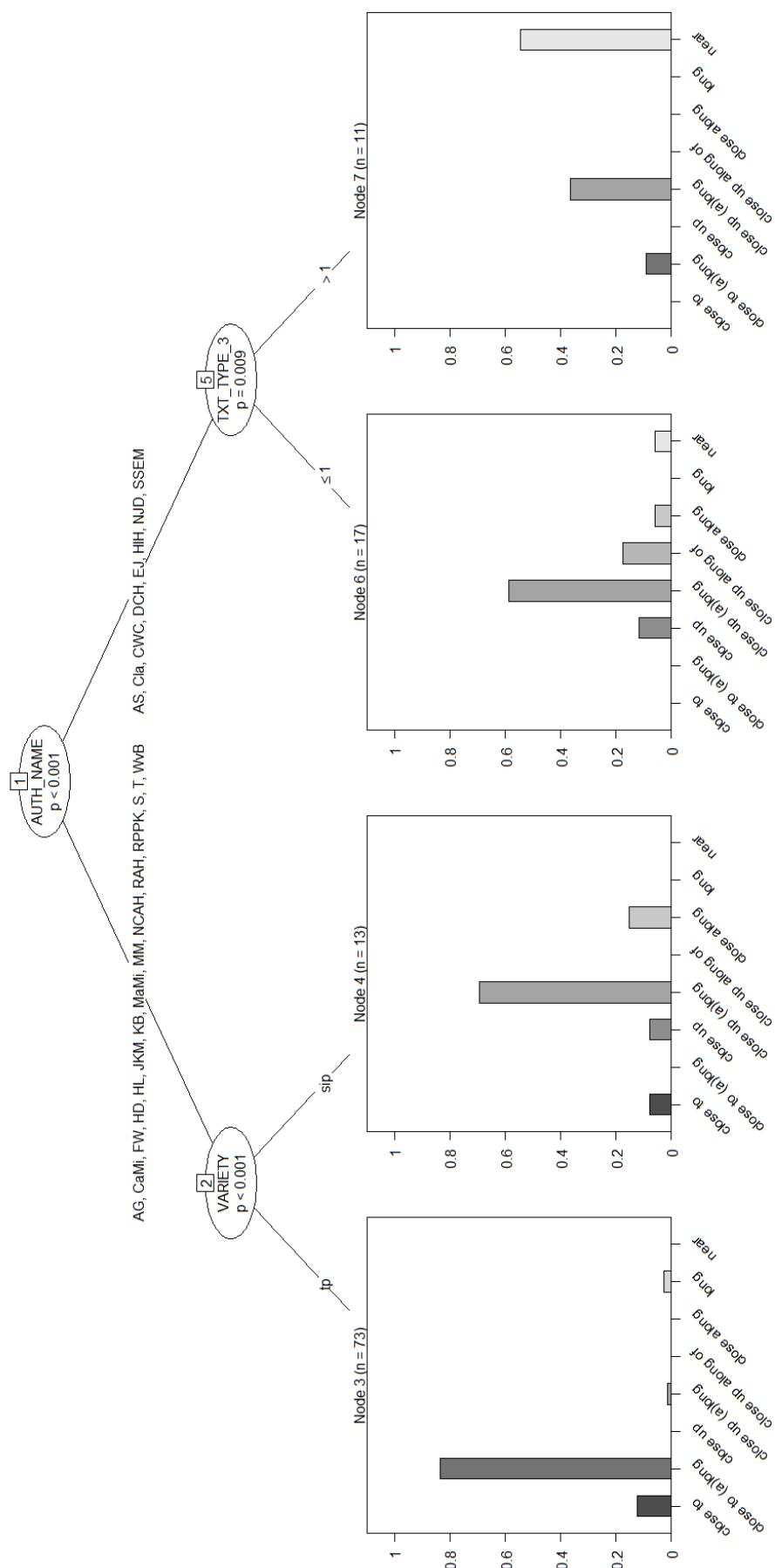


Figure 185: Conditional inference tree for adessive prepositions in the three MPE varieties $ST\_FORM \sim YEAR\_DET + VARIETY + TXT\_TYPE\_3 + AUTH\_NAME$

Appendix III: German summary

KAPITEL 1: THEMATISCHE EINFÜHRUNG UND ÜBERBLICK

Der melanesische Sprachraum zeichnet sich durch eine enorme Sprachenvielfalt aus. Allein auf Papua Neu Guinea gibt es mehr als 840 gesprochene Sprachen, während auf Vanuatu 110 Sprachen und auf den Salomonen Inseln 73 Sprachen gezählt werden (Ethnologue 2017; Glottolog 2021). Um sich in einem solch sprachvielfältigen Gebiet dennoch erfolgreich verständigen zu können, bedienen sich die Einwohner der Kontaktsprachen Tok Pisin (auf Papua Neu Guinea), Pijin (auf den Salomonen) und Bislama (auf Vanuatu), deren Entstehung auf das 19. Jahrhundert zurückzuführen ist.

Die drei genannten Kontaktsprachen stellen Varietäten des melanesischen Pidgin-Englisch dar, welche sich durch wechselseitige Verständlichkeit auszeichnen und deshalb von einigen Linguisten (vgl. z.B. Crowley 1991; Jourdan & Keesing 1997; Tyron & Charpentier 2004; Siegel 2008) auch als Dialekte des melanesischen Pidgin-Englisch bezeichnet werden. Zugleich besitzen die Varietäten jedoch individuelle phonologlogische, grammatikalische und lexikalische Merkmale, die in der Interkommunikation zwischen Sprechern unterschiedlicher Varietäten zu Fehlinterpretationen und Missverständnissen führen können.

Auch wenn erste Kontakte zwischen Europäern und Pazifik-Insulanern schon weitaus früher stattfanden, gilt das 19. Jahrhundert als ausschlaggebend für den wachsenden Kontakt und die zunehmende Kommunikation zwischen den Bevölkerungsgruppen. Während die Zusammentreffen anfänglich durch den Walfischfang, Sandelholz- und Trepang Handel sporadisch waren und zur Entwicklung von Handels-Jargons führten, wurde der Kontakt und Sprach austausch zwischen Melanesiern und Europäern, aber auch zwischen Melanesiern untereinander ab 1860, durch die Entstehung der ersten Plantagen in Queensland, Samoa, Neukaledonien und Fidji intensiviert. Um dem hohen Arbeiterbedarf auf den Plantagen gerecht zu werden, wurden Einheimische von den umliegenden pazifischen Inseln – unter anderem auch von den Salomonen Inseln, Vanuatu und Papua Neu Guinea – rekrutiert. Die Arbeitszeit war durch Arbeiterverträge auf eine Dauer von drei bis fünf Jahren begrenzt. Die auf den Plantagen aufeinandertreffenden Melanesier verfügten über diverse linguistische Hintergründe. Da eine Verständigung untereinander, sowie auch mit den europäischen Aufsehern unabdingbar war, entwickelten sich aus dieser intensiven Kontaktsituation heraus verschiedene Plantagen-Kontaktsprachen, wie beispielsweise das Queensland Plantagen Pidgin-Englisch (QPPE) und das Samoanische Plantagen Pidgin-Englisch (SPPE). Mit Ablauf der Arbeitsverträge wurden die Arbeiter zurück auf ihre Herkunftsinseln befördert und brachten die erlernte Pidgin-Varietät in ihre Heimat. Dies führt zur Annahme, dass die Varietäten, die um 1900 auf Papua Neu Guinea,

den Salomonen und Vanuatu importiert wurden, ein hohes Maß an Ähnlichkeit aufzeigten. Zugleich ist es jedoch auch möglich, dass sich das frühe QPPE vom SPPE unterschied. Wenn an einen Ort besonders viele Arbeiter aus Queensland und an einen anderen Ort viele Arbeiter aus Samoa zurückkehrten, dann könnten sich schon frühzeitig verschiedene Varietäten des Melanesischen Pidgin-Englisch herausgebildet haben. Zusätzlich ist es auch möglich, dass sich die Varietäten, die auf den Plantagen in Queensland und Samoa gesprochen wurden, unterschieden, es bei der Rückkehr der Arbeiter jedoch zu einer Durchmischung der Varietäten (= Levelling) und einer damit einhergehenden Homogenisierung kam.

Zu Beginn des 20. Jahrhunderts wurde der pazifische Arbeiterhandel beendet, wodurch der stetige Austausch und Kontakt der Arbeiter nicht länger stattfinden konnte. Die Beendigung des Arbeiterhandels und die daraus resultierende Isolierung der drei im Fokus stehenden Gebiete, begünstigte die individuelle Weiterentwicklung der Varietäten.

Unter Forschern (vgl. Mühlhäusler 1978; Clark 1979; Jourdan 1985; Keesing 1988; Baker 1993; Jourdan & Keesing 1997; Sankoff 2021+) herrscht infolgedessen weitestgehend Übereinstimmung, dass die Beendigung des Arbeiterhandels ausschlaggebend für die Spaltung der melanesischen Kontaktsprache in ihre Sub-Varietäten war. Bislang wurde diese Annahme jedoch ausschließlich mit extralinguistischen Ereignissen begründet – eine Vergleichsstudie basierend auf frühem Sprachmaterial aller drei Varietäten, die zur Klärung und Offenlegung des Sachverhalts unabdingbar ist, blieb bislang aus. Des Weiteren gibt es bislang keine Studien, die qualitative Methoden mit quantitativ-statistischen Methoden verknüpfen und diese auf das frühe Datenmaterial anwenden.

Auf Grundlage dieser Gegebenheiten war die Kernforschungsfrage der vorliegenden Dissertation, *wann* und *wie* sich die Varietäten Tok Pisin (TP), Bislama (BIS) und Solomon Islands Pijin (SIP) aus dem Melanesischen Pidgin-Englisch (MPE) in ihren individuellen Formen herausentwickelt haben und welche sprachstrukturellen Unterscheidungen im diachronen Verlauf vorhanden sind. Die Forschungsfrage wurde in vier Subfragen untergliedert:

Forschungsfrage #1:	Wie wurden die Sprachmerkmale in SIP, BIS und TP im Zeitverlauf realisiert?
Forschungsfrage #2:	Hat sich Variation hinsichtlich der untersuchten Sprachmerkmale in den einzelnen Varietäten, aber auch zwischen den Varietäten, verringert, stabilisiert oder vermehrt?
Forschungsfrage #3:	Basierend auf den Zeiträumen in denen Veränderungen sichtbar werden, welche externen Einflüsse könnten einen Einfluss auf die individuelle Entwicklung der Varietäten gehabt haben?
Forschungsfrage #4:	Was sagen die Ergebnisse über Prozesse und Mechanismen während der Auswahl, Entwicklung und Stabilisierung von Sprachmerkmalen im Allgemeinen aus?

Zur Analyse des historischen Sprachmaterials, welches im Rahmen der Studie in diversen Archiven in Deutschland, Australien und Neuseeland gesammelt wurde, wurden sowohl qualitative als auch quantitativ-statistische Methoden herangezogen.

Im ersten Kapitel der Dissertation erfolgt nach einer kurzen Einführung in die Entwicklungsgeschichte der melanesischen Kontaktsprachen ein Überblick über den derzeitigen Forschungsstand. Zudem wird ein kurzer Überblick über generelle Theorien zur Entstehung von Pidgin- und Kreolsprachen gegeben, da der Fokus der Arbeit zwar auf den melanesischen Kontakt-Varietäten und deren Entwicklung liegt, die Arbeit jedoch zugleich auch allgemeinere Theorien zur Entwicklung von Kontaktsprachen ergänzt.

Dies ist wichtig, da – nachzeitigem Forschungsstand – kontroverse Theorien zur Entstehung von Pidgin- und Kreolsprachen vorherrschen. McWhorter (2002) behauptet, dass Kreolsprachen eine eigene – synchron definierbare – typologische Klasse darstellen (= *Creole Exceptionalism*) und wird unter anderem von Bakker et al. (2011) und Daval-Markussen (2018) in dieser Ansicht unterstützt. Kreolisten wie Chaudenson (1992, 2001) und Mufwene (1996, 2001) sind der Ansicht, dass die Gründerpopulation (= *Founder Principle*) eine bedeutsame Rolle bei der Entwicklung von Kontaktsprachen gespielt hat. Basierend auf dem *Founder Principle* haben Kreolisten wie Aboh (2009), Ansaldo (2009a; 2009b), DeGraff (2014) und Yakpo (2021+) eine Theorie entwickelt, die besagt, dass das Sprachsystem einer Kontaktsprache von der sprachlichen Ökologie einer bestimmten Kontaktsituation abhängig ist (= *evolutionary account of creole formation*). Die Theorie ist zudem eng verknüpft mit Mufwenes Feature-Pool-Hypothese (2001, 2006), welche besagt, dass die sprachliche Ökologie einer Kontaktsituation zu einem einzigartigen Pool konkurrierender Sprachmerkmale führt, der die sprachlichen Systeme aller an der Kontaktsituation beteiligten Personen enthält. Da Sprecher in der Kontaktsituation interagieren, tauschen sie Äußerungen aus (vgl. Ansaldo 2009a: 275) und können auf den Feature-Pool zurückgreifen (vgl. auch Croft 2000). Wieder andere Theorien verweisen auf die Rolle der Mehrsprachigkeit (vgl. Muysken 2008; Ansaldo 2009b, 2010; Yakpo & Muysken 2017) oder auf Spracherwerbsprozesse (Arends 1989, 1992) und deren Einfluss auf die Kontaktsprachenentwicklung.

Die Anzahl der Erklärungsversuche ist vielfältig. Die vorliegende Arbeit liefert vertiefende historische Daten, um die divergierenden Theorien zu bewerten und die Prozesse bei der Bildung von Kontaktsprachen zu beleuchten. Am Ende des 1. Kapitels folgen generelle Anmerkungen zu den in der Arbeit verwendeten Terminologien, Beispielen, Zitaten und Sprachen.

KAPITEL 2: ERFORSCHUNG DER ENTSTEHUNG VON PIDGIN- UND KREOLSPRACHEN AUS EINER DIACHRONEN PERSPEKTIVE

In Kapitel 2 der vorliegenden Dissertation wird auf die Herausforderungen und Besonderheiten verwiesen, die bei der historischen Erforschung von Pidgin- und Kreolsprachen zu berücksichtigen sind. Ähnlich wie bei historischen Studien im Allgemeinen, auf die in Kapitel 2.1 eingegangen wird, stellt auch in der historischen Kreolistik die von Labov (1999) als *Bad Data Problem* bezeichnete Thematik eine Herausforderung dar. Historische Forscher können keine Daten passend zu ihrer Forschungsfrage sammeln, sondern müssen auf das überlieferte Datenmaterial zurückgreifen, welches meist fragmentiert und schwer zugänglich ist. Auch ist ein Großteil früher Daten in schriftlicher Form überliefert; frühe gesprochene Daten stellen eine Ausnahme dar. Dies bedeutet jedoch nicht, dass aus schriftlichen Überlieferungen keine Kenntnisse zu gesprochener Sprache gewonnen werden können, da die traditionelle Unterscheidung zwischen gesprochener und geschriebener Sprache längst überkommen ist (vgl. Koch & Oesterreicher 1985, 1994; Elspass 2012; Nevalainen & Raumolin-Brunberg 2012). Nichtsdestotrotz ist zu beachten, dass bei der Verschriftlichung der gesprochenen Sprache eine Vielzahl an Filtern die ursprünglichen Sprachäußerungen beeinflusst haben könnte, sodass es Aufgabe des Forschers/der Forscherin ist, diese Filter zu entfernen (vgl. Schneider 2013).

Koloniallinguistische Forschungen sind nochmals verstärkt mit diesen Problematiken konfrontiert, da die frühen Kontaktsprachen zunächst primär gesprochen und meist erst mit Beginn missionarischer Aktivitäten auch in schriftlicher Form als Ausdrucksmittel verwendet wurden. Für Kontaktsprachen, die ihren Ursprung in der vorkolonialen bzw. kolonialen Zeit haben und in Gesellschaften entstanden sind, in denen Lese- und Schreibkundigkeit wenig verbreitet war, müssen Sprachdaten unter anderem in Reise-, Regierungs- und Missionsberichten sowie in Gerichtsdokumenten und frühen Tagebüchern gesammelt werden. Die Quellen variieren in der Menge an enthaltenem Datenmaterial von keinen Belegen oder einzelnen Wörtern bis hin zu Sätzen oder Paragraphen in Pidgin-Englisch. Der Forscher/die Forscherin weiß jedoch nie vorab ob und in welcher Menge Sprachdaten in einer Quelle enthalten sein werden. Die meisten dieser Quellen wurden von Kolonisatoren in ihrer Heimatsprache verfasst, sodass das Pidgin-Englisch nur gelegentlich zitiert oder erwähnt wurde. Genau dies stellt eine Problematik bei der Rekonstruktion der frühen Entwicklungsstufen von Kontaktsprachen dar, da das überlieferte Datenmaterial meist von Europäern oder aus Superstrat-Perspektive niedergeschrieben wurde. Im Idealfall sollten jedoch Sprachbelege aller beteiligten Gruppen gesammelt werden, um eine einseitige, europazentrierte Analyse zu vermeiden (vgl. McWhorter 2000 und Roberts 2005).

Schriftliche Daten, wie z.B. Gerichtsverfahren und Briefe, welche längere Passagen in Pidgin-Englisch enthalten oder gar ganz in Pidgin-Englisch verfasst wurden, stellen eine Rarität dar. Die Kontaktsprachen dienten primär als mündliches Verständigungsmittel und fanden erst zu einem späteren Zeitpunkt auch in schriftlichen Dokumenten Verwendung. Letzteres hat jedoch den Vorteil, dass auch im Falle von Belegen schriftlicher Natur davon ausgegangen werden kann, dass es sich um die Dokumentation tatsächlich gesprochener Sprache handelt. Zugleich müssen jedoch andere Filter berücksichtigt werden, um Aussagen über die Verlässlichkeit der Quellen treffen zu können. In Kapitel 2.2.2 werden deshalb verschiedene Einflussfaktoren vorgestellt, die bei der Analyse von frühen Pidgin- und Kreolsprachen herangezogen werden sollten und welche sich in allgemeine und autorenbasierte Faktoren untergliedern lassen:

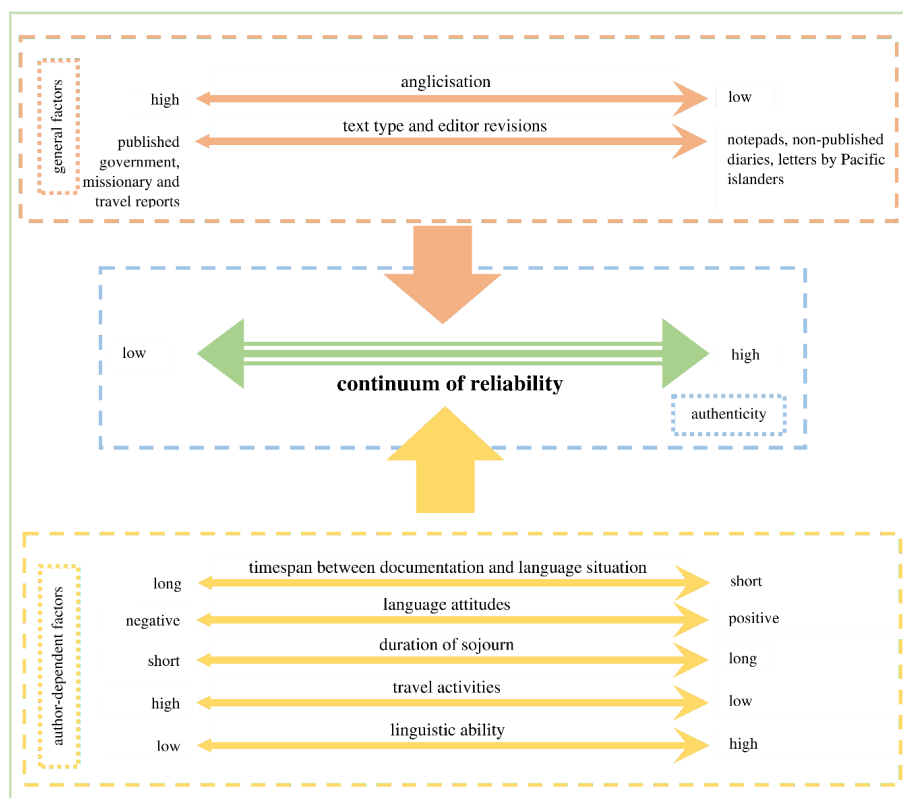


Abbildung 1. Kontinuum der Verlässlichkeit und Einflussfaktoren

Zu den autorenbasierten Faktoren zählen die Zeitspanne zwischen der Dokumentation und der tatsächlichen Sprachsituation, Sprachattitüden, die Dauer des Aufenthaltes und Reiseaktivitäten, sowie die linguistischen Fähigkeiten des Autors. Zu den allgemeineren Faktoren hingegen zählen editorische Überarbeitungen und eine generelle Anglisierung der Daten. Die einzelnen Aspekte lassen sich jeweils als Kontinuum darstellen. Ebenso kann auch die Verlässlichkeit der Daten als Kontinuum verstanden werden, welches verschiedene Grade der Zuverlässigkeit aufweist. Der einzige Einflussfaktor, der sich nicht als Kontinuum darstellen lässt, ist Authentizität, womit die

Originalität der frühen Daten gemeint ist, d.h. inwiefern die Sprachbeispiele tatsächlich vom Autor stammen oder gegebenenfalls von anderen Quellen kopiert beziehungsweise plagiiert wurden.

Die einzelnen Einflussfaktoren werden in Kapitel 2 näher erläutert und in Kapitel 4 erneut aufgegriffen, da sie eine wichtige Grundlage des methodologischen Rahmenmodells der vorliegenden Arbeit darstellen.

KAPITEL 3: URSPRUNG UND ENTWICKLUNG DES MELANESISCHEN PIDGIN-ENGLISCH UND DESSEN GEOGRAPHISCHE VARIATION

Sprachentwicklung kann niemals losgelöst von den sozialen Gegebenheiten analysiert werden (vgl. Soukoup 2017: 674). Insbesondere bei Kontaktsprachen tragen soziolinguistische Faktoren, wie beispielsweise die Interaktion und das Verhältnis zwischen Europäern und Nicht-Europäern, sowie zwischen den Einwohnern verschiedener Pazifikinseln dazu bei, die Entwicklung von Kontaktsprachen zu verstehen. Das dritte Kapitel fokussiert sich deshalb im ersten Teil auf die soziolinguistische Geschichte des Sprachkontaktes im melanesischen Sprachraum. Die darin geschilderten soziolinguistischen und historischen Hintergründe basieren primär auf der Analyse und Auswertung von vorkolonialen und kolonialen Quellen, die im Rahmen der Studie gesammelt wurden. Diese wurden durch neuere historische Literatur zur Geschichte Melanesiens ergänzt.

Die soziolinguistischen Hintergründe aus diachroner Perspektive zu beleuchten ist zudem notwendig, um die verschiedenen Theorien zu verstehen, die von Kreolisten hervorgebracht wurden, um die Entwicklung und Entstehung der drei Varietäten zu erklären. Wie zu Beginn erwähnt, entstanden ab 1860 die ersten Plantagen in Queensland, Samoa, Neukaledonien und Fidji. Die auf den Plantagen aufeinandertreffenden Melanesier hatten unterschiedliche Muttersprachen, sodass sich in den intensiven Kontaktsituationen verschiedene Plantagen-Kontaktsprachen (wie beispielsweise das QPPE und das SPPE) entwickelten. Mit Ablauf der Arbeitsverträge kehrten die Arbeiter auf ihre Herkunftsinseln zurück und brachten so die erlernte Pidginvarietät in ihre Heimat. Belege hierfür werden unter anderem in Kapitel 3.1.3.6 angeführt.

Während zu Beginn Einwohner von den Salomonen, Vanuatu und Papua Neu Guinea für alle vier Gebiete rekrutiert wurden, änderte sich dies im Jahre 1886, zwei Jahre nachdem der nordöstliche Teil von Neuguinea und der Bismarck Archipel von den Deutschen annektiert wurden. Der deutsche Beamte Gustav von Oertzen verkündete, dass fortan nur noch deutsche Plantagen ihre Arbeiter in Deutsch-Neuguinea rekrutieren durften. Da ein Teil Samoas von den Deutschen besetzt war, wurden Arbeiter aus Deutsch-Neuguinea weiterhin auf die Plantagen in

Samoa gebracht, während die Rekrutierung für die Plantagen in Queensland, Neukaledonien und Fidji beendet wurde. Auch die Jahre 1906, 1914, 1911 und 1929 sind von zentraler Bedeutung, da zu diesen Zeitpunkten der Arbeiterhandel in Queensland, Samoa, Fidji und Neukaledonien vollständig beendet wurde und somit auch die Interaktion zwischen Einheimischen aus den verschiedenen Regionen zu Ende ging. Weitere wichtige sozio-linguistische und sozio-historische Ereignisse, die möglicherweise einen Einfluss auf die individuelle Sprachentwicklung der drei Varietäten hatten und die deshalb im Kapitel 3.1 näher erläutert werden, sind beispielsweise die Etablierung von Plantagen in den Heimatregionen, die Aktivität und Sprachwahl von Missionen sowie der erneute Sprachkontakt mit Amerikanern und Australiern während des Zweiten Weltkrieges.

Basierend auf den geschichtlichen Gegebenheiten hat sich unter Kreolisten eine kontroverse Diskussion zur Frage entwickelt, inwiefern die Plantagen, im speziellen jene in Queensland und auf Samoa, die melanesischen Pidginvarietäten und ihre individuelle Entwicklung beeinflusst haben. Während Mühlhäusler (1978, 1985) die Behauptung aufstellt, dass der individuelle Charakter des Tok Pisin daher rühre, dass sich das Tok Pisin aus dem Pidgin-Englisch der samoanischen Plantagen herausentwickelt habe, und Bislama und Salomon Islands Pijin ihren Ursprung im Queensland Plantagen Pidgin-Englisch haben, sieht Baker (1993) den Ursprung aller drei Varietäten auf den Plantagen in Queensland und dem dort gesprochenen Pidgin-Englisch. Clark (1979) kommt – ähnlich wie Mühlhäusler – zu dem Entschluss, dass sich das Tok Pisin bereits um 1880 von den anderen Varietäten entfernt habe, während die individuelle Entwicklung von Bislama und Solomon Islands Pijin auf das Ende des Arbeiterhandels zurückgehe. Im Kontrast hierzu steht die Theorie von Keesing (1988), der den Ursprung der drei Varietäten in einem sich über den gesamten Pazifik ausdehnenden Seemannsjargon sieht. Im Gegensatz zu Mühlhäusler zweifelt er an, dass sich das Pidgin der Plantagen Samoas sowie jenes, welches auf den Plantagen in Queensland gesprochen wurde, unterschieden. Stattdessen führt er die individuelle Entwicklung des TP darauf zurück, dass in Deutsch-Neuguinea die deutsche sowie austronesische Sprachen als Substratsprachen die Entwicklung des Pidgin beeinflussten. Einen weiteren Ansatz bietet Drechsel (2014), der aufzeigt, dass das *Maritime Polynesische Pidgin* einen Einfluss auf die Entwicklung des melanesischen Pidgin-Englisch gehabt haben könnte.

Trotz der unterschiedlichen Ansichten in den obigen aufgeführten Theorien, herrscht in der Forschung weitestgehend Übereinstimmung, dass die Beendigung des Arbeiterhandels ausschlaggebend für die Spaltung der melanesischen Kontaktsprache in ihre Sub-Varietäten war (vgl. Mühlhäusler 1978; Clark 1979; Keesing 1988; Baker 1993; Jourdan & Keesing 1997;

Sankoff 2021+) – eine Annahme, die unter Berücksichtigung der extralinguistischen Ereignisse sinnvoll ist, bislang jedoch noch nicht basierend auf historischem Sprachmaterial geprüft wurde. Um diese Forschungslücke zu schließen und sich nicht von den extralinguistischen Ereignissen leiten zu lassen, war das Ziel der Arbeit, datenbasiert zu ermitteln, wann sich Unterschiede in den Varietäten bei der Kodierung ausgewählter Sprachmerkmale entwickelt haben und diese erst in einem zweiten Schritt mit extralinguistischen Faktoren in Verbindung zu bringen.

KAPITEL 4: EMPIRISCHE GRUNDLAGEN: DATEN UND METHODOLOGIE

Kapitel 4 gibt einen Überblick über die Datensammlung und Datenbasis der vorliegenden Untersuchung und erläutert die Methodologie, die zur Analyse der Daten herangezogen wurde. Auch wenn einige Kreolisten bereits Mitte der 1980er Jahre begonnen haben, historische Sprachbelege von Pidgin- und Kreolsprachen zu sammeln, liegen diese Sammlungen größtenteils verstreut bei einzelnen Forschern und Forscherinnen vor (vgl. Huber & Velupillai 2018: 133). Korpora oder Datenbanken, in denen frühe Sprachbelege von Tok Pisin, Bislama und Solomon Islands Pijin zur Verfügung stehen, gibt es bislang nicht. Infolgedessen wurde das Datenmaterial für die vorliegende Dissertation eigenständig in verschiedenen Archiven gesammelt:

Archives/Data Sources	SIP	BIS	TP
Deutsche Kolonialbibliothek	3	2	32
Online Archives	45	10	36
Philip Baker	0	38	0
Baker/Online Archives	0	47	0
National Library of Australia/Australian War Memorial	0	0	116
Pacific Research Archives, Canberra	23	1	11
National Library of Australia, Canberra	4	2	6
Western Pacific Archives, Auckland	102	70	0
Pacific Manuscript Bureau	324	46	12
Others (e.g. general libraries, Leibniz-Institut für Deutsche Sprache)	20	8	26
Total	521	224	239

Tabelle 1. Anzahl der Quellen pro Archiv und pro Varietät

Bei der Durchsicht der Archivdaten wurde Datenmaterial basierend auf den folgenden drei Aspekten exzerpiert: 1. Tatsächliche Sprachbeispiele in Pidgin-Englisch, 2. metalinguistische Aussagen über die Ähnlichkeit der Varietäten, und 3. sprachpolitische und extralinguistische Informationen, die für die Sprachentwicklung von Bedeutung sein könnten.

Der Zeitraum, für den die tatsächlichen Sprachbeispiele gesammelt wurden, deckt die frühesten auffindbaren Belege bis zum Jahr 1950 ab. Auch wenn Kreolisten das Ende des Arbeiterhandels im Pazifik als ausschlaggebend für die kulturelle Lokalisierung der drei Kontaktsprachen ansehen und somit unterstellen, dass die individuelle Sprachentwicklung

zwischen 1900 und 1930 eingetreten ist, umfasst das Datenmaterial eine größere Zeitspanne, da individuelle Entwicklungen, die bereits vor 1900 eingesetzt haben könnten und Entwicklungen, die sich durch einen Timelag erst später in den Daten erkennen lassen, ansonsten unentdeckt blieben. Die Sprachanalyse auf frühes Datenmaterial zu basieren, welches die Entstehung der Pidgin-Varietäten bis 1950 abdeckt, erlaubte es, sich nicht von den externen Ereignissen leiten zu lassen, sondern Sprachanalyse-basierend herauszufinden, wann die individuelle Entwicklung eingetreten ist und diese erst in einem zweiten Schritt mit externen Ereignissen in Verbindung zu setzen.

Die in der vorliegenden Studie tatsächlich verwendeten Sprachbeispiele stammen aus 984 verschiedenen Quellen. Bei den Quellen handelt es sich unter anderem um Reiseberichte, Regierungsberichte, Missionsberichte, Briefe und Gerichtsverhandlungen. Bei der Extraktion der Sprachdaten wurden zugleich auch – insofern möglich – sozio-biographische Sprecherinformationen (d.h. z.B. Gender, Alter, Herkunft) für das jeweilige Sprachbeispiel notiert. Gleichmaßen war es auch von Relevanz *kontextuelle* Informationen darüber zu sammeln, *wann* und *wo* die Sätze ursprünglich geäußert wurden. Die genaue Herangehensweise wird in Kapitel 4.1.2 näher erläutert. Des Weiteren wird im Kapitel 4 auch der Umgang mit Sprachbeispielen thematisiert, die kein ‚reines‘ Pidgin-Englisch darstellen, sondern bei denen direkt ersichtlich war, dass es zu einer Sprachmischung, beispielsweise durch editorische Überformung, gekommen ist.

Die methodologische Herangehensweise wird in Kapitel 4.3 näher erläutert, wobei zunächst auf die Auswahl der linguistischen Merkmale eingegangen wird. Es wurde davon abgesehen, Sprachmerkmale basierend auf gegenwärtigen Sprachstrukturen der drei Varietäten auszuwählen, da diese Unterschiede auch erst nach 1930 eingetreten und in anderen extralinguistischen Parametern begründet sein könnten. Die Auswahl der linguistischen Merkmale erfolgte datenbasiert. Um potentielle Unterschiede zwischen den frühen Formen der Varietäten festzustellen, wurden die frühen Daten zunächst auf Grundlage der *Leipzig Glossing Rules* (<http://www.eva.mpg.de/lingua/resources/glossing-rules.php>, letzter Zugriff 12 September 2021) analysiert. Entsprechend wurde jedes Morphem separat analysiert. Basierend auf den Ergebnissen der interlinearen Morphemanalyse konnten einige Unterschiede zwischen den Varietäten identifiziert werden, von denen – basierend auf drei Kriterien – schlussendlich *vier* Feature zur weiteren Analyse ausgewählt wurden. Die Merkmale sollten 1. unterschiedlich in den Varietäten kodiert sein, 2. mehrfach attestiert sein und 3. noch nicht vergleichend analysiert worden sein. Bei den ausgewählten Merkmalen, die in der Arbeit untersucht wurden, handelt es sich um adnominale und pronominale Demonstrativpronomen, Subjekt- und Objekt-Relativsätze,

Modalitätsmarker (Fähigkeit, Wollen/Wunsch, Erlaubnis und Vermutung) und ausgewählte Präpositionen (instrumental, komitativ, terminativ und adessiv).

Nach der Auswahl der Sprachmerkmale wurden die Daten kodiert, wobei zwischen sprachbasierter, quellenbasierter, kontextbasierter, autorenbasierter und sprecherbasierter Kodierung unterschieden wurde:

	Kodierung	Erklärung
sprachbasierte Kodierung	ID	Die ID ist eine individuelle numerische Identifikationsnummer für jedes Token.
	VARIETY	In der Spalte VARIETY wurden die Tokens eindeutig einer Varietät zugeordnet (<i>sip</i> , <i>bis</i> und <i>tp</i>).
	FEATURE	In der Spalte FEATURE wurde die grammatische Kategorie des Tokens aufgelistet (<i>prep_term</i> , <i>prep_inst</i> , etc.)
	TOKEN	In der Spalte TOKEN wurde das Morphem in seiner Originalform notiert, d.h. so wie es in der Ursprungsquelle erschien.
	ST_FORM	In der Spalte ST_FORM wurde das Morphem in standardisierter englischer Schreibweise gelistet, was für die Analyse in R notwendig war.
	STRUCTURE	In der Spalte STRUCTURE wurde der grammatikalische Kontext des Morphems analysiert.
	TRANSLATION	In der Spalte TRANSLATION befindet sich die englische Übersetzung des Tokens.
	GLOSS	In der Spalte GLOSS wurde die Gloss-Abkürzung der Morphemanalyse notiert.
quellenbasierte Kodierung	SENTENCE	In der Spalte SENTENCE wurde das Morphem in seinem Kontext, d.h. im kompletten Satz angezeigt.
	SOURCE	Um die Sprachbeispiele eindeutig zuzuordnen wurde in der Spalte SOURCE die Quelle angegeben, aus der das Sprachbeispiel entnommen wurde.
	TXT_TYPE	In der Spalte TXT_TYPE, wurde der Texttyp der Quelle bestimmt. Es wurde differenziert zwischen gesprochenen-sprachähnlichen Belegen, schriftlichen Belegen und intermediären Belegen.
	TXT_TITLE	In der Spalte TXT_TITLE wurde der Titel der Quelle, aus der das Sprachbeispiel extrahiert wurde, notiert.
kontextbasierte Kodierung	YEAR_PUBL	In der Spalte YEAR_PUBL ist das Publikationsdatum der Quelle notiert, aus der das Sprachbeispiel extrahiert wurde.
	YEAR_ATT	Die Spalte YEAR_ATT zeigt den Zeitpunkt an, auf den sich ein Sprachbeispiel bezieht und kann ein konkretes Datum, ein Jahr oder einen Zeitraum darstellen.
	YEAR_DET	Die Spalte YEAR_DET zeigt auch den Zeitpunkt an, auf den sich ein Sprachbeispiel bezieht. Da die Programmiersprache R jedoch nur genaue Daten verarbeiten kann, wurde für Zeiträume der mittlere Zeitpunkt gewählt (z.B. YEAR_ATT: 1894-1900 vs. YEAR_DET: 1896)
	LOCATION	Die Variable LOCATION gibt den Ort an, an dem ein Sprachbeispiel geäußert wurde (Plantage, Gericht, Hafen, etc.)
	PLACE	Die Variable PLACE gibt den Namen des Dorfes oder der Stadt an, an dem das Sprachbeispiel geäußert wurde.
	REGION/ISLAND	Die Variable REGION/ISLAND kennzeichnet den Namen der Insel/des Festlands, wo das Sprachbeispiel geäußert wurde.
	PROVINCE	Die Variable PROVINCE gibt die Provinz an, der eine Insel etc. zugeordnet ist.
	COORDINATE_1 COORDINATE_2	Die Variablen COORDINATE_1 und COORDINATE_2 geben die Koordinaten des Ortes an, wo das Sprachmerkmal attestiert wurde.
autoren- basierte Kodierung	AUTH_NAME	Die Spalte AUTH_NAME listet den Namen des Autors der Quelle auf.
	AUTH_ORIGIN	Die Spalte AUTH_ORIGIN zeigt die Herkunft des Autors.
	AUTH_ROLE	Die Spalte AUTH_ROLE identifiziert die Rolle des Autors (Missionar, Regierungsbeamter, Arzt, etc.)
sprecher- basierte Kodierung	SP_NAME	Die Spalte SP_NAME zeigt den Namen des Sprechers.
	SP_GENDER	Die Spalte SP_GENDER gibt das Geschlecht des Sprechers an.
	SP_ROLE	Die Spalte SP_ROLE gibt die Rolle des Sprechers an.
	SP_ORIGIN	Die Spalte SP_ORIGIN zeigt die Herkunft des Sprechers an.
	SP_AGE	Die Spalte SP_AGE zeigt das Alter des Sprechers.

Tabelle 2. Kodierungsvariablen mit Erläuterung

Ziel der vorliegenden Dissertation ist es, die Chronologie des Sprachwandels in TP, BIS, und SIP zu analysieren. Aus diesem Grund standen die Variablen ST_FORM (= Standard Form), das heißt, das Sprachmerkmal in standardisierter englischer Sprachweise, sowie YEAR_ATT und YEAR_DET, und somit der Zeitpunkt, auf den das Sprachbeispiel datiert war, im Vordergrund der quantitativen Analyse. Dies liegt darin begründet, dass Erstbelege in der historischen Kreolistik als Indikatoren dienen, um vergangene Sprachstufen zu rekonstruieren. Zugleich muss bei der Analyse von Erstbelegen immer berücksichtigt werden, dass diese von provisorischer Natur sind. Nicht nur die genaue Datierung der Sprachbelege stellt eine Herausforderung dar (vgl. Kapitel 4.1.2.1), es muss auch bedacht werden, dass diese Belege, die aus schriftlichen Quellen gewonnen wurden, lediglich Information darüber geben, wann Formen erstmals in geschriebenen Quellen vorkamen. Es kann also von einem *Timelag* zwischen der Innovation des Sprachmerkmals in gesprochener und geschriebener Sprache ausgegangen werden. Weitere Faktoren, die einen Einfluss auf die Größe des Timelags haben können, wie beispielsweise die Quantität und Qualität der Daten, die Gebräuchlichkeit des Sprachmerkmals in der Sprache, sowie die Verbreitung und Verwendung des Merkmals in verschiedenen Bevölkerungsgruppen (vgl. Abbildung 2), werden in Kapitel 4.3.4.1 näher thematisiert.

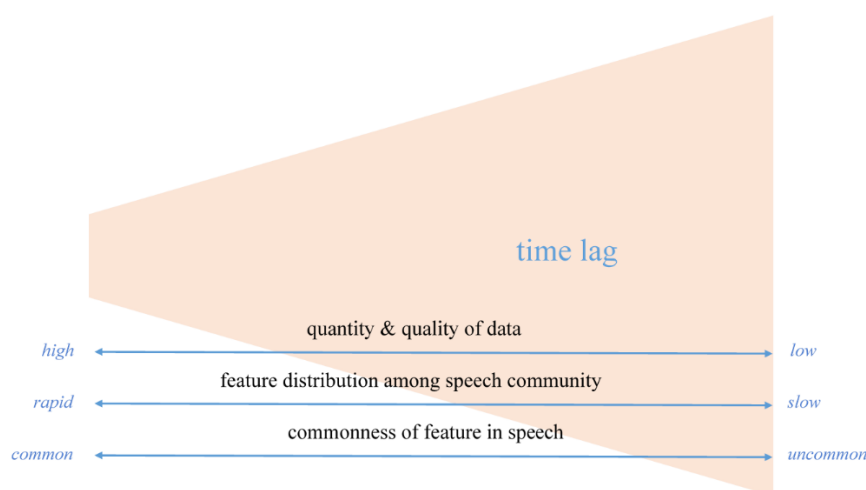


Abbildung 2. Einflussfaktoren auf die Größe des Timelags

Zur Veranschaulichung der Sprachmerkmale in ihrem Zeitverlauf wurden zwei unterschiedliche Darstellungsweisen gewählt. Zunächst wurde basierend auf den Variablen YEAR_ATT und ST_FORM ein Zeitstrahl mit Hilfe von Excel erstellt, um aufzuzeigen, ob in einem gewissen Jahr (YEAR_ATT) eine bestimmte Form (ST_FORM) in den Daten attestiert wurde oder nicht. Diese Herangehensweise ermöglichte es, einen ersten Eindruck über die Jahre bzw. Zeiträume zu gewinnen, in denen die einzelnen Varianten der untersuchten Variablen attestiert wurden. Die Häufigkeit, mit der eine bestimmte Variante in einem Jahr vorkommt, ebenso wie die

Datenverteilung, wurde bei dieser Darstellung jedoch nicht berücksichtigt, weshalb in einem zweiten Schritt Boxplots mit Hilfe der Programmiersprache R (R Core Team 2020) in RStudio (RStudio 2019) generiert wurden. Diese visualisieren die Verteilung der verschiedenen Varianten des untersuchten Sprachmerkmals (ST_FORM) im Verlauf der Zeit (YEAR_DET).

Da sowohl der mit Excel erstellte Zeitstrahl als auch die Boxplot-Darstellung keinen Aufschluss über die statistische Signifikanz von Veränderungen im Zeitverlauf geben, wurden die Daten im Anschluss mit Hilfe von *Conditional inference trees* (*Ctrees*) analysiert, welche Teil des *party* und *partykit* Pakets in R sind (Hothorn, Hornik, & Zeileis 2006). Auch wenn es in der historischen Linguistik ein weitverbreitetes Phänomen ist, mit vordefinierten Zeitabschnitten oder Jahr-für-Jahr-Analysen zu arbeiten, wurde dies in der vorliegenden Studie abgelehnt. Gries & Hilpert (2012: 136) haben aufgezeigt, dass die Anzahl und Länge der vordefinierten Zeitabschnitte einen großen Einfluss auf die Ergebnisse haben kann. Auch die Jahr-für-Jahr-Analyse ist als wenig sinnvoll zu bewerten, da für die berücksichtigten Variablen in den frühen Quellen nur eine geringe Anzahl an Belegen pro Jahr vorhanden ist. Um datenbasiert zu ermitteln, zu welchen Zeitpunkten signifikante Veränderungen ersichtlich werden, wurde deshalb der Algorithmus der *Ctrees* verwendet, welcher aus rekursiven binären Entscheidungen besteht, die sich als Baum darstellen lassen. Jeder Knoten (Split), der im Baum aufgezeigt wird, repräsentiert ein signifikantes Ergebnis.

Da sich grammatische Formen zu unterschiedlichen Zeitpunkten stabilisiert und grammatikalisiert haben könnten, wurden die ausgewählten Sprachmerkmale anhand von Fallstudien in vier separaten Kapiteln analysiert. Des Weiteren wurden die Sprachmerkmale varietätenspezifisch, non-komparativ analysiert, da auch die Stabilisierung in den verschiedenen Varietäten unterschiedlich datiert sein könnte. Am Ende eines jeden Kapitels wurden schließlich auch vergleichende Studien durchgeführt.

Auch wenn nicht alle Faktoren, die in Tabelle 2 aufgeführt sind, in der quantitativen Analyse berücksichtigt werden konnten, waren diese Informationen dennoch notwendig, um Rückschlüsse über die Verlässlichkeit der Daten ziehen zu können. Beispielsweise diene der Vergleich des Publikationsdatums und des Datums des Sprachbeispiels dazu, den Zeitraum zwischen Beleg und Dokumentation zu bestimmen. Je geringer dieser Zeitraum, desto verlässlicher erschien die Quelle. Die Rolle des Autors gab Rückschluss darüber, inwiefern der Autor gegebenenfalls politisch voreingenommen war und ob dies die Darstellung des Sprachbeispiels beeinflusst haben könnte. Weitere Strategien zur Prüfung der Verlässlichkeit der Daten werden in Kapitel 4.3.5 aufgeführt.

KAPITEL 5: META-LINGUISTISCHE ÄUßERUNGEN ÜBER DIE INDIVIDUALISIERUNG DER MELANESISCHEN VARIETÄTEN

Wie zuvor erwähnt wurden neben dem Sprachmaterial auch metalinguistische Belege gesammelt, die Aufschluss über die Ähnlichkeit bzw. Verschiedenheit der Varietäten im Verlauf der Jahre geben sollten. Die metalinguistischen Äußerungen, deren Analyse in Kapitel 5 präsentiert und evaluiert wird, zeigt, dass das Pidgin-Englisch im 19. Jahrhundert von den frühen Beobachtern als *eine* Varietät ohne regionale Unterschiede wahrgenommen wurde. Die frühen Beschreibungen heben hervor, dass es sich um eine allgemeingültige Händlersprache handelte, die überall verwendet wurde (vgl. z.B. Bridge 1886: 547; Mackellar 1912: 105). Auch nach 1900 wird das Pidgin-Englisch weiterhin als *Lingua Franca* beschrieben, welche zur Interkommunikation verwendet wurde und die überall *gleich* gewesen sei, unabhängig davon, ob man sich auf britischen, französischen oder anderen Besitzungen befand (vgl. z.B. Rivers 1914: 466-467; Alexander 1927: 213).

Der früheste Beleg in den mir vorliegenden Daten, der darauf hinweist, dass es regionale Unterschiede gab, ist auf das Jahr 1914 zurückzuführen (vgl. Jacomb 1914: 91). Ab 1926 wird in den metalinguistischen Äußerungen immer häufiger Bezug auf regionale Unterschiede genommen (vgl. z.B. Collinson 1926; Hogbin 1939; Reed 1943). Dabei ist auffällig, dass die Beschreibungen sich stets auf lexikalische – nicht auf morphosyntaktische – Unterschiede beziehen und die Varietäten bis Ende 1950 als gegenseitig verständlich charakterisiert werden. Dies bedeutet jedoch nicht, dass es keine morphosyntaktischen Unterschiede gab. Die metalinguistischen Äußerungen, die in Kapitel 5 untersucht werden, wurden von Europäern getätigt, die oft eine koloniale Haltung hatten und in der Regel keine SprachwissenschaftlerInnen waren. Es ist deshalb möglich, dass die europäischen BeobachterInnen – beispielsweise aufgrund von Übersetzungsfallen – in den Varietäten stärkere Ähnlichkeiten sahen, als tatsächlich vorlagen. Infolgedessen kann nicht davon ausgegangen werden, dass keine Variation vorlag. Nur dann, wenn in frühen Beobachtungen der Kontaktsprachen auf Variation hingewiesen wurde, gibt dies eindeutig Aufschluss über die Individualisierung der Varietäten.

KAPITEL 6: DEMONSTRATIVE

Kapitel 6 ist das erste von vier Fallstudien-Kapiteln. Jedes der Kapitel gibt einen kurzen theoretischen Hintergrund zum Sprachmerkmal, bevor auf die Realisierung und Kodierung in den heutigen Varietäten eingegangen wird. Des Weiteren wird jeweils die methodologische Vorgehensweise kurz erläutert, bevor dann die Ergebnisse aus qualitativer und quantitativer Perspektive präsentiert und diskutiert werden.

Kapitel 6 fokussiert sich auf die Formen, die zur Kodierung von Demonstrativen verwendet wurden. Insgesamt wurden 1.803 Demonstrativ-Tokens analysiert, wovon 1.401 Tokens (57,74%) SIP, 586 Tokens (32,50%) TP und 176 (9,76%) Tokens BIS zugeordnet werden konnten. Insgesamt wurden 15 verschiedene Demonstrativform-Varianten in den Daten identifiziert. Neben Formen, die dem Standardenglischen ähnlich sind, wie beispielsweise *this*, *that*, *these* und *those*, wurde auch die Form *them* verwendet, welche ihren Ursprung in nicht-standard- und regionalen Varietäten des Englischen haben könnte. Des Weiteren wurden u.a. *these fellow*, *them fellow*, *this fellow*, *that fellow*, *here* und *him here* attestiert, wobei die letztgenannten vier Varianten von besonderem Interesse sind, da es sich hierbei um die Formen handelt, die in den modernen Varietäten am häufigsten vorkommen. Es wurde zudem aufgezeigt, dass die Attestierung von standardenglisch-ähnlichen Formen nicht bedeutet, dass die Formen auch wie im Standardenglischen verwendet wurden. So gibt es z.B. Belege, die zeigen, dass *th*-Stopping verbreitet war, sodass die Formen *this* und *that* viel eher als *dis* und *dat* realisiert wurden (vgl. Cheesman 1933: 76; Methodist Mission 1935: 5). Die Formen fanden außerdem Verwendung in Kontexten, in denen der Plural im Standardenglischen benötigt wird.

Die Analyse der Daten hat gezeigt, dass es sich bei *this* mit 51,78% um die in den SIP Daten am häufigsten vorkommende Realisierung handelt, gefolgt von *this fellow* (22%) und *that* (17,2%). Für *that fellow* hingegen konnten lediglich zwölf Vorkommnisse in den Daten ermittelt werden (1,15%). Auch wenn die große Anzahl an *this* und *that* auf den ersten Blick auffällig wirkte und zunächst eine Anglisierung der Daten vermuten ließ, machte eine qualitative Analyse der Kollokationen deutlich, dass das häufige Vorkommen der Formen in den frühen Daten nicht abwegig ist, da auch im heutigen SIP *distaim/detaem* und Wörter wie *disaelan/desaelan* (dt. 'Insel') und *diskaen/deskaen* (dt. 'die Art von') verwendet werden. Dies belegt, dass es einen Zeitraum gegeben haben muss, in denen die Formen *this* und *that* verwendet wurden, sodass es zur Grammatikalisierung von *distaim/detaem* sowie zu Lexikalisierungen wie beispielsweise *disaelan/desaelan* und *diskaen/deskaen* kommen konnte.

Alle anderen Formen in den frühen SIP Daten kamen mit einer Häufigkeit von weniger als 2,5% vor. *Here* wurde beispielsweise nur in 1,63% der Tokens attestiert. Die Darstellungen anhand des Zeitstrahls sowie der Boxplots machten deutlich, dass fast alle Formen bis zum Ende des betrachteten Zeitraums Verwendung fanden.

Im BIS Datensatz repräsentierte *here* das am häufigsten vorkommende Demonstrativ (34,09%). *This* (22,73%) und *that* (18,75%) waren die zweit- und dritthäufigsten Varianten. Die Formen *this fellow* (6,25%) und *that fellow* (8,52%) hingegen konnten zwar ebenfalls attestiert werden, wurden jedoch vergleichsweise selten gebraucht. Der Zeitstrahl zeigte, dass *this* die

früheste Form in BIS repräsentierte. Sie wurde 1831 erstmals attestiert. In den späten 1860ern kamen die Formen *that*, *that fellow* und *here* hinzu, was darauf schließen lässt, dass die Formen nebeneinander existierten und Verwendung fanden. Der erste Beleg für die Form *this fellow* ist auf das Jahr 1871 datiert und liegt somit nach dem Erstbeleg von *here*. Während konkurrierende Formen bis 1929 vorherrschten, konnten ab 1930 nur noch die Formen *here* und *him here* ermittelt werden. Aufgrund der Datenlücke ab dem Jahr 1930 bis 1950 muss diese Beobachtung für BIS jedoch kritisch hinterfragt werden.

Im TP Datensatz dominierten Tokens, in denen das Demonstrativ mit der Form *this fellow* kodiert wurde (78,33%). Die zweithäufigste Variante repräsentierte *this*, welche in 8,87% der Tokens vorkam. *That* und *that fellow* fanden mit zehn bzw. acht Vorkommnissen vergleichsweise selten Verwendung. Die Form *here* wurde in 5,97% der Tokens identifiziert. Alle anderen Formen traten mit einer Häufigkeit von weniger als 2% auf. Der Zeitstrahl zeigte auf, dass es sich bei *this*, *this fellow* und *that fellow* um die frühesten Formen handelt, welche bereits in den 1880ern verwendet wurden. Dabei konnten *this* und *this fellow* bis Ende des betrachteten Zeitraums nachgewiesen werden. Die Formen *that* und *that fellow* wurden nach 1931 nicht mehr in den Daten attestiert. Nach 1920 kamen neue Formen wie beispielsweise *this one*, *here*, und *him here* hinzu, die auch in den 1940ern noch Verwendung fanden.

Die *Ctree*-Analyse machte deutlich, dass der Autor den stärksten Prädiktor für die Kodierung der Demonstrative in allen drei Varietäten darstellte, wobei jedoch auch der syntaktische Kontext einen Einfluss auf die Kodierung hatte. Wenn der Autor und der syntaktische Kontext außer Betracht gelassen wurden, konnte im Jahre 1899 eine signifikante Veränderung der TP-Daten bestimmt werden. Sprachmaterial, welches vor 1899 attestiert wurde, wurde mit *this*, *this fellow*, *that* und *that fellow* kodiert, während nach 1900 eine deutliche Präferenz von *this fellow* zu erkennen war. In Bislama wurde das Jahr 1892 als signifikanter Zeitpunkt identifiziert, da von da an *here* die wahrscheinlichste Form in pronominalen syntaktischen Kontexten darstellte. Im Kontrast zu BIS und TP ist eine Tendenz zu einem bestimmten Demonstrativ basierend auf der Variable *Zeit* in den SIP-Daten nicht erkennbar. Um die Entwicklung und Stabilisierung demonstrativer Partikel in den drei Varietäten vollständig zu rekonstruieren, sollten zukünftig Studien auch Daten berücksichtigen, die die Zeit nach 1950 mit einschließen.

Insbesondere die Vorkommnisse der Form *here* erschienen interessant. In allen drei Varietäten wurde die Form zunächst als lokatives Adverb verwendet. In BIS konnte die Form bereits 1869 als Demonstrativum attestiert werden. In SIP hingegen wurde die Form erstmals im Jahr 1895 und in TP im Jahr 1924 als Demonstrativ verwendet. Interessanterweise wurde die

Form *here* auf den Plantagen in Samoa und Queensland nur als demonstratives Adverb verwendet, sodass davon ausgegangen werden kann, dass die Verwendung von *here* als adnominales und pronominales Demonstrativ ihren Ursprung in Vanuatu hat. Die Grammatikalisierung von *here* könnte durch den Kontakt zum Französischen intensiviert worden sein, da das Französische die lokativen Adverbien *ici* (dt. 'hier') und *là* (dt. 'dort') verwendet, welche als postnominale Elemente Teil der Demonstrative sind. Es ist also denkbar, dass französische Sprachstrukturen einen Einfluss auf die Entwicklung und Stabilisierung dieses Markers hatten. Zugleich könnte die postnominale Form *here* in Bislama auch durch dessen Substratsprachen beeinflusst worden sein, da diese ebenfalls postnominale Demonstrative verwenden (Camden 1979: 76; Siegel 2008: 183-184).

KAPITEL 7: FALLSTUDIE RELATIVSÄTZE

Kapitel 7 untersucht die Entwicklung von Strategien zur Relativsatzbildung in SIP, BIS und TP. Basierend auf den Unterschieden, die während der interlinearen Morphemanalyse ersichtlich wurden, wurden zwei unterschiedliche Ansätze für die Analyse verwendet. So wurden nicht nur die morphosyntaktischen Strategien zur Relativsatzbildung untersucht (STRATEGY), sondern auch die konkreten Formen (ST_FORM), die zur Bildung der Relativsätze Verwendung fanden. Die Daten wurden auch dahingehend kodiert, ob der Relativsatz eine Subjektfunktion oder eine Objektfunktion hat, um festzustellen, ob es gegebenenfalls Unterschiede in der Relativform gibt.

Bisherige Studien zur Relativsatzbildung in den drei Varietäten, auf die in Kapitel 7.3 näher eingegangen wird, weisen noch einige Fragen auf, insbesondere bezüglich der Verbreitung des Partikels *where*. Während Mühlhäusler (1997: 174) annimmt, dass der Partikel nicht im frühen TP verwendet wurde und erst kürzlich den Weg in die Varietät fand, behauptet Siegel (1981: 30) dass der Partikel damals in allen drei Regionen verwendet wurde, in denen MPE gesprochen wurde. Auch Crowley (1990a: 330) unterstützt letztere Hypothese. Weder Siegel noch Crowley führen jedoch frühe Daten an, um diese Annahme zu belegen. Baker (1993) hingegen vermutet, dass der Partikel erst jüngst seinen Weg in das SIP fand.

Insgesamt konnten 815 Relativsätze aus den frühen Quellen extrahiert werden, wobei 493 (53,87%) in TP, 279 in SIP (34,23%), und 97 in BIS (11,90%) vorkamen. Bei der Analyse wurde ersichtlich, dass fünf unterschiedliche **Strategien** zur Relativsatzbildung in den diachronen Daten Verwendung fanden. Eine Möglichkeit der Relativsatzbildung war die Einleitung des Relativsatzes ohne *overtes* Element (= Nullmarkierung), wobei ein resumptives Pronomen optional folgen konnte (*zero+res* vs. *zero+gap*). Des Weiteren konnten Relativsätze durch

Relativpartikel eingeleitet werden. Ähnlich wie bei Nullmarkierungen konnte auf Relativpartikel optional ein resumptives Pronomen folgen (*rel.part+res* vs. *rel.part+gap*). Zudem gab es Relativsätze, in denen ein Relativpronomen verwendet wurde (*rel.pro*). Mit Fokus auf der **Form** (ST_FORM) konnten sechs verschiedene Varianten festgestellt werden. Neben Nullmarkierungen wurden die Relativpartikel *where* und *that*, sowie die Relativpronomen *who*, *which* und *who's that* attestiert.

Die Analyse zeigte, dass in allen drei Varietäten die *zero+gap*-Strategie am häufigsten und als früheste Strategie Verwendung fand. Während jedoch in den TP-Daten eine eindeutige Präferenz der Strategie zu erkennen war (86,10%) und diese auch im BIS-Datensatz dominierte (73,20%), waren nur 36,20% der Relativsätze in SIP nullmarkiert. SIP wies generell die größte Variation hinsichtlich der Relativsatz-Strategien auf.

Betrachtet man das Vorkommen der Strategien und Formen aus diachroner Perspektive, wird ersichtlich, dass die Nullmarkierung von Relativsätzen nur in TP als dominante Form im Verlauf der Zeit bestehen blieb. In SIP und BIS hingegen wurde 1908 bzw. 1913 erstmals der Relativpartikel *where* attestiert, welcher seinen Ursprung in dem Missionspidgin der *Queensland Kanaka Mission* (QKM) haben und durch die *South Sea Evangelical Mission* (SSEM) in die Regionen verbreitet worden sein könnte.

In den SIP Daten wurde zusätzlich ein hohes Vorkommen der Form *who* festgestellt, welche auch heute noch in der orthographischen Variante *hu* als Relativpronomen verwendet wird. Um aussagen zu können, ob die heutige Realisierung ihren Ursprung in der damaligen Zeit hat, ist es notwendig, Daten zu sammeln, die nach 1950 datiert sind. Die drei Varianten – *where*, *zero*, *who* – die gegen 1950 in SIP vorhanden sind, sind auch jene, die noch heute in der Varietät verwendet werden. Anhand des Ctree-Algorithmus wurden die Jahre 1908, 1933 und 1943 als Zeitpunkte identifiziert, zu denen die Daten signifikante Veränderungen aufwiesen. Allerdings sollte angemerkt werden, dass das Jahr der Attestierung sich nur dann als signifikanter Einflussfaktor erwies, wenn es als alleinige Variable berücksichtigt wurde. Sobald alle potentiellen Einflussvariablen gleichzeitig mit einbezogen wurden, hatte das Jahr der Attestierung keinen signifikanten Einfluss auf die Wahl der Form.

Ähnliche Beobachtungen konnten auch bei der Analyse der BIS Daten gemacht werden. Die BIS-Relativsätze waren vorrangig nullmarkiert. Der Relativpartikel *where* kam am zweithäufigsten vor und ist ab 1913 in den Daten vorzufinden. Die Analyse mit Hilfe von Ctrees identifizierte das Jahr 1914 als signifikant, was sich auch unter Berücksichtigung der potentiellen Einflussvariablen *Texttyp* und *Feature* bestätigte. Sobald jedoch der *Autor* als möglicher Einflussfaktor mit einbezogen wurde, erwies sich dieser als stärkster Prädiktor.

In den TP-Daten konnte kein signifikantes Jahr festgestellt werden, da von Beginn an und über den gesamten Zeitverlauf hinweg die Nullmarkierung die dominante Strategie darstellte.

Bezüglich der Fragestellung, wann die Varietäten sich voneinander entfernten, kann – basierend auf den vorliegenden Daten – somit keine konkrete Aussage getroffen werden. Die Ctree-Analysen zeigten, dass die zeitbasierten Knoten, die signifikante Veränderungen aufzeigen, nur dann in den SIP-Daten erkennbar waren, wenn der *Texttyp* und *Autor* nicht berücksichtigt wurden. In den BIS-Daten war das Jahr 1914 zwar signifikant, jedoch nur dann, wenn der Autor nicht als Einflussfaktor berücksichtigt wurde. In den TP-Daten war eine generelle Präferenz der Nullmarkierung – unabhängig vom Faktor Zeit – erkennbar.

Die Analyse der konkreten Formen zur Kodierung von Relativsätzen lässt jedoch nicht darauf schließen, dass das Ende des Arbeiterhandels den ausschlaggebenden Grund für die Verwendung von divergierenden Formen in den Varietäten darstellte. Auch wenn die Erstbelege des Relativpartikels *where* auf die Jahre 1908 in SIP und 1913 in BIS datiert werden können und diese somit nach dem Ende des Arbeiterhandels liegen, muss berücksichtigt werden, dass eine Erstattestierung in geschriebenen Daten nicht zwangsläufig mit dem ersten Aufkommen in der gesprochenen Sprache einhergeht. Es ist wahrscheinlich, dass *where* schon zuvor in der gesprochenen Sprache verwendet wurde und somit schon vor dem Ende des Arbeiterhandels in Gebrauch war. Ein Beleg hierfür ist, dass der Relativpartikel schon 1898-1899 in einem Brief eines Pazifikinsulaners in Queensland verwendet wurde. Auch wenn die Etablierung der Form durch das Ende des Arbeiterhandels womöglich gefördert wurde, scheint der Ursprung der verschiedenen Formen in den unterschiedlichen Plantagengeschichten und Arbeiterbewegungen zu liegen.

Die Analyse der Relativsatz-Marker führte zudem zu neuen Erkenntnissen hinsichtlich des Forschungsstandes zur Verbreitung des Relativpartikels *where*. Die Daten zeigen, dass der Relativpartikel sowohl in SIP als auch in BIS verwendet wurde, was Bakers (1993) Theorie widerspricht, dass der Partikel erst jüngst seinen Weg ins SIP fand. Des Weiteren lassen die Daten vermuten, dass der Partikel seinen Ursprung in SIP und nicht, wie bislang vermutet, in BIS hatte, und dass der Partikel erst nach 1950 seinen Weg ins TP fand.

KAPITEL 8: FALLSTUDIE MODALITÄTSMARKER

In Kapitel 8 wird über die Ergebnisse der qualitativen und der quantitativ-statistischen diachronen Analyse von Modalitätsmarkern berichtet. Der Fokus liegt hierbei auf Modalitätsmarkern, durch die *Wille*, *Fähigkeit*, *Erlaubnis* und *Spekulation* ausgedrückt werden, da diese in ihrer Realisierung in den frühen Daten Unterschiede aufwiesen. Da behauptet wird, dass es sehr

unwahrscheinlich für Pidginsprachen sei, spezifische Marker zur Kodierung von Modalitäten zu verwenden, war es von besonderem Interesse, zu analysieren, wann die Varietäten verbale Marker zur Kodierung von Modalitäten entwickelt haben, da dies als Indikator für ihre Stabilisierung zu Pidginkreolsprachen gesehen werden kann.

Als Grundlage für die Analyse wurde das Modell von Palmer (2001) verwendet, der zwischen Proportionalität- und Ereignis-Modalität unterscheidet. In Kapitel 8.1 wird das Modell zunächst näher erläutert, bevor in Kapitel 8.2 auf die Kodierung von Modalität in den heutigen Varietäten eingegangen wird. Da es sich hierbei um ein sehr komplexes Thema handelt, welches häufig von Kreolisten vermieden wird (vgl. Winford 2018: 202), existieren kaum Studien, die sich mit der Thematik befassen, geschweige denn einen diachronen Vergleich von Modalitätsmarkern in den drei Varietäten anstreben (vgl. Kapitel 8.3).

Betrachtet man die Ergebnisse der diachronen Analyse von Modalitätsmarkern (vgl. Kapitel 8.4-8.7) wäre es wünschenswert, dass mehr Daten aus der Zeit vor 1950 in den schriftlichen historischen Aufzeichnungen überlebt hätten. Die geringe Menge an verfügbaren Daten sowie die ungleichmäßige Verteilung der Datenpunkte über die Zeit hinweg erschweren es, die Entwicklung von Modalitätsmarkern in den drei MPE-Varietäten zu rekonstruieren. Dennoch konnten einige wichtige Beobachtungen gemacht werden:

Die Ergebnisse der *Willens*-Modalitätsmarker (539 Tokens) zeigten, dass TP – unabhängig von dem Faktor *Zeit* – eine eindeutige Präferenz für die Form *like* aufwies, während *want to*, welches die dominierende Form in den SIP und BIS Daten darstellte, nur sporadisch attestiert wurde. Da auf den Plantagen in Queensland sowohl Formen basierend auf dem Morphem *like*, als auch auf dem Morphem *want* koexistierten, während auf den samoanischen Plantagen lediglich Formen mit *like* vorzufinden waren, liegt die Vermutung nahe, dass die beobachteten Unterschiede zwischen den Varietäten auf die unterschiedlichen Plantagengeschichten zurückgeführt werden können und weniger mit dem Ende des Arbeiterhandels zu begründen sind. Zudem wurde ersichtlich, dass die Stabilisierung der heutigen in SIP und BIS verwendeten Formen bis zum Ende des Betrachtungszeitraums (= 1950) noch nicht abgeschlossen war.

Auch bei der Analyse von *Fähigkeit*-Modalitätsmarkern (474 Tokens) wurde deutlich, dass weiteres Datenmaterial benötigt wird, um deren Entwicklung vollkommen nachvollziehen zu können. Selbst für SIP, welches die höchste Anzahl an Tokens aufwies, waren gegen Ende der 1940er noch keine Präferenzen hinsichtlich der Kodierung erkennbar.

Eine interessante Beobachtung in SIP war, dass in nicht-negierten Kontexten Formen basierend auf dem Morphem *save* geringfügig häufiger vorkamen als solche, die auf dem

Morphem *can* basieren. Letztere waren jedoch in negierten Kontexten dominant. Formen wie *cannot*, *can't*, und *can't do it* erschienen zunächst auffällig, da sie eine große Ähnlichkeit zu standard-englischen Formen aufzeigen. Allerdings fiel beim Vergleich mit den heutigen in SIP gebräuchlichen Formen auf, dass diese, wenn auch weniger häufig, auch heute noch neben *no save* gebraucht werden (vgl. *kanot*, *kan* und *kanduit*).

In den BIS-Daten konnte mit Hilfe der *Ctrees*-Analyse das Jahr 1897 als signifikantes Jahr identifiziert werden, ab welchem für *save*-basierte Formen eine höhere Wahrscheinlichkeit ermittelt werden konnte als für solche, die auf *can* basierten. Sobald weitere Einflussfaktoren berücksichtigt wurden, wie beispielsweise der Texttyp oder der Autor, konnten keine Splits identifiziert werden. *Save*-basierte Formen dominierten folglich unabhängig vom Jahr der Attestierung, dem Texttyp, dem Autor und davon, ob die Form negiert war oder nicht.

In den TP-Daten wurde mit Hilfe des Algorithmus das Jahr 1934 als signifikant identifiziert. Während in und vor 1934 *save*-basierte Formen eine leicht höhere Wahrscheinlichkeit aufzeigten als *can*-basierte Formen, sind nach 1934 *can*-basierte Formen dominant. Zusätzlich wurden um 1940 erstmals Varianten basierend auf dem Morphem *enough* in den TP-Daten attestiert, welches der heute vorherrschende Marker in TP ist.

Die Analyse von Modalitätsmarkern, welche *Erlaubnis* und *Verbote* ausdrücken, zeigte, dass die Menge an gefundenen Tokens (248 Tokens) nicht ausreichte, um Aussagen über deren Entwicklung in den drei Varietäten machen zu können. Auch wenn die Nicht-Attestierung eines Merkmals nicht zwangsläufig bedeutet, dass das Merkmal nicht verwendet wurde, scheint es so, als haben sich *permissive* und *prohibitive* Marker erst relativ spät entwickelt. Sie eignen sich daher nicht, um die Frage zu beantworten, inwiefern das Ende des Arbeiterhandels für die individuelle Entwicklung der Varietäten verantwortlich war. Nichtsdestotrotz ist es auffällig, dass in den frühen TP-Daten ausschließlich *can*-basierte Formen attestiert wurden, während in SIP und BIS zudem auch *save*-basierte Formen vorkamen, wobei letztere auch die heutigen Formen zur Kodierung der permissiven und prohibitiven Modalität darstellen. Ein möglicher Grund, warum sich SIP und BIS in eine ähnliche Richtung entwickelten, könnte darin begründet sein, dass Fähigkeit- und Erlaubnis-Modalität eng miteinander verbunden sind und sich *permissive* und *prohibitive* Formen häufig aus Fähigkeitsmarkern grammatikalisieren.

Die Ergebnisse der Analyse der *spekulativen* Modalitätsmarker (180 Tokens) zeigten, dass – während alle drei Varietäten die Form *I think* zur Kodierung verwendeten – die Form *might* nur in SIP und BIS attestiert werden konnte. Am frühesten wurde *might* in den Salomonen attestiert, nämlich im Pidgin-Englisch eines Rekrutierers, der Arbeiter für die Plantagen in Queensland anwarb, sowie im Pidgin-Englisch eines QKM Konvertits. Auch wenn die Erstbelege

von *might* auf die Salomonen und nicht auf Queensland zurückgehen, ist es dennoch wahrscheinlich, dass *might* sich über Queensland nach Vanuatu ausbreitete, wo der Marker erstmals um 1907 bezeugt wurde. Dies würde auch erklären, warum es keine Vorkommnisse des Markers im TP-Datensatz gibt. Erneut scheint es so, als ob die frühe Abspaltung von TP einen großen Einfluss auf unterschiedliche Formpräferenzen in den Varietäten hatte.

KAPITEL 9: AUSGEWÄHLTE PRÄPOSITIONEN

Das finale Fallstudien-Kapitel konzentriert sich auf die Kodierung von ausgewählten Präpositionen. Auf Grundlage von Hagège (2010) semantischer Klassifizierung von Präpositionen wurden die in den frühen Daten vorkommenden 10.456 präpositionalen Tokens 51 verschiedenen semantischen Funktionen zugeordnet, um Unterschiede in der Kodierung der semantischen Funktionen zwischen den Varietäten aufzudecken. Auch wenn es noch weitere semantische Kategorien gab, in denen Unterschiede erkennbar waren, wurde der Fokus im finalen Analysekapitel auf Präpositionen gelegt, welche in *instrumentalen*, *komitativen*, *terminativen* und *adessiven* semantischen Kontexten verwendet wurden.

Die häufigsten Präpositionen in allen drei Varietäten waren *along* und *belong*. Die beiden Präpositionen stellen auch jene Formen dar, die frühe Reisende und Berichterstatter als typische und einzige Präpositionen beschrieben (vgl. beispielsweise Reed 1943: 281). Mit der Zeit wurden die beiden Präpositionsformen in einigen Kontexten jedoch ersetzt, um semantisch-funktionale Differenzierungen vorzunehmen.

Mit Fokus auf *instrumentale* und *komitative* Präpositionen (743 Tokens) konnte beispielsweise festgestellt werden, dass die Präposition *along* in den frühen TP-Daten zur Kodierung beider Kontexte verwendet wurde. Im Laufe der Zeit wurde (*a*)*long* in komitativen Kontexten durch die um 1930 erstmals attestierte Form *one time* ersetzt. Im frühen BIS und SIP wurden *along* und *with* als dominante Formen zur Kodierung instrumentaler und komitativer Kontexte identifiziert. In BIS und SIP wurde ab 1907 bzw. ab 1930 jedoch die verbale Präposition *with him* in komitativen Kontexten attestiert, welche auch in den heutigen Varietäten verwendet wird. Gegen Ende der 1940er Jahre stellte *with* immer noch die wahrscheinlichste Form in SIP dar, während in BIS die Verwendung von (*a*)*long* und *with him* gleichermaßen wahrscheinlich war.

Auffallend bei der Analyse der komitativen Präpositionsformen war, dass – während sich in TP eine komitative Präposition aus dem Wort *one time* grammatikalisierte – in SIP und BIS die Präposition ihren Ursprung im standardenglischen *with* und dem Transitivmarker *him* hat. Ein

möglicher Erklärungsansatz dafür, dass SIP und BIS die Form zur Kodierung von Komitativem teilen, ist, dass diese fast 30 Jahre länger miteinander verbunden waren als mit TP. Die Form *wetem* (< *with him*) kann heute zusätzlich auch zur Kodierung *instrumentaler* Kontexte in SIP und BIS verwendet werden, was basierend auf den Daten eine Entwicklung darstellt, die erst nach 1950 stattgefunden haben muss.

Die Analyse der *terminativen* Präpositionen (56 Tokens) zeigte, dass sich SIP und BIS auch in Bezug auf die Kodierung der terminativen semantischen Funktion ähneln. In SIP wurde das transitive Verb *catch him* erstmals 1927 zur Kodierung verwendet, während sich die Form in BIS bereits früher grammatikalisiert zu haben scheint. Sie wurde erstmals 1917 attestiert. Obwohl es sich bei beiden Erstbelegen um Daten handelt, die nach dem Ende des Arbeiterhandels liegen, muss wiederholt betont werden, dass die ersten Belege in den schriftlichen Daten nicht bedeuten, dass die Formen nicht schon früher in der gesprochenen Sprache verwendet wurden. Während den terminativen Präpositionen im frühen SIP ein redupliziertes *go* vorausgeht, kommt letzteres in den BIS-Daten nicht vor. Da die Form *catch him* im Queensland Plantagen Pidgin-Englisch als terminatives Verb verwendet wurde, ist es wahrscheinlich, dass die Form als Verb über QPPE ihren Weg ins SIP und BIS fand.

Im Gegensatz zu SIP und BIS wurde die Form *catch him* erst in den 1930er Jahren und auch nur als terminatives Verb mit der Bedeutung ‚erreichen‘ in TP attestiert. Es ist wahrscheinlich, dass die Form sich nicht grammatikalisierte, da *enough long*, welche in den TP Daten zur Kodierung der terminativen semantischen Funktion attestiert wurde, zu diesem Zeitpunkt bereits ihren Grammatikalisierungsprozess begonnen hatte. Unabhängig vom Faktor Zeit repräsentierte *enough long* die wahrscheinlichste Form in den TP-Daten.

Die Varietäten zeigten auch deutliche Unterschiede in der Kodierung von Präpositionen, welche eine *adessive* semantische Funktion haben (114 Tokens). Während sich in SIP *close up along* bis 1898 zur dominanten Variante entwickelte, wurden in den BIS Daten nur zwei adessive Tokens attestiert, die beide mit *close up (along)* kodiert waren und auf 1907-1914 datiert sind. In den TP Daten repräsentierte *close to long*, das erstmals 1917 attestiert wurde, unabhängig von der Zeitvariable die dominante Form.

Erneut zeigten SIP und BIS im Vergleich zu TP eine größere Ähnlichkeit zueinander. Allerdings qualifizierten sich nur die SIP-Daten für die Beantwortung der Frage, wann sich die adessiven Präpositionsformen stabilisierten. Da sich die Form zur Kodierung der adessiven Funktion in SIP bereits vor 1900 grammatikalisierte, kann das Ende des Arbeiterhandels nicht als entscheidender Faktor angesehen werden. Die TP-Daten waren spät datiert und zeigten keine Zeitpunkte mit signifikanten Veränderungen. Die Datenmenge für BIS war zu gering, weshalb

die Analyse der adessiven semantischen Präpositionen keine Erkenntnisse darüber liefert, ob das Ende des Arbeitshandels zur Divergenz der drei MPE-Varietäten geführt hat.

KAPITEL 10: ZUSAMMENFASSUNG UND AUSBLICK

Das letzte Kapitel fasst die Ergebnisse, die aus den einzelnen Fallstudien gewonnen wurden, abschließend zusammen (10.1), um daraufhin zu diskutieren, inwiefern das Ende des Arbeiterhandels tatsächlich als Grund für die individuelle Entwicklung von SIP, BIS und TP betrachtet werden kann (10.2). Dabei wird auch auf weitere Faktoren eingegangen, die einen Einfluss auf die individuellen Entwicklungen der Varietäten hatten und es wird aufgezeigt, welchen Mehrwert die Studie für den Erkenntnisstand der Varietätenforschung und Kreolistik hat (10.3). Abschließend wird auf Problematiken und Herausforderungen der Studie eingegangen, bevor ein Ausblick auf mögliche Anknüpfungspunkte für zukünftige Studien gegeben wird (10.4).

Mit Hilfe der Ctrees konnten zunächst 41 Zeitpunkte identifiziert werden, an denen die Daten signifikante Änderungen aufzeigten (vgl. Abbildung 3). Die Anzahl verringerte sich jedoch, sobald zusätzliche Einflussvariablen, wie z.B. der Texttyp und der Autor mit berücksichtigt wurden, sodass bei Berücksichtigung aller potentiellen Einflüsse nur noch neun signifikante Zeitpunkte verblieben (vgl. Abbildung 4).

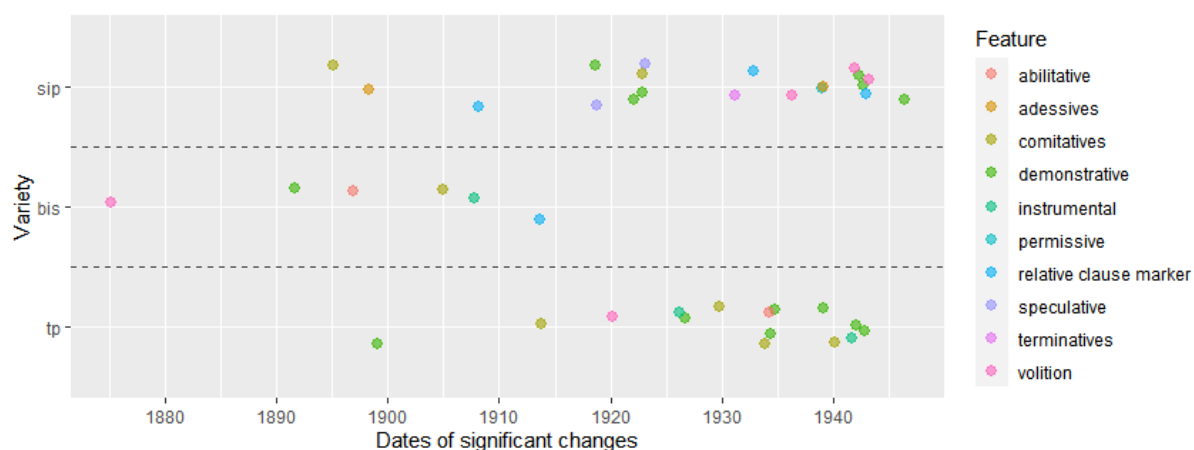


Abbildung 3. Signifikante Zeitpunkte die zu Veränderungen in den Daten führten (basierend auf den Variablen ST_FORM~YEAR_DET)

Die Ergebnisse der vorliegenden Studie haben zudem gezeigt, dass nicht per se generalisiert werden kann, wann die individuelle Entwicklung der drei Varietäten stattgefunden hat. Auch wenn die Ctree-Analyse ergab, dass die meisten signifikanten Veränderungen auf die erste Hälfte des 20. Jahrhunderts datiert werden können (vgl. Abbildung 3 und 4) und dies ein Hinweis darauf sein könnte, dass das Ende des Arbeiterhandels tatsächlich der Auslöser für die Individualisierung der Varietäten war, muss zugleich betont werden, dass die in Ctrees identifizierten Splits in erster

Linie auf signifikante Veränderungen in den Daten hinweisen und somit noch keinen Aufschluss darüber geben, inwiefern diese Veränderungen mit einer Verringerung von Variation und/oder neuen Formen einhergeht.

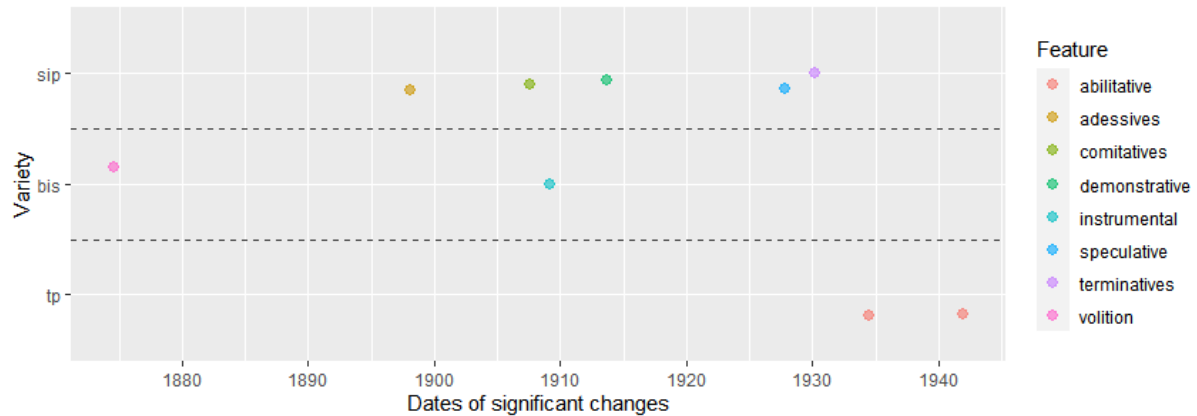


Abbildung 4. Signifikante Zeitpunkte die zu Veränderungen in den Daten führten (basierend auf den Variablen $ST\_FORM \sim YEAR\_DET + TYPE/FEATURE/STRUCTURE + TXT\_TYPE\_3 + AUTH\_NAME$)

Die Ergebnisse zeigen zudem, dass zwischen den Prozessen *Individualisierung* und *Stabilisierung* differenziert werden muss. Auch wenn sich die meisten Formen erst nach Ende des Arbeiterhandels stabilisierten, haben Formen wie *might* (dubitativ Modalität), *with him* (Komitativ), *catch him* (Terminativ), *close up long* (Adessiv), und *where* (Relativsatzpartikel), welche ausschließlich in SIP und BIS und nicht in TP vorkamen, gezeigt, dass die Rekrutierung auf überseeische Plantagen einen großen Einfluss auf divergierende Formentwicklungen hatte. Es ist wahrscheinlich, dass die genannten Formen über die Plantagen in Queensland und das dort gesprochene QPPE ihren Weg in SIP und BIS gefunden haben. Für die Formen *close up long*, *might* und *want* wurden Vorgängerformen im QPPE identifiziert.

Die Analyse hat somit gezeigt, dass Unterschiede zwischen den Varietäten schon vor dem Ende des Arbeiterhandels vorhanden waren und die unterschiedlichen Arbeiter- und Rekrutierungshistorien der Plantagen einen bedeutsamen Einfluss auf die Auseinanderentwicklung der Varietäten hatte – nicht jedoch auf ihre Stabilisierung. Auch wenn in allen drei Varietäten Einflüsse des QPPE erkennbar waren, was Bakers Theorie (1993) unterstützt, untermauert das Ergebnis der Analyse zugleich Mühlhäuslers (1978) Behauptung, dass der individuelle Charakter des TP daher rührt, dass sich TP schon früher von den anderen beiden Varietäten abgespalten hat.

Dies wird auch durch metalinguistische Äußerungen erkennbar, die deutlich machen, dass ehemalige Queensland- und Samoa-Arbeiter ihre Pidgin-Englisch-Kenntnisse in ihre Heimatregionen zurückbrachten. Während Hinweise dafür gefunden werden konnten, dass QPPE

und SPPE in alle drei Regionen überführt wurden, gibt es nur wenige Belege dafür, dass SPPE auf die Salomonen und Vanuatu zurückgebracht wurde, ebenso wie es nur wenig Belege dafür gibt, dass QPPE nach Deutsch-Neuguinea gebracht wurde.

Trotz der früh vorherrschenden Unterschiede in den Daten, zeigte die Auswertung der metalinguistischen Äußerungen über Ähnlichkeiten und Unterschiede in den Varietäten, dass die frühen Reisenden und Schreiber vor 1906 nicht über regionale Differenzen berichteten. Auch wenn sich dies nach 1906 änderte, beschränkte sich die Dokumentation auf lexikalische Unterschiede. Es muss jedoch berücksichtigt werden, dass die Mehrheit der Reisenden und Schreiber keine Linguisten waren und somit divergierende Formen gegebenenfalls nicht erkannt wurden.

Die Analyseergebnisse zeigen zudem, dass Missionen und ihre Pidgin-Englisch-Varietäten einen größeren Einfluss auf die Entwicklung von Kontaktsprachen und deren Auseinanderentwicklung gehabt haben könnten, als bislang vermutet. Auch wenn das Pidgin-Englisch, welches von der SSEM verwendet wurde, ein hohes Maß an standardenglisch-ähnlichen Formen aufwies und manchmal auch als ‘*Simple English*’ (Mühlhäusler & Mühlhäusler 2005) bezeichnet wurde, stellte es die Varietät dar, mit der die Missionsbesucher konfrontiert wurden, welche ihnen gelehrt wurde und mit der sie sich verständigten und die somit einen Einfluss auf das SIP hatte. Ein konkretes Beispiel repräsentiert der Relativpartikel *where*, der – basierend auf den analysierten Belegen – über die QKM/SSEM den Weg in das SIP und BIS fand. Auch Formen wie z.B. *kan* und *kannot*, die noch heute in SIP verwendet werden, kamen häufig in den frühen Quellen der SSEM vor, was dafür spricht, dass die Mission einen starken Einfluss auf die Entwicklung des SIP hatte.

Allerdings können nicht alle divergierenden Formen auf das Ende des Arbeiterhandels oder einen Einfluss der Missionen zurückgeführt werden. Während der Analyse wurde ersichtlich, dass es auch Sprachmerkmale gab, die in allen drei Varietäten verwendet wurden. Exemplarisch seien hier die Morpheme *save* und *can* genannt, die in den frühen Daten aller drei Varietäten zur Kodierung des Abilitativs verwendet wurden. Während in SIP und BIS *save* mit höherer Frequenz attestiert wurde, dominierte in den TP-Daten *can*, welches auch heute, neben *inap* verwendet wird, um Fähigkeiten auszudrücken. Auch wenn diese individuellen Entwicklungen sehr wahrscheinlich durch das Ende des Arbeiterhandels vorangetrieben wurden, erklärt dies nicht, warum die Varietäten unterschiedliche Merkmale behielten. Eine mögliche Erklärung hierfür könnte der Einfluss der Substratsprachen sein sowie das Maß, in welchem die Varietäten mit der Lexifier-Sprache in Berührung kamen. Dies lässt sich beispielsweise auch

feststellen, wenn man die Demonstrativformen und deren diachrone Entwicklung in den drei Varietäten betrachtet.

Insgesamt konnten in der vorliegenden Arbeit vier Faktoren als Gründe für die *Individualisierung* der Varietäten identifiziert werden:

- (a.) die unterschiedlichen Arbeiter-/Rekrutierungshistorien der Gebiete und der damit einhergehende Einfluss von Plantagenkontaktsprachen (SPPE und QPPE)
- (b.) der Einfluss von Missionen
- (c.) Substrateinfluss (welcher mit dem Ende des Arbeiterhandels zugenommen hat)
- (d.) die Menge des Kontaktes mit dem Lexifier

Als Gründe für die *Stabilisierung* wurden ersichtlich:

- (a.) das Ende des Arbeiterhandels
- (b.) die Verbreitung und Verwendung der Varietäten in den Heimatgebieten
- (c.) Substrateinfluss

Die Ergebnisse der Studie haben somit gezeigt, dass eine Kombination der Theorien von Baker (1993) und Mühlhäuser (1978) sowie das Ende des Arbeiterhandels am besten die Entwicklung und Individualisierung von SIP, BIS und TP erklären können. Es sei betont, dass dies einen möglichen Einfluss vorausgegangener Jargons und eines Maritimen Polynesischen Pidgins (Drechsel 2014) nicht ausschließt.

Die Ergebnisse der Dissertation bedeuten einen wichtigen Fortschritt für den Erkenntnisstand der Varietätenforschung und Kreolistik, da sie Aufschluss über den Ursprung und die Entwicklung von Kontaktsprachen, Grammatikalisierungen und Sprachuniversalien geben. So zeigen die Ergebnisse deutlich, dass die Prozesse, die zur Entstehung von Pidgin- und Kreolsprachen führten, weitaus komplexer waren, als einzelne Theorien vermuten lassen. Die Analyse der metalinguistischen und linguistischen Daten zeigt, dass es sehr wahrscheinlich ist, dass ein Zusammenspiel der verschiedenen Theorien und Mechanismen zur Entwicklung und Stabilisierung der Varietäten führte. Um dies zu verdeutlichen wurde der Versuch angestellt, ein Model zu entwerfen, welches basierend auf den Ergebnissen der vorliegenden Arbeit einen Erklärungsansatz und Einflussfaktoren aufzeigt, um die Entstehung und Entwicklung der Pidgin- und Kreolsprachen zu verstehen (vgl. Abbildung 5).

Die Analyse hat gezeigt, dass die Ökologie (d.h. die spezifische Situation, die involvierten Bevölkerungsgruppen, die Menge des Kontaktes, etc.) eine entscheidende Rolle in der Entwicklung und Stabilisierung der Kontaktsprachen spielt. Die individuellen Kontaktsituationen führten zu einzigartigen Pools konkurrierender Sprachmerkmale in den jeweiligen Regionen. Dies spricht für die Theorie eines *evolutionary account of creole formation* (cf. Ansaldo 2009a, 2009b; DeGraff 2014; Yakpo 2021+). So sind beispielsweise deutsche Lexeme nur in den frühen Daten Deutsch-Neuguineas vorhanden (siehe, z.B. *raus* ‘throw out’ in von Hesse-Wartegg 1902:

53). Zudem konnten Ähnlichkeiten und Unterschiede zwischen den Varietäten dadurch begründet werden, dass SPPE und QPPE in unterschiedlichen Maßen in die Heimatregionen zurückgebracht wurden. Die Ergebnisse der Analyse sprechen auch dafür, dass die Auswahl von Merkmalen aus dem Pool konkurrierender Merkmale (cf. Mufwene 2001, 2006; Croft 2000; Ansaldo 2009a) in der Entwicklung der Kontaktsprachen eine große Rolle spielte.

Gleichzeitig zeigt die Studie, dass die Entwicklung von Kontaktsprachen komplexer und länger zu sein scheint, als bisherige Modelle vermuten lassen. Mufwenes *Feature Pool Hypothese* (2001, 2005, 2006) erweckt den Anschein, als sei die Selektion von Sprachmerkmalen ein einfacher Prozess, der zu Beginn der Sprachentwicklung stattfindet und bei dem sich jene Merkmale stabilisieren können, die perzeptuell prominent sind oder sehr häufig vorkommen (2001, 2005, 2006). Die durch die Ctree-Analyse erhaltenen Daten zeigten jedoch deutlich, dass Pools konkurrierender Merkmale ihren Aufbau immer wieder verändern konnten und die Selektion und Stabilisierung von Sprachmerkmalen nicht nur zu Beginn stattfand (vgl. Abbildung 22 und 23 in Kapitel 6). Des Weiteren zeigte die Analyse, dass auch *Diffussion* einen bedeutenden Einfluss auf die Entwicklung der Varietäten hatte. Europäer reisten von einer pazifischen Region in die andere; auf den Handelsschiffen kamen multilinguale Mannschaften zusammen, die wiederum in Kontakt mit Küstenbewohnern oder auch neuen Schiffsmannschaften kamen, wenn sie die Schiffe wechseln mussten. Auch konnte aufgezeigt werden, dass die Arbeiter- und Missionsbewegungen zur Verbreitung von Sprachmerkmalen führten.

Im Zusammenhang mit Missionen muss auch die Bedeutung soziologischer Faktoren berücksichtigt werden. Auch wenn nur eine geringe Anzahl an Missionaren in den Regionen tätig war, fungierten diese als Lehrer und linguistische Vorbilder. Zudem verwendeten sie die Kontaktsprache nicht nur in gesprochener Form, sondern verschriftlichten sie auch, wodurch sie die Standardisierung beeinflussten (siehe z.B. Kapitel 7, SIP Relativartikel *where*).

Die Daten lassen zusätzlich vermuten, dass die Sprachökologie der *frühesten* Kontaktsituation einen größeren Einfluss auf die Sprachentwicklung hatte, als spätere linguistische Ökologien. Auch wenn vereinzelte Lexeme, die ihren Ursprung im Französischen haben, in Bislama verwendet wurden und auch deutsches Vokabular in den TP-Daten vorzufinden war, kann man den Einfluss des Französischen auf BIS und den Einfluss des Deutschen auf TP als sehr gering betrachten. Eine mögliche Erklärung hierfür liefert das *Founder Principle* (cf. Chaudenson 1992, 2001; Mufwene 1996, 2001).

Die vorliegende Studie lässt des Weiteren darauf schließen, dass nicht nur eine einzige linguistische Ökologie pro Region vorlag, sondern dass viele regional einzigartige linguistische Ökologien und somit auch Pools konkurrierender Merkmale nebeneinander existierten, die sich

gegenseitig durch Diffusion beeinflussten. Es wurde ersichtlich, dass Sprachmerkmale sich auch verbreiten konnten, obwohl sie sich noch nicht stabilisiert hatten. Das Pidgin, welches von der SSEM verbreitet wurde, unterschied sich beispielsweise von dem Pidgin, welches in anderen Regionen gesprochen wurde und war zudem in engerem Kontakt mit dem Lexifier Englisch, wodurch standardenglisch-ähnliche Formen verstärkt worden sein könnten.

Aus den Daten ging hervor, dass das Pidgin-Englisch für einige die Zweit- oder Drittsprache darstellte und viele zusätzlich Englisch erlernten. Dies macht deutlich, dass auch Spracherwerbsprozesse als mögliche Einflussfaktoren berücksichtigt werden müssen (vgl. Arends 1989, 1992). Die Daten zeigten zudem, dass Prozesse wie Code-Switching, Code-Mixing und Entlehnungen, die typisch für multilinguale Kontexte sind (vgl. Johanson 1992, 1993; Muysken 2008), die Sprachentwicklung beeinflussten. Exemplarisch sei hier das Code-Switching erwähnt, welches in den Gerichtsverfahren und in einigen Quellen der SSEM vorzufinden war.

Auch wenn es nicht Ziel der Studie war, zu analysieren, ob Pidgin- und Kreolsprachen eine eigene Sprachklasse mit spezifischen typologischen Merkmalen darstellen, wie von McWhorter (2002, 2005), Parkvall (2008) and Bakker et al. (2013) angenommen wird, hat die Studie deutlich gezeigt, dass es nicht ausreichend ist, die Entstehung von Kontaktsprachen ausschließlich basierend auf aktuellem Sprachmaterial zu rekonstruieren. Sprachmerkmale können existiert haben und dann verschwunden sein. Um also die linguistische Struktur von Pidgin- und Kreolsprachen zu verstehen und die Frage zu beantworten, ob sie eine typologisch differenzierte Sprachklasse darstellen, sollte in jedem Fall historisches Datenmaterial gesammelt und berücksichtigt werden.

Auch in Bezug auf Grammatikalisierungsprozesse konnten Erkenntnisse gewonnen werden. Sobald Kontaktsprachen sich ausdehnen und in mehr Domänen verwendet werden als in der ursprünglichen Kontaktsituation, entsteht die Notwendigkeit weitere grammatikalische Funktionen zu kodieren. Die Analyse der Daten zeigte, dass insbesondere Grammatikalisierung (vgl. *save, here, inap, wantaim, inap long and kitchim*), Degrammatikalisierung (vgl. *might*) und Reanalyse (vgl. *I think*) typische Mechanismen darstellen, die während der Entwicklung der Varietäten eine wichtige Rolle spielten. Es wurde festgestellt, dass die drei MPE Varietäten ähnliche Grammatikalisierungspfade aufweisen, wie andere Sprachen und sie somit mit den allgemeinen typologischen Tendenzen übereinstimmen. Dies könnte ein Argument gegen die *Creole Prototype* Theorie (McWhorter 2002, 2005) sein.

Abschließend lässt sich sagen, dass historisch-linguistisches sowie historisch-metalinguistisches Datenmaterial notwendig ist, um ein allumfassendes Verständnis über die Entstehung und Entwicklung von Pidgin- und Kreolsprachen zu erhalten. Auch wenn historische

Quellen unausgewogen sind, wurde ersichtlich, dass sie dennoch quantitativ-statisch ausgewertet werden können. *Ctrees* stellen ein hilfreiches Tool dar, um die Hauptindikatoren zu ermitteln, die für Veränderungen in den Daten verantwortlich waren. Auch wenn der Autor meist als signifikanteste Einflussvariable identifiziert wurde, bedeutet dies nicht, dass die Quellen oder Autoren unzuverlässig sind. Es weist lediglich darauf hin, dass aufgrund der Natur der Daten, Informationen über den Autor gesammelt und bei der Analyse der Daten berücksichtigt werden sollten. Der Algorithmus half dabei, Unregelmäßigkeiten in den Daten offenzulegen und die Jahre zu identifizieren, in denen signifikante Veränderungen in den Daten zu beobachten waren. Wie zuvor erwähnt, bedeuten diese Veränderungen nicht zwangsläufig eine Individualisierung oder Stabilisierung von bestimmten Merkmalen, sodass eine begleitende qualitative Analyse unabdingbar war. Auch wurde deutlich, dass Forscher über ein gutes Wissen über ihre historische Datengrundlage verfügen müssen, um Fehlinterpretationen zu vermeiden. Statistische Methoden können historisch-qualitative Studien zwar ergänzen, nicht jedoch ersetzen. In Zukunft sollten mehr Kollaborationen zwischen Statistikern und Kreolisten stattfinden, um die Expertisen aus beiden Bereichen zusammenzubringen und die historischen – meist unausgewogenen Daten – bestmöglich zu analysieren.

In der vorliegenden Studie lag der Fokus auf vier Sprachmerkmalen. Zukünftige Studien, die sich auf andere Sprachmerkmale konzentrieren, können dazu beitragen noch mehr über die Entstehung und Entwicklung der drei melanesischen Pidgin-Englisch-Varietäten zu lernen. Insbesondere sollte der Fokus auf Merkmale gelegt werden, in denen sich SIP und BIS unterscheiden. Auch sollten zukünftige Studien Substrat- und Adstratsprachen und deren Einfluss auf die Entwicklung der Varietäten berücksichtigen.

Es sollte zudem das Ziel verfolgt werden, frühe Datenbelege von Kontaktsprachen in online zugänglichen Datenbanken verfügbar und analysierbar zu machen. Während der Analyse der Daten wurde schnell erkennbar, dass die Varietäten in manchen Zeiträumen über- bzw. unterrepräsentiert waren. Da schon einige Forscher zuvor frühe Datenbelege gesammelt haben, sollten die einzeln verstreuten Datenbelege zusammengeführt und zugänglich gemacht werden.

Des Weiteren sollten auch Daten, die nach dem zweiten Weltkrieg datiert sind, gesammelt und analysiert werden. Manche Merkmale, die in der Studie analysiert wurden, weisen darauf hin, dass die Stabilisierung erst in der zweiten Hälfte des 20. Jahrhunderts stattgefunden hat. Es ist also notwendig, Korpora zu erstellen, die die Varietäten von ihrem Beginn bis zu ihrer jetzigen Form bestmöglich darstellen.

Dies hätte auch den Vorteil, dass analysiert werden könnte, wie erneuter Sprachkontakt (z.B. durch die Erbauung von Flughäfen, Studienaustauschprogramme, oder Radioprogramme,

welche Inhalte in TP, BIS, und SIP senden und in alle drei Regionen übertragen werden) die Entwicklung der Kontaktsprachen beeinflusst hat und noch immer beeinflusst. Zugleich ist nicht nur die Untersuchung eines erneuten Kontaktes der Varietäten untereinander, sondern auch mit dem Lexifier und/oder regionalen Varietäten des Englischen für zukünftige Projekte vielversprechend. Gleiches gilt auch für den Einfluss des Internets und der modernen Medien, da diese einen Anstieg an Pidgin-Englisch in geschriebener Form bedeuten und Grammatikalisierungsprozesse gegebenenfalls vorantreiben können.

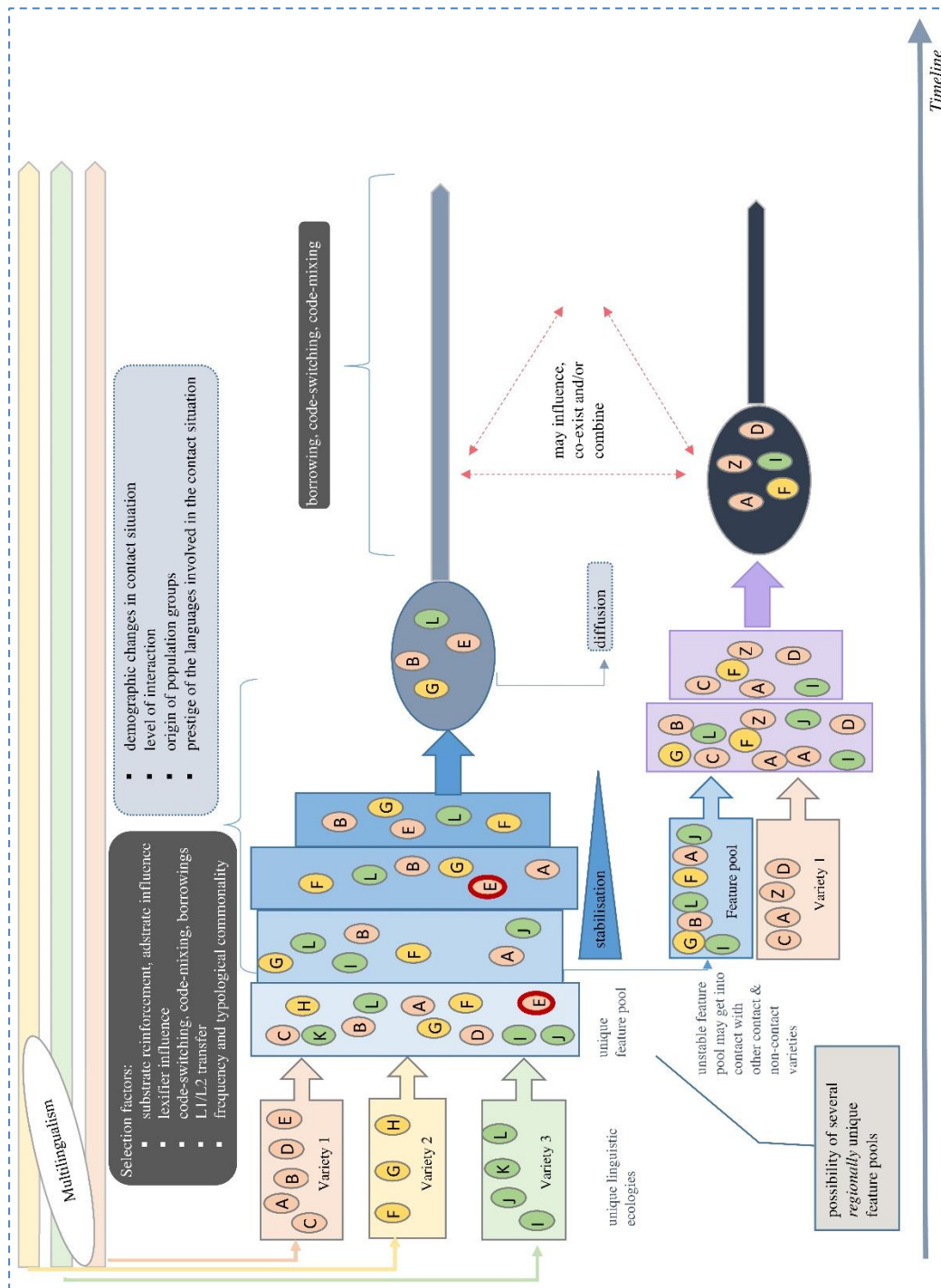


Abbildung 5. Komplexität der Entwicklung von Sprachmerkmalen in Kontaktsprachen (basierend auf Belegen der MPE Varietäten)