Received: 9 December 2019 | Revised: 27 April 2020 | Accepted: 3 May 2020

DOI: 10.1111/weng.12517

PAPER

WORLD ENGLISHES WILEY

# Pragmatic nativisation of thanking in South Asian Englishes

Nina Funke

Department of English, Justus Liebig University

**Correspondence**
Nina Funke, Department of English Language and Linguistics, Justus Liebig University Giessen, Otto-Behaghel-Strasse, 10C 35394, Giessen, Germany.
Email: Nina.S.Funke@anglistik.uni-giessen.de

## Abstract

As South Asian Englishes have been studied regarding their structural levels but only few studies have focused on their pragmatics, this study investigates thanking strategies in the spoken parts of the British, Indian, and Sri Lankan components of the International Corpus of English. The paper answers the questions: Do VARIETY, speaker AGE and GENDER, context FORMALITY, POSITION within the speaker turn, and the presence of a BENEFACTOR, INTENSIFIER, and REASON influence the choice between *thank* and *thanks*; and what implications do these findings have for the notion of pragmatic nativisation? In a multifactorial analysis, a conditional inference tree identifies AGE, BENEFACTOR, INTENSIFIER, and VARIETY as significant predictors while a random forest highlights the importance of variable interactions with VARIETY.

## 1 | INTRODUCTION

Thanking is a frequently used 'acknowledgement of one's having benefited from the actions of another person' (Norrick, 1978, p. 285). In their analysis of gratitude in American English, Eisenstein and Bodmann (1993) conclude that 'thanking is a speech act that is mutually developed. It can involve a complex series of interactions and encodes cultural values and customs' (p. 74). Coulmas (1981) contrastively investigates thanks and apologies in Japanese and finds that, although thanking strategies exist across cultures, 'it is obvious that the pragmatic considerations of their implementation are culturally defined' (p. 89), thus showing that the choice of thanking strategies is strongly linked to cultural norms of linguistic behaviour. This culture-specificity of how people use language to achieve particular communicative goals has been referred to as pragmatic nativisation as defined by Bamgbose (1998): '[P]ragmatic and creative nativization […] fall largely within the scope of behavioural norms. An appeal to behavioural norms in native English is

entirely useless in determining what is appropriate and what is not in non-native English, since the very existence of non-native behavioural norms is one of the defining characteristics of a non-native variety of English' (p. 2).

The present study is concerned with the use of thanking formulae across different varieties of English and offers a contribution to the emerging field of variational pragmatics (Schneider & Barron, 2008), which, however, has so far had a clear focus on 'native varieties' of English, although noteworthy exceptions exist (for example, Revis & Bernaisch, 2020). This general lack of research into the pragmatics of Outer Circle varieties is also reflected in the descriptions available for thanks. Analyses of thanking strategies have so far been restricted to Inner Circle varieties such as British and American English (Cheng, 2010; Schauer & Adolphs, 2006) and learner varieties of English (Azima & Hesabi, 2015; Hinkel, 1994). Cheng (2010) finds *thank* to be more frequently used than *thanks* and intensification to be the most frequent modification of thanking formulae in British and American English. Schauer and Adolphs (2006) compare thanking formulae in the CANCODE corpus (Cambridge and Nottingham corpus of Discourse in English) to thanking formulae in a discourse completion task. They find that thanking formulae in the corpus are generally longer than those in the discourse completion task. When it comes to learner varieties, Azima and Hesabi (2015) analyse the use of thanking strategies of Iranian learners of English and conclude that female speakers choose their strategies more cautiously than male speakers and that the level of proficiency also plays an important role in the process of strategy selection. Hinkel (1994) looks at choices of thanking formulae that second-language speakers at college level make in the United States and finds their choices to differ from those that 'native speakers' make. Although not quantitatively analysing the use of thanking strategies, several studies take a culturally-based approach to thanking and find that speakers from South Asian cultures do not tend to express thanks overtly toward their family members (Apte, 1974; Kachru, 2008) or in 'situations involving exchange of goods' (Apte, 1974, p. 69) because 'only the proper behaviour without expectations is counted as a merit towards final salvation' (Apte, 1974, p. 81). Apte (1974) and Kachru (2008) argue that the 'linguistic and cultural contact with English' (Kachru, 2008, p. 357) has led to an increased use of thanking strategies. In contrast to that, Kachru (1992) argues that 'culture-bound styles are *transcreated* in L2 with close approximation to the style range available to a writer in his L1' (Kachru, 1992, p. 319) which indicates that the frequency of use of thanking strategies in South Asian Englishes should resemble that of the indigenous languages in South Asia. The present study thus provides the first investigation of the frequencies and structural realisations of, as well as influential factors for, different thanking formulae in two second-language varieties of English, that is Indian (IndE) and Sri Lankan English (SLE). These varieties have been studied with regard to their pronunciation (Fuchs, 2015; Gunesekera, 2005), lexis (Meyler, 2007; Nihalani, Tongue, & Hosali, 2004), syntax (Lange, 2012; Schilk, 2011; Sedlatschek, 2009), and semantics (Gueldenring, 2017; Werner & Mukherjee, 2012). So far, only few studies have focused on pragmatic features in these varieties (Revis & Bernaisch, 2020).

With regard to Schneider's (2003, 2007) model, which describes five phases of the evolution of postcolonial Englishes, previous research has found that SLE has passed – with differential perspectives for individual structural levels – the nativisation phase and now displays several characteristics of endonormative stabilisation (Bernaisch, 2015). However, British English (BrE) – the historical input variety – does not account for all linguistic influences on SLE. In addition to influences from language contact with Sri Lanka's indigenous languages Sinhala and Tamil and more global trends potentially lead by American English (Mair, 2013), IndE, an endonormatively stabilised variety in its own right (Mukherjee, 2007), has been found to exert an epicentral influence on English-speaking countries in South Asia (Gries & Bernaisch, 2016; Heller, Bernaisch, & Gries, 2017; Leitner, 1992, 225). The analysis of BrE and IndE in addition to SLE thus considers its geohistorical context and aims at investigating whether these varieties exhibit different pragmatic developments with regard to the use of thanking strategies. Meyler (2007, p. 261), for example, lists *thanking you* as another formulaic sequence at the end of letters in SLE, which in BrE would rather be *thank you* or *with thanks*, providing first anecdotal indications that said differences in the developments of the culture- and variety-specific thanking systems might exist.

The term *pragmatic nativisation* refers to both the indigenisation of the pragmatics of English as described in the third phase of Schneider's (2003, 2007) model of postcolonial Englishes and the performance of certain actions via language by local speakers of English. The present study thus follows Hoffmann, Blass, and Mukherjee (2017)) analytic

operationalisation of the concept by investigating how thanking formulae are structurally realised in SLE and IndE as opposed to their historical input variety BrE. Considering the theoretical background of *pragmatic nativisation*, the present study establishes how linguistic factors and sociobiographic attributes of British, Indian, and Sri Lankan speakers influence the choice of thanking formulae in the respective locales. After a description of the relevant terminology and the functions of thanks, the structural and sociobiographic factors assumed to influence thanking strategies are presented. Section two then starts with a presentation of the ICE components under scrutiny and their structure followed by a description of different criteria for extracting thanking formulae, the descriptive statistics which highlight the distribution of thanking formulae across the levels of predictors, and the introduction of the two statistical methods applied to the data. The results of these analyses are presented in section three followed by a discussion in section four. Finally, the conclusion in section five will summarize the most important points that were made in this study and give directions for further research.

## 1.1 | Functions and terminology

Thanking formulae fulfil different functions depending on the conversational context in which they are uttered. Jautz (2013, pp. 69–70) gives a detailed five-fold distinction of possible functions of thanking strategies:

- Organising discourse: Thanking formulae which acknowledge a speaker's contribution to a conversation or thank an audience for listening. However, a thanking strategy can also mark an interruption of a speaker's turn or the intention to change the topic.
- Phatic communion: When serving the phatic communion, thanking formulae are concerned with the interlocutors' relationship and are prominently used to gently decline an offer or as an answer to greetings, compliments, and good wishes. Highly routinised expressions such as *Thank God* or *Thank Goodness* belong to this category as 'they tell the addressee something about the speaker's emotions and their mental state' (Jautz, 2013, p. 70).
- Responding to material goods and services: The most common case in which thanking formulae are used is in response to when one is given material goods and services.
- Responding to immaterial goods and interpersonal support: In addition to material goods and services, a speaker's gratitude may also be directed at immaterial goods or interpersonal support, for example advice or information.
- Thanking formulae used jokingly or ironically: In these cases, the thanking routine is often exaggerated to show that the speaker is actually not grateful 'but plays with the concept of gratitude to express the opposite' (Jautz, 2013, 70).

(1) Thank you for the question Sir (ICE-GB-S1B-025#43:1:A)

(2) They did a very good job thank goodness (ICE-GB-S1A-046#127:1:B)

(3) Thank you for the cup of tea Vicky (ICE-GB-S1A-040#455:1:E)

(4) Thanks a lot Kaveetha for your time and your opinion (ICE-SL-S1A-092#85:1:A)

These functions are not mutually exclusive as, for example, a thanking formula expressing gratitude for interpersonal support may also serve the phatic communion and vice versa.

In previous research, *thank* and *thanks* have been identified as the most frequent realisations of thanking strategies (Cheng, 2005, 2010; Farnia & Suleiman, 2009; Schauer & Adolphs, 2006). Studies have, so far, mainly looked into how structural aspects influence the choice of thanking formulae. The present study adds to this research in taking up the factors they profiled (as described in what follows) as important, but also introduces novel perspectives via the inclusion of speaker- and genre-related factors in addition to the focus on South Asian Englishes.

## 1.2 | Structural factors

Structural factors particularly concern the presence of optional elements of thanking formulae such as BENEFACTORS, INTENSIFIERS, and REASONS. Additionally, research has been conducted into the POSITION of thanking routines in conversations. In the following, an overview will be given of these text-internal factors that can influence the choice of thanking formulae.

In general, optional elements have been shown to be able to lift thanking formulae from their routinisation (Norrik, 1978, p. 285). Okomato and Robinson (1997), in their analysis of thanking strategies in BrE, find those formulae that include modifiers and additional phrases to be more polite than non-modified formulae.

An addressee is always part of a thanking routine since thanking is always – at least implicitly – directed at a BENEFACTOR. A BENEFACTOR can be a syntactically obligatory element of the thanking formula, as is the case with *you* in *thank you*, or an optional modification, like *to you* in *I am grateful to you*. Jautz (2013) focuses on explicitly expressed optional elements and finds names to be the most frequent realisations of BENEFACTORS followed by institutions, such as the *Central Institute of Foreign Language*, and endearments, like *sweetheart*.

INTENSIFIERS have been found to be the default modification of thanking formulae in BrE and 'the most frequently used [INTENSIFIER], as can be expected, is *very much*' (Cheng, 2010, p. 267). In Holmes' (1984) analysis of the modification of illocutionary force, she establishes that boosting a speech act with a positive illocutionary force such as thanking formulae by using an INTENSIFIER 'can be interpreted as an expression of friendliness' (p. 350). Brown and Levinson (1978) similarly find the boosting of an illocutionary force to serve as a social 'accelerator' increasing social interaction. The speaker may also 'use devices to boost the illocutionary force of the speech act asserting the proposition, expressing great certainty or conviction concerning its validity' in case the addressee is 'doubtful or hesitant' (Holmes, 1984, p. 348) about the speaker's sincerity. Jautz (2013, p. 92) finds that INTENSIFIERS are more frequently used with *thank*, as in *thank you very much*, than they are used with *thanks* across British and New Zealand English.

The last modifying element of thanking formulae that will be discussed throughout this paper is the naming of a REASON for the speaker's gratitude. Jautz (2013) finds constructions that include *for* and either a noun phrase, as in *thank you for your help*, or a verb phrase, such as *thanks for coming*, to be by far the most frequent realisations. She also finds that the construction including *for* and a noun phrase slightly favours *thank* over *thanks* while there is a marginal, while not significant, favouring of *thanks* if the construction includes *for* and a verb phrase.

The final structural aspect to be investigated in this section is the POSITION of the thanking formula in a conversation. Jautz (2013) differentiates conversation-initial, conversation-internal, and conversation-final expressions of gratitude, but finds no statistically significant differences between BrE and New Zealand English datasets in the distribution of thanking formulae in conversations. However, the present study will not investigate the distribution of expressions of gratitude in conversations but in single speaker turns because the database used for this study, that is the International Corpus of English as described in section 2.1, is a sample-text corpus providing access not to entire conversations, but rather snippets of them.

## 1.3 | Speaker-related factors

So far, the influence of AGE on the choice of thanking formulae has not been studied and only a few analyses take the influence of the speakers' GENDER into account. In this context, Azima and Hesabi (2015) have found female Iranian EFL learners to use significantly more optional modifications of thanking strategies such as intensifications and mentioning a REASON than male EFL learners.

## 1.4 | Genre-related factors

Genre-related factors have not yet been analysed with regard to their influence on thanking strategies either. The present study will thus investigate whether the level of FORMALITY of the context in which the expression of gratitude is uttered has a significant influence on the choice of thanking formula.

**TABLE 1** Corpus design

| | | GENRE | NUMBER OF TEXTS IN ICE GENRE |
|---|---|---|---|
| **Dialogues** | **Private** | Face-to-face Conversations | 90 |
| | | Phone Calls | 10 |
| | **Public** | Classroom Lessons | 20 |
| | | Broadcast Discussions | 20 |
| | | Broadcast Interviews | 10 |
| | | Parliamentary Debates | 10 |
| | | Legal Cross-examinations | 10 |
| | | Business Transaction | 10 |
| **Monologues** | **Unscripted** | Spontaneous Commentaries | 20 |
| | | Unscripted Speeches | 30 |
| | | Demonstrations | 10 |
| | | Legal Presentations | 10 |
| | **Scripted** | Broadcast News | 20 |
| | | Broadcast Talks | 20 |
| | | Non-broadcast Talks | 10 |
| | | TOTAL | 300 |

In light of earlier research into thanks and the research gaps profiled, the present paper seeks to investigate the following research questions:

- Are there VARIETY-specific differences in the choice of thanking formulae?
- To what extent do structural and sociobiographic factors other than VARIETY influence the choice between thanking formulae?
- To what extent does the choice between *thank* and *thanks* in South Asian Englishes align with that in BrE and what implications do the findings have for the notion of pragmatic nativisation?

## 2 | METHODOLOGY

### 2.1 | Corpus

This analysis is based on a dataset drawn from three spoken components of the International Corpus of English (ICE; Greenbaum, 1991), that is ICE-Great Britain (ICE-GB), ICE-India (ICE-IND), and ICE-Sri Lanka (ICE-SL). The ICE project compiles corpora of different varieties of English, each of which contains 500 texts of approximately 2,000 words. Of these texts, 60 per cent, that is 300 texts, are spoken and 40 per cent, that is 200 texts, are written. The spoken components of the ICE corpus cover monologues and dialogues, as well as scripted and unscripted speech, as shown in Table 1.

### 2.2 | Data extraction and annotation

To prepare the dataset for extraction, extra-corpus material, that is text that is marked with < X > , editorial comments, that is text marked with < & > , untranscribed text, that is text marked with < O > , and unclear text, that is text marked with < unclear > was deleted. While previous research combines *thank* and *thanks* to form one variant and compares

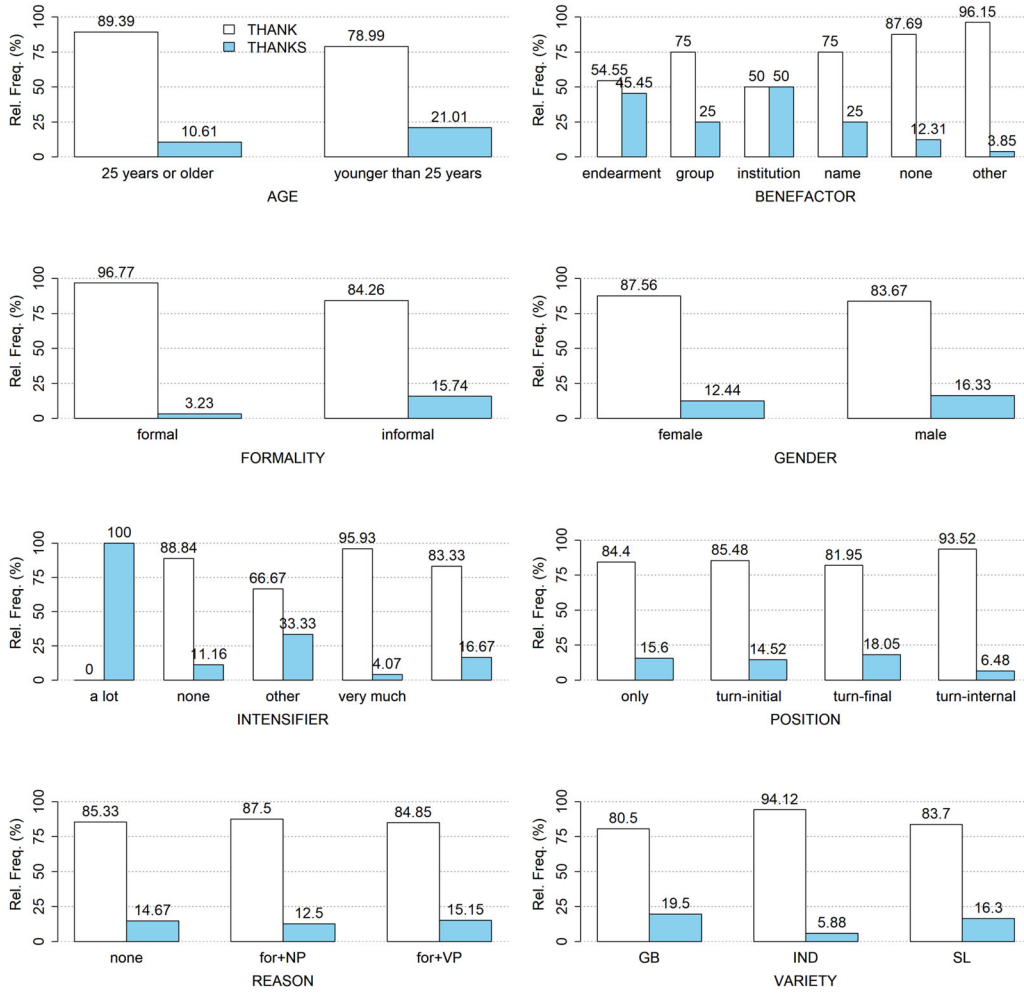


**FIGURE 1** Choice of thanking formula by each predictor [Colour figure can be viewed at wileyonlinelibrary.com]

this one variant to other expressions of gratitude such as *I am grateful*, *I appreciate*, the present study seeks to investigate influences on the choice between *thank* and *thanks* disregarding other expressions of gratitude as these other variants make for too small a percentage of all thanking strategies in the ICE components under scrutiny. The two variants *thank* and *thanks* were extracted via Antconc (Anthony, 2017) and all instances in which the speech act of thanking is performed were identified. Only those instances were retained where speakers express their gratitude directly to an interlocutor, that is a) they do not speak on somebody else's behalf and b) there is a clear reference to the time of speaking, meaning the utterance is not a report of having been grateful at some point in the past. In this process, 547 instances of *thank* and 93 instances of *thanks* were extracted and they are distributed across ICE-GB, ICE-IND, and ICE-SL as shown in the last panel of Figure 1.

While all three varieties show higher frequencies for *thank*, this observation is more pronounced in ICE-IND (94.1 per cent) and ICE-SL (83.7 per cent) than it is in ICE-GB (80.5 per cent), as shown in the eighth panel of Figure 1. In addition to VARIETY, all instances were then manually annotated for the independent variables AGE, BENEFACTOR, FORMALITY, GENDER, INTENSIFIER, POSITION, and REASON.

The present study operationalizes AGE as a binary distinction between speakers who are 25 years old or younger and speakers who are older than 25 because the speaker's AGE is specified in different intervals in the metadata for the

three ICE components. Information on the speaker's AGE was given for 364 of the 640 extracted instances. The first panel in Figure 1 shows that *thank* is used relatively more often by speakers older than 25 (89.39 per cent) than by the younger AGE group (78.99 per cent).

Following Jautz (2013), the addressee or BENEFACTOR of the speech act of thanking was operationalized with a six-fold distinction that differentiates between thanking formulae without a BENEFACTOR and five different categories of BENEFACTORS, that is terms of endearments, groups, institution, names, and others, if they were optional and explicitly named (5), which excludes the obligatory BENEFACTOR after *thank* (6). Endearments (7) include kinship terms and other affectionate terms used 'if benefactor and beneficiary know each other well' (Jautz, 2013, p. 66). Groups of people (8) comprise terms that refer to several animate BENEFACTORS at once. If the BENEFACTOR is an institution (9), 'gratitude is often expressed for some kind of sponsoring of an event' (Jautz, 2013, p. 66). If a BENEFACTOR is called by their name, this was indicated as such (5) and every BENEFACTOR that did not fall into one of the categories mentioned above was coded as *other*. The second panel in Figure 1 illustrates that *thank* is relatively most frequently used when a BENEFACTOR coded as *other* is present (96.15 per cent) followed by thanking strategies that do not include a BENEFACTOR (87.69 per cent), names (75 per cent), and groups (75 per cent). Terms of endearment (54.55 per cent) and institutions (50 per cent) are the categories in which *thank* is least frequently used.

(5) Alright cool thanks a lot **Badrani** (ICE-SL-S1A-093#192:5:A)

(6) I want to thank **Mr Renton** very much (ICE-GB-S1B-022#103:1:C)

(7) Thank you **sweetheart** (ICE-GB-S1A-028#168:1:A)

(8) I like to thank you **all** very much for giving me such a warm welcome (ICE-IND-S1B-071#96:1:C)

(9) I express my sincere thanks to **the Central Institute of Foreign Language** (ICE-IND-S2A-033#24:1:A)

The operationalization of the extra-linguistic variable FORMALITY was based on Xiao's (2009) multidimensional analysis of register variation in five ICE components. Xiao specifies linguistic features that are associated with interactive and casual discourse and contrasts these with markers of informative elaborate discourse. He then analyses the different registers covered by the ICE components with regard to these features and concludes that linguistic elements associated with interactive and casual speech predominate in private and public spoken discourse, as well as in unscripted monologues. Features associated with informative elaborate discourse, in contrast, predominate in scripted monologues. FORMALITY was, thus, operationalized as a binary distinction between formal speech, that is scripted monologues, and informal speech, that is private and public discourse and unscripted monologues. Besides the fact that informal speech favours the use of thanking formulae, with 90.31 per cent of thanking formulae being uttered in informal speech, the data show that *thank* is relatively more frequently used in formal speech (96.77 per cent) than in informal speech (84.26 per cent) as illustrated in the third panel of Figure 1.

In regards to speakers' GENDER, information was available for 617 of the 640 extracted speech acts of thanking.[1] Of those 617 speech acts, women uttered 36.47 per cent and men uttered 63.53 per cent. Although both men and women favour *thank* over *thanks*, the fourth panel of Figure 1 shows that, in the present dataset, female speakers have a slightly higher tendency to use *thank* (87.56 per cent) than male speakers (83.67 per cent).

The presence, as in example (10), or absence, as in example (11), of an INTENSIFIER, with *very much* and *a lot* being the most frequently found variants, was also annotated. The data show that, while *a lot* is exclusively used with *thanks*, *thank* is relatively most frequently used with *very much* (95.93 per cent), followed by instances that do not include an INTENSIFIER (88.84 per cent), *very much indeed* (83.33 per cent), and others (66.67 per cent) as illustrated in the fifth panel of Figure 1.

(10) Thank you **very much** for giving me this opportunity (ICE-SL-S1B-039#29:1:B)

(11) Thanks for your sharing your thoughts on this topic (ICE-SL-S1B-023#59:1:G)

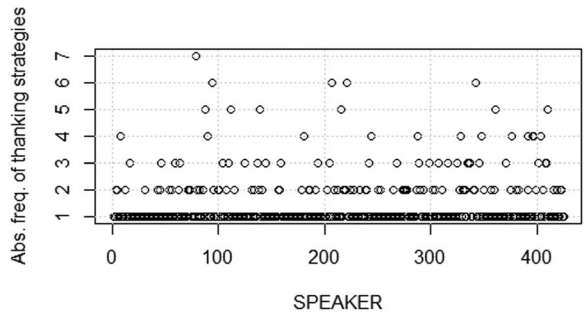


**FIGURE 2** Absolute frequency of speech acts of thanking per speaker

Unlike previous research, the present study does not investigate the distribution of expressions of gratitude in conversations, but in single speaker turns because of the sample-text nature of ICE as explained above. The POSITION of a thanking formula in a speaker turn was operationalized with a four-fold distinction, differentiating thanking formulae that are turn-initial, turn-internal, or turn-final and speaker turns that comprise only the speech act of thanking. The sixth panel in Figure 1 illustrates that, in the present dataset, *thank* is relatively most often used turn internally (93.52 per cent) while *thanks* is relatively most often used turn finally (18.05 per cent).

The last variable for which the data was annotated is REASON. A REASON for the act of thanking was considered present if the thanking formula included *for* followed by a gerund as in (12) or a noun phrase as in (13). The seventh panel in Figure 1 highlights that, for this variable, the data show a slight tendency for *thank* being more frequently used with a REASON that takes the form of *for* followed by a noun phrase (87.5 per cent) than without a REASON (85.33 per cent) or with a REASON that consists of *for* followed by a verb phrase (84.85 per cent).

(12) Thank you **for being with us** (ICE-IND-S1B-043#139:2:A)

(13) Thank you Nisha **for your extremely kind words** (ICE-SL-S2B-036#1:1:A)

## 2.3 | Statistical analysis

After the annotation, the data were checked for how many speech acts of thanking were contributed by the individual speakers.[2] As the mean number of speech acts of thanking per speaker is 1.51 and no speaker in the data produces more than seven acts of thanking, as illustrated in Figure 2, it is, in this case, not conducive to apply a mixed-effects model to partial out potential speaker-specific variation.

Conditional inference trees and random forests were used as statistical modeling techniques for the data at hand. Conditional inference trees split the dataset according to 'a series of binary questions about the values of predictor variables' (Tagliamonte & Baayen, 2012, p. 159). The formula used to compute the conditional inference tree reported in this study is THANK/-S ~ AGE + BENEFACTOR + FORMALITY + GENDER + INTENSIFIER + POSITION + REASON + VARIETY.

Random forests as implemented in the randomForest package (Liaw & Wiener, 2002) consist of several of these decision trees and – in addition to trends evident from the conditional inference tree – show a ranking of exogenous variable importance for a given endogenous variable. The variable importance values reported in the random forest are measured as mean decrease in predictive accuracy in case the variable concerned was left out of the model. In the construction of the random forest, the number of variables randomly sampled at each split in a decision tree was set to 3 and the number of trees grown in the random forest was set to 500 as these settings resulted in an optimal random forest for the data at hand.

Following Gries (under revision), the random forest reported in this study was – in comparison to the conditional inference tree – complemented by several predictors that represent all two-way interactions of the predictors with

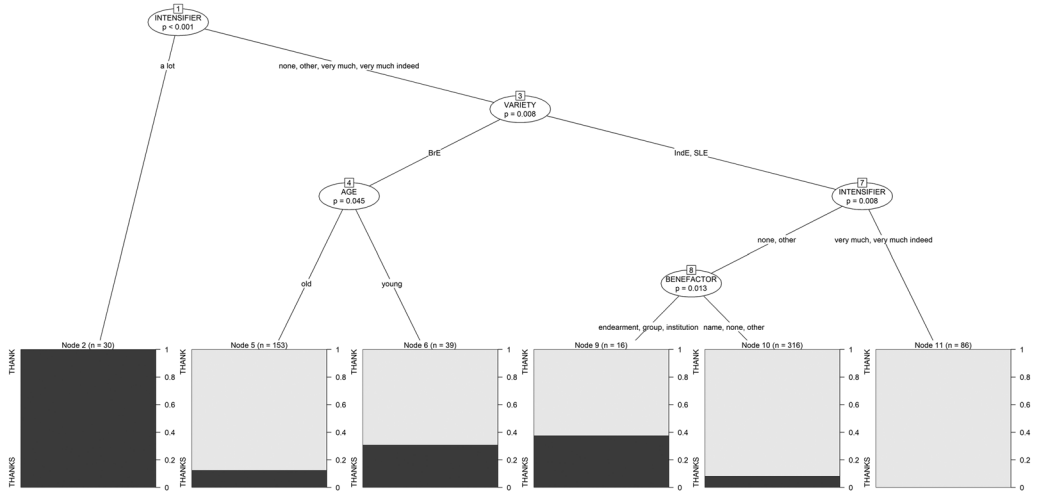


**FIGURE 3** Conditional inference tree

VARIETY since 'tree-based approaches can in fact be less good at detecting interactions than is commonly assumed' (pp. 15–16). These predictors were manually created and added to the formula for the random forest: THANK/-S ~ BENEFACTOR + FORMALITY + GENDER + INTENSIFIER + POSITION + REASON + VARIETY + VARIEY:BENEFACTOR + VARIETY:FORMALITY + VARIETY:GENDER + VARIETY:INTENSIFIER + VARIETY:POSITION + VARIETY:REASON.

## 3 | RESULTS

### 3.1 | Conditional inference tree

The conditional inference tree has a classification accuracy of 90.16 per cent, which is significantly better ($p < 0.001$) than a baseline model which predicts the more frequent variant of the endogenous variable – *thank* – all the time with a classification accuracy of 85.47 per cent.

Each node in this tree constitutes a binary split of one independent variable. The relevant levels of the respective variable are shown on the branches that lead from one node to the next until the lowest branches terminate in stacked bar plots that illustrate the relative frequencies of *thank* and *thanks*, that is the levels of the dependent variable, for the different level combinations of the independent variables that have led to these terminal nodes.

The model illustrated in Figure 3 features the independent variables INTENSIFIER, VARIETY, AGE, and BENEFACTOR. These variables are thus relevant for a speaker's choice between *thank* and *thanks*. The most important split depicted in the model is that by INTENSIFIER between *a lot* vs. *none, other, very much,* and *very much indeed*. *A lot* only occurs with *thanks*, as in (14), while *thank* is the default choice for the other levels of this variable.

(14) **Thanks** a lot for the compliment. (ICE-IND-S1A-062#185:1:A)

The next split in the model is that by VARIETY between BrE vs. IndE and SLE. For BrE, the data is then split by AGE resulting in terminal nodes illustrating that British speakers who are older than 25 years use *thank* relatively more frequently than younger British speakers if the level of INTENSIFIER is not *a lot*. For IndE and SLE the data is then again split by INTENSIFIER. The terminal node on the right end of the model shows that *very much* and *very much indeed* occur in IndE and SLE only with *thank*, as in (15).

(15) **Thank** you very much for a really interesting and detailed presentation (ICE-SL-S1B-024#1:1:A)

**TABLE 2** Mean decrease in accuracy

| Predictor | Var. Imp. |
|---|---|
| VARIETY:INTENSIFIER | 22.264768 |
| VARIETY:POSITION | 18.737376 |
| INTENSIFIER | 17.030451 |
| VARIETY:BENEFACTOR | 15.719108 |
| VARIETY:AGE | 15.315544 |
| VARIETY:REASON | 11.886116 |
| AGE | 9.273623 |
| VARIETY:FORMALITY | 7.163825 |
| POSITION | 6.898200 |
| VARIETY:GENDER | 6.889850 |
| VARIETY | 4.619325 |
| GENDER | 4.393648 |
| FORALITY | 3.279052 |
| BENEFACTOR | 2.985687 |
| REASON | 2.208129 |

Finally, the model splits the data by BENEFACTOR between *endearments*, *groups*, and *institutions* vs. *names*, *none*, and *others*. These branches terminate in nodes that combine thanking strategies by SLE or IndE speakers without or with unspecified INTENSIFIER and the just mentioned BENEFACTOR categories, respectively. In these cases, the terminal nodes highlight that IndE and SLE speakers will choose *thanks* relatively more frequently if no or an unspecified INTENSIFIER is part of the thanking formula and the BENEFACTOR takes the form of an *endearment*, a *group*, or an *institution* (16) than if – ceteris paribus – the BENEFACTOR is a *name*, not present, or categorized as *other* (17).

(16) **Thanks** man (ICE-SL-S1A-091#69:1:B)

(17) Welcome to the studio and **thank** you for being with us (ICE-IND:S1B-043#139:2:A)

The results of the conditional inference tree are illuminating in that they highlight the statistical significance of the variables AGE, BENEFACTOR, INTENSIFIER, and VARIETY. Conversely, the factors FORMALITY, GENDER, POSITION, and REASON do not seem to significantly influence the choice between *thank* and *thanks* in the varieties studied.

## 3.2 | Random forest

The random forest resulting from the above analysis performed well on the data. It obtained a prediction accuracy of 92.29 per cent which is significantly better ($p < 0.001$) than the baseline model predicting the more frequent variant *thank* which obtained a prediction accuracy of 85.95 per cent. The variable importance values reported in Table 2 were measured as mean decrease in predictive accuracy in case the variable concerned was left out of a random forest model.

Table 2 shows that the most important variable is a variable interaction with VARIETY, that is VARIETY:INTENSIFIER, whereas VARIETY as a main effect is one of the less important predictors in the model. Overall, the interaction effects rank higher than their respective main effects in terms of mean decrease in accuracy. In what follows, more detailed results will be illustrated by partial dependence plots of the interaction variables that show the effect on the predicted outcome of the variant *thank* (Figure 4).

The variable interaction of VARIETY and INTENSIFIER has the highest mean decrease in accuracy value in this random forest. The most obvious result here is that the probability of *thank* being chosen over *thanks* is highest in BrE and IndE if no INTENSIFIER is present while it is lowest for *a lot* and other INTENSIFIERS in BrE and IndE.

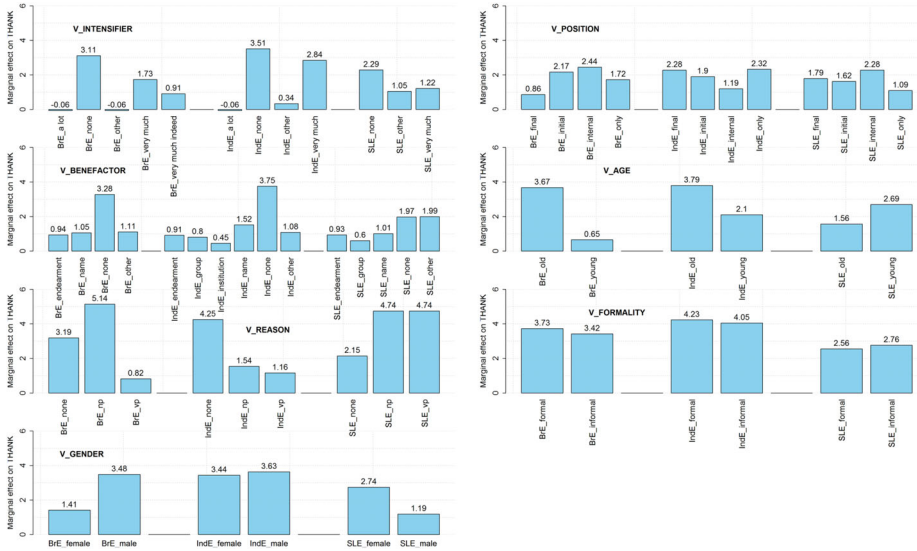


**FIGURE 4** Partial dependence plots for each of the predictor variables with VARIETY [Colour figure can be viewed at wileyonlinelibrary.com]

The interaction variable of VARIETY and POSITION has the second highest variable importance value reported in the random forest. The partial dependence plot of this interaction variable illustrates that, in BrE, *thank* is most likely at the beginning or in the middle of a speaker turn. In IndE, it is most likely to occur at a turn final position but also in a speaker turn that only comprises the thanking formula, and, in SLE, *thank* is most likely in the middle or at the end of a speaker turn.

The third panel in Figure 4 illustrates that, while *thank* is most likely to be chosen over *thanks* in BrE and IndE if no BENEFACTOR is part of the thanking formula, it is least likely to be chosen in IndE if the BENEFACTOR is an institution and in SLE if the BENEFACTOR is a group of people.

The partial dependence plot that illustrates the interaction effect of VARIETY and AGE, shows that the probability of *thank* being chosen over *thanks* is highest for the older AGE groups in BrE and IndE. In SLE however, the probability for *thank* being chosen is higher in the younger than in the older AGE group.

The interaction of VARIETY and REASON is shown in the fifth panel of Figure 4. In British and IndE, *thank* is least likely if the REASON takes the form of a verb phrase, as in *thanks for coming*. In SLE, however, the probability for *thank* being chosen over *thanks* is highest if a REASON is present regardless of the form it takes.

The interaction of VARIETY and FORMALITY is illustrated in the sixth panel of Figure 4. The most striking result is that *thank* is least likely in formal contexts in SLE and most likely in formal contexts in BrE and in IndE.

The last interaction to be discussed is represented in the seventh panel of Figure 4, showing that females in BrE and males in SLE are least likely to choose *thank* over *thanks*. In BrE, *thank* is more likely with male speakers whereas, in SLE, it is by far more likely with female speakers.

## 4 | DISCUSSION

The two statistical analyses have shed light on those factors that guide speakers' choices between *thank* and *thanks* in the three varieties under investigation. While VARIETY is part of the conditional inference tree and has proven to be statistically significant in this model, it shows a rather low Mean Decrease in Accuracy value in the random forest model. However, the new statistical modelling technique adapted from Gries (2019) highlights that speakers

of different varieties behave differently across the levels of the other factors this study controlled for. The structural and sociobiographic factors significantly influencing the choice of thanking formulae are AGE, BENEFACTOR, and INTENSIFIER, all of which are part of the conditional inference tree with INTENSIFIER constituting the first split in the tree and thus being the most important variable in this model. If the INTENSIFIER is *a lot*, the speaker will always choose *thanks* – as in *thanks a lot* – and if another or no INTENSIFIER is part of the thanking formula, the speaker will choose *thank* over *thanks*. In the random forest model, however, these factors on their own turn out to be less important than their interaction variables with VARIETY, which are among the most important predictors in the random forest. Without looking at the random forest model, one would have assumed that FORMALITY, GENDER, POSITION, and REASON do not affect the speaker's choice between *thank* and *thanks*. However, the random forest highlights the importance of their interaction variables. Especially, the interaction terms of POSITION with VARIETY is among the most important predictors, which would not be expected by looking at the conditional inference tree that a) cannot explicitly account for variable interaction and b) does not profile POSITION as a significant predictor. This interaction shows that BrE speakers will choose *thank* mostly turn-initially or -internally, IndE speakers will choose *thank* turn-finally and if the speech act only comprises the thanking formula, and SLE speakers are most likely to choose *thank* over *thanks* turn-internally. The complementation of the conditional inference tree and the random forest modelling thus highlights different sets of variables that are of importance for the choice between *thank* and *thanks*.

With regard to the third research question concerned with the extent to which the choice between *thank* and *thanks* in South Asian Englishes aligns with that in BrE and related implications for the notion of pragmatic nativisation, the statistical analyses applied to the data have highlighted significant differences in the choice of thanking strategies across the three varieties under scrutiny. The conditional inference tree has illustrated the statistical significance of the main effect of VARIETY, while the random forest has accentuated the influences of other variables on the effect that VARIETY has on the choice of thanking formulae. Speakers of different varieties behave, for example, differently depending on the POSITION of the thanking formula within the speaker turn or depending on the form of INTENSIFIER that is part of the formula. These interaction effects will now be further investigated by looking at the partial dependence plots that illustrate the predicted outcome of the variant *thank*.

The partial dependence plots highlight that, in comparison to BrE speakers, SLE speakers will use *thank* more often turn-finally, and if a REASON takes the form of a verb phrase. Furthermore, *thank* is more frequently used by young speakers and by females in SLE than in BrE. IndE speakers use *thank* more frequently if the thanking formula is intensified using *very much*, if it occurs turn-finally, if the speaker turn only consists only of the thanking formula, or if no REASON is part of the thanking strategy if compared to BrE speakers. If the use of thanking strategies by SLE speakers is compared to the use by IndE speakers, the partial dependence plots highlight that SLE speakers use *thank* less often if no BENEFACTOR is part of the thanking formula. In addition to that, SLE speakers use *thank* more often if it is used turn-internally, or if a REASON is part of the formula. Female speakers in Sri Lanka also show a tendency to use *thank* more frequently than speakers in India do.

Both the conditional inference tree and the random forest highlight differences between BrE and the two Asian varieties of English under scrutiny. The conditional inference tree divides BrE from IndE and SLE in its second split while the partial dependence plots created based on the random forest highlight behavioural profiles for IndE and SLE speakers that differ from those of BrE speakers. Previous research has analysed these two varieties in terms of Schneider's (2003, 2007) model and found IndE to be more advanced in its nativisation process than SLE (Bernaisch, 2015; Mukherjee, 2007). Revis and Bernaisch (2020) have corroborated these findings for pragmatic nativisation of filled and unfilled pauses in IndE and SLE. The conditional inference tree in this study highlights similarities of the South Asian Englishes by splitting them from BrE. However, a closer look into the behavioural profiles of IndE and SLE speakers reveals differences in these varieties that do not uniformly support the argument that IndE is more advanced in its nativisation process than SLE. Bernaisch's model of the emergence of distinctive structural profiles of semiautonomous varieties of English illustrates how exonormative and endonormative tendencies can have unequally strong influences on a variety depending on the subject or structure of investigation (Bernaisch, 2015, p. 215).

For SLE speakers, the partial dependence plots presented in Figure 4 highlight very distinct patterns for females and young people. The plots highlight that those speakers use *thank* more frequently than male and older speakers do, which is the case for neither BrE nor IndE. This phenomenon constitutes an interesting basis for a diachronic study of thanking formulae in SLE, as young females have been found to drive linguistic change (Lange, 2012). Although *thank* is an established variant of thanking strategies, a diachronic analysis might yield insights into a potential ongoing linguistic change in SLE.

The conditional inference tree and the random forest reported in this paper highlight that the three varieties under scrutiny show distinct patterns for the choice of thanking formulae. These distinct usage patterns provide enough evidence to argue that SLE has developed localised norms for thanking strategies. This can be regarded as one facet – next to many others – of pragmatic nativisation in SLE. However, interestingly, IndE speakers exhibit usage patterns of thanking formulae that are more similar to those in BrE than in SLE, as the usage patterns for AGE and BENEFACTOR highlight.

## 5 | CONCLUSION

The present paper has analysed the choice between *thank* and *thanks* in SLE using a conditional inference tree and a random forest analysis. Theoretically, this study adds to the description of the choice of thanking formulae by identifying AGE, BENEFACTOR, INTENSIFIER, and VARIETY as statistically significant variables and by highlighting the importance of interaction variables with VARIETY that show higher variable importance values than their respective main effects. From a methodological perspective, the random forest analysis complements the conditional inference tree in that it adds variable interactions to the analysis and thus provides a more fine-grained perspective on how VARIETY affects thanking choices in the varieties covered. Furthermore, these findings contribute to the description of pragmatic nativisation in SLE as the choice of thanking strategies in SLE shows a distinct, VARIETY-specific pattern that is mirrored by neither BrE, the historical input variety, nor by IndE as a likely linguistic epicentre in South Asia

As the ICE components used in this study were too small to consider other variants of thanking strategies than *thank* and *thanks* – for example appreciate, gratitude – future research into thanking strategies in SLE might work with larger datasets such as GloWbE (Davies & Fuchs, 2013) to take these variants into account and possibly obtain complementary results from a large, but sociobiographically poorer database with regard to the two variants considered in this study. However, even an analysis including other English variants such as *appreciate* does not consider the multilingual background of Sri Lankan and Indian English speakers and thus misses thanking strategies motivated by structures in the indigenous languages of India and Sri Lanka. Furthermore, the metadata available with the ICE component was relatively limited. This is especially true of information on the speaker's AGE, which was only given for 364 of 640 extracted expressions of gratitude; this resulted in an exclusion of more than 200 data points in the statistical analyses, which is why datasets with consistent metainformation from speakers constitutes a desideratum. Future research might also apply the present analysis to other South Asian varieties of English in order to compare the development of pragmatic nativisation of thanking strategies in varieties of English with regional proximity.

### ACKNOWLEDGEMENTS

I thank Tobias Bernaisch for his help and advice throughout the process of writing this paper.

Open access funding enabled and organized by Projekt DEAL.

### CONFLICT OF INTEREST STATEMENT

There is no conflict of interest to declare.

## NOTES

[1] Information on AGE and GENDER were annotated with the help of biodata scripts for R written by Martin Schweinberger (http://www.martinschweinberger.de/blog/resources/).

[2] The number of thanking formulae per speaker was extracted using *exact.matches.2* (Gries, 2016) for R.

## ORCID

*Nina Funke* https://orcid.org/0000-0002-4140-7136

## REFERENCES

Anthony, L. (2017). *AntConc (Version 3.5.0) [computer software]*. Tokyo, Japan: Waseda University. Retrieved from http://www.laurenceanthony.net/software/antconc/

Apte, M. L. (1974). 'Thank you' and south Asian languages: A comparative sociolinguistic study. *Linguistics*, *136*, 67–90. https://doi.org/10.1515/ling.1974.12.136.67

Azima, M., & Hesabi, A. (2015). Speech Acts of Thanking: A Contrastive Analysis among Iranian EFL learners in terms of gender and level of proficiency. *International Letters of Social and Humanistic Sciences*, *59*, 76–84. https://doi.org/10.18052/www.scipress.com/ILSHS.59.76

Bamgose, A. (1998). Torn between the norms: Innovations in world Englishes. *World Englishes*, *17*, 1–14. https://doi.org/10.1111/1467-971X.00078

Bernaisch, T. (2015). *The lexis and lexicogrammar of Sri Lankan English*. Amsterdam: John Benjamins Publishing. https://doi.org/10.1075/veaw.g54

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press. Retrieved from www.cambridge.org/9780521308625

Cheng, S. W. (2005). An exploratory cross-sectional study of interlanguage pragmatic development of expressions of gratitude by Chinese learners of English (Doctoral Dissertation). University of Iowa, Iowa. https://doi.org/10.17077/etd.5mjl4i0y

Cheng, S. W. (2010). A corpus-based approach to the study of speech act of thanking. *Concentric: Studies in Linguistics*, *36*, 257–274. Retrieved from http://www.concentric-linguistics.url.tw/issues_detail.php?nid=25

Coulmas, F. (1981). 'Poison to your soul!' Thanks and apologies contrastivley viewed. In F. Coulmas (Ed.), *Conversational routine. Explorations in standardized communication situations and prepaterned speech* (pp. 69–91). The Hague: Mouton.

Davies, M., & Fuchs, R. (2013). *Corpus of global web-based English*. Brigham: Brigham Young University.

Eisenstein, M., & Bodmann, J. W. (1986). 'I very appreciate': Expressions of gratitude by native and non-native speakers of American English. *Applied Linguistics*, *7*, 167–185. https://doi.org/10.1093/applin/7.2.167

Eisenstein, M., & Bodmann, J. W. (1993). Expressing gratitude in American English. In G. Kasper & S. Blum-Kulka (Eds.), *Interlanguage pragmatics* (pp. 64–81). Oxford: Oxford University Press.

Farnia, M., & Suleiman, R. R. R. (2009). An interlanguage pragmatic study of expressions of gratitude by Iranian EFL learners: A pilot study. *Malaysian Journal of ELT Research*, *5*, 108–140. https://doi.org/10.1016/j.sbspro.2014.03.624

Fuchs, R. (2015). *Speech rhythm in varieties of English: Evidence from educated Indian English and British English*. Singapore: Springer. https://doi.org/10.1007/978-3-662-47818-9

Greenbaum, S. (1991). The development of the International Corpus of English. In K. Aijmer & B. Altenberg (Eds.), *English corpus linguistics: Studies in honour of Jan Svartvik* (pp. 83–91). London: Longman.

Gries, S. Th. (2016). *Quantitative corpus linguistics with R. A practical introduction* (2nd ed.). New York: Routledge.

Gries, S. Th. (2019). On classification trees and random forests in corpus linguistics: Some words of caution and suggestions for improvement. *Corpus Lingistics and Linguistic Theory*. [available ahead of publication].

Gries, S., & Bernaisch, T. (2016). Exploring epicentres empirically: Focus on south Asian Englishes. *English World-Wide*, *37*, 1–25. https://doi.org/10.1075/eww.37.1.01gri

Gueldenring, B. A. (2017). Emotion metaphors in new Englishes: A corpus-based study of ANGER. *Cognitive Linguistic Studies*, *4*, 82–109. https://doi.org/10.1075/cogls.4.1.05gul

Gunesekera, M. (2005). *The postcolonial identity of Sri Lankan English*. Colombo: Katha Publishers.

Heller, B., Bernaisch, T., & Gries, S. (2017). Empirical perspectives on two potential epicenters: The genitive alternation in Asian Englishes. *ICAME Journal*, *41*, 111–144. https://doi.org/10.1515/icame-2017-0005

Hinkel, E. (1994). Pragmatics of interaction: Expressing thanks in a second language. *Applied Language Learning*, *5*, 73–91.

Hoffmann, S., Blass, A.-K., & Mukherjee, J. (2017). Canonical tag questions in Asian Englishes. In M. Filppula, J. Klemola, & D. Sharma (Eds.), *The Oxford handbook of world Englishes* (pp. 697–714). New York: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199777716.013.025

Holmes, J. (1984). Modifying illocutionary force. *Journal of Pragmatics*, *8*, 345–365. https://doi.org/10.1016/0378-2166(84)90028-6

Hothorn, T., & Zeileis, A. (2015). Partykit: A modular toolkit for recursive partytioning in R. *Journal of Machine Learning Research*, *16*, 3905–3909.

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, *15*, 651–674.

Jautz, S. (2013). *Thanking formulae in English: Explorations across varieties and genres*. Philadelphia: John Benjamins. https://doi.org/10.1075/pbns.230

Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: The English language in the outer circle. In R. Quirk & H. G. Widdowson (Eds.), *English in the world: Teaching and learning the language and literatures* (pp. 11–30). Cambridge: Cambridge University Press.

Kachru, B. B. (1992). Meaning in deviation: Toward understanding non-native English texts. In B. B. Kachru (Ed.), *The other tongue: English across cultures* (pp. 301–326). Urbana: University of Illinois Press.

Kachru, Y. (2008). Language in social and ethnic interaction. In B. B. Kachru, Y. Kachru, & S. N. Sridhar (Eds.), *Language in South Asia* (pp. 345–360). Cambridge: Cambridge University Press. Retrieved from https://doi-org.ezproxy.uni-giessen.de/10.1017/CBO9780511619069

Lange, C. (2012). *The syntax of spoken Indian English*. Amsterdam: John Benjamins. https://doi.org/10.1075/veaw.g45

Leitner, G. (1992). English as a pluricentric language. In M. Clyne (Ed.), *Pluricentric languages: Differing norms in different nations* (pp. 179–237). Berlin: Mouton. https://doi.org/10.1515/9783110888140.179

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*, 18–22.

Mair, C. (2013). The world system of Englishes: Accounting for the transnational importance of mobile and mediated vernaculars. *English World Wide*, *34*, 253–278. https://doi.org/10.1075/eww.34.3.01mai

Meyler, M. (2007). *A dictionary of Sri Lankan English*. Colombo: Mirisgala.

Mukherjee, J. (2007). Steady states in the evolution of new Englishes: Present-day Indian English as an equilibrium. *Journal of English Linguistics*, *35*, 157–187. https://doi.org/10.1177/0075424207301888

Nihalani, P., Tongue, R. K., & Hosali, P. (2004). *Indian and British English: A handbook of usage and pronunciation* (2nd edition). Oxford: Oxford University Press.

Norrick, N. R. (1978). Expressive illocutionary acts. *Journal of Pragmatics*, *2*, 277–291. https://doi.org/10.1016/0378-2166(78)90005-X

Okomato, S., & Robinson, W. P. (1997). Determinants of gratitude expressions in England. *Journal of Language and Social Psychology*, *16*, 411–433. https://doi.org/10.1177/0261927İ970164003

Revis, M., & Bernaisch, T. (2020). Pragmatic nativisation in Asian Englishes? Evidence from filled and unfilled pauses. *World Englishes*, *39*, 135–153. https://doi.org/10.1111/weng.12387

Schauer, G. A., & Adolphs, S. (2006). Expressions of gratitude in corpus and DCT data: Vocabulary, formulaic sequences, and pedagogy. *System*, *34*, 119–134. https://doi.org/10.1016/j.system.2005.09.003

Schilk, M. (2011). *Structural nativization in Indian English lexicogrammar*. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.46

Schneider, E. W. (2003). The dynamics of new Englishes: From identity construction to dialect birth. *Language*, *79*, 233–281. Retrieved from https://www.jstor.org/stable/4489419

Schneider, E. W. (2007). *Postcolonial English: Varieties around the world*. Cambridge: Cambridge University Press.

Schneider, K. P., & Barron, A. (2008). *Variational pragmatics: A focus on regional varieties in pluricentric languages*. Amsterdam: Benjamins. https://doi.org/10.1075/pbns.178

Sedlatschek, A. (2009). *Contemporary Indian English: Variation and change*. Amsterdam: John Benjamins. https://doi.org/10.1075/veaw.g38

Tagliamonte, S., & Baayen, H. (2012). Models, forests and trees of York English: "Was/were" variation as a case study for statistical practice. *Language Variation and Change*, *24*, 135–178. https://doi.org/10.1017/S0954394512000129

Werner, J., & Mukherjee, J. (2012). Highly polysemous verbs in new Englishes: A corpus-based study of Sri Lankan and Indian English. In S. Hoffmann, P. Rayson, & G. Leech (Eds.), *English corpus linguistics: Looking back, moving forward: Papers from the 30th international conference on English language research on computerized corpora (ICAME 30), Lancaster, UK, 27–31 May 2009* (pp. 249–266). Amsterdam: Rodopi. https://doi.org/10.1163/9789401207478_018

Xiao, R. (2009). Multidimensional analysis and the study of world Englishes. *World Englishes*, *28*, 421–450. https://doi.org/10.1111/j.1467-971X.2009.01606.x

**How to cite this article:** Funke N. Pragmatic nativisation of thanking in South Asian Englishes. *World Englishes*. 2022;41:136–150. https://doi.org/10.1111/weng.12517