UNIVERSITÄT AUGSBURG
Universitätsbibliothek

OPEN ACCESS
OPUS AUGSBURG

The existential there construction in spoken South Asian Englishes

Kathrin Kircili, Julia Degenhardt, Tobias Bernaisch, Sandra Götz

Angaben zur Veröffentlichung / Publication details:

Kircili, Kathrin, Julia Degenhardt, Tobias Bernaisch, and Sandra Götz. 2025. "The existential there construction in spoken South Asian Englishes." *Asian Englishes* 27 (1): 105–22. https://doi.org/10.1080/13488678.2024.2390240.

Nutzungsbedingungen / Terms of use:

CC BY 4.0

Dieses Dokument wird unter folgenden Bedingungen zur Verfügung gestellt: / This document is made available under these conditions:
CC-BY 4.0: Creative Commons: Namensnennung
Weitere Informationen finden Sie unter: / For more information see:
https://creativecommons.org/licenses/by/4.0/deed.de

Asian Englishes

ISSN: 1348-8678 (Print) 2331-2548 (Online) Journal homepage: www.tandfonline.com/journals/reng20

The existential there construction in spoken South Asian Englishes

Kathrin Kircili, Julia Degenhardt, Tobias Bernaisch & Sandra Götz

To cite this article: Kathrin Kircili, Julia Degenhardt, Tobias Bernaisch & Sandra Götz (2025) The existential there construction in spoken South Asian Englishes, Asian Englishes, 27:1, 105-122, DOI: 10.1080/13488678.2024.2390240

To link to this article: https://doi.org/10.1080/13488678.2024.2390240

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

Published online: 27 Aug 2024.

Submit your article to this journal

Article views: 938

View related articles

View Crossmark data

Citing articles: 4 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=reng20

ASIAN ENGLISHES
2025, VOL. 27, NO. 1, 105–122
https://doi.org/10.1080/13488678.2024.2390240

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

The existential there construction in spoken South Asian Englishes

Kathrin Kircili[a], Julia Degenhardt[b], Tobias Bernaisch[b] and Sandra Götz[a]

[a]Department of English and American Studies, Philipps-Universität, Marburg, Germany; [b]Department of English, Justus-Liebig-Universität, Gießen, Germany

ABSTRACT

The study of South Asian Englishes has long focused on the largest and oldest local English as a second language variety, Indian English (IndE). Methodologically, research on South Asian Englishes has predominantly relied on descriptive statistics and monofactorial analyses while multifactorial analyses – although statistically superior – have remained rare. Against this background, the present article compares the use of the existential there construction (ExT) in spoken IndE and spoken Sri Lankan English (SLE) along with British English (BrE) as a reference variety, considering a variety of sociobiographic and structural predictors. The latter are shown to guide the speakers' choice of either initial or non-initial ExTs. Likewise, the findings indicate a highly significant difference between the proportional uses of said choice across the three varieties, with IndE employing non-initial ExTs more frequently than SLE and BrE.

ARTICLE HISTORY
Received 23 October 2023
Accepted 5 July 2024

KEYWORDS
Existential there; South Asian varieties of English; multifactorial analysis; spoken English

1. Introduction

The history of the English language in South Asia commenced at the beginning of the seventeenth century as the result of British colonial activities. Ever since, numerous second language varieties – shaped both by indigenous languages and by cultural customs and traditions – have evolved. Today, within the Outer Circle varieties of English (see Kachru, 1992), speakers of South Asian Englishes – including, for example, Indian English (IndE), Sri Lankan English (SLE), Pakistani English and Nepali English – constitute the largest number of English as a Second Language (ESL) speakers worldwide. In fact, according to Bolton and Bacon-Shone (2020), the total number of what they refer to as 'English-knowing' people in Asia, that is, those who either speak or acquire the language, is assessed to exceed 800 million (Bolton & Bacon-Shone, 2020, p. 49).

According to Schneider's dynamic model, there are various developmental stages that second language varieties pass through (see Schneider, 2007) before reaching the point of differentiation, which allows for variety-internal variation. In the context of South Asian varieties, and in particular those that are of interest to the present article, namely IndE and SLE,[1] the centuries-long localized employment and development have

CONTACT Kathrin Kircili kathrin.kircili@uni-marburg.de Department of English and American Studies, Philipps-Universität Marburg, Wilhelm-Röpke-Straße 6 , Room 00016, Marburg 35032, Germany

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

already allowed them to enter into the linguistically most diverse and interesting stages, namely *nativization*, as the stage during which structural patterns start to develop independently from the British norm, and *endonormative stabilization*, which entails the recognition and increasing codification and thus acceptance of these new structures within the varieties.

As Figure 1 illustrates, SLE, which has seen the development of local standards on all structural levels of language, including, for example, pronunciation, lexis and grammar (see Bernaisch, 2015), still leans towards the nativization stage, while IndE, as a 'semiautonomous variety' (Mukherjee, 2007, p. 182) that is about 200 years older and thus more established than SLE, is presumed to have evolved farther into the phase of endonormative stabilization than SLE. Its special status as both the oldest and the largest post-colonial ESL variety has, in fact, also led to a pronounced focus on IndE in previous research endeavours, which have revealed, among other things, that while 'all the languages that have been spoken in South Asia have [. . .] influenced each other so thoroughly that we can speak of a South Asian *Sprachbund*' (Witzel, 1995, p. 108; original emphasis), as suggested by researchers like Jakobson (1962) or, more recently, Thomason (2001) and Bernaisch and Lange (2012), IndE can be considered unique in that it has been found to occupy the role of a regional epicentre. This implies that, as a result of geographical proximity as well as language contact, it may exert an influence on neighbouring varieties of English (see, for example, Gries & Bernaisch, 2016; Leitner, 1992).

The present study intends to shed light on the structural properties of one of these geographically close varieties, namely SLE, which, even though it is the only South Asian variety with a similar degree of institutionalization to IndE, has so far only received limited scholarly attention, with the majority of studies providing qualitative accounts (see, for example, Gunesekera, 2005; Meyler, 2007). Other examples include the studies by Werner and Mukherjee (2012) with a focus on semantics, by Kraaz and Bernaisch (2020) as well as Degenhardt and Bernaisch (2022) who took quantitative approaches to pragmatics, by Meyler (2007) and Bernaisch (2015) on lexis and by Bernaisch, Gries, and Mukherjee (2014) as well as Gries and Bernaisch (2016) on lexico-grammar. As for studies on syntax in general, and the investigation of syntactic nativization processes in SLE in particular, there is only a limited body of research available – especially with regard to quantitative statistical approaches. Laudable exceptions include Bernaisch (2015) and Götz (2017). Both researchers take an interest in written language, however, with the latter making use of the South Asian varieties of English (SAVE) corpus (Bernaisch, Koch, Mukherjee, & Schilk, 2011) and the former employing both the SAVE corpus and the International Corpus of English

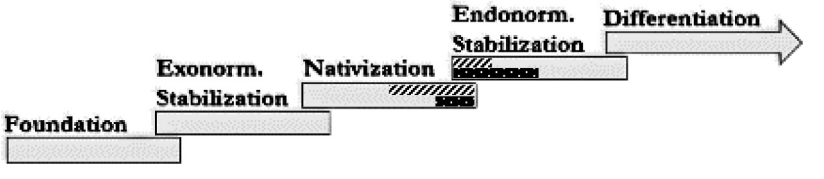

Figure 1. Possible localization of IndE (dotted bars) and SLE (striped bars) within the five stages of Schneider's Dynamic Model.

(ICE) (Körtvelyessy, Bernaisch, Mukherjee, & Mendis, 2012) along with the Google Advanced Search Tool.

In this study, we would like to contribute to the academic discourse by reporting the results of a multifactorial analysis on spoken SLE or, to be more precise, the use of one particular feature, namely the existential there construction (ExT), as in 'There are many people in this room', which poses a common feature in conversational speech (see Biber, Johannson, Leech, Conrad, & Finegan, 1999, p. 953). The next section focuses on a detailed description and definition of this non-canonical sentence pattern before giving an account of established theories and previous research into the topic. We then zoom in on our own study. As a first step, we provide the research gaps and research questions as well as a detailed description of the database and methodology, before presenting and discussing the quantitative and qualitative results. The final section is dedicated to the conclusion and outlook, also including potential starting points for future research.

2. The existential there construction

2.1. Defining the existential there construction

Following the seven basic clause patterns identified by Quirk, Greenbaum, Leech, and Svartvik (1985), the English language follows a so-called canonical S_n–V–X constituent order and has, in fact, been described as having a fairly rigid sentence structure (see Givón, 2001, p. 235). In spite of this, however, there are multiple linguistic as well as extralinguistic factors – including information structural concerns or interactional contexts – that either allow for or even demand the manipulation of, and thus the deviation from, this basic structure. One example of a sentence pattern that entails such a deviation is the ExT, which, in its traditional form, is formed by employing an expletive or dummy 'there', whose use is neither deictic nor locative and which functions as the grammatical subject (S_g), followed by a form of BE and the notional subject (S_n), carrying the semantic weight.[2] The latter commonly constitutes an indefinite noun phrase (see, for example, Huddleston & Pullum, 2002; Prado Alonso, 2008; Quirk, Greenbaum, Leech, & Svartvik, 1985) with varying degrees of complexity based on which a distinction can be made between *bare* and *extended* ExTs. While the former contain a simple, indefinite noun phrase with or without postmodification – illustrated by Examples 1 and 2 in the following – or an optional adverbial (Example 3), the latter may be composed of a noun phrase plus a locative (Example 4), a complement (Example 5) or an infinitive-clause (Example 6), a participle-clause (Example 7) or a relative-clause extension (Example 8) (see Huddleston & Pullum, 2002, p. 1393; Quirk, Greenbaum, Leech, & Svartvik, 1985, pp. 1406–1408):[3]

2.1.1. Bare existentials

(1) There is a problem. (vs. *A problem is ø.)
(2) There is a code of conduct. (vs. *A code of conduct is ø.)
(3) There is a problem again. (vs. *A problem is again.)

2.1.2. Extended existentials

(4) There is a strange person outside. (vs. A strange person is outside.)
(5) There were many people absent. (vs. Many people were absent.)
(6) There are some students to take a resit. (vs. Some students are to take a resit.)
(7) There are some kids playing cards. (vs. Some kids are playing cards.)
(8) There was one boy that kept complaining. (vs. One boy kept complaining.)

As Examples 1–8 illustrate, a clear distinction between the two structural groups can be made based on the possibility of constructing a grammatically complete canonical counterpart to the non-canonical variant (indicated in parentheses), which is only possible with extended ExTs.

2.2. Existing theories and previous findings on the ExT

As briefly mentioned in Section 2.1, use of the ExT is inextricably intertwined with versatile (information) structural concerns and concepts. Apart from the previously discussed realizations as bare or extended ExTs, aspects such as subject–verb concord, that is whether the copular BE is used in the singular or the plural, and also in relation to *tense* or *polarity*, that is whether an ExT is negated (using 'no' or 'not') or whether it is used in the affirmative, as well as *definiteness* (see, for example, Collins, 2012; Crawford, 2005; Martínez Insua & Palacios Martínez, 2003; Winkle, 2015), have been of as much interest to researchers as the pattern's relation to well-known principles like that of end-weight (Behaghel, 1909). The latter implies that syntactically weightier constituents are placed in sentence-final position in order to ease cognitive efforts in both production and perception (see Arnold, Losongco, Wasow, & Ginstrom, 2000; Breivik, 1981; Martínez Insua, 1998). Hence, in the case of the ExT, the commonly longer notional subject is moved to a post-verbal position while the dummy 'there' stands proxy as the grammatical subject.

Another principle that is associated with the ExT concerns the information flow within a given message – be it spoken or written. In the English language, messages are usually structured in a way that – in line with what was discussed earlier concerning the end-weight principle – is cognitively most efficient. Consequently, new or previously unknown information that is introduced to the ongoing discourse is usually 'weightier' and succeeds information that has already been established and is therefore old (see, for example, Clark & Haviland, 1977; Prince, 1981). As previous research has confirmed (see, for example, Biber, Johannson, Leech, Conrad, & Finegan, 1999; Birner & Ward, 1998; Ward & Birner, 2004), the ExT poses a fruitful means to adhere precisely to this given–new progression, given that the notional subject of the ExT commonly constitutes an indefinite noun phrase which – in turn – indicates the introduction of information the addressee is not yet familiar with. This is also confirmed by Huddleston and Pullum (2002), according to whom the construction 'is characteristically used to introduce addressee-new entities into the discourse, and for this reason the displaced subject [noun phrase] is usually indefinite. In many cases, the presence of an indefinite [noun phrase] makes the existential pragmatically obligatory in that the corresponding non-existential is infelicitous' (Huddleston & Pullum, 2002, p. 1396).

Naturally, the adherence to the principles of end-weight and the given–new structure also goes hand in hand with another major information structural principle, namely that of end-focus (see Huddleston & Pullum, 2002, p. 1371). This refers to the tendency of important information to be placed at the end of a construction (see Behaghel, 1909, p. 138; Biber, Johannson, Leech, Conrad, & Finegan, 1999; Huddleston & Pullum, 2002; Quirk, Greenbaum, Leech, & Svartvik, 1985).

Turning from this rather general account of previous research endeavours concerning the characteristics and purposes of the ExT in the major standards of the English language, namely BrE and American English, to the South Asian context, all of these aspects have already been scrutinized for IndE (see, for example, Balasubramanian, 2009; Collins, 2012; Lange, 2012; Winkle, 2015). The findings suggest that the variety exhibits both interesting deviations from its historical input variety and innovative realizations.

Collins (2012), for example, who investigated singular concord (e.g. 'There **is** *a lot of discussions* about the ongoing conflict') in a total number of nine spoken components of ICE, found that it was markedly favoured by Inner Circle varieties, with the highest frequency being determined for Australian and American English (52.7% and 44.4%, respectively), while the British participants used it in 25.4% of the cases. This frequency is, in fact, higher than the highest percentage determined for any of the Outer Circle varieties, namely 20.7% determined for Hong Kong English. IndE speakers opted for singular concord in only 7.0% of the cases. Another interesting observation by Collins (2012) is the fact that there is a strong correlation between the singular agreement variable and that of tense. He reports a general tendency across the investigated varieties to attest singular concord with present tense. Winkle (2015), who investigated a variety of non-canonical patterns in nine spoken ICE components, among them IndE, found that IndE exhibited the lowest percentage of singular concord with plural notional subjects and that, while the determiner type did not yield significant results when taken as a single predictor, the consideration of individual types of determiners revealed that quantifiers with the indefinite determiner 'a' correlate with the singular agreement predictor. In addition, her analyses show that IndE speakers clearly favour bare ExTs. Of the 643 instances that were detected, 377 (i.e. 58.6%) were, in fact, used in bare constructions in the variety while the highest percentage of all English as a Native Language backgrounds is reported for the BrE data, with 46.0% (Winkle, 2015, p. 173). In contrast, Lange's (2012) analysis shows that both IndE and BrE speakers clearly favour bare over bi-clausal ExTs – a somewhat narrower notion adopted from Lambrecht (1994).

One example of an innovative realization whose attestations serve as clear indicators of ongoing structural nativization processes is the use of the so-called non-initial or Indian ExT (see, for example, Balasubramanian, 2009; Lange, 2012) as illustrated in the following:

(9a) Initial (traditional): There$^{[Sg]}$ are$^{[V]}$ many questions$^{[Sn]}$.
(9b) Non-initial (Indian): Many questions$^{[Sg,\ Sn]}$ are$^{[V]}$ there$^{[Cg]}$.

As can be seen, in comparison with the traditional ExT (Example 9a), the non-initial or Indian ExT (Example 9b), which has previously been described as a feature of spoken rather than written language (Balasubramanian, 2009), represents a particularly intriguing structure for various reasons. First and foremost, the need for a grammatical subject is fulfilled by the notional subject itself, which moves back into sentence-initial

position (see Example 9b). At the same time, however, the necessity of the copular BE to be succeeded by a subject complement to render the construction grammatical is met by means of the dummy 'there' which now fulfils the function of a grammatical complement (see the final constituent in Example 9b).

Against the backdrop of what has been discussed concerning the information structural principles of relevance to the use of the ExT, namely that of end-weight and end-focus as well as the given–new progression, this newly identified variant clearly opposes each: the weighty component comes first and focus is placed on the initial element which (in the majority of cases) contains new information. However, even though still the majority, Lange (2012) determined that as many as 43.3% of the notional subjects in non-initial ExTs were actually discourse-old. When taking a closer look at their functions, she found that they, for example, served to repeat 'the immediately preceding topic NP' (Lange, 2012, p. 106) as well as to sum up or refer back 'to what has been said before' (Lange, 2012, p. 107). However, even with discourse-new notional subjects, the contexts of use of this variant of the ExT offer some interesting insights. Lange, for example, discovered that it was frequently employed in situations in which a topic was introduced to the discourse but actually represented 'readily recoverable background knowledge' (Lange, 2012, p. 110) or information that was somewhat 'active in the speakers' consciousness' (Lange, 2012, p. 111).

Based on these accounts, it also seems reasonable to assume principles like that of task urgency (Givón, 1985) at play, which is employed to '[highlight] the main purpose of the utterance' (Biber, Johannson, Leech, Conrad, & Finegan, 1999, p. 904). This claim remains to be investigated further, however.

Due to the overarching influence that has been attested for IndE and the fact that, so far, no research has been conducted into whether SLE also exhibits tendencies to employ comparable structures, we are, among other things, interested in the extent to which similarities and differences can be observed concerning the use of the traditional versus the Indian ExT in both IndE and SLE. The precise nature of the research gaps identified and our research questions will be discussed in the following section.

3. The existential there construction in spoken Sri Lankan English

3.1. Research gaps and research questions

There are two major research gaps that have been identified for the present project. Firstly, there is a lack of multifactorial analyses of SLE given that, as mentioned previously, the majority of studies provide qualitative accounts or, if quantitative, most of them focus on descriptive statistics and monofactorial analyses. Secondly, with the few notable exceptions highlighted earlier, we have observed a lack of empirically reliable studies into syntactic structures of SLE in general and spoken SLE in particular.

In an attempt to contribute to filling these research gaps, we intend to answer the following research questions:

(1) Can quantitative differences between the three varieties (BrE [control], SLE and IndE) be attested for the different types of the ExT?

(2) (a) Which predictors guide structural choices of the ExT in both South Asian Englishes as well as in BrE and
(b) to what extent are the effects of said predictors cross-varietally distinct?
(3) Is the structural evidence of the ExT compatible with the more advanced evolutionary status of IndE compared to SLE, given that IndE as a more evolved variety is to be expected to feature more localized forms of the ExT?

3.2. Database and methodology

The present study works with the spoken parts of the BrE, IndE and SLE components of the ICE. Each component includes private and public dialogues as well as scripted and unscripted monologues, contributing a total of 300 texts with approximately 2000 words each. The aforementioned components were searched for instances of 'there', using AntConc (Anthony, 2018). A total of 11,567 tokens were extracted from the subcorpora and, as a first step, manually disambiguated based on their immediate co-text. Instances of locative 'there' or presentational 'there' were excluded from the dataset. In cases in which a speaker produced an utterance including more than one 'there', for instance due to repetitions, the final occurrence of 'there' was the one to be retained and analysed – all other instances were disregarded. After the disambiguation process, a total of 5754 ExTs were annotated for the structural, contextual and socio-biographic variables presented in Table 1.

All of these variables are believed to have an influence on the speakers' choice of the response (or the dependent) variable, namely whether speakers realize the ExT initially or non-initially. In the light of the third research question, we want to outline which factors determine the choice between indigenized and common core forms of the ExT. Given the complex nature of the annotation scheme, some of the aforementioned independent variables need further explanation. Concerning definiteness (SUBJ_DEF) and structure (SUBJ_STR) of the notional subject, it should be added that, in cases in which it was not entirely clear whether the notional subject represented a compound or a premodified noun, the *Oxford English Dictionary* (OED Online) was consulted to see whether the respective subject featured there. If it did, it was treated as a head in its entirety (for example, 'hay fever'), whereas subjects without an entry in the OED were considered cases of premodification followed by the head (such as 'pollen allergy'). Following Quirk, Greenbaum, Leech, and Svartvik (1985), a distinction was made between premodifications and different kinds of determiners which precede the head. Premodifiers commonly constitute adjectives or adverbs while determiners fall into three categories: pre-determiners (e.g., 'all', 'both', 'half', 'whole'; multipliers ['double', 'twice']; fractions ['one-third', 'three-quarters']), central determiners (definite and indefinite articles, e.g., 'the', 'a', 'an'), and post-determiners (cardinals, numerals, quantifiers, e.g., 'one', 'two', 'three', 'first', 'second', 'third', 'plenty of', 'a lot of', 'a little', 'some') (see Quirk, Greenbaum, Leech, & Svartvik, 1985, pp. 251–252). Decisions concerning the definiteness of the notional subject were based on Aarts, Chalker, and Weiner (2014), according to whom a distinction can be made between a definite and indefinite NP depending on whether the noun phrase contains an 'identifiable particular or exclusive reference' (p. 110). As previously mentioned, the ExT is more commonly associated with indefinite noun phrases which, in turn, generally constitute new information. For this reason, the present study also accounts for the discourse accessibility

Table 1. Overview of the annotated variables.

Variable	Levels	Explanation
POSITION_EX	*initially* (I) or *non-initially* (N)	Position of 'there' within the existential construction
META_ID		An identifier for each speaker in the corpus data
SUBJ_LENGTH_LOG		The log of the length of the notional subject in words
SUBJ_DA	*given* or *new*	The discourse accessibility of the subject, that is whether the head(s) of the phrase(s) functioning as a subject have been introduced in the pretext
SUBJ_COMPL	*bare* (B), *locative* (L), *complement* (C), *infinitive* (I), *participle* (P) or *relative clause* (R) (as explained above and based on Huddleston & Pullum, 2002, p. 1393; Quirk, Greenbaum, Leech, & Svartvik, 1985, pp. 1406–1408)	The complexity of the notional subject
SUBJ_HEAD_COMPL	*simple* (S) or *complex* (C)	Whether 'there' is associated with one or more than one noun phrase heads of the notional subject
SUBJ_DEF	*definite* (D) or *indefinite* (I) (cases in which both definite and indefinite entities constitute the subject were categorized as *definite*)	Definiteness of notional subject
SUBJ_STR	*head* (H) only, *premodifier* and *head* (PH), *head* and *postmodifier* (HP) or other modification structures of the head (OTHER)	The structure of the notional subject
SUBJ_ANIMACY	*animate* (AN) or *inanimate* (IN) (cases in which both animate and inanimate entities constitute the subject were categorized as *animate*)	The animacy of the notional subject
SUBJ_V_AGREE	*true* (T) or *false* (F)	Whether or not notional subject and verb within the existential construction agree in number
VERB_FORM	*complex* (C) (with the auxiliary preceding a form of *BE* or another main verb) or *simple* (S) (form of BE without an auxiliary)	The realization of the verb phrase
SENT_TYPE	*declarative* (D) or *non-declarative* (NON-D), with the latter covering *exclamative*, *imperative* and *interrogative* sentence types	The type of sentence the existential occurs in
SC_MC	*main clause* (MC) or *subordinate clause* (SC)	The type of clause the ExT is used in
SENT_NEG	*yes* (Y) or *no* (N)	Whether the sentence in which the existential occurs is negated
TENSE	*present* (PR), *past* (PA) or *future* (FU)	The tense of the sentence
VARIETY	*British* (ICE-GB), *Indian* (ICE-IND) or *Sri Lankan English* (ICE-SL)	The variety of English
GENRE_COARSE	*private* or *public dialogue*, or *scripted* or *unscripted monologue*	Broad record of genre in which the existential construction was used
GENDER	*female* or *male*	The speaker's gender
TOPIC	*court*, *government*, *humanities*, *legal*, *news*, *personal*, *politics*, *research* and *schooling*	The overall topic the speaker talks about in their contribution established through Latent Dirichlet allocation topic modelling (Grün & Hornik, 2011)
DIALOGUE	*dialogue* or *monologue*	Whether the ExT occurred in dialogic or in monologic situations

of the notional subject which also helped differentiate between ExT cases and locatives. Given that some researchers state that the so-called definiteness effect or definiteness restriction (see, for example, Birner & Ward, 1998; Milsark, 1977) is not a mandatory criterion for the use of the ExT and in order to ensure a consistent analysis, decisions concerning definiteness in the present study were based on lists (Table 2) provided by Röthlisberger (2018, pp. 66–67).

Although researchers often differentiate between discourse accessibility and hearer accessibility (Birner & Ward, 1998; Prince, 1981), in the present study the data were

Table 2. Indicators of definite and indefinite noun phrases.

Definite	Noun phrases that constitute proper nouns or pronouns or heads that are preceded by one of the following: the, this, that, those, these, her, his, its, my, our, their, your, all, both, each, either, every, most, neither, last and next
Indefinite	Common nouns in plural (without determiner) or heads that are preceded by one of the following: a, an, another, any, enough, few, fewer, half, less, little, little or no, lots of, many, so many, more, much, so much, no, no more, no such, such, none, one-third, half, one, one or more, ones, plenty of, several, some, twice and so

automatically annotated and, therefore, focus on discourse accessibility (SUBJ_DISC_ACC) only. The head or the coordinated heads of the subject as well as the maximally 500-character-long context to the left of the ExT instance under scrutiny were lemmatized with the help of the *textstem* package (Rinker, 2018) in R (R Core Team, 2023). In case (one of) the lemmatized subject head(s) was present in the lemmatized pretext, the discourse accessibility of the subject was coded as given; if not, discourse accessibility was set to new.

The concept of animacy refers to whether the notional subject constitutes a sentient being or not. Quirk, Greenbaum, Leech, and Svartvik (1985, p. 314), in the context of their discussion of gender classes, provide a useful overview of the animacy status of various entities as well, which was adopted for the present article (SUBJ_ANIMACY). Furthermore, following Zhang's (2013) approach, instances of notional subjects that would usually be considered inanimate, but stood representative of animate subjects, were counted as animate by metonymy. For example, in the utterance '[. . .] *there* are many governments which are formed <,> uh by many parties <,>' (ICE-IND S1A–005#13:1:A) the term 'government' stands representative of people forming the government. With people being animate, 'government' was annotated as animate as well.

Topic modelling (TOPIC) as implemented in the R package *topicmodels* (Grün & Hornik, 2011) does not automatically establish the optimal number of topics for the texts concerned, but allows users to specify the number of topics. To inform the number of topics to be modelled empirically, semantic topic coherence scores have been calculated for different numbers of topics and the maximum topic coherence was achieved with nine topics for the texts at hand. Subsequently, individual words from the texts analysed were – as a starting point in assigning texts to a particular topic – randomly allocated to these nine topics with the goal of iteratively re-assigning them to increasingly more fitting topics. This iterative re-assignment is based on the word's pervasiveness in different topics as well as the dominance of said topics in the texts concerned to ultimately establish the most adequate topic out of the nine available for a given text.

4. Results

A global perspective on the distribution of initial and non-initial ExTs across BrE, IndE and SLE is shown in Figure 2. Although initial positioning of 'there' is the default form of ExTs in each of the varieties covered, it is apparent that – in comparative terms – IndE features by far the largest share of non-initial ExTs (14.27%) as illustrated in Example 10. Example 11 exemplifies an instance of non-initial forms in SLE, which occur notably less frequently in SLE (1.68%) than in IndE. Non-initial forms are – with the exception of the instances in Examples 12

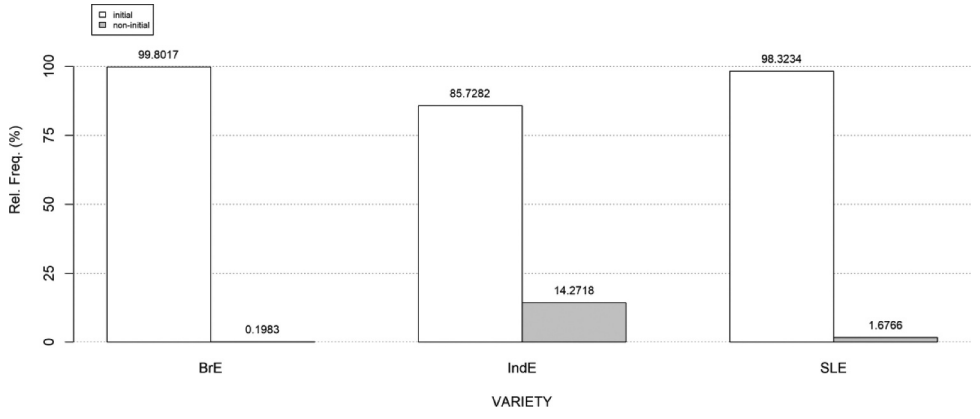

Figure 2. Relative frequencies of initial and non-initial existentials in BrE, IndE and SLE.

and 13 – absent from BrE (0.2%). In the light of these findings, IndE appears to feature non-initial ExT variants notably more frequently than its neighbour variety SLE, whose usage rather approximates and only lies slightly above the low level of non-initial ExTs evident from the BrE data. Although a feature attested across IndE and SLE, and thus a characteristic of South Asian Englishes, the non-canonical non-initial ExT appears significantly more frequently in IndE than in SLE ($X^2 = 183.91$, $df = 1$, $p < 0.001$).

(10) <ICE-IND:S1A-046#69:1:B> <{> <[> Egoism is there at a very high rate </[> <$A>

(11) <ICE-SL:S2A-027#59:1:B>And it is very important that if you are gonna get things right from education that inclusiveness is there

(12) <ICE-GB:S1B-024#134:1:F> Uhm whips are there to put pressure on people

(13) <ICE-GB:S1B-048#89:1:B> It was <,> the the compulsion to write stories was always there too

In order to gain an understanding in which contexts ExTs – independent of whether they are realized in their initial or non-initial form – are used more or less frequently in the varieties covered, a conditional inference tree has been constructed on the basis of the following model formula: EX_PER1000 ~ GENRE_COARSE + VARIETY + GENDER + TOPIC + DIALOGUE. In terms of model quality, the mean standard error (mse) of the conditional inference tree (25.874) is statistically significantly better ($p < 0.01$) than the mse of a baseline model always predicting the mean of the dependent variable (27.9862). Under consideration that EX_PER1000 represents the normalized frequency of ExTs per speaker in the corpora analysed, a conditional inference tree model is a suitable statistical choice since each speaker only contributes one data point to the data set and random effects for speaker-internal variation thus cannot be explicitly modelled.

With a view to the independent variables passed to the conditional inference tree algorithm, each of the independent variables has a strong enough effect on the frequencies of ExTs to be represented as significant in the tree (see Figure 3). Most discriminatory power is assigned to GENRE_COARSE at the top of the tree, which separates private dialogic and unscripted monologic communicative contexts, that is comparatively informal ones, from public

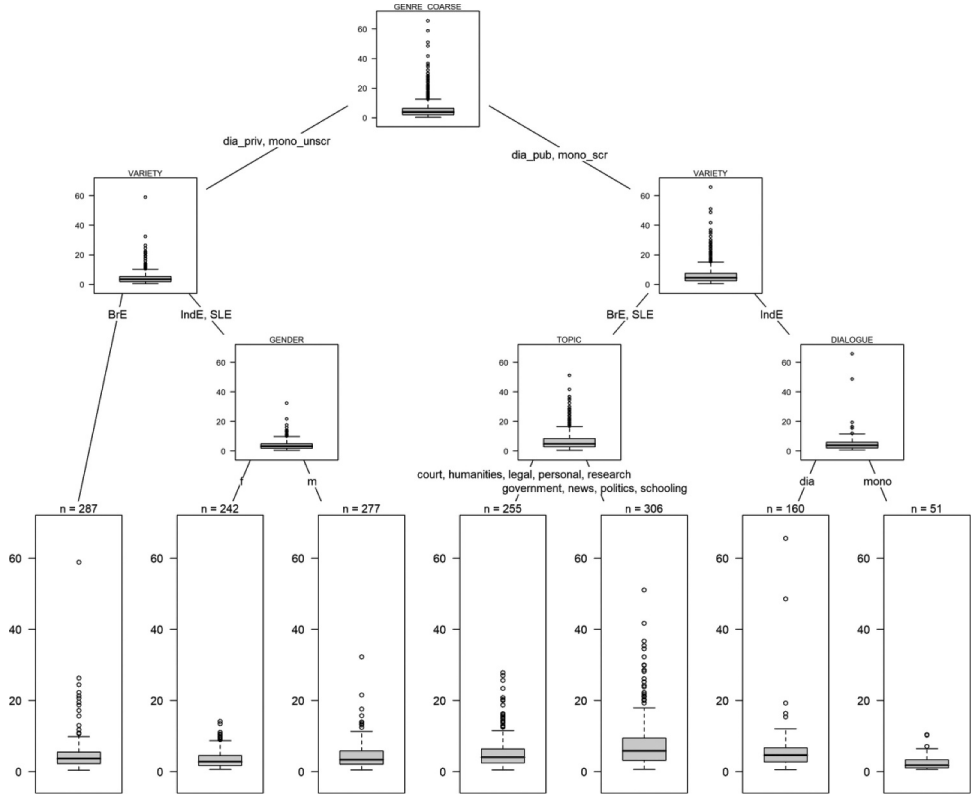

Figure 3. Conditional inference tree of normalized frequencies of existentials.

dialogues and scripted monologues, that is relatively formal settings. The more formal settings depicted in the right branch of the tree feature – with the exception of monologues in IndE which yielded the lowest frequencies of ExTs compared to all other terminal nodes – generally higher frequencies of ExTs than the more informal contexts in the left branch of the tree. This observation might not come unexpectedly since more informal contexts are often characterized by an immediately shared context as well as topics pertaining to this immediately shared context to which interlocutors can refer deictically, reducing the need to allude to abstract concepts with the help of ExTs. With the branch to the left depicting informal conversational set-ups, BrE speakers are shown to use more ExTs in these relaxed contexts than Indian or Sri Lankan speakers, with Indian and Sri Lankan men using slightly more ExTs than Indian and Sri Lankan women. In the formal contexts on the right of the tree, VARIETY separates BrE and SLE from IndE. On average, BrE and SLE speakers use ExTs more often in these formal settings – particularly when topics like government, news, politics or schooling are discussed, yielding the highest frequencies across all terminal nodes in the tree. In formal contexts in India, dialogues feature notably more ExTs than monologues. Across formal and informal settings, ExTs thus appear slightly more often per speaker in BrE than in IndE or SLE, and Indian men appear more inclined towards ExTs in formal settings than Indian women.

Still, it needs to be mentioned critically that the dependent variable EX_PER1000 collates individual examples of ExTs used by a given speaker in normalized frequencies per 1000 words. For this reason, only the contextual predictors, which are stable across the

individual ExT examples produced by a particular speaker, are used to predict the frequencies of ExTs. This approach undoubtedly causes the loss of important structural information that would have to be integrated in subsequent analyses where non-initials figure prominently, not just in one variety but a number of varieties.

With a view to the Indian data, where non-initial realizations are chosen in approximately one out of seven ExTs, a more fine-grained analysis of the choice between initial and non-initial ExTs is called for and promises novel insights into where Indian speakers of English tend to choose the indigenized non-initial variant as opposed to the internationally more widely used initial form. In contrast to the data modelled in the conditional inference tree shown in Figure 3, which is based on one summarized data point of normalized ExT frequency per speaker and consequently on one speaker contributing only one data point, an investigation of initial versus non-initial ExTs needs to be based on observations whether IndE speakers choose one or the other variant in a given communicative context. These observations should not be summarized per speaker as some IndE speakers regularly contribute more than one example of individual speaker choices, thus constituting a random effect in the data, which conditional inference trees cannot account for. Against this background, we opted for another tree-based approach that can explicitly separate random from fixed effects in the modelling process, namely generalized linear mixed-effects model trees (glmertrees; Fokkema, Smits, Zeileis, Hothorn, & Kelderman, 2018). Just like conditional inference trees, glmertrees iteratively divide observations into two groups with the aim of making the resulting groups similar in terms of the observed outcomes – in this case, the choice of initial or non-initial ExTs – while simultaneously considering random effects in the data. The glmertree algorithm consequently allows using the entire suite of contextual, structural and socio-biographic independent variables annotated along with a random effect for speaker, which, in the subsequent model formula – with the choice of initial and non-initial ExTs as the dependent variable – is represented via META_ID: POSITION_EX ~1 |META_ID | SUBJ_HEAD_COMPL + SUBJ_LENGTH_LOG + SUBJ_DEF + SUBJ_STR + VERB_FORM + SENT_TYPE + SC_MC + SENT_NEG + TENSE + SUBJ_V_AGREE + SUBJ_ANIMACY + GENRE_COARSE + GENDER + SUBJ_DA + TOPIC.

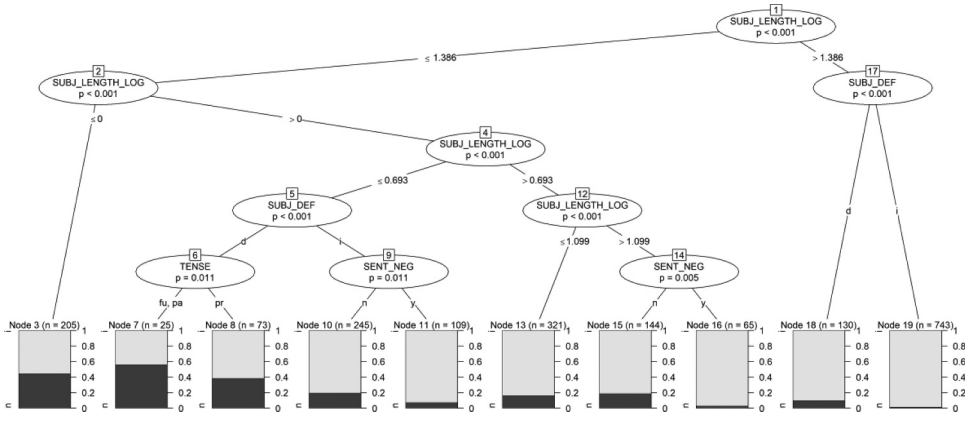

Figure 4. Glmertree of initial versus non-initial existentials in IndE.

The classification accuracy of the glmertree shown in Figure 4 is 90.34%, which is statistically significantly ($p < 0.001$) better than a baseline model always predicting the more frequent surface structure, namely initial ExTs, with a classification accuracy of 85.73%.

With a view to the total of 15 independent variables for which we annotated each example, only SUBJ_DEF, SUBJ_LENGTH_LOG, SENT_NEG and TENSE, each of which is structural in nature, are profiled as significant while contextual (for example, GENRE_COARSE, TOPIC) or socio-biographic predictors (GENDER) have not been included in the model. Among the significant structural independent variables, SUBJ_LENGTH_LOG is most important for ExT ordering choices in that it constitutes the first split in the tree and occurs repeatedly in later splits in the tree. The first split separates ExTs with subjects equal to four words in length or shorter – the natural log[4] of 4 is 1.386 and these logs are used and shown in the glmertree, but we will generally report antilogs (the reverse operation of logarithms) for more accessibility – shown in the left branch of the tree from ExTs with subjects of five words or more represented in the right branch of the tree. In this light, initial ExTs following the order 'there'+ BE + subject_{n} become more likely as subjects grow more complex (for example, nodes 18 and 19, where initial constructions dominate), whereas non-initials composed of $\text{subject}_{n,g}$ + BE + 'there' figure most prominently in comparative terms when subjects are one word (node 3) or two words (nodes 7 and 8) long, as exemplified in Example 14. In fact, the context in which non-initial ExTs occur most often is characterized by definite two-word-long subjects in clauses with future or past reference (node 7), an example of which is provided in Example 15. With indefinite two-word-long subjects (nodes 10 and 11), a notable decrease in non-initial constructions can be observed – with the decrease being more pronounced in negated contexts (node 11). The tendency of negated contexts to feature slightly more non-initial constructions than affirmative ones is also evident from a comparison of nodes 15 and 16.

(14) <ICE-IND:S1A-089#142:1:B> Competition is there but but this competition is not healthy <,> often it is manipulated competition <,> <$A>

(15) <ICE-IND:S2A-046#50:1:A> They wanted to become Australians or Canadians and New Zealanders <,>
<ICE-IND:S2A-046#51:1:A> So that problem <,> was there

5. Discussion

This study aimed to investigate the use of the initial and non-initial ExTs in IndE, SLE and BrE. Our first research question, enquiring whether quantitative differences between the three varieties can be attested, must clearly be answered affirmatively, as we were indeed able to uncover highly significant differences in the proportional uses of the two variants of the construction across the three varieties. While all three varieties clearly prefer using the initial ExT, BrE does so almost exclusively, SLE makes use of the non-initial variant very rarely (in less than 2% of the cases), whereas IndE uses this variant in almost 15% of the cases. While this seems to be confirmation that the construction can be considered institutionalized/nativized in IndE (see also Lange, 2012), our initial hypothesis – according to which this innovative construction might have already spread out from India to its

neighbouring varieties, for example through epicentral influences (see Section 2) – cannot be confirmed for the case of Sri Lanka. Although we were able to attest the feature in SLE, the frequencies are still too low to make any general assumptions. However, it occurs with higher frequencies in SLE than in BrE, which may allude to the fact that the feature has only just been introduced into SLE and might establish itself within the next few decades. It would therefore make for an interesting case to keep observing in the future in order to be in a position to trace its development diachronically.

Turning to our second research question, due to the low frequencies of the ExT in SLE and BrE, we were not able to perform a cross-varietal comparison to investigate (a) which predictors guide structural choices of ExTs in both South Asian Englishes as well as BrE and (b) to what extent the effects of said predictors are cross-varietally distinct. Instead, we zoomed in on IndE to uncover the factors that may influence the use of an initial versus non-initial ExT. Here, our model revealed that structural features constitute the most important factors in predicting the choice between the two variants. More specifically, subject length (SUBJ_LENGTH_LOG), the subject definiteness (SUBJ_DEF), the tense of the verb (TENSE) and whether or not the verb is negated (SENT_NEG) are the only significant predictors in our model, whereas socio-biographic speaker variables (like GENDER, as reported in Lange, 2012, for instance) do not figure at all in our model. This observation highlights the fact that, while monofactorial studies are indispensable for hypothesis building and for uncovering hugely important connections between variables, it seems to be absolutely worthwhile to conduct follow-up studies that are multifactorially controlled for by structural features. This study shows that the use of non-initial ExTs is inextricably intertwined with (information) structural concerns and concepts. More specifically, the length and definiteness of the subject play a great role in predicting the choice between the use of an initial or a non-initial ExT. Non-initials are notably more likely to occur with shorter than with structurally more complex subjects, probably instantiating effects from syntactic-cognitive ordering mechanisms like the principle of end-weight (Behaghel, 1909) or the EasyFirst principle (MacDonald, 2013), stating that less complex elements with lower cognitive loads tend to precede more complex ones with higher cognitive processing demands. This might be an indication that the non-initial ExT has become truly nativized in IndE, because its use can be predicted on the basis of information-structural and cognitive principles with an extremely high accuracy across speakers and genres.

Our third research question, namely whether the structural evidence of the ExT is compatible with the more advanced evolutionary status of IndE compared to SLE, cannot be clearly answered by our analysis. While we seem to be in a position to assign the status of grammatical nativization to the non-initial ExT in IndE (which might, in turn, be considered to be an indicator for having entered the endonormative stabilization phase), this cannot be inferred to the same extent for SLE. While the low number of occurrences could well be explained by SLE being less advanced than IndE in terms of its alleged evolutionary status (see Section 2), first language (L1) influence of this feature in SLE seems unlikely because there are no directly equivalent structures in Sinhala, the L1 predominantly spoken in Sri Lanka. We might well have seen that the feature has only just entered SLE, potentially due to epicentral influences or simply

contact with the IndE neighbour, for example through travelling or (social) media. One hypothesis we could build from this observation is that once SLE has emerged further on the evolutionary scale, the feature might also start figuring more frequently in SLE. However, definite assessments would require future investigations in different corpora as well as a diachronic reassessment of the evolution of this feature in 20–30 years.

6. Conclusion

The present study has conducted one of the first multivariate analyses of the ExT in IndE and SLE as compared to BrE. Our approach proved to be extremely beneficial to attest significant varietal differences between the two South Asian varieties and the control corpus. The non-initial ExT is virtually non-attested in BrE. Building on previous valuable work that has uncovered the very existence of the Indian ExT (see Section 2), our new and multivariate take on the construction allowed for a fine-grained and holistic description of the linguistic profile of the pattern in IndE. In doing so, we were able to present a very detailed account of one clear instance of grammatical nativization in IndE, whose use is mainly governed by structural, linguistic variables that are very much in line with cognitive principles and are valid across speakers and registers. It thus seems to be a distinctively developed feature of IndE and we are encouraged to investigate other grammatical features of IndE with the same level of granularity, as this approach seems to be valuable towards developing a grammar of IndE.

Even though we were not able to present the same level of detail in describing the construction in SLE due to data sparsity, we still managed to attest the use of the non-initial ExT in the variety. Future analyses on larger corpora and a few decades down the line might reveal a similarly thorough account of the construction. Regarding the present data set, SLE seems less established than IndE regarding the ExT, which stands in line with the evolutionary more advanced status of IndE. Only time will tell whether or not SLE will be emancipating further from BrE in the use of the construction and whether or not it will gain similar popularity in SLE as it has in IndE. Observing the evolution of this feature in SLE might also contribute to explaining general nativization processes of varieties of English. According to our knowledge, Sinhala, as the dominant L1 in Sri Lanka, has no direct grammatical equivalent to the structure in that it does not employ dummy pronouns. Instead, the language makes use of existential predicates (namely we- for literary contexts and in- for animate referents as well as tiye- for inanimate referents in colloquial contexts; see Paolillo, 2000, pp. 221–222), and it will be particularly interesting to see whether the ExT still has an equal chance of becoming nativized as it has in IndE, where equivalent structures do exist in several L1s spoken in India.

The present study cannot make very broad claims because it was only concerned with the ExT in the spoken mode. In order to be in a position to make a more global and reliable assessment of the status of grammatical/structural nativization in both IndE and SLE, future investigations of different non-canonical features in different modes and genres need to follow.

Notes

1. For a sociohistorical description of SLE, see de Silva (1981); for an overview on the linguistic situation in Sri Lanka, see Coperahewa (2009).
2. Note that while the focus in this project is on the ExT, we are aware of the distinction between the existential and the presentational constructions, in the case of which a verb other than be is employed, for example There *followed* a long silence.
3. In the examples, the notional subject is underlined with a solid line whereas optional adverbials (as in Example 3) and extensions (see Examples 4–8) are indicated using broken lines.
4. The natural logarithm of a numeric value is the power to which the mathematical constant *e* has to be raised to be the same as that numeric value.

Acknowledgements

The authors would like to thank the Forschungscampus Mittelhessen for funding this project (project number: 2021_1_1_3). Furthermore, they thank Alina Schmischke, Caitlin Grigoleit and Ferhat Yasar for their support with regard to data annotation.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was supported by the Forschungscampus Mittelhessen [2021_1_1_3].

ORCID

Kathrin Kircili http://orcid.org/0009-0008-0543-6889
Julia Degenhardt http://orcid.org/0000-0003-0060-1802
Tobias Bernaisch http://orcid.org/0000-0003-0122-1975
Sandra Götz http://orcid.org/0000-0003-1474-5612

References

Aarts, B., Chalker, S., & Weiner, E. (2014). *The oxford dictionary of English grammar* (2nd ed.). Oxford: Oxford University Press.

Anthony, L. (2018). *AntConc* (version 3.5.0) [computer software]. Tokyo, Japan: Waseda University. Retrieved from http://www.laurenceanthony.net/software/antconc/

Arnold, J. E., Losongco, T., Wasow, A., & Ginstrom, R. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, *76*(1), 28–55. doi:10.1353/lan.2000.0045

Balasubramanian, C. (2009). *Register variation in Indian English*. Amsterdam: John Benjamins.

Behaghel, O. (1909). Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen*, *25*, 110–142.

Bernaisch, T. (2015). *The lexis and Lexicogrammar of Sri Lankan English [varieties of English around the world G54]*. Amsterdam: John Benjamins.

Bernaisch, T., Gries, S. T., & Mukherjee, J. (2014). The dative alternation in South Asian englishes: Modelling predictors and predicting prototypes. *English World-Wide*, *35*(1), 7–31. doi:10.1075/eww.35.1.02ber

Bernaisch, T., Koch, C., Mukherjee, J., & Schilk, M. (2011). *Manual for the South Asian varieties of English (SAVE) corpus: Compilation, cleanup process, and details on the individual components*. Giessen: Justus Liebig University, Department of English.

Bernaisch, T., & Lange, C. (2012). The typology of focus marking in South Asian englishes. *Indian Linguistics*, *73*, 1–18.

Biber, D., Johannson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Pearson Education.

Birner, B., & Ward, G. (1998). *Information status and noncanonical word order in English*. Amsterdam: John Benjamins.

Bolton, K., & Bacon-Shone, J. (2020). The statistics of English across Asia. In K. Bolton, W. Botha, & A. Kirkpatrick (Eds.), *The handbook of Asian englishes* (pp. 49–80). Hoboken: Wiley Blackwell.

Breivik, L. E. (1981). On the interpretation of existential *there*. *Language*, *57*(1), 1–25. doi:10.1353/lan.1981.0048

Clark, H. H., & Haviland, S. E. (1977). Comprehension and the Given-New contract. In R. O. Freedle (Ed.), *Discourse production and comprehension* (pp. 1–40). Norwood: Ablex Publishing Corporation.

Collins, P. C. (2012). Singular agreement in *there*-existentials: An intervarietal corpus-based study. *English World-Wide*, *33*(1), 53–68. doi:10.1075/eww.33.1.03col

Coperahewa, S. (2009). The language planning situation in Sri Lanka. *Current Issues in Language Planning*, *10*(1), 69–150. doi:10.1080/14664200902894660

Crawford, W. J. (2005). Verb agreement and disagreement: A corpus investigation of concord variation in existential there + *be* constructions. *Journal of English Linguistics*, *33*(1), 35–61. doi:10.1177/0075424204274001

Degenhardt, J., & Bernaisch, T. (2022). Apologies in South Asian varieties of English: A corpus-based study on Indian and Sri Lankan English. *Corpus Pragmatics*, *6*, 201–223. doi:10.1007/s41701-022-00117-8

de Silva, K. M. (1981). *A history of Sri Lanka*. Berkeley: University of California Press.

Fokkema, M., Smits, N., Zeileis, A., Hothorn, T., & Kelderman, H. (2018). Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects Model trees. *Behavior Research Methods*, *50*(5), 2016–2034. doi:10.3758/s13428-017-0971-x

Givón, T. (1985). Iconicity, isomorphism, and non-arbitrary coding in syntax. In J. Haiman (Ed.), *Iconicity in syntax* (pp. 187–220). Amsterdam: John Benjamins.

Givón, T. (2001). *Syntax: An introduction*. Amsterdam: John Benjamins.

Götz, S. (2017). Non-canonical syntax in south-Asian varieties of English: A corpus-based pilot study on fronting. *ZAA - Zeitschrift für Anglistik und Amerikanistik*, *65*(3), 265–281. doi:10.1515/zaa-2017-0028

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

Grün, B., & Hornik, K. (2011). Topicmodels: An R package for fitting topic models. *Journal of Statistical Software*, *40*(13), 1–30. doi:10.18637/jss.v040.i13

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

Huddleston, R., & Pullum, G. K. (2002). *The cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Jakobson, R. (1962). Über die phonologischen Sprachbünde. (Travaux du Cercle Linguistique de Praque 4, 234–240). Reprinted in *Selected writings* (Vol. I, pp. 137–143). The Hague: Mouton.

Kachru, B. (1992). *The other tongue*. Urbana: University of Illinois Press.

Körtvelyessy, M., Bernaisch, T., Mukherjee, J., & Mendis, D. (2012). *Manual to the written component of the International corpus of English – Sri Lanka*. Giessen: Justus Liebig University, Department of English.

Kraaz, M., & Bernaisch, T. (2020). Backchannels and the pragmatics of South Asian englishes. *World Englishes*, *41*(2), 224–243. doi:10.1111/weng.12522
Lambrecht, K. (1994). *Information structure and sentence form: Topic, focus and the mental representations of discourse referents*. Cambridge: Cambridge University Press.
Lange, C. (2012). *The syntax of spoken Indian English*. Amsterdam: John Benjamins.
Leitner, G. (1992). English as a pluricentric language. In M. G. Clyne (Ed.), *Pluricentric languages: Differing norms in different nations* (pp. 179–237). Berlin: Mouton de Gruyter.
MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*, 1–16. doi:10.3389/fpsyg.2013.00226
Martínez Insua, A. E. (1998). Existential-there construction in present day English: A corpus-based study. *Atlantis*, *20*(2), 107–131.
Martínez Insua, A. E., & Palacios Martínez, I. M. (2003). A corpus-based approach to non-concord in present day English existential *there*-constructions. *English Studies*, *84*(3), 262–283. doi:10.1076/enst.84.3.262.16852
Meyler, M. (2007). *A dictionary of Sri Lankan English*. Colombo: Mirisgala.
Milsark, G. L. (1977). Toward an explanation of certain peculiarities of the existential construction in English. *Linguistic Analysis*, *3*(1), 1–29.
Mukherjee, J. (2007). Steady states in the evolution of new englishes: Present-day Indian English as an equilibrium. *Journal of English Linguistics*, *35*(2), 157–187. doi:10.1177/0075424207301888
Paolillo, J. C. (2000). Formalizing formality: An analysis of register variation in sinhala. *Journal of Linguistics*, *36*(2), 215–259. doi:10.1017/S0022226700008148
Prado Alonso, J. C. (2008). *Inversion in written and spoken contemporary English* (Doctoral dissertation). Universidade de Santiago de Compostela.
Prince, E. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical pragmatics* (pp. 223–255). New York: Academic Press.
Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. Harlow: Longman.
R Core Team. (2023). *R: A language and environment for statistical computing*. Vienna, Austria. URL: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/
Rinker, T. W. (2018). *Textstem: Tools for stemming and lemmatizing text*. Version 0.1.4. Available at http://github.com/trinker/textstem
Röthlisberger, M. (2018). *Regional variation in probabilistic grammars: A multifactorial study of the English dative alternation* (Doctoral dissertation). KU Leuven.
Schneider, E. W. (2007). *Postcolonial English: Varieties around the world*. Cambridge: Cambridge University Press.
Thomason, S. T. (2001). *Language contact: An introduction*. Edinburgh: Edinburgh University Press.
Ward, G., & Birner, B. (2004). Information structure and non-canonical syntax. In L. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 153–174). Malden: Blackwell.
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.), *Corpus linguistics: Looking back – moving forward* (pp. 249–266). Amsterdam: Rodopi.
Winkle, C. (2015). *Non-canonical structures, they use them differently: Information packaging in spoken varieties of English* (Doctoral dissertation). Albert–Ludwigs-Universität Freiburg.
Witzel, M. (1995). Early Indian history: Linguistic and textual parametres. In G. Erdosy (Ed.), *The Indian-aryans of ancient south asia: Language, material culture and ethnicity* (pp. 85–125). Berlin and New York: Walter de Gruyter.
Zhang, J. (2013). Animacy hierarchy effects on the second language acquisition of attributive psych adjectives. *Applied Psycholinguistics*, *36*(2), 275–298. doi:10.1017/S0142716413000234