



**Mathilde Hennig, Nilüfer Cakmak-Niesen,  
Philipp Meisner**

## **DFG-Langfristvorhaben an der JLU Gießen: Syntaktische Grundstrukturen des Neuhochdeutschen.**

### **Zur grammatischen Fundierung eines Referenzkorpus Neuhochdeutsch**

#### **Motivation**

Bei der Lektüre von Texten der jüngeren deutschen Kulturgeschichte machen wir alle immer wieder die folgende Erfahrung: Ab und zu müssen wir vielleicht ein Wort nachschlagen, das in der Gegenwartssprache nicht mehr zu unserem Wortschatz gehört; im Großen und Ganzen können wir aber die Sprache der Zeitgenossen von Gryphius, Goethe und Fontane gut verstehen. Im Gegensatz zu mittelalterlichen Texten ist das Deutsche der Frühen Neuzeit, Aufklärung und Klassik keine Fremdsprache für uns. Dennoch unterscheidet sich die Sprache von unserer heutigen: Manche Form klingt antiquiert, an der einen oder anderen Stelle würden wir heute einen anderen Satzbau wählen.

Angesichts der hohen Bedeutung des Zeitraums vom 17. bis zum 19. Jahrhundert für die deutsche Kulturgeschichte und der hohen Aufmerksamkeit, die die Schriftzeugnisse dieser Zeit in verschiedenen Wissenschaftsdisziplinen und in der Öffentlichkeit genießen, ist das Wissen über Wandelphänomene im Bereich der Grammatik in der jüngeren Sprachgeschichte des Deutschen als erstaunlich gering einzustufen (vgl. Ágel 2000). Dabei handelt es sich ausgerechnet um die sprach- und kulturhistorisch zentrale Epoche der Herausbildung der modernen Schriftsprache und Standardsprache, der Entstehung von modernen Umgangssprachen und Regionalsprachen und der Ausdifferenzierung von Textsorten, Textstilen und literarischen Gattungen. Während also etwa das 18. und das 19. Jahrhundert literatur- und philosophiegeschichtlich sehr gut erforscht sind, wäre eine fundierte grammatische Analyse von zentralen literarischen und philosophischen Texten des 18. und

19. Jahrhunderts mangels Untersuchungen zu syntaktischen Grundstrukturen des Neuhochdeutsch gar nicht erst möglich. Grammatische Strukturen sind aber keine reinen Formen, sie sind nicht bedeutungslos, sondern tragen entscheidend zur semantischen Interpretation von Texten bei.

Das DFG-Langfristvorhaben „Syntaktische Grundstrukturen des Neuhochdeutschen. Zur grammatischen Fundierung eines Referenzkorpus Neuhochdeutsch“ will Abhilfe schaffen, indem es Quellentexte aus dem Zeitraum vom 17. bis zum 19. Jahrhundert syntaktisch erschließt und die auf diese Weise grammatisch aufgearbeiteten Texte als Korpus der Fachöffentlichkeit zur Verfügung stellt. Das Projekt, das im Herbst 2016 begonnen hat und auf eine Laufzeit von 12 Jahren angelegt ist, ist ein Kooperationsprojekt der JLU Gießen und der Universität Kassel. Es wird in Gießen von Prof. Dr. Mathilde Hennig geleitet und in Kassel von Prof. Dr. Vilmos Ágel. Mit der syntaktischen Analyse von Quellentexten aus den Bereichen Wissenschaftssprache, Alltagssprache, Gebrauchsliteratur und Belletristik schafft das Projekt eine Grundlage für die systematische Dokumentation von syntaktischem Wandel und Variation im Neuhochdeutschen. Mit der Erstellung eines syntaktisch annotierten Referenzkorpus des Neuhochdeutschen arbeitet das Projekt mit den Methoden der Korpuslinguistik und folgt damit einem – wenn nicht dem – zentralen linguistischen Trend unserer Zeit. Unter einem Korpus versteht man in der Linguistik Sprachmaterial (in der Regel Texte), das, durch weitere Informationen angereichert, „annotiert“ ist. Welcher Art diese Informationen sind, hängt von der Zielsetzung des Korpus ab. So gibt es für den Zeitraum des Neuhochdeutschen mit dem „Deutschen Textarchiv“ (= DTA) eine

groß angelegte Textsammlung aus Schriftzeugnissen der letzten Jahrhunderte, bei der es vor allem darum geht, die Texte verschiedenen Wissenschaftsgebieten digital verfügbar zu machen und den Nutzern Informationen zur Textgestalt und zu sogenannten „Metadaten“, d.h. Informationen über den Text anzubieten. Außer einer sehr groben, maschinell erstellten Wortartanalyse und einer Zuweisung von Satzgrenzen auf der Basis von Interpunktion bietet das DTA keine tiefere linguistische Erschließung der Texte. Das auf einem Projekt der Berlin-Brandenburgischen Akademie der Wissenschaften basierende Korpus hat auf diese Weise in den letzten 13 Jahren 6500 Texte im stattlichen Umfang von 320 Millionen Wortformen erschließen können.

Das Korpus GiesKaNe (= GießenKasselNeuhochdeutsch) hingegen wird nach seiner Fertigstellung ein kleines Korpus sein, weil es hier um eine sogenannte tiefe Mehrebenenannotation geht, d.h., es erfolgt eine sehr detaillierte Analyse syntaktischer Grundstrukturen. Da die Sätze in den älteren, häufig auch durch konzeptionelle Mündlichkeit geprägten Texten nicht alle einem einheitlichen, gar aus der Gegenwartssprache bekannten Muster folgen und die Interpunktion keine zuverlässige Größe für die syntaktische Strukturierung ist, erfolgt die Analyse und somit auch Annotation im Wesentlichen manuell. Deshalb kann das Korpus nach seiner Fertigstellung nicht mehr als schätzungsweise eine Million Wortformen umfassen. Geplant ist dafür die Analyse von 72 Texten à ca. 30–40 Druckseiten.

### **Eine Grammatik des gesamten Sprachgebrauchs**

Eine Grammatik des Neuhochdeutschen kann man sich einerseits als eine Grammatik vorstellen, die die grammatische Kunst der literarischen und intellektuellen Größen dieser Zeit würdigt, andererseits aber auch als eine Grammatik, die die gesamte Bandbreite des Sprachgebrauchs in den Blick nimmt. Das ist der Weg, den das GiesKaNe-Projekt gehen möchte. Damit folgen wir den Trends zu einer soziopragmatischen Sprachgeschichtsforschung (von Polenz 1995), d.h. einer Sprachgeschichtsforschung, die nicht nur die Entwicklung der

sprachlichen Formen als solcher zu ihrem Gegenstand macht, sondern die sich – salopp formuliert – auch dafür interessiert, welche Menschen in welchen Kontexten welche Art von Sprache verwenden. Im Sinne einer – sich an Entwicklungen in der Geschichtsforschung orientierenden – „Sprachgeschichte von unten“ (Elspaß 2005) erscheint es uns gerade aufschlussreich, die syntaktischen Grundstrukturen des elaborierten (etwa: wissenschaftlichen und/oder intellektuellen) Sprachgebrauchs mit den Strukturen in Texten von einfachen und ungeübten Schreibern zu vergleichen. So haben wir es ja mit demjenigen Zeitraum in der deutschen (Kultur-)Geschichte zu tun, in dem etwa durch einen rasanten Anstieg der Verbreitung von Schriftzeugnissen, insbesondere durch eine starke Entwicklung des Zeitungs- und Zeitschriftenwesens, durch „Massenalphabetisierung“ (im 19. Jahrhundert; Knoop 1994: 869; Elspaß 2005: 77) und „Leserevolution“ um 1770 (von Polenz 1994: 34ff.) eine „Verschriftlichung des Lebens“ zu verzeichnen ist (Hennig 2009: 41ff.). Die Schreibenlässe für ungeübte Schreiber wiederum werden durch die historischen Verhältnisse vorgegeben: Ein Söldner berichtet von seinen Erlebnissen im Dreißigjährigen Krieg („Söldnerleben“), ein Kannengießer von seinen Reisen im 17. Jahrhundert („Güntzer“), ein durch den Deutsch-Französischen Krieg in der zweiten Hälfte des 19. Jahrhunderts getrenntes Ehepaar schreibt sich Briefe („Briefwechsel“), Auswanderer berichten ihrer Familie aus der neuen amerikanischen Heimat („Zimmer“). Die aus der Verschriftlichung des Lebens folgende „Verschriftlichung des Sprechens“ (Hennig 2009: 41; Koch/Oesterreicher 1994: 587; Oesterreicher 1993: 272) ist für die Sprachwandelforschung ausgesprochen spannend: Das Spannungsfeld von Mündlichkeit und Schriftlichkeit verschiebt sich in diesem Zeitraum fundamental. Wir haben es nun mit einem mehr oder weniger gleichberechtigten Nebeneinander von Mündlichkeit und Schriftlichkeit zu tun, was natürlich mit reziproken Auswirkungen verbunden ist. Hinzu kommt die Umschichtung im Varietätengefüge, die Etablierung der geschriebenen Standardsprache als Leitvarietät. Diese „Vertikalisierung“ (Reichmann 1988), also Aus-

richtung nach oben, wird von „immer intensiveren Normierungs- und Kodifizierungsbestrebungen“ begleitet (Ágel 2000: 1857). Warum aus diesen Beobachtungen das Mantra „eine Grammatik des Neuhochdeutschen muss den gesamten Sprachgebrauch abbilden“ folgt, dürfte nun klar sein: Nur eine solche Gesamtgrammatik kann die kulturgeschichtlich interessanten Entwicklungen in der Periode des Neuhochdeutschen abbilden.

### Zur Textprofilbildung grammatischer Muster

Wenngleich das primäre Arbeitsfeld unseres Projekts durch die grammatische Analyse von Sprachmaterial ausgezeichnet ist, erschöpft sich ihr Ertrag nicht in grammatiktheoretischen Fragestellungen. Da die Analyseobjekte authentische Texte darstellen, werden durch die Abbildung der Sprachoberfläche im Korpus

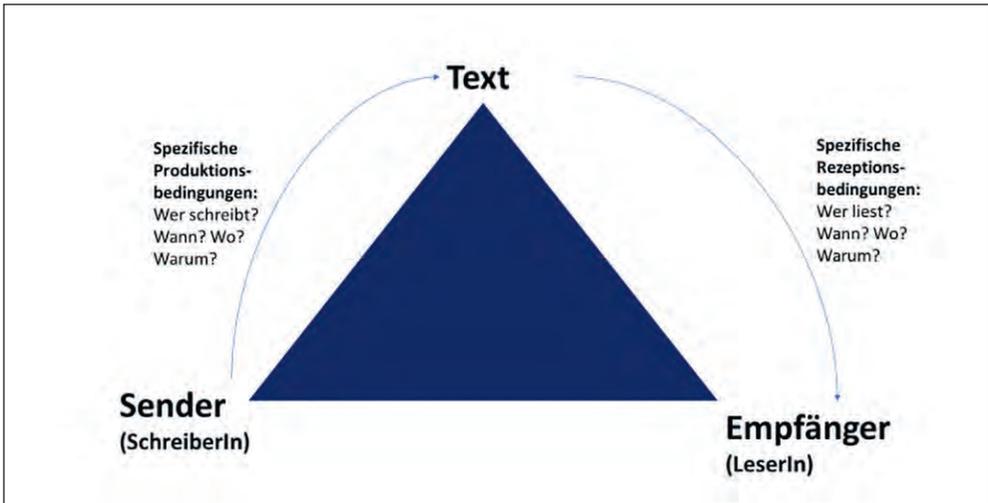


Abb. 1: Grundmodell der Textproduktion <sup>1</sup>

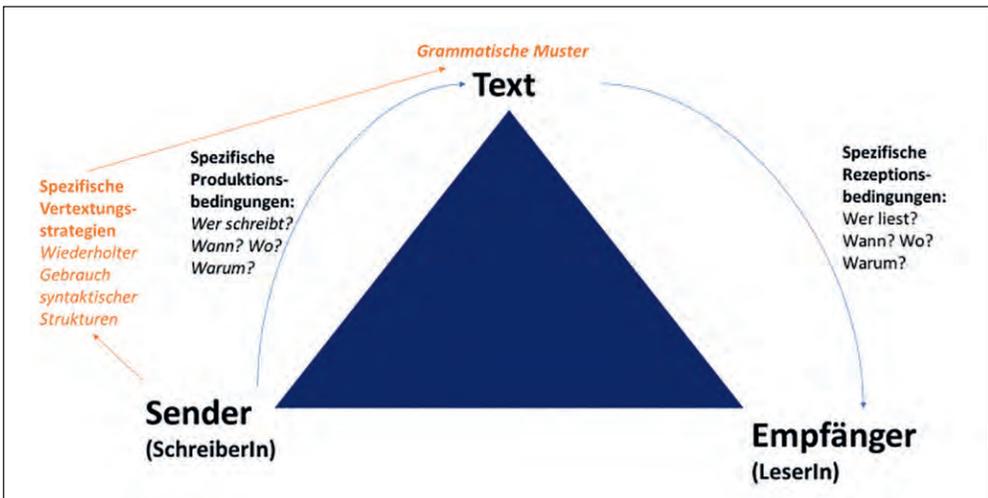


Abb. 2: Zur Rolle grammatischer Muster an der Ausbildung eines Textprofils

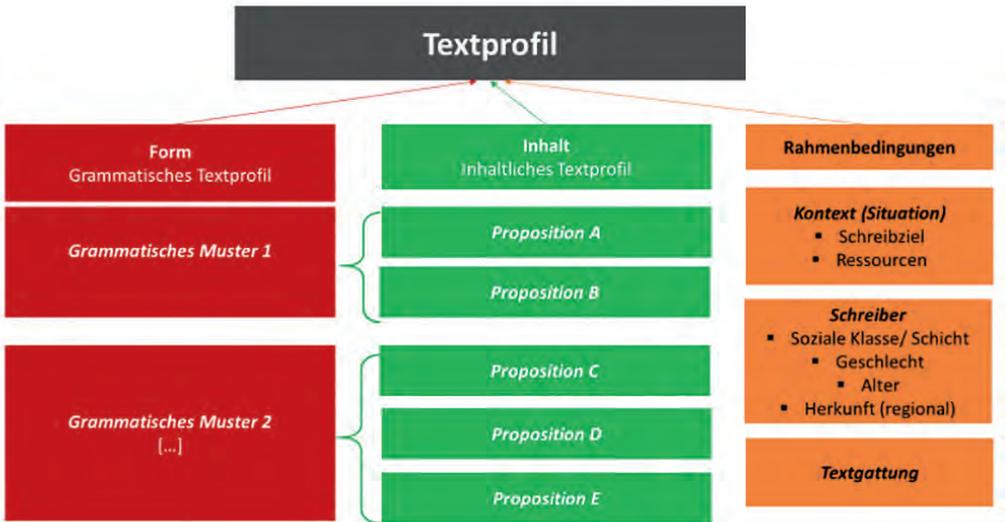


Abb. 3: Textprofil als Interaktion formaler und inhaltlicher Merkmale auf der Basis bestimmter Rahmenbedingungen <sup>2</sup>

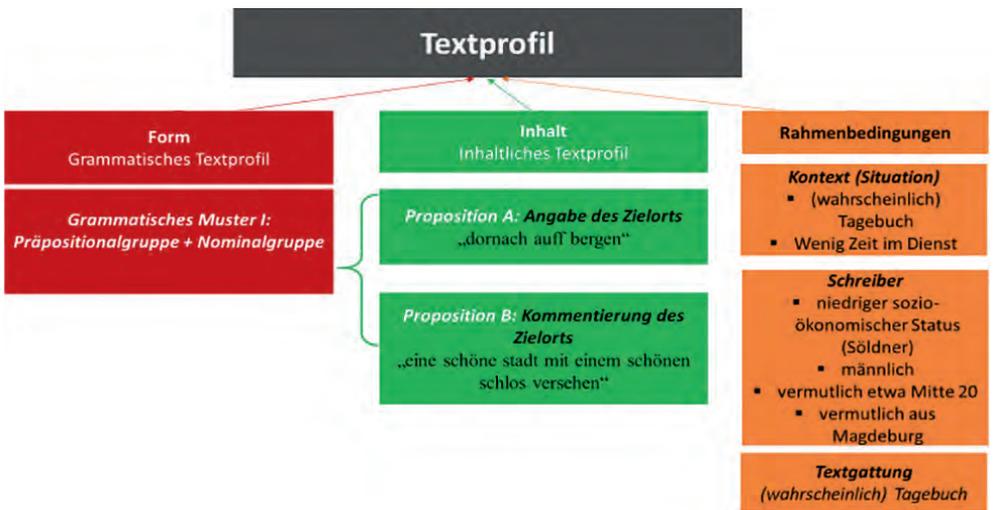


Abb. 4: Exemplarische Realisierung eines Textprofils am Beispiel einer Passage aus dem Text „Söldnerleben“

Perspektiven für interdisziplinäre Forschungsfragen eröffnet. In diesem Kapitel soll exemplarisch demonstriert werden, inwiefern die gewonnene syntaktische Analyse aus soziopragmatischen Gesichtspunkten fruchtbar gemacht werden kann. Die übergeordnete Einheit unserer Analysekatoren stellt der jeweilige Text dar. Der Text

kann als Ausdruck bestimmter kommunikativer Bedürfnisse des jeweiligen Schreibers verstanden werden. Der Schreiber verfolgt durch den Text ein kommunikatives Ziel, sei es, dass er andere überzeugen will, sei es, dass er den Text als Ausdrucksmedium für seine Emotionen nutzt. Prinzipiell sind an der Kommunikationssituation also drei Größen beteiligt: der Text-

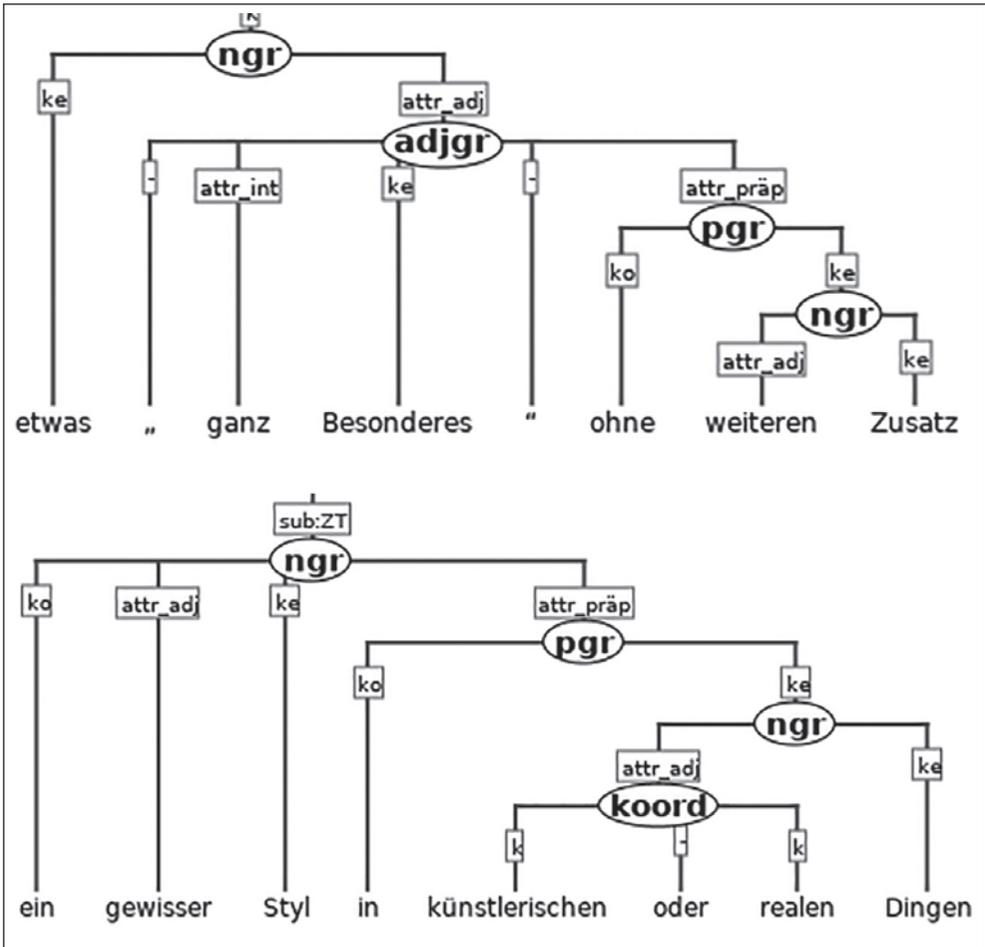


Abb. 5: Komplexe Nominalgruppe in „Simmel“

produzent, der Text selbst und schließlich der Rezipient.

Der Schreibprozess unterliegt seitens des Textproduzenten spezifischen Produktionsbedingungen. Relevant für die Konstitution des Textes ist stets, wer zu welchem Zeitpunkt und an welchem Ort für welchen Zweck schreibt.

Der Schreiber selbst (das „Wer“) ist beispielsweise durch seine jeweilige Schicht-/Klassenzugehörigkeit und seinen damit verbundenen Bildungsstand, sein Alter und sein Geschlecht in seinem Text gespiegelt. Der Entstehungsort des Textes kann Aufschluss über regionale Varietäten geben. Der Zweck des Schreibens kann ei-

ne gewisse Adressatenorientierung bedingen. Rahmenbedingungen wie die Verfügbarkeit von Ressourcen (Zeit, Papier) können einen bestimmten Schreibstil forcieren. So kann beispielsweise anhand der Schreibsituation erklärt werden, warum der anonyme Söldner in seinem Tagebuch während seines Diensts im Dreißigjährigen Krieg (*Söldnerleben*) mehrheitlich kurze Parataxen und angebrochene Sätze schreibt, jemand mit ausreichend Zeit für Sorgfalt dagegen längere, ausschweifende Perioden formulieren kann. Damit soll hier nicht behauptet werden, dass man aus jeder Textoberfläche immer auf sämtliche soziopragmatischen Rahmenbedin-

gungen der Textproduktion geschlossen werden kann. Prinzipiell ist aber davon auszugehen, dass es kein Zufall ist, wer wie schreibt, sondern dass sich die Rahmenbedingungen der Textproduktion in der Auswahl der sprachlichen Mittel niederschlagen.

In Texten werden Inhalte organisiert. Die Organisationseinheiten stellen Sätze dar – wie Lötischer treffend formuliert, werden „in Sätzen [...] Inhalte in formale Strukturen gegossen“ (2016: 365). Der wiederholte Gebrauch bestimmter formaler Strukturen führt schließlich zur Bildung bestimmter Muster.

Sofern bestimmte syntaktische Strukturen repetitiv gebraucht werden, lassen sich grammatische Muster erkennen. Wenn bestimmte Inhalte wiederholt werden, können auch diese musterhaft werden. Im Nebeneinander von Form und Inhalt sowie den äußeren Faktoren entsteht in Summe ein Textprofil, das vom Rezipienten zusammengesetzt werden kann. Das folgende Modell erhebt keinen Anspruch auf Vollständigkeit; es kann als erste Annäherung zu einem möglichen Ansatz für *text profiling* verstanden werden.

Obwohl ein Textprofil aus grammatischen Mustern erst in der Gesamtheit des Textes erschlossen werden kann, lässt bereits der erste Blick auf einen Textausschnitt Annahmen zum erwarteten Textprofil zu. Beispielsweise können wir auf den ersten Blick große Unterschiede in den folgenden Textsequenzen feststellen und Vermutungen zu den o.g. Parametern eines Textprofils anstellen:

A *Schon die Sprache läßt die „Seltenheit“ zugleich als Vorzüglichkeit und etwas „ganz Besonderes“ ohne weiteren Zusatz, als etwas ganz besonders Gutes gelten, während das Gemeine, d. h. das dem weitesten Kreise Eigene, Unindividuelle, zugleich das Niedrige und Wertlose bezeichnet.“* (Simmel 1890)

B *„dornach auff bergen, eine schöne stadt mit einem schönen schlos versehen“ Alhir gehet das venedieger landt Recht an, dornach auff pressa, eine schöne stadt, Alhir habe Ich mich lassen vnterhalten vnter die venediegern, In Ihren dienst, da sindt wir aus einer stadt In die ander gezogen, In die qartir sein, Als von pressa Auff biskere am gardtsch, eine vberaus schöne festung, [...] Alhir bin Ich krank worden, den ein gar hiedzieger wein, In diesen landt, bin krank gewessen 2 Monat, doch hat mich der almechtige wieder geholffen, da sonsten viel deutzen gestorben sindt“* (Söldnerleben 1625–1649)

Hinter A lässt sich anhand des gebrauchten Wortschatzes ein gebildeter Schreiber vermuten. Die Verwendung von Nebensätzen und erweiterten Nominalgruppen (d.h. Wortgruppen, deren Kern ein Substantiv darstellt, z.B. *das dem weitesten Kreise Eigene; etwas ganz besonders Gutes*) lässt auf einen distanzsprachlichen (schriftsprachlichen) Text schließen. Inhaltlich geht es um die Abgrenzung der Konzepte des Seltenen und Gemeinen.

Ausschnitt B lässt einen weniger gebildeten Schreiber erwarten. Der Text macht einen nähe-

Tabelle 1: Komplexitätsmaße in zwei Texten im Vergleich

|                    | A (Simmel) | B (Söldnerleben) |
|--------------------|------------|------------------|
| Sätze              | 10         | 10               |
| Teilsätze          | 28         | 10               |
| Teilsatz/Satz      | 2,8 TS/S   | 1 TS/S           |
| Wörter             | 242        | 103              |
| Wörter/Satz        | 23,1 W/S   | 10,3 W/S         |
| Wörter/Teilsatz    | 7,9 W/TS   | 10,3 W/TS        |
| Komplexe Attribute | 6          | 0                |

sprachlichen Eindruck (klingt wie gesprochen). Es gibt kein Prädikat, die Sätze erscheinen unvollständig. Inhaltlich geht es darum, dass der Schreiber im Rahmen seines Einsatzes als Söldner in die Stadt Bergen gezogen ist, die er als schön bewertet. Das vorgeschlagene Textprofil zu Ausschnitt B könnte wie folgt ausgefüllt werden:

Die unter *Rahmenbedingungen* aufgeführten Elemente sind weitestgehend nur unter Einbezug weiterer Textteile und entsprechend nicht allein durch den Ausschnitt zu beantworten. Man benötigt hierfür sogenannte „Metainformationen“ über den Autor und den Kontext des Entstehens des Textes. In einem Korpus werden solche Metainformationen mit erfasst, damit man in der Analyse die Informationen zu den Rahmenbedingungen systematisch mit den sprachlichen Merkmalen korrelieren kann.

Jedenfalls können wir nach der ersten Lektüre der beiden Ausschnitte mit großer Wahrscheinlichkeit behaupten: Das sind zwei unterschiedliche Schreiber und zwei unterschiedliche Texte. Worin liegt es nun begründet, dass der Leser mit den beiden Textausschnitten instinktiv ganz unterschiedliche Textprofile mit verbindet? Der augenscheinliche Unterschied könnte u.a. mit dem Begriff der Komplexität benannt werden: Sowohl im Hinblick auf den Inhalt als auch auf die sprachliche Gestaltung kann Ausschnitt A im Vergleich zu B als komplexer beschrieben werden. Einfach gesagt: A verfügt über mehr Wörter, größere Wortgruppen, einen größeren Wortschatz und verschachtelte Sätze. Komplexität ist ein Indikator zur Bestimmung der einzelnen Parameter im *text profiling*. Für die Messbarkeit von Komplexität auf der Formseite ist die Analyse der Sprachoberfläche unerlässlich. Hierfür gibt das syntaktisch tiefenannotierte Korpus dem Forscher die Messinstrumente zur Hand.

Die von uns bereits nach der ersten Lektüre der Ausschnitte festgestellten Unterschiede lassen sich auch in Zahlen ausdrücken. Mit Hilfe der von Lötscher (2016: 387) entwickelten Hilfsvariablen zur Determinierung von Komplexität sollen die beiden Texte nun empirisch nebeneinandergestellt werden. Lötscher (ebd.) hat in unterschiedlichen Texten jeweils 100 Sätze nach den in ihnen vorkommenden Teilsätzen, Wörtern

und komplexen Nominalgruppen untersucht. Mehr Teilsätze, mehr Wörter und mehr komplexe Nominalgruppen bedeuten jeweils mehr Komplexität.

Wir wollen uns exemplarisch die ersten zehn (grammatischen) Sätze der jeweiligen Texte anschauen.

Zunächst fällt auf, dass im Text A mehr als doppelt so viele Wörter als im Text B auf 10 Sätze kommen. Durchschnittlich kommen in A 23 Wörter auf einen Satz, in B sind es lediglich 10 Wörter pro Satz. Außerdem sind in A fast dreimal so viele Teilsätze enthalten wie in B. A hat also mehr Wörter und mehr Nebensätze; in B gibt es keinen einzigen Nebensatz.

Im Text *Söldnerleben* lassen sich in den ersten zehn Sätzen auch keine komplexen Nominalgruppen im Sinne Lötschers<sup>3</sup> feststellen, d.h. Wortgruppen werden einfach aufgebaut (z.B. durch *Artikel+Adjektiv+Substantiv: eine schöne Stadt*). Die geringe formale Komplexität stimmt mit der inhaltlichen Seite überein: Im Text werden die Kriegszüge in kurzen Phrasen nachgezeichnet,

„1625. von Vrssenowe, wieder auff pressa“

der Dienst und die Tätigkeiten des Söldners werden in Kurzform festgehalten;

„von daaus, sindt wir auff den musterplatz gezogen,  
nach die ober Marggraffschaff baden,  
Aldort In quartier gelehgen“

die Tode der Kameraden werden nüchtern dokumentiert.

„also habe Ich meinen camerradt verloren,  
weis auch noch nicht wo er hin kommen Ist“

Die grammatischen und inhaltlichen Muster sind am Kommunikationsbedürfnis des Schreibers ausgerichtet. So scheint es dem anonymen Söldner maßgeblich darauf anzukommen, seine Erlebnisse im Krieg gebündelt und ohne Ausschweife zu dokumentieren.

Ganz anders verhält es sich in Text A, einem wissenschaftlichen Artikel des Philosophen Georg Simmel im Rahmen seiner Monographie „Über soziale Differenzierung“ aus dem Jahre 1890. Der argumentative Aufbau des Textes dient im zitierten Abschnitt der Definition und Konzeptualisierung des sozialen Niveaus. In diesem Text lässt sich das „Streben nach ikonischer Abbildung logischer Zusammenhänge[...] in Fachtext-

Tabelle 2: Grammatische Ebenen, Analyseeinheiten und Werte in GiesKaNe nach Ágel (2017: 26)

| Ebene      | Analyseeinheit       | Grammatischer Wert |
|------------|----------------------|--------------------|
| Makroebene | Text                 | Textglied          |
| Mesoebene  | Satz                 | Satzglied          |
| Mikroebene | Wortgruppe/Wortarten | Wortgruppenglied   |

ten“ (Lötscher 2016: 370) nachvollziehen, in dem die hohe sprachliche Komplexität zunächst als „Reflex eines Bemühens um sprachliche Präzision“ (ebd.) erscheint. Auf der sprachlichen Oberfläche wird dieses Bemühen um die bildhafte Abbildung von Logik in Passagen wie der folgenden besonders greifbar, in der für diesen Zweck mathematisch-logische Formeln versprachlicht werden:

*„so völlig falsch es logischerweise ist, dafs,  
weil alle  $a = b$  sind, nun auch alle  $b = a$  sein  
sollen, so begeht doch das thatsächliche  
Denken und Fühlen unzähligemal  
diesen Fehlschlufs“*  
(Simmel 1890)

Die Komplexität erstreckt sich in diesem Fachtext von der Makroebene (ausgedehnte Sätze, viele Wörter) über die Mesoebene (viele Satzglieder, viele Gliedsätze) bis hin zur Mikroebene (erweiterte Wortgruppen, komplexe Nominalgruppen i.S. Lötschers). Die komplexen Nominalgruppen dienen nicht zwingend einfach nur der Unterbringung zusätzlicher Informationen oder von „Parallelhalten“, sondern können in der jeweiligen Aussage als „eigenständige Textthemen thematisiert“ werden (Lötscher 2016: 371). Auf der Formseite bedeutet das wiederum, dass sie grammatisch nicht nur als kontextmodifizierende Adverbiale, sondern auch als direkte Prädikatsergänzungen gebraucht werden können. Die attributive Präzision dient der Erörterung, Reflexion und Diskussion der formulierten eigenständigen Textthemen. Die mehrschichtige Textthematik übersetzt sich in die mehrschichtigen syntaktischen Strukturen.

Zur Verdeutlichung der syntaktischen Mehrstufigkeit des Textes sollen exemplarisch zwei komplexe Nominalgruppen graphisch visualisiert werden:

Die Darstellungsform der Beispiele wird im folgenden Kapitel näher erläutert. Die Abhängigkeiten sind hier als sogenannte Konstituentenstrukturen in einem Baum abgebildet. Beide Beispiele repräsentieren je eine Nominalgruppe (d.h. Wortgruppen mit einem nominalen Kern: *etwas, Stil*) mit komplexer interner Struktur, d.h. beide Nominalgruppen verfügen über Attribute zweiten Grades, d.h. Wortgruppenglieder, die den anderen zwei Stufen untergeordnet sind. So werden die Wortgruppen länger und „tiefer“, ergo komplexer.

Dass mit der syntaktischen Komplexität auch eine ausgewählte Leserschaft i.S. eines recipient design zum Wunschatressaten erklärt wird, kann an dieser Stelle nicht ausgeweitet werden. Es kann aber festgehalten werden, dass Ausschnitt A ein höher gebildeteres Publikum als Ausschnitt B adressiert. Schließlich sind verschachtelte Nebensätze und mehrstufige Wortgruppen für den Leser schwieriger zu entschlüsseln und somit zu verstehen.

Wie an den beiden Texten gezeigt werden konnte, sind die inhaltlichen Aspekte nicht getrennt von den sprachlichen zu betrachten. Sprachoberfläche und -inhalt sind eng miteinander verwoben. Die sprachliche Gestalt wird als Vermittlungsmedium für Inhalte nach den jeweiligen Kommunikationsbedürfnissen des Sprechers ausgerichtet.

Sprachliche Oberfläche und Inhalt stehen folglich in einem kodependenten, reziproken Verhältnis.

## Datenaufbereitung

Wie weiter oben bereits beschrieben wurde, erfolgt die Annotation der sprachlichen Daten aufgrund des komplexen Mehrebenenmodells zu großen Teilen manuell. Als Werkzeug dafür

| GKN_Token | Normalisierung | Lemmatisierung | Wortart (Klasse) | Wortart (Subklasse) | Makro-glieder | Teilsätze | Mesoebene 0 | Semantik 0 | Mikroform Ebene A | Wert/Funktion Ebene A |          |  |  |     |
|-----------|----------------|----------------|------------------|---------------------|---------------|-----------|-------------|------------|-------------------|-----------------------|----------|--|--|-----|
| Uunder    | Unter          | Unter          | pr               |                     | s             | hs        | dirum       |            | pgr               | ko                    |          |  |  |     |
| solche    | solche         | solch          | d                | denu                |               |           |             |            |                   |                       |          |  |  |     |
| beße      | böse           | böse           | adj              |                     |               |           |             |            |                   |                       |          |  |  |     |
| Buben     | Buben          | Bube           | n                |                     |               |           |             |            |                   |                       |          |  |  |     |
| undt      | und            | und            | j                | kon                 |               |           |             |            |                   |                       |          |  |  |     |
| Merdter   | Mörder         | Mörder         | n                |                     |               |           |             |            |                   |                       |          |  |  |     |
| bin       | bin            | sein           | v                | lv                  |               |           |             |            |                   |                       |          |  |  |     |
| ich       | ich            | ich            | p                | pers                |               |           |             |            |                   |                       |          |  |  |     |
| auch      | auch           | auch           | j                | ap                  |               |           |             |            |                   |                       |          |  |  |     |
| gefallen  | gefallen       | fallen         | v                | vv                  |               |           |             |            |                   |                       |          |  |  |     |
| undt      | und            | und            | j                | kon                 | kg            |           |             |            |                   |                       |          |  |  |     |
| nahe      | nahe           | nahe           | adv              |                     | s             | hs        | kom_neg     | VORG       | vgr               | ke                    |          |  |  |     |
| ermordtet | ermordtet      | morden         | v                | vv                  |               |           |             |            |                   |                       |          |  |  |     |
| worden    | worden         | werden         | v                | lv_pass             |               |           |             |            |                   |                       |          |  |  |     |
| -         |                |                |                  |                     |               |           |             |            |                   |                       |          |  |  |     |
| so        | so             | so             | j                | subj                |               |           |             |            |                   |                       |          |  |  |     |
| mich      | mich           | ich            | p                | rf                  |               |           |             |            |                   |                       |          |  |  |     |
| mein      | mein           | mein           | d                | poss                |               |           |             |            |                   |                       |          |  |  |     |
| Gott      | Gott           | Gott           | n                |                     |               |           |             |            |                   | ns                    | adv_kond |  |  | ngr |
| nicht     | nicht          | nicht          | ptk              | neg                 |               |           |             |            |                   |                       |          |  |  |     |
| erhalten  | erhalten       | erhalten       | v                | vv                  |               |           |             |            |                   |                       |          |  |  |     |
| hette     | hätte          | haben          | v                | lv                  |               |           |             |            |                   |                       |          |  |  |     |

Abb. 6: Auszug einer Annotation des Guntzer in Excel

| GKN_Token | Form 1 | Funktion 1 | Form 2 | Funktion 2     | Form 3 | Funktion 3 | Form 4 | Funktion 4 |              |     |
|-----------|--------|------------|--------|----------------|--------|------------|--------|------------|--------------|-----|
| Uunder    | s      | dirum      | pgr    | ko             | ngr    | ke         | koord  |            |              |     |
| solche    |        |            |        |                |        |            |        |            |              |     |
| beße      |        |            |        |                |        |            |        |            |              |     |
| Buben     |        |            |        |                |        |            |        |            |              |     |
| undt      |        |            |        |                |        |            |        |            |              |     |
| Merdter   |        |            |        |                |        |            |        |            |              |     |
| bin       |        |            |        | prädi_vk:VORG  |        |            |        | vgr        | ko           |     |
| ich       |        |            |        | sub:VT         |        |            |        |            |              |     |
| auch      |        |            |        |                |        |            |        |            |              |     |
| gefallen  |        |            |        | <prädi_vk:VORG |        |            |        | <vgr       | ke           |     |
| undt      | kg     |            |        |                |        |            |        |            |              |     |
| nahe      | s      | adv_kond   | ns     | kom_neg        | ngr    | ke         | ke     |            |              |     |
| ermordtet |        |            |        | prädi_vk:VORG  |        |            |        | vgr        | ke           |     |
| worden    |        |            |        |                |        |            |        |            | ko           |     |
| -         |        |            |        |                |        |            |        |            |              |     |
| so        |        |            |        |                |        |            |        |            |              |     |
| mich      |        |            |        |                |        |            |        |            | ob_akk:HG    |     |
| mein      |        |            |        |                |        |            |        |            | sub:HT       |     |
| Gott      |        |            |        |                |        |            |        |            |              |     |
| nicht     |        |            |        |                |        |            |        |            | kom_neg      |     |
| erhalten  |        |            |        |                |        |            |        |            | prädi_vk:HAN | vgr |
| hette     |        |            | D      |                | ko     |            |        |            |              |     |

Abb. 7: Annotation der Baumstruktur in Excel

diert das verbreitete Tabellenkalkulationsprogramm Microsoft Excel. Der zu verarbeitende Text gestaltet sich somit als eine umfangreiche Tabelle, die die syntaktischen Informationen beinhaltet. Jede Zeile entspricht einem Token, also

einer Wortform bzw. einem Satzzeichen, des Texts, jede Spalte repräsentiert eine Annotations-ebene des Korpus. In Excel lassen sich Zellen verbinden, was sich für Spannotationen nutzen lässt, die für die syntaktische Analyse eine

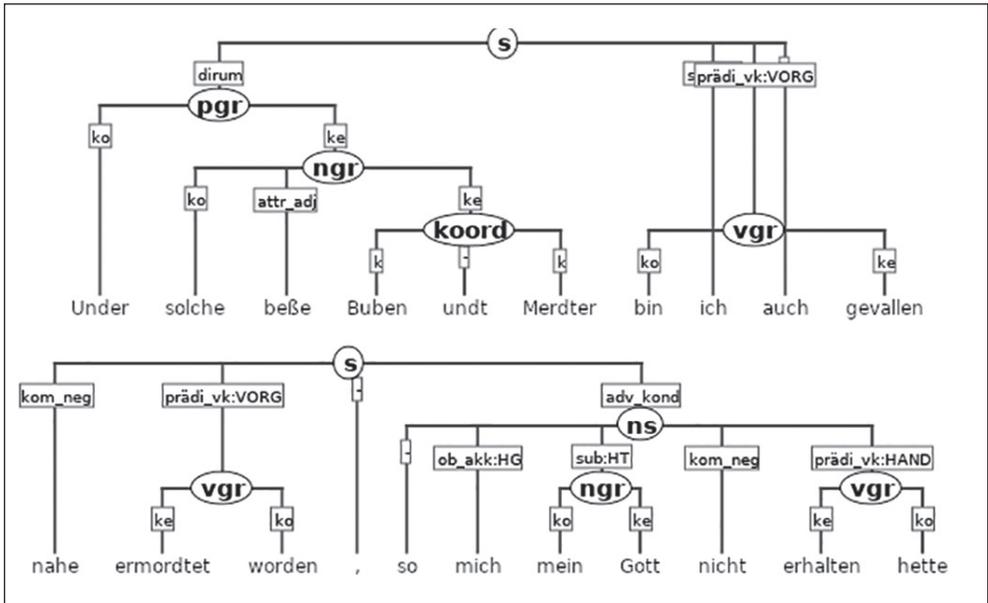


Abb. 8: Darstellung der Syntax im Baumformat in ANNIS

wichtige Grundlage bilden. So lassen sich Strukturen erfassen, die über mehrere Token hinaus reichen. In Abb. 6 sieht man am Beispiel eines Auszug aus Güntzer, wie eine solche Tabelle aussehen kann.<sup>4</sup>

Wie bei linguistischen Korpora üblich, enthält das GiesKaNe-Korpus eine Normalisierung und Lemmatisierung (grün), bei denen die Token an das Gegenwartsdeutsche angepasst bzw. ihre Grundform erfasst wird, was spätere Suchanfragen vereinfacht. Darüber hinaus umfasst es eine Annotation der Wortarten (orange). Das Herzstück des Korpus, die syntaktischen Informationen, gliedert sich in drei übergeordnete Ebenen – die Makro-, Meso- und Mikroebene. Terminologisch lehnt sich diese Segmentierung an Ágel (2017) an. Auf der Makroebene (blau) werden die größten syntaktischen Einheiten, die Textglieder, erfasst, es handelt sich dabei etwa um Sätze oder Nichtsätze (Ellipsen). Zusätzlich werden auf der Makroebene auch Teilsätze wie Haupt- oder Nebensätze getaggt (= als Analysekatgorie dem Text zugewiesen). Eine Ebene tiefer, auf der Mesoebene, erfolgt die Annotation der Satzglieder (pink), auf der Mikroebene werden

schließlich die kleinsten Einheiten, die Wortgruppen, dokumentiert (gelb).

In GiesKaNe ist die wichtigste Ebene des Mehrebenenmodelles eine Baumbank. Baumbanken bestehen aus sogenannten syntaktischen Bäumen, die sich für die Darstellung komplexer hierarchischer Strukturen wie Syntagmen in der Linguistik etabliert haben. „Ein Baum definiert sich durch seine Wurzel (engl. "root"), die über beliebig viele Verzweigungen (Knoten; engl. "node") mit den sogenannten terminalen Knoten, den Token bzw. Wortformen verbunden sind“ (Hirschmann 2019: 60). Auf dieser Form der Visualisierung aufbauend entstanden und entstehen umfassende Korpora, die Sätze im Baumformat abbilden, so auch ein Teil von GiesKaNe. Dies sei am Beispiel des ersten Satzes des Güntzer-Auszugs aus Abb. 6 kurz illustriert: *Under solche beße Buben undt Merdter bin ich auch gefallen*.

Dieser Satz ist die Wurzel eines Baums, dessen Komponenten die Knoten darstellen. Die Baumidee fortführend könnten die Knoten auch als Äste bezeichnet werden. Diese wachsen zunächst als Satzglieder. Unser Satz gliedert sich in eine direktionale lokale Bestimmung (*under sol-*

*che beßen Buben undt Merdter*), ein Prädikat (*bin gevallen*) und ein Subjekt (*ich*). Sowohl die direktionale Bestimmung als auch das Prädikat bestehen aus mehreren Wörtern, sie sind also Wortgruppen. Diese Gruppen lassen sich abermals zerlegen, da ihre Bestandteile wiederum Funktionen innerhalb der Wortgruppe einnehmen. Um im Bild zu bleiben, könnte man sich diese Funktionen als Zweige vorstellen, die von den Ästen sprießen. In der Verbalgruppe, die das Prädikat bildet, tritt das Hilfsverb *bin* in diesem Fall als sogenannter Kopf, das Partizip *gefallen* als sogenannter Kern der Wortgruppe auf. An dieser Stelle wären somit die Blätter des Baums erreicht, die Tokenebene bildet wie beschrieben den Abschluss der Analyse. Je komplexer der Satz und seine Wortgruppen sind, desto stärker verzweigt ist folglich der Baum.

Im Laufe der Annotation „verwachsen“ so die Makro-, Meso- und Mikroebene miteinander zu einem gemeinsamen Baum. So stellt Abb. 2 die rohe Form der Bäume in Excel dar, hier stehen Funktionen wie die Satzglieder, die vorher auf der Mesoebene erfasst wurden, neben ihren formalen Gegenstücken der Wortgruppen, die zunächst auf der Mikroebene annotiert wurden.<sup>5</sup>

Abbildung 8 zeigt schließlich die finale Visualisierung der Sätze im Baumformat. Durch eine Überführung der in Excel gespeicherten Daten in ein spezielles XML-Format wird GiesKaNe zur öffentlichen Ressource. Diese Bäume sind somit neben den weiteren im Korpus gesammelten Daten online über ANNIS abrufbar, wo das Korpus digital bereitgestellt wird. Als ein browserbasiertes Suchwerkzeug, das an der Humboldt-Universität zu Berlin entwickelt wurde, dient ANNIS als Plattform für vielschichtig annotierte linguistische Korpora (Krause/Zeldes 2016, siehe auch <https://corpus-tools.org/annis/>).

Die Veröffentlichung von Korpora ist im Bereich der Korpuslinguistik ein gängiges Verfahren, da sie sich so für die weiterführende Forschung nutzen lassen. Des Weiteren können Forschungsergebnisse auf diese Weise nachverfolgt und überprüft werden. ANNIS unterstützt dazu verschiedene Darstellungsformate, so auch die besagte Baumstruktur. Zur Erforschung sprachlicher Phänomene verfügt ANNIS zudem über eine umfangreiche Abfragespra-

che, die die Suche nach interessanten Entwicklungen innerhalb der Periode des Neuhochdeutschen im GiesKaNe-Korpus ermöglicht.

### Anmerkungen:

<sup>1</sup> Die Kommunikationsbeteiligten können je nach Textorte unterschiedlich stark ausgeprägt sein. So ist in einem Tagebucheintrag der Empfänger i.d.R. weniger wichtig und somit weniger präsent als in einem Brief. Die pragmatischen Bedingungen spiegeln sich in der Sprachoberfläche: Imperative, Anredenominative und der Gebrauch von Personalpronomina sowie Verben in der zweiten Person (Singular) machen den Empfänger und Dialogteilnehmer im Brief auf der Sprachoberfläche sichtbar; der Verzicht auf diese Mittel und der kumulative Gebrauch der 1. Person Singular kennzeichnen die Abwesenheit eines intendierten Rezipienten und die monologische Sprechsituation im Tagebuch.

<sup>2</sup> Mit der hier vorgestellten vereinfachten Darstellung soll nicht impliziert werden, es handle sich um rein additive Bausteine der Textprofilbildung. Genau genommen bestehen zwischen den einzelnen Bausteinen hierarchische Beziehungen: Die Rahmenbedingungen bedingen die Ausgestaltung von Form und Inhalt. Die einzelnen Parameter müssen theoretisch skalar gedacht werden. Im Rahmen dieses Artikels verzichten wir auf eine genauere Modellierung.

<sup>3</sup> „mindestens zwei (einfache oder komplexe) nominale Attribute in pränuklearer oder in postnuklearer Position, koordiniert oder parallel; mindestens ein komplexes Attribut (NG oder AdjG mit zusätzlicher Einbettung); zwei einfache Adjektive werden nicht gezählt“, ebd.

<sup>4</sup> Die Abbildung zeigt eine stark vereinfachte Form der Analyse, die für den Zweck der Anschaulichkeit angepasst wurde. Im Normalfall sind die Analysetabellen um einiges umfangreicher.

<sup>5</sup> Dass sich diese finale Form der Annotation erst im späteren Prozess ergibt, hängt mit unterschiedlichen Arbeitsabläufen zusammen, so umfassen die unterschiedlichen Ebenen etwa noch weitere grammatische Informationen, die über den Baum hinausgehen und daher in unterschiedlichen Schritten erfasst werden müssen.

### Literatur:

*GiesKaNe:*

<https://gieskane.com/>

[https://korpling.org/annis3/#\\_c=R2llc0thTmVfMC4x](https://korpling.org/annis3/#_c=R2llc0thTmVfMC4x)

*Primärquellen:*

Simmel = Simmel, Georg (1890): Über sociale Differenzierung: Sociologische und psychologische Untersuchungen. Leipzig: Duncker & Humblot.

Briefwechsel = (1871–72/1999): „Wenn doch dies Elend ein Ende hätte“: ein Briefwechsel aus dem Deutsch-Französischen Krieg 1870/71. Hrsg. von Isa Schikorsky. Köln et al.: Böhlau 1999 (Selbstzeugnisse der Neuzeit 7).

Güntzer = Güntzer, Augustin (1657/2002): Kleines Biechlin von meinem ganzen Leben. Die Autobiographie eines Elsässer Kannengießers aus dem 17. Jahrhundert. Hrsg. von Fabian Brändle und Dominik Sieber. Köln/Weimar: Böhlau 2002 (Selbstzeugnisse der Neuzeit 8).

Söldnerleben = (1625–1649/1993): Ein Söldnerleben im Dreißigjährigen Krieg. Eine Quelle zur Sozialgeschichte. Hrsg. v. Jan Peters. Berlin: Akademie-Verlag 1993 (Selbstzeugnisse der Neuzeit 1).

Zimmer = (1861–64/2001): Michael Zimmer's Diary. Ein deutsches Tagebuch aus dem Amerikanischen Bürgerkrieg. Hrsg. von Jürgen Macha und Andrea Wolf. Frankfurt/Main et al.: Lang 2001 (Sprachgeschichte des Deutschen in Nordamerika 1).

#### *Sekundärliteratur:*

Ágel, Vilmos 2000: Syntax des Neuhochdeutschen bis zur Mitte des 20. Jahrhunderts. In: Besch, Werner/Betten, Anne/Reichmann, Oskar/Sonderegger, Stefan (Hrsg.): Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung. Bd. 2. 2. Aufl. Berlin/New York: de Gruyter (Handbücher zur Sprach- und Kommunikationswissenschaft 2.2), 1855–1903.

Ágel, Vilmos 2017: Grammatische Textanalyse. Textglieder, Satzglieder, Wortgruppenglieder. Berlin/Boston: de Gruyter.

Elspaß, Stephan 2005: Sprachgeschichte von unten. Untersuchungen zum geschriebenen Alltagsdeutsch im 19. Jahrhundert. Tübingen: Niemeyer (Reihe Germanistische Linguistik 263).

Hennig, Mathilde 2009: Nähe und Distanzierung. Verschriftlichung und Reorganisation des Nähebereichs im Neuhochdeutschen. Kassel: University Press.

Hirschmann, Hagen 2019: Korpuslinguistik. Eine Einführung. Stuttgart: J.B. Metzler.

Knoop, Ulrich 1994: Entwicklung von Literalität und Alphabetisierung in Deutschland. In: Günther, Hartmut/Ludwig, Otto (Hrsg.): Schrift und Schriftlichkeit. Writing and Its Use. Berlin/New York: de Gruyter (Handbücher zur Sprach- und Kommunikationswissenschaft 10.1), 859–872.

Koch, Peter/Oesterreicher, Wulf 1994: Schriftlichkeit und Sprache. In: Günther, Hartmut/Ludwig, Otto (Hrsg.): Schrift und Schriftlichkeit: ein interdisziplinäres Handbuch internationaler Forschung = Writing and its use. Berlin/New York: de Gruyter (Handbücher zur Sprach- und Kommunikationswissenschaft 10), 587–604.

Krause, Thomas/Zeldes, Amir 2016: ANNIS3: A new architecture for generic corpus query and visualization. In:

Digital Scholarship in the Humanities 2016 (31). <http://dsh.oxfordjournals.org/content/31/1/118>

Lötscher, Andreas 2016: Komplexe Attributierung als Element von Textstilen im diachronen Vergleich. In: Hennig, Mathilde (Hrsg.): Komplexe Attribution. Ein Nominalstilphänomen aus sprachhistorischer, grammatischer, typologischer und funktionalstilistischer Perspektive. Berlin/Boston: de Gruyter (Linguistik – Impulse & Tendenzen 63), 353–390.

Oesterreicher, Wulf 1993: Verschriftung und Verschriftlichung im Kontext medialer und konzeptioneller Schriftlichkeit. In: Schaefer, Ursula (Hrsg.): Schriftlichkeit im frühen Mittelalter. Tübingen: Narr (Script Oralia 53), 267–292.

Polenz, Peter von 1994: Deutsche Sprachgeschichte vom Spätmittelalter bis zur Gegenwart: Band II: 17. und 18. Jahrhundert. Berlin/New York: de Gruyter (de Gruyter Studienbuch).

Polenz, Peter von 1995: Sprachsystemwandel und sozio-pragmatische Sprachgeschichte in der Sprachkultivierungsperiode. In: Gardt, Andreas/Mattheier, Klaus/Reichmann, Oskar (Hrsg.): Sprachgeschichte des Neuhochdeutschen: Gegenstände, Methoden, Theorien. Tübingen: Niemeyer (Reihe Germanistische Linguistik 156), 39–67.

Reichmann, Oskar 1988: Zur Vertikalisierung des Varietätenspektrums in der jüngeren Sprachgeschichte des Deutschen. In: Munske, Horst, Haider et al. (Hrsg.): Deutscher Wortschatz. Lexikologische Studien. Ludwig Erich Schmitt zum 80. Geburtstag von seinen Marburger Schülern. Berlin/New York: de Gruyter, 151–180 [unter Mitwirkung von Christiane Burgi, Martin Kaufhold und Claudia Schäfer].

#### *Kontakt:*

Prof. Dr. Mathilde Hennig  
Justus-Liebig-Universität Gießen  
Institut für Germanistik  
Otto-Behaghel-Straße 10B  
35394 Gießen



Terminvereinbarung  
erwünscht

Einrahmungen  
Restaurationen  
Künstlerbedarf



Ludwig-Rinn-Straße 14–16 | Haus D  
(im Rinn & Cloos Gewerbepark)  
35452 Heuchelheim bei Gießen  
☎ (0641) 13 18 51 69 | [www.artikum.com](http://www.artikum.com)

Mo.–Fr. von 15–18 Uhr  
jeden ersten + letzten Samstag von 10–13 Uhr