



Apologies in South Asian Varieties of English: A Corpus-Based Study on Indian and Sri Lankan English

Julia Degenhardt¹ · Tobias Bernaisch²

Received: 5 May 2021 / Accepted: 3 February 2022 / Published online: 25 February 2022
© The Author(s) 2022

Abstract

Despite notable recent exceptions (for instance Funke, 2020), corpus-based research into South Asian Englishes has so far concentrated on structural features. Thus, empirical pragmatic research in said varieties is still sparse, although there has been increasing interest in variational pragmatics in world Englishes (Schneider & Barron, 2008). For instance, requests have been examined in both native and non-native varieties of English. However, studies on apologies have largely focussed on first-language varieties of English (Deutschmann, 2003). Against this background, the present study investigates apology patterns in two South Asian second-language varieties of English, Indian and Sri Lankan English, and their historical input variety British English. With the help of the spoken parts of the respective components of the International Corpus of English, multifactorial analyses—including an improved form of random forests that explicitly takes interactions between several predictors into account—model the choice of *sorry* as opposed to other apology forms. Findings suggest quantitative differences in the use of *sorry* which are influenced by factors such as type of apology, topic and age or combinations of said factors. In sum, this study suggests that apology forms and frequencies are sensitive to the speakers' regional background and sociobiographic factors as well as to structural and contextual parameters.

Keywords South asian englishes · Indian english · Sri Lankan english · Apologies · Random forest · Multifactorial analysis

✉ Julia Degenhardt
Julia.Degenhardt@anglistik.uni-giessen.de
Tobias Bernaisch
Tobias.J.Bernaisch@anglistik.uni-giessen.de

¹ Department of English, Justus Liebig University Giessen, Otto-Behaghel-Str. 10B, Room B404, 35394 Giessen, Germany

² Department of English, Justus Liebig University Giessen, Otto-Behaghel-Str. 10B, Room B415, 35394 Giessen, Germany

Introduction

Apologies are needed when “social norms have been violated” (Trosborg, 1995, p. 373) in a communicative context. When speakers have offended their interlocutor in any way, they feel the need to “set things right” (Kitao & Kitao, 2013, p. 1). In this light, apologies can be regarded as routinized means of redressive actions and have been researched extensively in the English language, comparing it to other languages and cultures (Blum-Kulka et al., 1989; Ogiermann, 2009). Although corpus-linguistic perspectives on apology routines in English are available (Jucker, 2018 on historical American English (AmE) or Lutzky & Kehoe, 2017 on blog data), these have largely focussed on native speakers of English (with notable exceptions such as Barron, 2019 on German learners of English) while accounts of apologies as used in New Englishes have so far been lacking. Nevertheless, investigating apologies in first-language varieties of English as compared to apologies in Postcolonial Englishes is essential to an understanding of pragmatic routines in world Englishes. Cultural differences might crystallize more easily with pragmatic routines than with structural elements, say, at the lexis-grammar interface. Therefore, this paper takes into account both a pragmalinguistic and a sociopragmatic point of view on apologies in Indian English (IndE) and Sri Lankan English (SLE), and thus aims to fill this research gap by employing multifactorial statistical approaches, shedding light on the predictors influencing the choice of apology strategies in both South Asian Englishes (SAEs) compared to their historical input variety British English (BrE).

With the aim of establishing connections with earlier findings on apologies reported in the ‘[Apologies in Varieties of English](#)’ section, the following research questions address central aspects to be empirically explored in the ‘[Analysis](#)’ section of this paper:

- (1) Are variety-specific (BrE, IndE or SLE on their own) or pan-South Asian (IndE and SLE as compared to BrE) apology patterns regarding both frequency and form present in the varieties studied?
- (2) Are frequencies of apologies generally sensitive to factors such as age, gender or regional background of the speakers?
- (3) Which (other) factors influence the choice of one apology over another and how and to what extent do these factors influence apology choice?

In the ‘[Apologies in Varieties of English](#)’ section, previous research on structural realisations of apologies in English and influential factors including sociobiographical characteristics such as age, gender or regional background is presented. The ‘[Corpus Data and Data Extraction and Annotation](#)’ section describes the corpus data and methods for data extraction and annotation. In the ‘[Analysis](#)’ section, we present results from analyses involving a conditional inference tree and a random forest, complemented with qualitative perspectives. The ‘[Discussion](#)’ section discusses selected findings and caveats of the present study and suggests avenues for future research.

Apologies in Varieties of English

Apologies are “redressive speech act[s] for a face-threatening act (FTA)” (Kasanga & Lwanga-Lumu, 2007, p. 65), meaning they mitigate a possible violation of social norms. Amongst others, speakers apologise to express regret for a previous action and apologies can consequently be viewed as expressive speech acts (Searle, 1979); they “can be defined as compensatory action for an offence committed by S [speaker] which has affected H [hearer]” (Márquez Reiter, 2000, p. 44). Thus, an apology’s main function is to save a communicative situation following an act usually causing a disruptive offence.

For Brown and Levinson (1978), apologies are a prime example of a negative politeness strategy: they redress the degree of imposition, thus the risk of face-loss, on the hearer’s individual right to go about unaffectedly. For Leech (2014), apologies are part of positive politeness since “intensifying modifiers can be added or further intensified to increase the degree of (pragmalinguistic) politeness” (Leech, 2014, p. 120). As such, apologies represent “concern for others” (Leech, 2014, p. 132).¹ Hence, as Spencer-Oatey et al. (2008) point out, both Brown and Levinson’s and Leech’s views on apologies ignore the fact that an apology can also be a face-threatening act for the speaker since the speaker needs to admit to having caused an offence. So-called on-record apologies are often mitigated to decrease the risk of face loss for the speaker and the perception of apologies might also be further influenced depending on whether speakers are part of individualist or collectivist cultures (see the ‘[The Sociopragmatic View on Apologies](#)’ section). In this light, Leech (1983, 2014) suggests investigating politeness-sensitive speech acts such as apologies from both a pragmalinguistic and a sociopragmatic point of view, which we present in the following.

The Pragmalinguistic View on Apologies

The pragmalinguistic view on apologies is primarily concerned with possible structural realisations of apologies. A prominent research project in this context is Blum-Kulka et al.’s (1989) Cross-Cultural Study of Speech Act Realization Patterns (CCSARP). With the help of discourse completion tasks (DCTs), they extract three main strategies to realise the illocutionary force indicating device (IFID) of apologies (Blum-Kulka et al., 1989):

- (1) Expressing regret.
- (2) Offering an apology.
- (3) Requesting forgiveness.

¹ Pos-politeness, as described by Leech (2014: 12), “gives or assigns some positive value to the addressee [...]: the speaker makes a positive gesture to cancel out an imbalance favo[u]ring the speaker or disfavo[u]ring the hearer”.

Expressing regret is by far the most common strategy, usually involving a declarative sentence structure as in *I'm sorry I hurt you* or simply *Sorry* (Leech, 2014, 125ff.). Offering an apology can include performative verbs, such as in *I beg your pardon*. Requesting forgiveness is usually realised by means of a highly conventionalised and thus accepted imperative construction (*Excuse me* or *Pardon me*) (Leech, 2014, p. 127).

Apologies can also come in much longer speech sequences. For instance, in DCTs filled in by native speakers of English Blum-Kulka and Olshtain (1984) find that “the act of apologizing can take one or two basic forms, or a combination of both” (Blum-Kulka & Olshtain, 1984, 206ff.):

- (1) an explicit IFID, including a performative verb such as (*be*) *sorry*, *excuse*, *apologize*, *forgive*, *regret*, *pardon* (also called “head act” by Leech (2014, p. 116))
- (2) utterance(s) relating to one or more of the following four potential strategies (also called “satellite speech events” (SSEs) by Leech (2014, p. 116)):
 - (a) explaining the cause of the offence
 - (b) expressing one’s responsibility for the offence
 - (c) offering repair
 - (d) promising forbearance

While forms with explicit IFIDs include formulaic expressions that are accepted means of apologising (Leech, 2014, p. 125ff.), the four strategies for SSEs can come in “an open-ended variety of utterances” (Blum-Kulka & Olshtain, 1984, p. 207). Individual speaker decisions on how to realise apologies based on the taxonomy above are influenced by “cultural, personal and contextual elements” (Blum-Kulka & Olshtain, 1984, p. 209).

Furthermore, apologies can come with so-called supporting moves which include, for instance, intensifiers (such as *very*, *so* (*much*) or *terribly*) to signal a deeper regret on the speaker’s side (Márquez Reiter, 2000, p. 54). With the help of role-plays among speakers of Uruguayan Spanish and BrE, she finds that the combined usage of an IFID and an expression of responsibility is cross-culturally independent of the situation in which the apology occurs while other possible combinations of apology strategies are dependent on the situation (Márquez Reiter, 2000, p. 179). For example, BrE speakers “show a marked preference for the ‘I’m sorry’ lexical phrase in its intensified form”, which “seems to be a convention representing a ritualised Anglo-Saxon conflict avoidance strategy aimed at redressing the hearer’s ‘negative’ face” (Márquez Reiter, 2000, p. 167). Olshtain and Cohen (1990, p. 57), too, stress the importance of intensification by arguing that any part of an apology that lacks appropriate intensification might sound insincere.

Corpus-based research into apologies was undertaken by, for instance, Deutschmann (2003) and Kitao and Kitao (2013). The latter look at apologies in the American sitcom *Modern Family* by eliciting IFIDs from a subtitle corpus. They find that *sorry* is used almost exclusively compared to other IFIDs and—in

Table 1 Possible realisations of apologies, adapted from Deutschmann (2003, 53ff.)

Syntactic form	Examples
“Detached” apologies	
A1. Detached, ellipsed form, and/or simplest possible form	A1. Sorry(?!), Pardon (?!), Excuse me (?!)
A2. Partially/fully expanded, detached	A2. I’m sorry (?), I beg your pardon (?)
“Detached” apologies + markers	
B. Interjection (such as downtoner or hesitation marker) + apology	B. Well, pardon me!/Erm, sorry.
C. Explicit apology + proper name or epithet	C. Sorry, Bob./Sorry, darling.
D. Intensifier/emphatic “do” + apology	D. I’m really sorry./I do apologise.
E. Apology + <i>please</i>	E. Forgive me, please!
Syntactically complex forms	
For example <i>Sorry + to</i> + verb phrase or request form	Sorry to say this. Would you excuse me?

about a quarter of the cases—it is used without any SSEs (Kitao & Kitao, 2013, p. 6). Once a SSE accompanies the IFID, intensification is used in about 15% of the apologies.

Deutschmann (2003) offers a fine-grained taxonomy on possible apology realisations (Table 1). His prototypical apology contains a recognition of the offence, an acceptance of responsibility and, most importantly, an explicit “expression of regret”, that is an IFID (Deutschmann, 2003, p. 45). Searching for these “explicit apologies which appeared in the form of illocutionary force indicating devices” (Deutschmann, 2003, p. 17), he conducted a corpus-based study of apologies in the dialogue part of the spoken component of the British National Corpus (BNC). However, not all apologies take on this arguably prototypical form. Therefore, he offers a three-part non-prototypical categorisation, comprising “formulaic”, routinized apologies for minimal offence situations, “formulaic apologies with added functions”, which involve “request cues” or “attention cues”, and “face attack” apologies (Deutschmann, 2003, p. 46). It transpires that about half of the cases studied are “‘formulaic’ apologies [...] often consist[ing] of simple, syntactically detached IFIDs” (Deutschmann, 2003, p. 204), namely apologies consisting of the explicit head act only.

Moreover, Deutschmann (2003, p. 206) makes the more general observation that speakers’ apology choices are indicative of their sociolect. Therefore, investigations of apology behaviour should also include sociopragmatic perspectives.

The Sociopragmatic View on Apologies

The sociopragmatic view considers “various scales of value that make a particular degree of politeness seem appropriate or normal in a given social setting” (Leech, 2014, p. 14). Hence, apologies can be considered sensitive to the speech communities in which they are used.

Generally, there is a difference between individualist and collectivist cultures. The former “have an independent concept of self”, the latter “have an interdependent concept of self” (Bhawuk, 2017, p. 1) and there is a tendency in Western cultures towards individualism and in most Asian or African cultures towards collectivism (Bhawuk, 2017, p. 3). Incidentally, in a study on politeness in South African Indian English, Bharuthram (2003, p. 1525) finds that speakers are rather concerned with group identity than with individualism.

Ever since Schneider and Barron (2008) introduced the field of variational pragmatics, there has been an increasing interest in the roles of sociobiographic and contextual factors in pragmatic variation (Barron, 2008; Barron & Schneider, 2009). In the search of triggers for different realisations of speech acts, Barron (2005) suggests that factors such as age and gender play a role in the choice of strategies and their frequencies in varieties of English, though “intra-lingual pragmatic variation does not generally affect the inventory of strategies nor the modification devices available for use” (Barron, 2005, p. 530). Although limitations of space do not permit extensive discussion of (the development of) earlier research in variational pragmatics, we want to acknowledge that situational variables, such as social distance, the degree of imposition or topic (Barron, 2017b; Staley, 2018), are without a doubt important factors to take into account at the interface of (variational) pragmatics and world Englishes (Barron, 2017a).

Already in 1989, Blum-Kulka et al.’s CCSARP concludes that a similar set of apology strategies is available in all the languages and cultures they studied;² yet the appropriateness of various strategies in certain situations, the amount of apologies uttered and whether the apologies were intensified turned out to be context- and culture-dependent.

Based on DCTs, Ogiermann (2009) studies apologies in English and two Slavic languages and finds that “apologies serving the function of negative politeness strategies have broader applicability in British culture” than in both Slavic speech communities, indicating that “the potential of a particular strategy to serve as an apology is culture-specific” (Ogiermann, 2009, pp. 261–262).

With regard to age and gender differences in apologies, earlier research has led to rather mixed insights. While some researchers found a more frequent use of apologies in male speech (Mattson Bean & Johnstone, 1994), other studies detected no significant gender difference at all (Aijmer, 1995; Deutschmann, 2003) and yet other studies found that, throughout various settings, women apologise more frequently than men (Tannen, 1994), especially to other women, while men do not apologise as often, neither to other men nor to women (Holmes, 1989). This might be because men perceived apologies as self-oriented FTAs (which led them to apologise less often) whereas women perceived them as hearer-oriented “ways of facilitating social harmony” (Holmes, 1989, p. 208). However, qualitative differences are not necessarily apparent (Márquez Reiter, 2000, p. 165).

² For more information on the languages and cultures investigated, please refer to Blum-Kulka et al. (1989).

In his BNC study already introduced above, Deutschmann (2003) also investigates the influence of social variables such as age, gender and social class of the speakers. Despite finding an interaction effect between gender and formality, stating that female speakers apologise more often the more formal the situation is perceived to be while the opposite is the case with male speakers (Deutschmann, 2003, p. 205), gender only plays a minor role. Regarding age and social class, “younger speakers in the corpus apologi[se] far more frequently than older speakers, as [do] middle-class speakers compared to working-class speakers” (Deutschmann, 2003, p. 205).

Apologies in Indian and Sri Lankan English

While research on apologies in varieties of English has increased in the last decades, SAEs—with notable exceptions introduced in the following—have not been studied extensively in that respect.

Sridhar (1991) finds that IndE speakers who are exposed to English through education and media (Westernized Indians in his terminology) realise speech acts more similarly to speakers of BrE than non-Westernized Indians. Moreover, while IndE speakers are more hearer-oriented in their apology choice, BrE speakers are more interlocutor-oriented, meaning that BrE speakers are more likely to save the interlocutor’s negative face despite uttering a possibly face-threatening act to the speaker’s own face when apologising.

With data from Indian novels, Tinkham (1993, p. 245) finds that non-Westernized Indians do not use *I’m sorry* at all. Moreover, Indians bilingual in English and a local language, presumably with greater exposure to the Western world, tend to use *sorry* even while communicating in their local L1s (Sailaja, 2009, p. 89), which is compatible with more Westernized IndE speakers showing a more frequent use of *sorry*.

In a theoretical account of “How to be polite in Indian English”, Mehrotra (1995, 99ff.) states that acceptable pragmatic behaviour in IndE differs considerably from BrE. While IndE features fewer fixed forms of polite articulation than BrE, IndE speakers often opt for longer utterances as they are locally more acceptable than shorter utterances. As apologies are at the core of polite behaviour, one could also assume that a longer apology would be more acceptable, which is why the current study anticipates localised facets in IndE apology strategies.

Corpus-based research into pragmatics in SAEs has lately been on the increase, for example on the nativisation of thanking strategies (Funke, 2020), realisations of backchannels (Kraaz & Bernaisch, 2020) and request patterns (Degenhardt, 2020). While results of research on IndE apologies are provided above, apologies have not been studied in SLE, and up to this point, apologies have generally not yet been investigated corpus-linguistically across SAEs. Brown and Levinson (1978, p. 190), however, mention that speakers of Tamil, which is spoken in India and Sri Lanka, often apologise by using the Tamil equivalent of *you should/must forgive me*. This has implications for possible search expressions because IndE or SLE speakers with Tamil as their L1 may adopt strategies transferred from their first language. More

Table 2 ICE corpus sections used

Private dialogues (100)	Face-to-face conversations (90) Phone calls (10)
Public dialogues (80)	Classroom Lessons (20) Broadcast Discussions (20) Broadcast Interviews (10) Parliamentary Debates (10) Legal cross-examinations (10) Business Transactions (10)

details on search expressions of the present study will be introduced in the next section.

Corpus Data and Data Extraction and Annotation

This study uses the spoken parts of three components of the International Corpus of English (ICE): ICE-Great Britain (ICE-GB) with 624,342 words, ICE-India (ICE-IND) with 648,414 words and ICE-Sri Lanka (ICE-SL) with 584,276 words, where extra-corpus material, editorial comments as well as untranscribed and unclear text have not been counted respectively. As apologies are more likely to occur in actual exchanges between speakers rather than in monologues, this study resorts to the private and public dialogue sections in ICE (Table 2).

The ICE components represent an unprecedented database for the empirical comparison of (apologies in) BrE, IndE and SLE—also in the light of the socio-biographic speaker information they offer, but a corpus-based approach to the study of apologies requires pre-defined search strings, which is challenging in that apologising cannot always be clearly separated from other speech acts. Yet, previous research (Blum-Kulka & Olshtain, 1984; Deutschmann, 2003) documents relevant IFID search strings, the use of which ensures—to a certain degree—comparability between former results and the results reported here.

Leech's (2014) terminology of head act and SSEs may be interpreted as an indication that apologies are constituted by their combination and that SSEs on their own should not be considered apologies, but this is not the case. As mentioned in the 'The Pragmalinguistic View on Apologies' section, SSEs open up a nearly infinite number of possible utterances that cannot be accounted for within a corpus-based approach, but in a given context, utterances such as "My car broke down" can serve as fully acceptable apologies even without a head act. Consequently, we understand apologising "as covering a particular '*illocutionary territory*' with internal variation as well as contrastive relations with other speech events" (Leech, 2014, p. 119, italics in original). Therefore, the data from ICE were extracted using relevant IFIDs for apologies based on predefined single-word search terms. To account for potentially variety-specific means of apologising, we also read a sample of ten files of each of the components in order to identify possible localised apology strategies not covered by the predefined list

of IFIDs. On the one hand, this led to the inclusion of the IFIDs laid out in the theoretical part of this paper (*apologize, excuse, forgive, pardon, regret, sorry*). On the other hand, we also searched for *apologise, apologies, apology* to cover all variants, and – based on our close reading of the data – we included *accident, afraid, fault, mean(t) to* and *mistake* (examples (1) and (2)).

- (1) <ICE-GB: S2B-002#56:1:E> I've heard it suggested that the markets have already discounted a one or two per cent reduction in interest rates <,>
<ICE-GB: S2B-002#57:1:E> I'm afraid I don't buy that <,,>
- (2) <ICE-GB: S1A-022#175:1:D> I wore this one then <,> Naomo Naomo
<ICE-GB: S1A-022#176:1:D> I meant to give it to you earlier
<ICE-GB: S1A-022#177:1:D> It's a bit late now <,,>

Arguably, some researchers may consider instances involving *accident, afraid, fault, mean(t) to* and *mistake* rather SSEs, and thus indirect means of apologising, as opposed to head acts. Still, as these utterances illustrated in examples (1) and (2) are locally produced and accepted as apologies, they certainly cover the “illocutionary territory” (Leech, 2014, p. 199) of apologising, which is why we opted to include rather than to discard them. Moreover, though example (2) shows an instance of what one could term an indirect apology, it can nevertheless constitute the core part of an apology, which can be followed by other SSEs. The latter, then, only derive their meaning as a SSE in the context of the core apology. The inclusion of instances of apologies including *accident, afraid, fault, mean(t) to* and *mistake* leads to a more thorough analysis of apology strategies that can be found based on common search terms—independent of whether they lead to a direct or indirect apology. By corpus-linguistic necessity, the data analysed in the present study have been extracted with the help of word forms expressing apologies within the core apology act since the corpora used have not been pragmatically annotated with regard to apologetic speech acts, hindering a non-manual speech act-based mode of data extraction. Still, for each example extracted, its context has been checked to ensure that the forms function as apologies in the concrete utterances.

Consequently, instances of, for example, “I was afraid of him [...]” (<ICE-SL:S1A-072#52:1:B>) were excluded from the data. Moreover, though we agree that a routinized *Sorry?*, *Pardon?* or *Excuse me?* can be viewed as a non-prototypical apology, we believe that a difference needs to be made between IFIDs that—in their specific communicative context—rather instantiate a request than an apology, for example in the sense of requesting the repetition of a previously unintelligible utterance, which we discarded as in (3), as opposed to IFIDs that represent sincere apology behaviour.

- (3) <ICE-IND:S1A-038#31:1:A> They haven't come over here you see
<ICE-IND:S1A-038#32:1:B> Pardon <,>

Table 3 Absolute and relative frequencies of apologies in ICE-GB, ICE-IND and ICE-SL

	ICE-GB		ICE-IND		ICE-SL	
	Abs.	Rel. (%)	Abs.	Rel. (%)	Abs.	Rel. (%)
<i>Sorry</i>	204	72.08	90	69.23	200	86.21
Other	79	27.92	40	30.77	32	13.79

<ICE-IND:S1A-038#33:1:A> They they haven't come here no <,>

In all three components, *forgive* occurred ten times in total, but its contexts as in (4) suggested that these cases should rather be considered requests than apologies. Consequently, we did not consider the examples with *forgive* at hand in our analyses.

- (4) <ICE-GB:S1B-022#131:1:F> I think one of the main areas of crisis is
 <ICE-GB:S1B-022##132:1:F> and the minister will forgive me saying this
 uhm
 <ICE-GB:S1B-022##133:1:F> but local authorities put in as m put as much
 into the arts in London as do uh central government

This procedure led to the extraction of a total of 645 apologies (Table 3). As the paper focusses on the concrete apology form a speaker chooses in a given communicative setting, each apology in an utterance was recorded as a separate apology in case a speaker's utterance contained multiple apologies.

As previous research has shown, *sorry* is by far the most frequent apology realisation. Therefore, we investigated apologies including *sorry* versus apologies including any other possible realisation including *afraid*, *apology/apologies/apologise/apologize*, *accident*, *excuse*, *fault*, *mean(t) to*, *mistake*, *pardon* and *regret* to capture the variability between the dominant *sorry* and other apology forms.

The independent variables included in this study cover various (socio-)linguistic variables. Their levels show the respective relative frequencies for *sorry*:

- VARIETY: the variety of English spoken, either BrE (GB; 72.08%), IndE (IND; 69.23%) or SLE (SL; 86.21%),
- GENDER: the speaker's gender, either *female* (78.97%) or *male* (76%),
- GROUPGENDER: the gender composition of other speakers present, either the *same* (80.69%) as the speaker or *different* (77.17%) from the speaker's gender,
- AGE: the speaker's age, either *younger* (< 26; 78.57%) or *older* (≥ 26; 77.71%),³
- INTENSIFIER: whether (*yes*; 82.61%) or not (*no*; 76.37%) the apology is intensified,
- SETTING: the communicative context, either *private* (79.19%) or *public* (74.33%),

³ Please note that there is a considerable time gap between the compilations of ICE-GB/ICE-IND on the one hand and the more recently released ICE-SL on the other hand, meaning that the data representing BrE and IndE are not as up-to-date as they could be.

- TYPE: the type of the apology, categorised as either single word unit (*swu* (93.37%), such as *sorry* or *pardon*) or multiword unit (*mwu* (50.59%), such as *I must apologise*); examples like *very sorry* were categorised as intensified *swu* and *I am very sorry* were counted as intensified *mwu*,
- TOPIC: the overall topic of the apologisee's text: 1) *courtroom* (64.86%), 2) *government* (37.5%), 3) *humanities* (81.82%), 4) *legal* (78.69%), 5) *news* (67.74%), 6) *personal* (77.91%), 7) *political* (58%), 8) *research* (92.21%), 9) *school* (79.12%),⁴
- TTR: the type-token ratio (number of types divided by number of words multiplied by 100) of the apologisee as a measure of lexical diversity: 1) *low* ($\leq 33.33\%$; 78.75%), 2) *medium* ($> 33.33\%$ and $< 66.66\%$; 73.45%) or 3) *high* ($\geq 66.66\%$; 82.35%).

The predictor `GROUPGENDER` serves as an approximation to the gender of the apologisee because some previous studies have attested a difference in apology use in same-sex and different-sex interactions. All interlocutors who—according to the metadata—took part in the conversation are collapsed into `GROUPGENDER` to account for whether gender homogeneity/heterogeneity influences apology choices.

Similarly, the predictor `AGE` can give valuable insights into language use of younger and older speakers. Still, the cut-off point between the speakers the present study treats as younger as opposed to older can be criticised in that the cut-off point is not a balanced product of age distributions across the datasets, but a pragmatic choice resulting from the fact that the age of 26 was the only shared boundary featured in the metadata for age for ICE-GB and ICE-IND, which report speaker age in bands instead of exact numeric values.

The data is also coded for whether the apology appears in a private or a public setting. We believe that speakers apologise differently (maybe even less) in private conversations as the risk of face loss might be lower among friends or family members than, for example, amongst business partners.

Analysis

To address the research questions listed at the end of the 'Apologies in Varieties of English' section, the frequencies of apologies are modelled via different statistical techniques. Descriptive statistics using normalised frequencies are presented to gain a relatively global understanding of how apologies are distributed across the

⁴ Topic modelling resorted to Latent Dirichlet Allocation (LDA) available in the R package `topicmodels`. After the specification that nine different topics be modelled in the light of maximum semantic topic coherence, each word in the data was first grouped into a particular topic randomly. Subsequently, the topic assignment of each word was iteratively adjusted under consideration of the word's pervasiveness across the topics and the pervasiveness of the topics in the data (https://cbail.github.io/SICSS_Topic_Modeling.html). In a last step, topic labels were derived from prominent words of each topic. Generally, `TOPIC` captures the content of speaker texts more readily than the rather language-external ICE-based genre distinctions. We are grateful to Benedikt Heller for his support with regard to topic modelling.

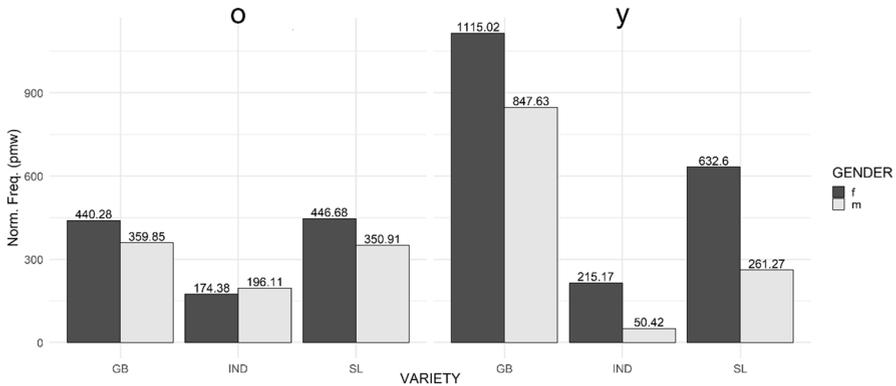


Fig. 1 Normalised frequencies (pmw) of apologies according to AGE, GENDER and VARIETY

varieties concerned. Focussing on the choice between the apology *sorry* as opposed to any other structural realisation of an apology such as *excuse me*, *pardon*, etc., a conditional inference tree (CIT) is used to profile which speaker groups have a tendency towards employing *sorry* instead of other apology forms. This multifactorial perspective on the two choices is complemented by a (recently improved type of) random forest (RF) analysis, which allows gauging variety-specific differences in the importance of the predictor variables for apology choice and provides measures of how different levels of a predictor influence the frequency of said choice.

The normalised frequencies of apologies per million words (pmw) are visualised in Fig. 1 and provide a first descriptive impression of how apologies are distributed across the age and gender groups across the varieties. In the left half of the plot, the frequencies of apologies by older speakers are shown while those for younger speakers are plotted on the right. For each age group, two bars are shown per variety with dark bars representing the frequencies of female speakers and light bars representing the frequencies of male speakers in the VARIETY-and-AGE groups concerned.

Apology frequencies range from 50.42 (pmw) for younger IndE males to 1115.02 (pmw) for younger BrE women. Regarding overall age differences in their frequencies of occurrence, apologies are generally more frequent with younger speakers across varieties and genders to the exception of IndE and SLE males. The bars also visually suggest that—on average—apologies are more frequent in BrE than in the SAEs, which is also borne out by the normalised apology counts per VARIETY independent of AGE and GENDER with BrE (453.28 pmw) featuring more apologies than IndE (200.49 pmw) and SLE (397.07 pmw). When it comes to speaker GENDER, women apologise more frequently than men across all VARIETY-and-AGE groups except for older IndE speakers.

While these global perspectives may serve as a first portrait of the variability in the apology data, they should be complemented with more robust multifactorial statistical approaches that consider all the predictors coded and model their (potential) effects on the response variable jointly. When only those instances for which speaker information on AGE and GENDER are documented are retained, 380 examples are

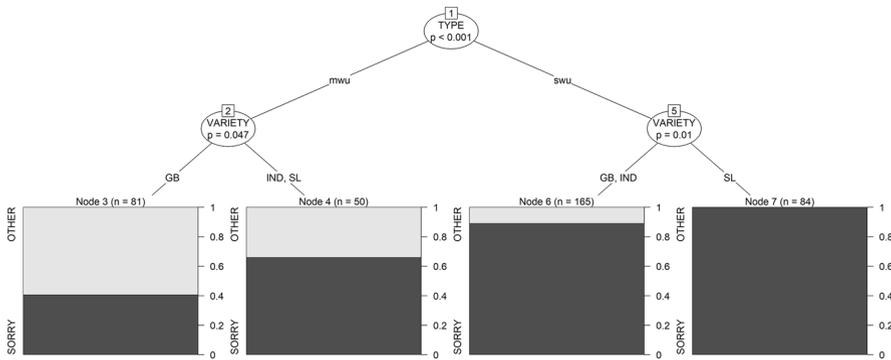


Fig. 2 CIT for *sorry* vs. *other* apology

available for analysis. As the apology *sorry* is used in 297 (78.16%) of these cases, and the remaining 21.84% of apologies are shared among *I'm afraid*, *I apologise*, etc. it is theoretically meaningful and statistically conducive to ask which speakers groups tend towards using *sorry* as opposed to one of the other apology forms. To model such a choice, a CIT can be constructed.

With their origin in the seminal work by Morgan and Sonquist (1963) on automated interaction detection, CITs (see, for example Levshina, 2015 for a practical introduction) are applicable to a wide range of phenomena in that they can be used for ratio-scaled, ordinal, binary and categorical response variables, meaning every kind of variable a corpus linguist might want to deal with. The underlying algorithm tests whether any of the predictors significantly influences the response variable. If that is the case, the predictor with the strongest association with the response is selected and a binary split is introduced for the selected predictor. This process is repeated until the addition of another split no longer makes the next model significantly better than the former. The introduction of this statistical threshold is one of the aspects that render CITs superior to earlier Classification and Regression Trees (CARTs) because these CART approaches cannot differentiate a significant from an insignificant improvement and thus tend to split the data more often than is statistically warranted—an issue referred to as overfitting (Mingers, 1987). True, for earlier CARTs, it needed to be checked whether the number of splits in the tree needed to be reduced, in other words whether the tree needed pruning, to arrive at the optimal tree for the data modelled, but CITs devise this optimal tree via statistical checkpoints as the tree grows, which is computationally more efficient (Hothorn et al., 2006, 652f.). CITs thus avoid overfitting and improve earlier CARTs further in that they also circumvent a selection bias in the variables used for splitting the data based on the “conditional distribution of statistics measuring the association between responses and covariates” (Hothorn et al., 2006, p. 652). Earlier CARTs contained a bias towards predictors with many possible splits and/or many missing values (Hothorn et al., 2006, p. 651; Kim & Loh, 2001), for which they have repeatedly been criticised (Breiman et al., 1984; Segal, 1988). The resulting CIT model is visualised as a decision tree with nodes splitting the data according to the variable

in the node to create subgroups that are maximally homogeneous with regard to the occurrence of the response variable. From each node two branches with the relevant levels shown on top of them lead to the next nodes until the branching terminates in stacked bar plots, where percentages of the response variable for the factor combinations having led to the respective bar plots are shown.

For the apology data at hand, the response variable covers whether speakers chose *sorry* or another apology in a particular communicative context and the predictors are those listed in the ‘Corpus data and data extraction and annotation’ section. Consequently, the model formula for the CIT for apologies reads: $\text{APOLOGY} \sim \text{AGE} + \text{GENDER} + \text{GROUPGENDER} + \text{INTENSIFIER} + \text{SETTING} + \text{TOPIC} + \text{TYPE} + \text{TTR} + \text{VARIETY}$. The resulting CIT in Fig. 2 has a classification accuracy of 82.11% and is significantly better ($p < 0.05$) than a baseline model always predicting the more frequent level of the response variable (here: *sorry* occurring with a frequency of 78.16%).

Out of the nine independent variables available for the prediction of *sorry* vs. another apology (dark grey vs. light grey), only TYPE and VARIETY are modelled as significant for apology realisation. The first split separates single-word apologies (*swu*) from structurally more complex multiword apologies (*mwu*) with the former featuring *sorry* notably more frequently than the latter. Both branches emerging from node 1 are further split up according to VARIETY in nodes 2 and 5. With *swu*, BrE and IndE speakers appear to behave similarly in that, despite their strong preference for *sorry*, they still use other apology forms while SLE speakers are modelled to categorically opt for *sorry* in *swu* (node 7). It could be assumed that SLE speakers use *sorry* as a “highly routinised all-purpose token” (Barron, 2019, p. 15) which decreases planning pressure and is, thus, preferred by English L2 speakers in Sri Lanka. In contrast, with a view to *mwu*, where *sorry* is not as dominant as with *swu*, SAE speakers are different from BrE speakers in that they opt for *sorry* more often.

While it is certainly insightful to observe that unique varietal profiles in apology choice sensitive to the structural complexity of the apology exist, it is still necessary to determine the relative importance of all predictors available for the choice between *sorry* and another apology and how these predictors may affect this choice. For this, CITs are not ideal because the importance of certain predictors cannot be evaluated in case they do not meet the statistical threshold for inclusion in the tree. In this light, RFs (Breiman, 2001, p. 5) are a promising complementation. In contrast to CITs, RFs (see, for instance, Levshina, 2015 or Bernaisch, accepted for practical introductions) grow a large number of CARTs—the default is 500—based on different subsets of examples and predictors (also referred to as bootstrapping). This selection of examples and predictors for each CART in the forest has the advantage that particularly important predictors will be profiled accordingly because they will assume high-order positions in the CARTs of which they are part, but they will also be left out of a number of CARTs, which allows showing how the less important variables not featured in a CIT affect the response variable. Still, tree-based approaches like CITs or RFs have also been shown to occasionally miss important interactions between predictors (Bernaisch et al., 2014), which leads Gries (2019, p. 15) to suggest that the set of predictors should be complemented with interaction predictors one considers relevant for the classification trees concerned. As the present paper focusses on potential cross-varietal differences in apologies, interaction predictors

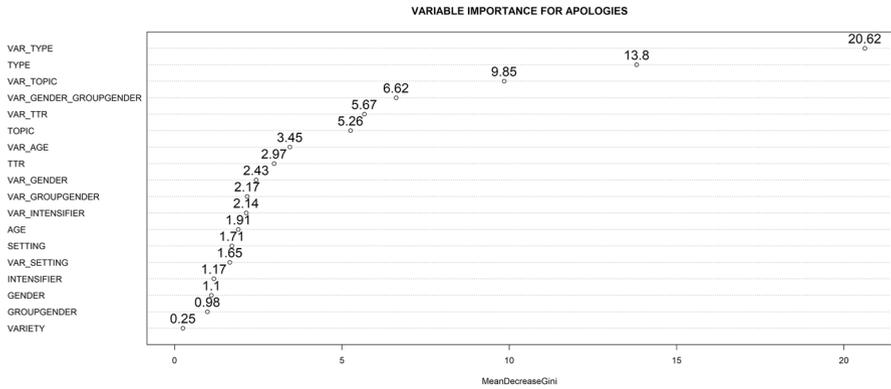


Fig. 3 Variable importance plot for *sorry* vs. *other* apology

with VARIETY are created for each of the predictors except for VARIETY itself. As it could also be of interest how apology choice is affected by GENDER under consideration of the gender composition in the group where an apology is uttered, a three-way interaction predictor featuring VARIETY, GENDER and GROUPGENDER is added. Consequently, the model formula for the RF reads: APOLOGY ~ AGE + GENDER + GROUPGENDER + INTENSIFIER + SETTING + TOPIC + TTR + TYPE + VAR_AGE + VAR_GENDER + VAR_GENDER_GROUPGENDER + VAR_GROUPGENDER + VAR_INTENSIFIER + VAR_SETTING + VAR_TOPIC + VAR_TTR + VAR_TYPE + VARIETY. The resulting RF model has a classification accuracy of 88.95%, which is highly significantly ($p < 0.001$) better than a baseline model always predicting the more frequent level of the response variable. In Fig. 3, the variable importance scores for the predictors of the choice between *sorry* and other forms of apologies are shown.

To some extent, the variable importance scores of the RF echo the variables selected as significant in the CIT in Fig. 2 in that the interaction predictor VAR_TYPE is profiled as the most important variable in the RF for apology choice, although VARIETY on its own is least relevant for said choice. As the analytical focus is on detecting possible cross-varietal differences in how speakers apologise, it is revealing that the interaction predictors VAR_TOPIC, VAR_GENDER_GROUPGENDER and VAR_TTR as well as VAR_AGE play central roles in determining the form of an apology while the other predictors and interaction predictors appear to be more marginal. From a methodological angle, the interaction predictors consistently receive higher variable importance scores than single predictors—only to the exception of SETTING having a higher variable importance score than VAR_SETTING at the lower end of the scale. This can be considered a testament to the benefits of including said interaction predictors in RF models—particularly because they also allow zooming in on how the

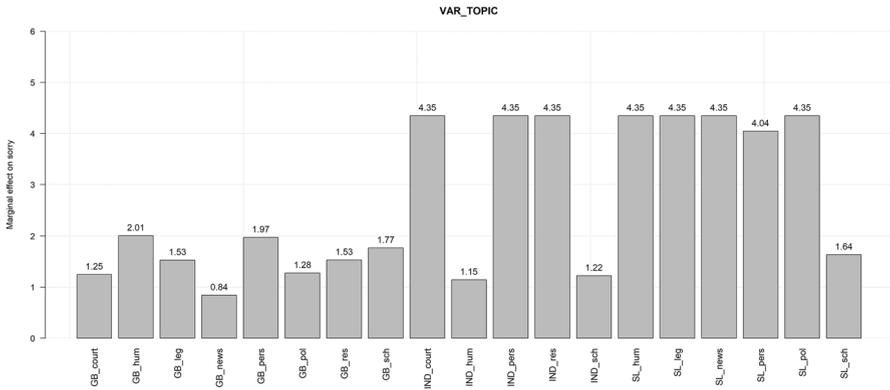


Fig. 4 PDP of VAR_TOPIC for *sorry*

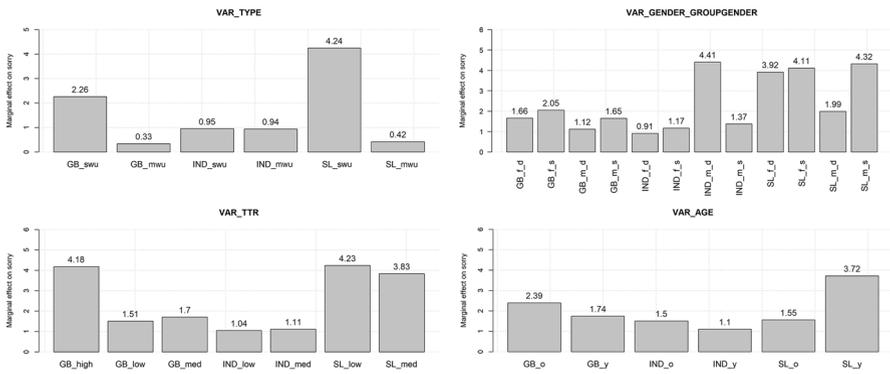


Fig. 5 PDPs of VAR_TYPE, VAR_GENDER_GROUPGENDER, VAR_TTR and VAR_AGE for *sorry*

predictors affect apology choices in the respective varieties. The partial dependence plot (PDP) for VAR_TOPIC in Fig. 4 illustrates with which topics speakers have a higher or lower tendency towards using *sorry* as opposed to another apology form.⁵

With BrE speakers, *sorry* occurs most often with topics related to the humanities, personal matters and school while IndE speakers show a reversed tendency. The IndE pattern is largely reconcilable with that of SLE, where *sorry* figures prominently across all topics except for school. The PDPs for VAR_TYPE, VAR_GENDER_GROUPGENDER, VAR_TTR and VAR_AGE are given in Fig. 5.

When it comes to whether the apology concerned is realised as a *swu* or a structurally more complex *mwu*, it can be seen in the top left panel of Fig. 5 for VAR_TYPE that *sorry* is more likely to occur in *swu* than in *mwu*, which holds uniformly across

⁵ Please note here that the second most important interaction predictor VAR_TOPIC is plotted in a separate figure to increase readability and that the PDP of the predictor with the highest variable importance score, VAR_TTR, is in the top left panel of Fig. 5.

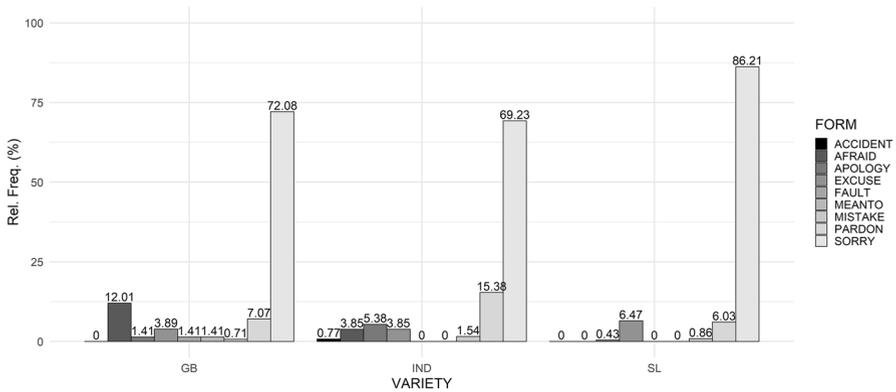


Fig. 6 Relative frequencies of different apology forms in BrE, IndE and SLE

BrE, IndE and SLE. This difference is particularly notable in SLE, present in BrE, but almost absent from IndE. In the plot for VAR_GENDER_GROUPGENDER in the top right corner of Fig. 5, BrE and SLE speakers behave relatively similarly in that *sorry* is uttered more frequently when apologies are offered in groups of speakers with the same sex. While this is also true for Indian women, Indian men tend to offer apologies with *sorry* less frequently in same-sex than in mixed-sex groups. Regarding VAR_TTR in the bottom left corner of Fig. 5, BrE speakers with high TTRs use *sorry* most often. SLE speakers with low TTRs tend to use *sorry* more than speakers with medium TTRs and with IndE speakers, there is a marginal difference in that speakers with a medium TTR employ *sorry* slightly more often than speakers with a low TTR.⁶ In terms of speaker age, older speakers use *sorry* as an apology more often than younger speakers in BrE and IndE while the reverse applies to SLE speakers. In sum, the distribution of *sorry* is sensitive to a number of (sociobiographic) speaker variables and interactions between them, but a closer look at the alternatives to *sorry*—although numerically consistently in the minority—also provides relevant insights. The relative frequencies of all of the apology forms in the data are visualised in Fig. 6 and grouped according to the varieties they occur in.⁷

Although the dominance of *sorry* becomes apparent yet again with BrE (72.08%), IndE (69.23%) and SLE speakers (86.21%), there are still statistically highly significant ($p < 0.001$ (Fisher’s exact test)) differences in the frequencies of apology forms across the varieties. In BrE, the second most frequent apology form is *afraid* as illustrated in (5) and *pardon* as in (6) ranks third. In IndE, *sorry* is followed by *pardon* as in (7) and *apologise* as shown in (8) and SLE has a slightly stronger preference for apologies with *excuse* as exemplified in (9) than for *pardon* in (10).

⁶ None of the speakers who used an apology in the Indian or Sri Lankan datasets had a high TTR, which is why the respective bars are absent from Fig. 5.

⁷ While the CIT and the RF are based on only those apology examples where full sociobiographic speaker information is documented, the normalised and relative frequencies in Figs. 1 and 6 display all apologies extracted independent of whether full speaker information is available.

- (5) <ICE-GB:S1A-032#221:2:D> Does a choir does a choir have a gig
 <ICE-GB:S1A-032#222:2:C> Did the Baroque Singers
 <ICE-GB:S1A-032#223:2:B> I don't know
 <ICE-GB:S1A-032#224:2:B> I'm afraid I really don't know
- (6) <ICE-GB:S1B-063#157:1:C> Very minor
 <ICE-GB:S1B-063#158:1:B> Very <unclear-word>
 <ICE-GB:S1B-063#159:1:A> Very minor I beg your pardon I must 've misheard you
- (7) <ICE-IND:S1B-023#54:1:D> Bold initiative was needed and bold initiative has been taken
 <ICE-IND:S1B-023#55:1:B> Uh pardon me for interrupting Mr Mukharjee now
 <ICE-IND:S1B-023#56:1:B> Uh Dileep made a point and he says that [...]
- (8) <ICE-IND:S1B-022#27:1:C> Did you expect to stand it in one go so to speak
 <ICE-IND:S1B-022#28:1:B> Uh Dileep I <, > must apologise for not responding to the statement [...]
- (9) <ICE-SL:S1A-002#184:1:A> You know what I mean saying
 <ICE-SL:S1A-002#185:1:C> Yeah I think I do
 <ICE-SL:S1A-002#186:1:B> Excuse me please I have a minor problem
 <ICE-SL:S1A-002#187:1:C> Yeah
- (10) <ICE-SL:S1B-018#146:1:M> If the patient has other comorbidities
 <ICE-SL:S1B-018#147:1:A> Pardon
 <ICE-SL:S1B-018#148:1:M> If the patient has other comorbidities we have to mention this also

More generally, we observed that BrE speakers use the widest range of different apology forms followed by IndE and SLE speakers. For BrE speakers, each apology form except for those with *accident*, of which one form uniquely occurs in the Indian data, could be attested. IndE does not feature apologies with *fault* or *meant to*, which is also the case in SLE, where in addition apologies with *accident* and *afraid* are also absent.

Discussion

To inform the subsequent discussion, the main findings are summarised here in the light of the research questions that guided this study. With earlier research on apologies in general and in SAEs in particular suggesting that there should be differences in apology strategies chosen by BrE speakers as opposed to IndE and SLE speakers, the present study can support some of the earlier findings, but contradicts others. Though it is hard to argue that variety-exclusive forms of apologies became evident, with regard to the overall distribution of apology forms, it could, nevertheless, be seen that there are in fact variety-specific preferences with BrE speakers employing a larger set of apology forms than IndE and SLE speakers and that there are different quantitative preferences across the varieties with regard to these apology forms. Against the background of previous research, this study suggests that apologies are sensitive to sociobiographic factors of the speakers as the overall frequencies of apologies change in the light of speaker's age, gender and regional background. Consequently, the most important findings of this study concern the predictors AGE, GENDER and VARIETY and the interaction predictor VAR_GENDER_GROUPGENDER on the choice of *sorry* over other apology forms. Still, it needs to be pointed out that the structural complexity of an apology as well as the TOPIC under discussion, the speakers' TTRs and the group composition with regard to GENDER play important roles in the choice between *sorry* vs. other forms of apologies as well.

It appears that male speakers largely independent of AGE do not apologise as often as women in the varieties studied (to the exception of older IndE males, whose apology frequencies are notably similar to those of older IndE females). We can only assume that there is a connection between men perceiving apologies as self-oriented FTAs (as suggested by Holmes, 1989, p. 208) and therefore they do not apologise as often as women—unless they find themselves in a communicative situation with at least one woman, in which case they apologise more often than in same-sex situations. Nevertheless, this does not necessarily mean that SAE speakers are less polite; it could simply indicate that other forms of apologising, for example kinesics, are a more frequently used and acceptable means of apologising in the given cultural context. Non-verbal apologetic behaviour, however, can hardly be investigated in the corpus data at hand. A more specialised corpus including non-verbal politeness marker annotation could shed light on this particular aspect in the future.

Regarding the realisation of an apology as either a *swu* or a *mwu*, our study points out that SAE speakers opt for *sorry* more often than for other expressions. Since, for instance, local Indian languages do not include a word equivalent to the meaning of *sorry*, the more frequent use of *sorry* in English-speaking situations could be a sign of a so-called lexical teddy bear, that is a word that L2 speakers feel comfortable using and thus use frequently as a dominant variant (Hasselgren, 1994). However, due to the huge amount of local L1 languages in India, stating in how far apology strategies generally derive from local L1 influences is outside the scope of this paper and could be subject of future research.

Moreover, in the majority of cases, *sorry* is used as a single word unit in all varieties, meaning that IndE speakers do not opt for longer utterances as often as

suggested by previous research. With a difference between the effect of *swu* or *mwu* on *sorry* being largely absent in IndE (see Fig. 5), we argue that IndE speakers make no politeness distinction between one-word utterances and longer apologies. This could also indicate that there is no clear rule or pattern as to when to use either of the two options in IndE—which is in line with earlier findings stating that IndE speakers show “a lower percentage of [...] fixed forms” than BrE speakers (Mehrotra, 1995, p. 99). However, it should be noted that this study did not investigate SSEs that occurred in a new utterance preceding or following the IFID. Therefore, future studies could also take into account more than the immediate context of the IFID to include possible apology combinations.

In order to account for the acceptability of an uttered apology in the different varieties of English investigated here, future research could include responses to the apologies the speakers received to see if the chosen apology strategy was a successful one. Moreover, the type and seriousness of the offence that caused the speaker to apologise could be other factors that determine the chosen apology strategy.

Further, this study mainly took into account the head act of an apology and introduced a binary distinction between *sorry* and other apology forms. While this operationalisation was conducive with regard to the focus on apology forms of the present paper, further research into other distinctions such as different types of detached apologies as suggested by Deutschmann (2003, 53ff.) would be welcome complements to the present study. In terms of the apologies analysed, the paper came to the conclusion that South Asian varieties do not feature longer apologies and future research, possibly featuring a more specialised corpus, could investigate the total length of apology, including its SSEs to establish to what extent the observations reported here are also applicable in this extended context of apologies.

Acknowledgements We would like to thank the German Research Foundation (BE 5812/2-1) for funding this project. Furthermore, we thank Nina Funke and Michelle Kraaz for their support in data extraction and annotation.

Author contributions Both authors worked jointly on each step of the study and also collaborated during the writing process of each of the sections of the paper. Both authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. This paper was developed in the context of the project “Pragmatic nativisation in spoken Sri Lankan English: a corpus-based study”, which was funded by the German Research Foundation (BE 5812/2-1).

Data Availability This study used three components of the International Corpus of English (ICE). ICE-Great Britain is based at the Survey of English Usage, University College London. ICE-India is coordinated jointly by Professor S.V. Shastri (Shivaji University, Kolhapur, India) and Professor Dr. Gerhard Leitner (Freie Universität Berlin, Germany). ICE-Sri Lanka is based at the University of Giessen, Germany. All data can be obtained from the respective compilers.

Code Availability The R-code used in this study included the R-packages *caret*, *partykit* and *randomForest*.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose. This paper

has not been published elsewhere.

Ethical Approval Not applicable

Informed Consent Not applicable

Consent for Publication Not applicable

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aijmer, K. (1995). Do women apologise more than men? In G. Melchers & B. Warren (Eds.), *Studies in anglistics* (pp. 55–69). Stockholm: Almqvist and Wiskell International.
- Barron, A. (2005). Variational pragmatics in the foreign language classroom. *System*, 33(3), 519–536. <https://doi.org/10.1016/j.system.2005.06.009>
- Barron, A. (2008). Contrasting requests in inner circle Englishes: A study in variational pragmatics. In M. Pütz & J. Neff-van Aertselaer (Eds.), *Developing contrastive pragmatics: Interlanguage and cross-cultural perspectives* (pp. 335–402). New York: De Gruyter.
- Barron, A. (2017a). The speech act of “offers” in Irish English. *World Englishes*, 36(2), 224–238. <https://doi.org/10.1111/weng.12255>
- Barron, A. (2017b). Variational pragmatics. In A. Barron, Y. Gu, & G. Steen (Eds.), *The Routledge handbook of pragmatics* (pp. 91–104). New York: Routledge.
- Barron, A. (2019). Using corpus-linguistic methods to track longitudinal development: Routine apologies in the study abroad context. *Journal of Pragmatics*, 146, 87–105. <https://doi.org/10.1016/j.pragma.2018.08.015>
- Barron, A., & Schneider, K. P. (2009). Variational pragmatics: Studying the impact of social factors on language use in interaction. *Intercultural Pragmatics*, 6(4), 425–442. <https://doi.org/10.1515/IPRG.2009.023>
- Bernaish, T. (accepted). Comparing generalised linear mixed-effects models, generalised linear mixed-effects model trees and random forests: Filled and unfilled pauses in varieties of English. In O. Schützler & J. Schlüter (Eds.), *Data and methods in linguistics: Comparative approaches*. Cambridge University Press.
- Bernaish, T., Gries, S. Th., & Mukherjee, J. (2014). The dative alternation in South Asian Englishes: Modelling predictors and predicting prototypes. *English World-Wide*, 35(1), 7–31. <https://doi.org/10.1075/eww.35.1.02ber>
- Bharuthram, S. (2003). Politeness phenomena in the Hindu sector of the South African Indian English speaking community. *Journal of Pragmatics*, 35, 1523–1544. [https://doi.org/10.1016/S0378-2166\(03\)00047-X](https://doi.org/10.1016/S0378-2166(03)00047-X)
- Bhawuk, D. P. S. (2017). Individualism and collectivism. In K. Young Yun (Ed.), *The international encyclopedia of intercultural communication* (pp. 1–9). New York: Wiley.
- Blum-Kulka, S., House, J., & Kasper, G. (Eds.). (1989). *Cross-cultural pragmatics: Requests and apologies*. Norwood, NJ: Ablex.
- Blum-Kulka, S., & Olshtain, E. (1984). Requests and apologies: A cross-cultural study of speech act realization patterns (CCSARP). *Applied Linguistics*, 5(3), 196–213. <https://doi.org/10.1093/applin/5.3.196>

- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. New York: Chapman and Hall/CRC.
- Brown, P., & Levinson, S. (1978). Politeness – Some universals in language use. In J. J. Grumperz (Ed.), *Studies in interactional sociolinguistics 4*. Cambridge University Press.
- Degenhardt, J. (2020). Requests in Indian and Sri Lankan English. *World Englishes*. <https://doi.org/10.1111/weng.12573>
- Deutschmann, M. (2003). *Apologising in British English*. Ume Universitet.
- Funke, N. (2020). Pragmatic nativisation of thanking in South Asian Englishes. *World Englishes*. <https://doi.org/10.1111/weng.12517>
- Gries, S. Th. (2019). On classification trees and random forests in corpus linguistics: Some words of caution and suggestions for improvement. *Corpus Linguistics and Linguistic Theory*, 16(3), 617–647. <https://doi.org/10.1515/cllt-2018-0078>
- Hasselgren, A. (1994). Lexical teddy bears and advanced learners: a study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics*, 4(2), 237–260. <https://doi.org/10.1111/j.1473-4192.1994.tb00065.x>
- Holmes, J. (1989). Sex differences and apologies: One aspect of communicative competence. *Applied Linguistics*, 10(2), 194–213. <https://doi.org/10.1093/applin/10.2.194>
- Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, 15(3), 651–674. <https://doi.org/10.1198/106186006X133933>
- International Corpus of English (ICE). <http://www.ucl.ac.uk/english-usage/ice>
- Jucker, A. (2018). Apologies in the history of English: Evidence from the corpus of Historical American English (COHA). *Corpus Pragmatics*, 2, 375–398. <https://doi.org/10.1007/s41701-018-0038-y>
- Kasanga, L. A., & Lwanga-Lumu, J. C. (2007). Cross-cultural linguistic realizations of politeness: A study of apologies in English and Setswana. *Journal of Politeness Research*, 3, 65–92. <https://doi.org/10.1515/PR.2007.004>
- Kim, H., & Loh, W. Y. (2001). Classification Trees with unbiased multiway splits. *Journal of the American Statistical Association*, 96, 589–604. <https://doi.org/10.1198/016214501753168271>
- Kitao, S. K., & Kitao, K. (2013). Apologies, apology strategies, and apology forms of non-apologies in a spoken corpus. *Journal of Culture and Information Science*, 8(2), 1–13.
- Kraaz, M., & Bernaisch, T. (2020). Backchannels and the pragmatics of South Asian Englishes. *World Englishes*. <https://doi.org/10.1111/weng.12522>
- Leech, G. (1983). *Principles of pragmatics*. London: Longman.
- Leech, G. (2014). *The pragmatics of politeness*. Oxford University Press.
- Levshina, N. (2015). *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins. <https://doi.org/10.1075/z.195>
- Lutzky, U., & Kehoe, A. (2017). ‘Oops, I didn’t mean to be so flippant’. A corpus pragmatic analysis of apologies in blog data. *Journal of Pragmatics*, 116, 27–36. <https://doi.org/10.1016/j.pragma.2016.12.007>
- Márquez Reiter, R. (2000). *Linguistic politeness in Britain and Uruguay: A contrastive study of requests and apologies*. Amsterdam: John Benjamins. <https://doi.org/10.1075/pbns.83>
- Mattson Bean, J., & Johnston, B. (1994). Workplace reasons for saying you’re sorry: Discourse task management and apology in telephone interviews. *Discourse Processes*, 17(1), 59–81. <https://doi.org/10.1080/01638539409544859>
- Mehrotra, R. (1995). How to be polite in Indian English. *International Journal of the Sociology of Language*, 116, 99–110. <https://doi.org/10.1515/ijsl.1995.116.99>
- Mingers, J. (1987). Expert systems—Rule induction with statistical data. *Journal of the Operations Research Society*, 38, 39–47. <https://doi.org/10.1057/jors.1987.5>
- Morgan, J. N., & Sonquist, J. A. (1963). Problems in the analysis of survey data, and a proposal. *Journal of the American Statistical Association*, 58, 415–434. <https://doi.org/10.2307/2283276>
- Ogiermann, E. (2009). *On apologising in negative and positive politeness cultures*. Amsterdam: John Benjamins. <https://doi.org/10.1075/pbns.191>
- Olshain, E., & Cohen, A. (1990). The learning of complex speech act behaviour. *TESL Canada Journal*, 7(2), 45–65.
- Sailaja, P. (2009). *Indian English*. Edinburgh University Press.

- Schneider, K. P., & Barron, A. (Eds.). (2008). *Variational pragmatics: A focus on regional varieties in pluricentric languages*. Amsterdam: John Benjamins.
- Searle, J. R. (1979). *Expressions and meaning: Studies in the theory of speech acts*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511609213>
- Segal, M. R. (1988). Regression trees for censored data. *Biometrics*, *44*, 35–47. <https://doi.org/10.2307/2531894>
- Spencer-Oatey, H., Ng, P., & Li, D. (2008). British and Chinese reactions to compliment responses. In H. Spencer-Oatey (Ed.), *Culturally speaking: Culture, communication and politeness theory* (pp. 95–117). London: Continuum.
- Sridhar, K. K. (1991). Speech acts in an indigenized variety: sociocultural values and language variation. In J. Cheshire (Ed.), *English around the world* (pp. 308–318). Cambridge University Press.
- Staley, L. (2018). *Socioeconomic pragmatic variation*. Amsterdam: John Benjamins.
- Tannen, D. (1994). *Gender and discourse*. Oxford University Press.
- Tinkham, Th. (1993). Sociocultural variation in Indian English speech acts. *World Englishes*, *12*(2), 239–247. <https://doi.org/10.1111/j.1467-971X.1993.tb00024.x>
- Trosborg, A. (1995). *Interlanguage pragmatics: Requests, complaints and apologies*. New York: De Gruyter. <https://doi.org/10.1515/9783110885286>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.