

The subjunctive alternation in Indian English

Karola Schmidt 

Department of English, Justus Liebig
University Giessen, Giessen, Germany

Correspondence

Karola Schmidt, Justus Liebig University
Giessen, Otto-Behaghel-Str. 10B, 35394
Giessen, Germany.

Email:

karola.schmidt-1@anglistik.uni-giessen.de

Funding information

Deutsche Forschungsgemeinschaft,
Grant/Award Numbers: DFG BE 5812/3-1, GR
5308/1-1

Abstract

The study at hand is an exploration of the alternation between the mandative subjunctive and its equivalent modal construction with the verb *should* in Indian English. The study complements the growing body of research on the morphosyntax of the variety and it enhances our understanding of the relatively under-researched alternation. It adds a special focus on the short-term diachronic development of the subjunctive alternation in Indian English by using data from the *South Asian Varieties of English* corpus and its 2020 update. The following research questions are being investigated in the paper: Which factors influence the alternation between subjunctive and *should* in Indian English; are there short-term diachronic adjustments to the subjunctive paradigm? A multifactorial model was fitted on 508 extracted data points with the following predictors: LEXICAL DIVERSITY, NEWSPAPER, WORD COUNT, DISTANCE, GENDER, LINKING WORD, NEGATION, SUBJECT NUMBER, SUBJECT PERSON, TIME and TRIGGER LEMMA. The results show minor diachronic adjustments and highlight well-known predictors of the alternation like VOICE as important factors.

1 | INTRODUCTION

The paper at hand is an exploration of the subjunctive alternation in Indian English (IndE), that is the different realisations of the subjunctive paradigm with, either, the base form of the verb or via the addition of a modal verb as exemplified by the following sentence from an Indian newspaper. Following a suasive verb, the author foregoes the

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2024 The Authors. *World Englishes* published by John Wiley & Sons Ltd.

inflectional -s that signifies agreement with the third-person singular subject in the subordinate clause in favour of the base form of the verb, indicating a subjunctive:

- (1) It also asked that Corbyn **[should] meet** the High Commissioners of both India and Pakistan. (SAVE2020 India: 4338991; my emphasis and insertion)

The paper is, therefore, clearly situated in two well-established domains of the study of English. Within world Englishes as a research area, the English varieties in South Asia have continuously attracted interest over the course of the last decades. IndE in particular is certainly the best-known and most researched of the South Asian varieties. Starting with work by scholars like Braj Kachru (e.g. Kachru, 1965), IndE has paved the way for other Outer Circle varieties to receive the academic attention they deserve. Similarly, (morpho-)syntactic alternations in the Labovian sense like in the subjunctive paradigm are popular features in recent linguistic research.

Particularly when it comes to world Englishes, analysing alternations seems to be a fruitful endeavour given their saliency and embeddedness in all varieties of English (see, e.g., Bernaisch et al., 2014; Tamaredo et al., 2020; compare also Bernaisch et al., 2022 for their discussion of this). Change in and adjustments to the choices influencing the realisation of alternating forms can provide valuable insights into processes of nativisation and stabilisation driving Outer Circle varieties.

In this vein, the following research attempts to fill two gaps. On the one hand, it looks to add to our knowledge of the structure of IndE aided by a short-term diachronic focus when examining the subjunctive so that possible variety-internal adjustments to the paradigm over time may be uncovered and discussed. To this end, the research draws on the *South Asian Varieties of English* corpus (Bernaisch et al., 2011) and its 2020 update (Bernaisch et al., 2021), which has been used for a number of similar projects (e.g. Bernaisch et al., 2014 on the dative alternation). On the other hand, although a sizable body of work tracing the history of the English subjunctive exists, multifactorial statistical treatment of the subjunctive-alternation remains relatively rare (Subsection 2.2 will discuss the notable exceptions in more detail). It seems prudent to combine this, in terms of newer techniques of statistical modelling, relatively under-researched alternation with an often-discussed variety to introduce novel alternation-based perspectives to well-explored IndE and add to the growing body of research on the subjunctive.

To this end, the paper will first introduce the variety in question and then the main theoretical avenue, namely, IndE in Subsection 2.1, followed by the subjunctive mood and its different forms in Subsection 2.2. Subsection 2.2 also includes a brief note on the theoretical controversy surrounding the status of the subjunctive paradigm in the English language, namely, whether it can and should be considered a grammatical mood or not. Research questions are formulated at the end of the theoretical section. Subsection 3.1 introduces the (sub-)corpora used before subsequently detailing how the data were extracted and prepared for statistical analysis, which is the subject of Subsection 3.2. Following this, Section 4 illustrates the results which are then discussed in Section 5. Finally, Section 6 concludes the paper with a summary and an outlook with suggestions for future research avenues.

2 | THEORY

2.1 | Indian English

IndE is a well-researched Outer Circle variety of English, which is likely due to the number of speakers of the variety and socioeconomic importance of the country itself. However, given the sheer size of India and the accompanying wealth of different speaker backgrounds, determining how many Indians speak English and with what degree of fluency has, so far, posed somewhat of a difficult task. Sharma (2012) and Sridhar (2020) point out that the Indian census is not reliable here because there is no way to operationalise the concept of fluency employed in the census surveys in a way that would allow comparable linguistic judgements by the Indian census informants. In particular, Sridhar

(2020) discusses the importance of literacy over actively speaking English in India. He argues that Indians are more likely to hear and read English than they are to speak it and that the degree to which they speak it may vary considerably between individual speakers due to their specific language needs (p. 244). By extension, census-based speaker counts could be too conservative if the question is whether people *speak* English and even if the question is worded towards *using* English, ostensibly similar speakers may provide different self-assessments. Consequently, Bolton and Bacon-Shone's (2020) estimate of 280 million speakers of English in India in 2020, which came to roughly 20% of the population at that point in time (p. 53), should be taken with a grain of salt, as they themselves point out (pp. 65–69). These issues are subject for other studies, however, and I iterate them and the ensuing speaker count only to underline the relative importance of IndE not just in terms of Outer Circle varieties but with regard to the English language as a whole. Even allowing discrepancies between reported and actual numbers, it is likely that IndE is both the largest Outer Circle variety of English in the world and that it may even surpass British English (BrE) in terms of first language speakers.

I will briefly go over the history of English in India in order to better situate the research object and research gap later on in the paper. The English language began to spread in India with the arrival of the East India Company in the early 17th century (Gargesh, 2020, p. 109) after Elizabeth I granted the corresponding charter on 31 December 1600 (Schneider, 2007, p. 162). Although the activities of the East India Company were initially restricted to a few trading factories, namely, Surat, Madras, Bombay and Calcutta (Gargesh, 2020, p. 109), English gradually gained in importance due to its standing in education. The much-cited *Minute on Indian Education* as the culmination of the debate between Anglicists and Orientalists was passed on 2 February 1835. It decreed English the official language of education in India (Gargesh & Sailaja, 2017, p. 426). This went against the (initial) policies of the East India Company, which was in favour of the Orientalist approach (Evans, 2002, p. 262), but was eventually upended by the British Crown that took away its monopoly in 1813 before issuing the *Act for the Better Government of India* in 1858, transferring all power still held by the Company to the Crown (Hickey, 2009, pp. 538–539). Following this, colonial administration in British India, including modern Pakistan and Bangladesh up until Partition, was restructured. In the course of this restructuring, new government offices opened up across the country, which was another decisive factor increasing contact between Indians and the English language (Sedlatschek, 2009, p. 15).

English continued to spread like this for almost a century until India liberated itself from its colonial oppressors in 1947 with the *Indian Independence Act*. The role of English in this new nation was contested, considering that English had been considered the language of economic opportunity since decades before independence, but at the same time, this language-political development naturally did not change the colonialist origins of the language in India. Effectively, English was the language of the oppressors. This complicated limbo is duly reflected in the implementation of the *Three Language Formula*. Bhatia (2022) summarises the situation as follows:

after Indian independence, the Indian constitution had also made room for English, the colonial exocentric (foreign) language, as a tradition language to be replaced by Hindi as the sole national language within 15 years. However, recognising that English had gained more and more appropriateness, acceptance, power and vitality in India, the Three Language Formula required the teaching of English and Hindi nationwide along with the regional language of the area. (pp. 65–66)

Originally modelled after the language policies in the Soviet Union in an attempt to accommodate the rich linguistic diversity of India, the Three Language Formula quickly became an easy political target and festered resentment, with non-Hindi speakers feeling discriminated by the legal prominence given to Hindi. Consequently, both Hindi and English as highly privileged languages linked to economic well-being were and still are being contested (Bhatia, 2022, pp. 65–66). The, for now, final piece to this complicated linguistic and political puzzle then is the (amended) Official Languages Act from 1967, which 'finally assigned English the permanent status of associate official language of the Indian Union that was to be used for the proceedings in the Supreme Court [...], in the High Courts [...] and in Parliament' (Sedlatschek, 2009, p. 19).

Today, English serves as a link language in India. As Lange (2017) points out, there occur 'occasional reflexes of a colonial cringe with respect to English, but the importance of the language for the Indian communicative space remains unchallenged' (p. 27). There are 22 regional languages in India that are 'scheduled', meaning officially recognised and part of the Three Language Formula. Beyond that, there are numerous more languages and regional dialects also in use in the country (Bhatia, 2022, p. 63). Therefore, English is a logical choice as language of communication for people with different first languages as, despite its contested character, it is more neutral than Hindi in that regard (Lange, 2017, p. 21). This echoes Schilk's (2011) prediction that as English fills this specific and frequent communicative need in India, it 'may well retain and consolidate its status' (p. 7). In fact, Bhatia (2022) even argues that the Indian government is actively violating the Three Language Formula by not only promoting English but restricting Hindi and regional languages at the same time (p. 70). It seems that one can no longer separate India from the English language and vice versa. English is the de-facto language of communication in institutes of higher education across the country in almost all subjects (Sridhar, 2020, p. 251), increasing the demand for English-medium instruction at primary and secondary school levels. Education in English is (considered) synonymous with higher status and better economic opportunities (Gargesh & Sailaja, 2017, p. 427). Proficiency in English is a marker of the social divide in India: 'English-medium education in India effectively acts as a caste marker, a given for the affluent upper middle class but largely out of reach for the majority of the population, especially in rural areas' (Lange, 2017, pp. 19–20).

The position of IndE in Schneider's Dynamic Model of the Evolution of Postcolonial Englishes (Schneider, 2003, 2007) is well-documented and generally agreed upon. Mukherjee (2007) details the trajectory of IndE through the different phases proposed by Schneider before concluding that 'the process of nativization [...] is more or less over; the variety is now largely endonormatively stabilized' (p. 170), which has since been echoed numerous times (e.g. Gries & Bernaisch, 2016; Schilk, 2011). To arrive at this conclusion, IndE has been studied extensively on all levels of language, including phonology (e.g. Chand, 2010), lexis (e.g. Balasubramanian, 2016), lexicogrammar (e.g. Sedlatschek, 2009), syntax (e.g. Lange, 2012) and pragmatics (e.g. Degenhardt, 2021). For a paper on the subjunctive, syntactic studies are of most immediate relevance. Previous research focused extensively on, first, the linguistic butterflies that characterise IndE (e.g. Bernaisch & Lange, 2012 on focus marking; Koch et al., 2016 on the 'intrusive-*as*' construction), and, second, underlying syntactic structures and knowledge, embodied, in particular, in linguistic alternations like the one under the microscope in this paper (e.g. de Cuypere & Verbeke, 2013 on the dative alternation). I will turn to those studies that have covered the subjunctive in IndE in the next section after a theoretical introduction of the subjunctive in the first place.

2.2 | The mandative subjunctive

Present-Day English has lost almost all of its inflectional system, but the mandative subjunctive is one of the syntactic paradigms that retains both an inflectional and a periphrastic realisation of the same form, not unlike the possessive. With a view to its history, Traugott (1992) explains that the subjunctive had a wide range of uses in Old English, for example, to indicate reported speech (p. 240). This widely used subjunctive came, as was the general standard for Old English, with a rich set of inflections that was lost as English evolved over the next centuries. Additionally, modal verbs and the indicative developed into meaning-equivalent alternatives over the same time (e.g. Moessner, 2007). By the 18th century, the mandative subjunctive had become a low-frequency variant in both AmE and BrE (Hundt, 2009). Coupled with this, the form was associated with high-formality contexts and thus took on the role of a politeness marker, at least in British society (Auer & González-Díaz, 2005). By the 20th century, the subjunctive was so uncommon that it was described largely negatively in the prescriptive grammars of the time. On the other hand, however, Aarts (2012) speaks of rising frequencies of the subjunctive in the 20th century (p. 3). This development seems to have held up in so far as that a century later, English retains at least the option of using the (mandative) subjunctive.

The envelope of variation for the mandative subjunctive has not reached similar levels of academic consensus as more extensively studied alternations.¹ The minimal consensus is that we speak of the mandative subjunctive

when discussing the verb form in subordinate clauses that were introduced by deontic / suasive verbs, that is verbs expressing a request or command. A minimal requirement that follows from this is the presence of an overt subject as is generally needed in a subordinate clause. In example (2), both the deontic verb and the following verb in the subordinate clause are highlighted in bold, the subject is italicised:

(2) He **demanded** that *the government* **withdraw** the turnover tax. (SAVE India: IN_SM_2005-01-06_c)

In this realisation of the subjunctive, the verb remains in the base form (Hoffmann, 1997, p. 6). That includes that there is 'no concord with the subject, no backshifting of tense depending upon the superordinate verb [...], and no do-periphrasis in negative constructions' (Johansson & Norheim, 1988). It follows from this that the subjunctive is only overtly marked as a syntactic category in very limited contexts, that is chiefly in the third-person singular present tense when the base form foregoes the inflectional -s (see example (2)). An additional context is when the main verb in the relative clause is *be* or in the passive so that the auxiliary *be* is present and in both cases is not realised fully inflected (see example (3)). Finally, the subjunctive is also realised overtly when the subordinate clause is either negated, so lacking *do*-support (Johansson & Norheim, 1988; see example (4)), or the deontic/suasive verb introducing the relative clause would ordinarily trigger a backshifting of tense (Berg et al., 2020; see example (5)). Naturally, these contexts can also appear together (Hundt, 2018, p. 224). In all examples, the respective forms are in bold.

(3) Local contractors are demanding that they **be** allotted some portion of the work as well. (SAVE India: IN_SM_2003-05-30_c)

(4) The Pakistani inquiry commissioner recommended that he **not** be appointed Pakistan captain again. (SAVE India: IN_TI_37763_c)

(5) I have asked officers **get** the copy. (SAVE2020 India: 4508261)

Beyond this consensus, the status of the mandative subjunctive is heavily debated. In terms of finiteness, grammars generally tend to agree that there is no *inflectional* subjunctive in English, but cannot come to a consensus on whether the subjunctive itself and its container clause are finite or non-finite (Aarts, 2012, pp. 2–3). Anderson (2001) explains that cross-linguistically subjunctives

are syntactically nonfinite, their presence being licensed by other elements [...]; but they are often described as 'finite' partly because of their (relatively) unreduced morphology, as well as their ability to take an unmarked (nominative) subject. However, this latter too is not a necessary property of finites. (p. 160)

In addition to finiteness of the realising verb, the status of the subjunctive itself is also in question. Grammars tend to classify the subjunctive as a grammatical mood (e.g. Quirk et al., 1985), which is often picked up in the research into the subjunctive alternation (e.g. Hoffmann, 1997). More theoretically minded studies have proposed other classifications and terminologies, however. Aarts (2012), for example, describes the subjunctive as a specific clause type of the rank of declaratives, interrogatives, imperatives, and exclamatives (p. 12). I mention these types of discussion here because an introduction of the mandative subjunctive would certainly not be complete without pointing out the many theoretical issues that accompany it. However, for the study of the respective alternation, these considerations only play a peripheral role because the different options speakers of English have at their disposal to realise the subjunctive are in the limelight here.

The first alternant is the base form of the verb as described above. Perhaps somewhat counter-intuitively, as alternative forms also technically realise the mandative subjunctive, I shall refer to this alternant as the *subjunctive* from here on, as this is the preferred term in the literature. There are several alternatives to this realisation. Naturally, in some of the subjunctive contexts, the relative clause could be foregone entirely, in favour of indisputably non-finite

forms like a *to*-infinitive (see example (6)). As this choice is relatively restricted in terms of triggers, however, it tends to be disregarded for studies of the subjunctive. More applicable are indicative and modal alternatives. The subjunctive, as demonstrated in example (7), is generally interchangeable with the indicative, but, to my knowledge and likely because of low frequencies (compare Collins et al., 2014, p. 265), is rarely considered in the literature (see Berg et al., 2020 for a notable exception, albeit with a different theoretical focus). More frequently, the subjunctive is contrasted with modal constructions. *Should* is a viable alternative to the subjunctive without appreciable differences in meaning: 'it is clear that *should* is a (periphrastic) marker of subjunctiveness; and as such it has distinctive semantics from other uses of *should*' (Anderson, 2001, p. 163). In this light, the paper at hand only considers constructions featuring the modal verb *should* as another alternant of the mandative subjunctive (following, e.g., Berg et al., 2020; Deshors & Gries, 2020; see example (8)). This construction will henceforth be referred to as *should*-subjunctive. Although other modal verbs like *could* can in theory be used in the same manner, their frequencies were negligible with the data at hand and their exclusion appears to be the standard in the study of the subjunctive alternation.

- (6) The district administration has asked them to **stay back** in the national capital. (SAVE2020 India: 4636529)
- (7) The groups are also demanding that SCs, STs, OBCs, minorities and women **are included** for recruitment. (SAVE2020 India: 4338059)
- (8) The Congress workers from Uttar Pradesh demanding that the party **should** project Priyanka Gandhi as its chief ministerial candidate. (SAVE2020 India: 4335152)

Although comparatively under-researched multifactorially, several linguistic and extra-linguistic factors have been shown to play a role in the choice between *subjunctive* and *should*-subjunctive. VARIETY is among the most-discussed factors due to the long-lasting argument that American English (AmE) favours the *subjunctive*, whereas BrE displays high frequencies of the *should*-subjunctive (e.g. Johansson & Norheim, 1988). However, Deshors and Gries (2020), based on their large-scale GloWbE data, argue that varietal influence on the subjunctive may have been overstated (p. 227). Hundt (2018) uses, among others, IndE data from the *International Corpus of English*. She finds the variety to be what she calls conservative, that is leaning towards the modal alternant (pp. 235–236). This is echoed by Deshors and Gries (2020) although they also show huge fluctuations of that preference between deontic verbs triggering the subjunctive paradigm (p. 227). In fact, the lemma of said trigger verb is one of the most important factors overall, with the different verbs displaying varying degrees of preference for one of the alternants (e.g. Hundt, 2018). According to Deshors and Gries (2020), who investigate the same triggers as this study, *ask*, *request* and *demand* show a strong preference towards the *subjunctive*, whereas *require*, *recommend* and *order* also prefer the *subjunctive* but comparatively less so than the previous three, and *insist*, *propose* and *suggest* generally pattern towards the *should*-subjunctive (p. 227).

Less influential overall but still considered in most studies are the following factors: LINKING WORD, NEGATION and VOICE. LINKING WORD describes the presence or absence of a linking *that* between main clause and the relative clause containing the subjunctive realisation. Hundt (2018) finds no relevant effect of LINKING WORD, but Deshors and Gries (2020) do describe it in interaction with the trigger verb. Although most verbs are not affected by the absence or presence of *that* in their data, they find that *order* and *recommend* prefer the *subjunctive* when there is no linking word and vice versa, whereas *ask* displays the opposite pattern, strongly preferring the modal alternant *should* when there is no linking word and vice versa. Next, NEGATION refers to whether the verb in the subordinate clause where the alternation is realised is negated or not. With descriptive statistics only, negated verbs have been reported to prefer the *should*- over the *subjunctive* (Hundt, 2018, p. 231; Johansson & Norheim, 1988, p. 30), but, to my knowledge, there is no multifactorial proof of this tendency. Further, Deshors and Gries (2020) show that in passive constructions, the subjunctive is slightly more likely, but this effect of VOICE once again does not hold steady across trigger lemmas but only appears in that interaction. Similar attention has been devoted to the number and person of the subordinate clause's subject, henceforth SUBJECT NUMBER and SUBJECT PERSON. Deshors and Gries (2020) summarise the different person/number constellations in one variable and find that this factor, too, plays a role but, once again, depends on TRIGGER LEMMA. Only for the combination of *first person* and *singular* do they find a steady effect in that this kind of subject generally prefers the verb in the *subjunctive*.

The last variable to discuss here is *DISTANCE*. Special attention has been devoted to it by Berg et al. (2020), who use it in the context of their exploration of range, which is the time a construction remains activated in a speaker's mind. They propose *DISTANCE* as an influential factor as it is an approximation for the processing effort between trigger and subjunctive construction. In their paper, *DISTANCE* is operationalised as the distance in words between the trigger verb and the verb in the target verb, excluding the modal verb *should* if present and the negator *not* as both of these already mark the alternants (2020, pp. 242–243). Berg et al. (2020) also test and show that there is no visible difference between counting *DISTANCE* in words or in morphemes (pp. 242–243; see also Ehret et al., 2014 for a similar test including also number of characters for the length of genitive constituents). Their study shows that the larger the distance between trigger verb and verb in the subordinate clause, or target verb, the more likely the modal alternant becomes (Berg et al., 2020).

The following research questions appear relevant against the background of existing knowledge of the subjunctive alternation:

1. Which factors govern the alternation between *subjunctive* and *should-subjunctive* in contemporary IndE newspaper language?
2. Can we observe short-term diachronic adjustments with the subjunctive alternation?

After the exploration of subjunctive and its various realisations from a theoretical standpoint, Section 3 will now discuss the extraction and annotation of the subjunctive data as well as their statistical handling.

3 | METHODS

3.1 | Data and annotation

The data come from the Indian components of the *South Asian Varieties of English* corpus (SAVE, Bernaisch et al., 2011) and its 2020-update (SAVE2020, Bernaisch et al., 2021). SAVE India and SAVE2020 India consist of roughly 3 million words of newspaper English from circa 2003 and circa 2019 respectively and were sampled from articles from the *Times of India* and the *Statesman*, two of India's largest English-medium newspapers. Using newspaper data like SAVE and SAVE2020 can be both a drawback and an advantage. SAVE comes with in-text headers containing each text's title, author and date of publication where available. The metadata for SAVE2020 is somewhat more substantial as it also includes the column in which an article was published and the author's gender, although both again only where available online at the time of compilation. In comparison to spoken data elicited for the purpose of compiling a corpus, SAVE and SAVE2020 are relatively unsuited for studies with a sociolinguistic emphasis due to their sparse amount of sociobiographic information on the contributors of the data. However, Hundt (2006), for example, argues that newspaper language is an ideal source for the exploration of Outer Circle varieties of English given the language's strong association with the written medium in these contexts. In addition, in the absence of detailed and large-scale corpora differentiating between all potential variants of IndE, the acrolectal nature of newspaper language may allow to draw conclusions on a 'national standard', thus fitting in with the umbrella term 'Indian English' standing in for all versions of IndE to currently exist.

Due to the relative rarity of the subjunctive, no further sampling of the available files was done. It is relatively difficult to narrow down a search context for realisations of the subjunctive alternation given the number of triggers. Nouns, verbs and adjectives alike can trigger a subjunctive form as long as they show deonticity to some degree. Crawford (2009), for example, lists 47 verbs, 38 nouns and 23 adjectives capable of being deontic. However, for feasibility the study at hand is restricted to the nine verbs that most commonly act as a subjunctive trigger identified by Hoffmann (1997): *ask*, *demand*, *insist*, *order*, *propose*, *recommend*, *request*, *require* and *suggest*. This selection has the added benefit of matching the sampling in Deshors & Gries (2020) and thus ensuring comparability with their study.

The extraction was automatically done in *R* (Posit team, 2023) using all verb forms of the chosen triggers as search terms. Naturally, this process resulted in a significant number of false positives of different kinds. For one, the target verb need not match the restricted realisations specified for the variable context; it can appear in the indicative, in a non-finite form, or with a modal verb other than *should*. Additionally, the trigger verbs can all be deontic but can also appear in other semantic contexts without deonticity. Finally, while the form of the search term will match, the result may not be a verb at all. Consider the verb *demand* and its corresponding noun, for example. These false positives were filtered out manually to ensure the validity of all data points.

Additionally, Subsection 2.2 discussed the difficulties in identifying the *subjunctive* as specific parameters need to be met for the construction to be unambiguous. Again to increase comparability, in particular with Deshors & Gries (2020) and Schmidt & Funke (accepted), the study at hand includes all those cases that were ambiguous as per the parameters above but where a subjunctive reading would be constructionally licensed (see example (9)).

(9) Will the government insist that all Naga groups ink any such accord? (SAVE India: IN_SM_2003-10-11_c)

In the same vein, all *should-subjunctives* were retained, regardless of whether the alternant would be an ambiguous *subjunctive* or not (but see Hundt, 2018 for a different approach). From a theoretical standpoint, this decision is motivated by the deonticity of the trigger. I propose that the trigger demands a *subjunctive* and so the indicative would represent an unlikely interpretation of the construction. As for the *should-subjunctive*, I further argue that the modal itself is a choice marking the *subjunctive/should* paradigm and, therefore, need not fulfil further requirements.

The ensuing data were annotated for the variables discussed in Subsection 2.2. Where possible, annotation was done automatically in *R*—if not, additional annotation was added by hand. Notable deviations from the approaches presented above are the *DISTANCE* variable, which was calculated in characters instead of words and the operationalisation of person and number of the subject in the subordinate clause, which was computed as two individual variables instead of one. In addition, several other variables were annotated to try and account for as many data idiosyncrasies as possible. Although not a control variable in the traditional sense, I included *GENDER*, which refers to the gender of the newspaper article author, although it is not available for every text. Both *NEWSPAPER* and *WORD COUNT* act as controls for corpus-internal differences. *NEWSPAPER* refers to the make-up of SAVE and SAVE2020 as laid out above and *WORD COUNT* consists of the length of each newspaper article sampled for the subcorpora and counted in words. Finally, *LEXICAL DIVERSITY* is an aggregate of different measures that indicate a text's lexical diversity.² Table 1 shows all annotated variables, their levels as well as their type.

The resulting data were then further inspected for statistical analysis, the details of which are described in the following section.

3.2 | Statistics

To prepare the data for statistical treatment, I chose to mathematically transform several variables. Both *DISTANCE* and *WORD COUNT* showed a strong skew and were thus logged to their natural base, meaning that the natural logarithmic function \ln was applied to each observation of the two variables, creating the more symmetric predictors *DISTANCE.LOG2* and *WORDCOUNT.LOG2*. Additionally, due to data sparsity, I conflated *SUBJECT PERSON*, combining two of its three attested levels, to create *SUBJECTPERSON.CONFL*. The new variable is binary with the levels *first.second* and *third* to account for the high frequency of 3rd person subjects in the subordinate clauses. Following Gries and Deshors (2020), I explicitly coded interaction terms, that is pairwise combined effects of predictors, but chose to only do so for interactions with *TIME* as the sociolinguistic focus of the study. To compute interactions with the numeric, so measure- or count-based, variables these were binned, meaning that all observations were grouped based on value, and the ensuing categorical predictors appear in the interaction terms.

TABLE 1 Overview of the variables included in the study.

Variable	Type	Levels
CONSTRUCTION	Response	<i>Should, subjunctive</i>
LEXICAL DIVERSITY	Control	Numeric
NEWSPAPER	Control	<i>Statesman, Times of India</i>
WORD COUNT	Control	Numeric
DISTANCE	Main	Numeric
GENDER	Main	<i>Female, male</i>
LINKING WORD	Main	<i>Absent, present</i>
NEGATION	Main	<i>Absent, present</i>
SUBJECT NUMBER	Main	<i>Singular, plural</i>
SUBJECT PERSON	Main	<i>First, second, third</i>
TIME	Main	<i>Early, late</i>
TRIGGER LEMMA	Main	<i>Ask, demand, insist, order, propose, recommend, request, require, suggest</i>
VOICE	Main	<i>Active, passive</i>

TABLE 2 Distribution of CONSTRUCTION across TIME.

TIME/CONSTRUCTION	Should	Subjunctive	Sum
SAVE (2000s)	98 (28.6%)	245 (71.4%)	343 (100%)
SAVE2020 (2010s)	62 (37.6%)	103 (62.4%)	165 (100%)

The final data consist of 508 data points split into 160 *should-subjunctives* and 348 *subjunctives*. Table 2 shows the distribution of the data across the two levels of TIME, namely, the earlier data from SAVE and the later data from SAVE2020. As the table shows, there are fewer occurrences of mandative subjunctives overall in the later data—fewer than half the number of occurrences in the early 2000s. At both points in time, the *subjunctive* is preferred although that tendency seems to incur a downward trend. While 71.4% of subjunctives were realised in the bare form in the earlier data, only 62.4% of subjunctive in the later data took the bare form.

With all variables shown in Table 1 and their interaction terms with TIME mentioned above, I fit a random forest using the function *randomForest::randomForest* (Liaw & Wiener, 2002). Fitting a random forest allows the user to specify the number of trees the model grows and the number of sampled variables at each split. To achieve the best settings for these hyper-parameters, I looped over different combinations of the two numbers to fine-tune and settled for the combination that results in the lowest possible out-of-bag error rate (see also Schmidt & Funke, *accepted*, for the same procedure). The final forest, therefore, grows 2500 trees and randomly samples two variables with each split. The forest will be compared to a null model that always chooses the *subjunctive* as the most frequent construction, which achieves a baseline accuracy of 68.5%.

4 | RESULTS

The random forest detailed in Subsection 3.2 achieves a classification accuracy of 72.44% which is significantly better than the baseline ($p = 0.03$). The model performs well when predicting the *subjunctive* with a precision of 73.21% and a recall of 94.25%. In other words, of the observations the model predicts to be a *subjunctive*, 73.21% are, indeed,

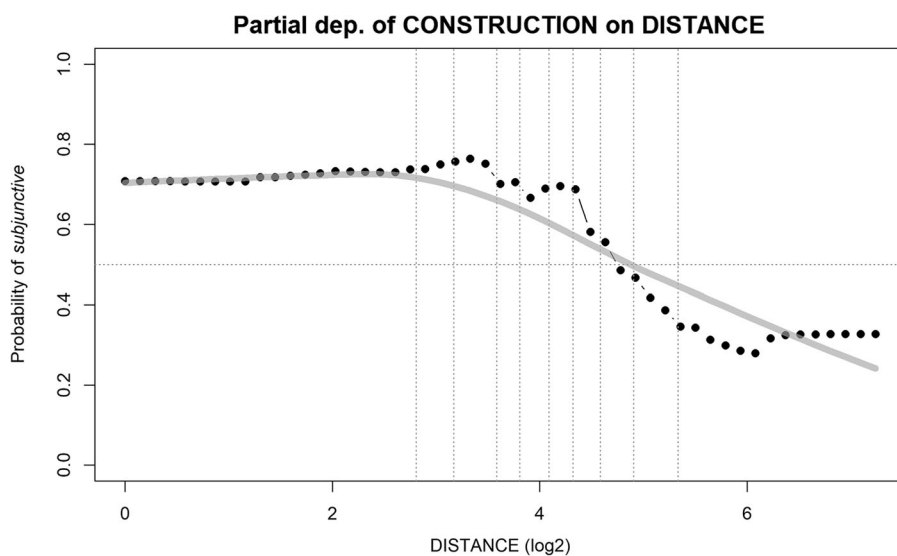


FIGURE 3 Partial dependence plot of DISTANCE (logged to the base of 2).

whereas the dark grey signifies NEGATION as *present*. Coloured correspondingly, the dashed lines represent the main effect of NEGATION for *absent* and *present*.

It is important to note that the probability of the *subjunctive* never falls below 50% in this case. In other words, none of the shown combinations prefer the *should-subjunctive* and no changes in tendency between predictor levels actually culminate in a dispreference of the *subjunctive*. The main effect of NEGATION would suggest that negated constructions as in example (10) are much less likely to take a *subjunctive* than affirmative ones as in example (11). The interaction reveals a significant diachronic adjustment, however. In the earlier data, the preference for the *subjunctive* is stronger with non-negated contexts at a 73.45% probability versus the lower 69.46% when there is negation. However, between 2005 and 2019 this trend completely reverses so that the *subjunctive* is now more likely in negated contexts (65.95%) than in positive ones (59.14%). That being said, we still observe a general trend towards the *should-subjunctive*. Although in 2019 a *subjunctive* is indeed more likely in negated clauses, the probability of a *subjunctive* in these clauses did go down from 2005 to 2019. To put it differently, we observe a general downwards trend, which is just much more pronounced in constructions where there is no negator.

(10) Slamming the leaders who demanded that the Electronic Voting Machines (EVMs) should **not be used** in the upcoming general elections. (SAVE2020 India: 4339459)

(11) Focin demanded that the state government **take over** the Railway lands for rehabilitation. (SAVE India: IN_SM_2004-01-21_c)

The second variable to be discussed is DISTANCE.LOG2, so the distance in characters between the trigger verb and the following verb in the subordinate clause, logged to the base of 2. In Figure 3, the probabilities of the *subjunctive* are again plotted on the y-axis with DISTANCE.LOG2 on the x-axis. The light-grey line is a smoother, highlighting the curve of the data points. The vertical grey lines are a decile grid representing where most of the data are by splitting it into 10 equally large subsections. The middle 50% of these data sit between DISTANCE.LOG2 values of 3.32 and 4.71. For the unlogged variable, this means that 50% of all subordinate clauses in the data set have between 10 and 26.25 characters between trigger verb and target verb. The horizontal grey line sits at a height of 0.5 to better highlight when the probability drops below 50%.

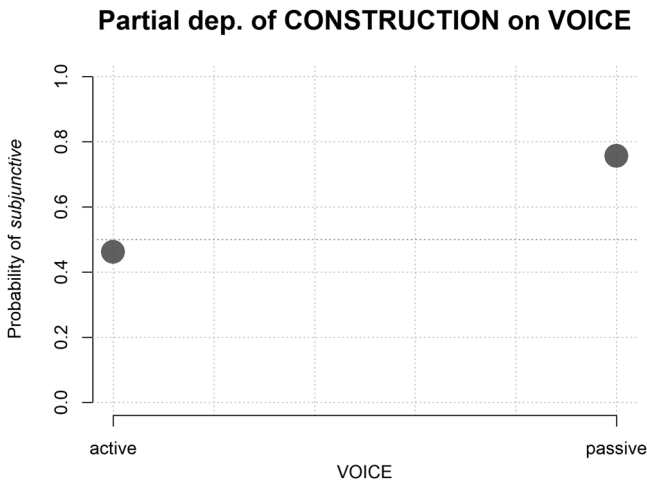


FIGURE 4 Partial dependence plot of VOICE.

There is a clear downwards trend visible so that the larger *DISTANCE.LOG2*, the lower the possibility of a *subjunctive*. The probability holds relatively steady, even displaying a slight upwards trend towards and just about crossing a 75% chance of a subjunctive between logged distance of 0 and 3.33 characters. Beyond this range, there is the trend that the larger the distance between the two verbs grows, the higher the possibility of a *should-subjunctive* becomes. In fact, from a logged length of 4.78 characters, the *should-subjunctive* is continuously preferred, peaking when *DISTANCE.LOG2* is 6.08. Example (12) shows one of the corpus utterances with the smallest possible value of *DISTANCE*, namely, 1, whereas example (13) has the greatest value of *DISTANCE* in the entire data set, namely, 151 characters.

(12) She insisted I wear a hat. (SAVE India: IN_TI_37372_c)

(13) It is also learned that the commission has recommended that about 385 villages, covering a population of nearly 3 lakhs residing within 6 km of the IB in the three districts within 6 km of the IB in the three districts of Jammu, Samba and Kathua, be granted 3% reservation. (SAVE2020 India: 4339205)³

The last term to be discussed here is *VOICE*, namely, *active*, as exemplified in example (14) or *passive* as shown in example (15). Figure 4 follows the schematics of the previous figures and shows the probability of a *subjunctive* on the y-axis and the predictor *VOICE* on the x-axis. Matching the grid, there is an additional horizontal line plotted in the figure at a height of 0.5 for better orientation towards the *subjunctive/should-subjunctive* threshold. Active voice slightly disprefers the *subjunctive* but only barely so at a probability of 46.18%. Passive constructions, on the other hand, show a clear preference for the *subjunctive* at 75.59%.

(14) Courtesy, however, demanded that even a senior leader **should visit** a junior if the latter is in trouble. (SAVE India: IN_SM_2003-09-23_c)

(15) The Adyar Residents Welfare Association has requested that recharge wells **be dug** on both sides of the road. (SAVE2020 India: 4504029)

Section 5 will now discuss these results and their implications with regard to the two research questions formulated in Subsection 2.2.

5 | DISCUSSION

The first research question was geared towards the importance of the individual predictors on the choice between *subjunctive* and *should-subjunctive*. Before we turn to the predictors presented in Section 4, let us first discuss the absence

of another predictor among the most important ones. TRIGGER LEMMA, both on its own and in the interaction with TIME is ranked comparatively low in Figure 1, which may seem surprising given Deshors and Gries' (2020) results that showcase it as one of the strongest predictors. However, this difference should not be overstated but treated with caution given that due to the focus on uncovering possible diachronic effects, only one interaction with TRIGGER LEMMA was plotted. In fact, looking into the main effect of TRIGGER LEMMA (see Figure A1 in the Appendix with the trigger verbs sorted as per the findings in Deshors and Gries, 2020) reveals strong differences between the individual verbs that seem in line with Deshors and Gries' (2020) results and preceding work (e.g. Hoffmann, 1997). *Propose* and (just barely) *suggest* are outliers in that they display the opposite preference in these data, namely, for the *subjunctive* instead of the *should-subjunctive*. Other than that, while with different strengths, all other verbs match the previous findings in their preferences for either construction.

In terms of the featured predictors, VOICE, then, might be surprising not for an absence but for its presence as it has not been discussed much in previous (multifactorial) studies due to a lack of significance. However, the tendency matches previous assumptions to a degree. Hornoiu (2015) shows the *subjunctive* to co-occur frequently with the passive voice in BrE as collected in LOB, although not in AmE as per Brown (2015, p. 72; see also Hundt, 1998 for an earlier iteration of these results). This preference seems to even out over time as in F-LOB BrE active-voice constructions outnumber passive-voice constructions with the *subjunctive* (Leech et al., 2009, p. 59), but the preference is clearly evident for the Indian newspaper English at hand. The combination of the passive voice and the *subjunctive* may be tied to the use of *to be* to create the passive in English. Although not statistically significant, Hundt (2018) posits that the verb '*be* is a stronghold for the *subjunctive*' (p. 230). The *subjunctive* in BrE has often been argued to be a form marked for formality although Hornoiu (2015) hedges that, at least in BrE, this degree of formality may be decreasing (p. 72). A decrease does not yet equal a total loss of formality, however, and so, potentially, the relative formality of newspaper language together with the presence of *be* in the passive as the only verb in English that is clearly identifiable as a *subjunctive* for all combinations of person and number may explain this tendency then. This line of thought may further be aided by the consideration that standardised, or even formal, BrE continues to be the (outspoken) target in English-language education in India and so the somewhat outdated penchant of BrE for the *subjunctive* in the passive may be emulated particularly in newspapers.

The high importance of DISTANCE supports Berg et al.'s (2020) insights as does the general tendency for the *should-subjunctive* with long distances between trigger and target verb. Berg et al. (2020) propose the following explanation for this result:

Implicationally, the syntactic level has a wider range than the morphological level [...]. Owing to this difference in range, syntax can support the production of a modal auxiliary at greater distances more easily than morphology can support the production of the indicative or subjunctive; hence there is an increasing probability of choosing modals at larger distances. (p. 252)

Although not entirely applicable here as their research includes the indicative, Berg et al.'s (2020) general reasoning appears to track. The additional processing efforts of longer distances between trigger and target verb are easier to compensate for when the target verb is supported by a modal verb instead of undergoing a (zero-) process of morphological change. In other words, one might suggest that the *subjunctive* is the more marked of the two alternants. The longer the subject of the subordinate clause and any additional material, the harder it is to process what the deontic trigger is directed at. If the periphrastic construction is preferred in these cases, one could argue that the *subjunctive* with its unusual (zero-) morphology is dispreferred because it adds to the processing effort. It seems that where, as with the subject of the subordinate clause, long constituents cannot be relegated to the end of the clause, processing effort is compensated for in the verb phrase.

The effect of TIME.NEGATION as the most important predictor in the model is an important notion for both research questions formulated earlier. If we zero in on NEGATION within the interaction, the diachronic adjustment when the verb is not negated is markedly more pronounced than when it is. This is somewhat surprising given the fact that one might consider the absence of negation to be the unmarked form, also reflected in the fact that it is much more

frequent in the present data, and would thus expect it to hold relatively steady. The large adjustment could then be indicative of the strength of the diachronic development the subjunctive alternation has undergone and is undergoing in IndE, which segues neatly into the second research question to do with whether we can observe a diachronic effect in the present data. Several of the interactions with TIME exceed its main effect in terms of accuracy added to the model. However, of the four interaction terms in the top half of variable importance, three are outranked by their main terms, namely, VOICE ranks higher than TIME.VOICE, SUBJPERSON.CONFL ranks higher than TIME.SUBJPERSON and DISTANCE.LOG2 outranks TIME.DISTANCECAT, meaning the main tendencies of these predictors are more important than how they develop over the roughly 15 years in question. In sum, there are definite diachronic adjustments that the model unearths. The subjunctive alternation is in flux with continuous changes to the predictor constellations determining the surface structures manifesting in language. However, these adjustments are not as frequent as one might expect with a postcolonial variety as developed as IndE. Some of the predictors are not susceptible to change and so do not resonate a general tendency of development.

In terms of larger implications, these findings are somewhat hard to place also because of how the study is set up. The relative weakness of diachronic adjustments may be somewhat unexpected but could have different reasons. It is entirely possible that the subjunctive alternation presents relatively stable in the last two decades overall which would make this not a finding unique to IndE. Schmidt and Funke (*accepted*) find that the alternation in Pakistani English during the same time span was much more in flux and underwent relatively strong changes with some of the predictors influencing the choice between *subjunctive* and *should-subjunctive*. However, without a concrete comparative study including more varieties from South Asia and generally, trying to categorise the development of the subjunctive alternation in IndE is typical or atypical is founded on speculation. Nevertheless, the results, as would likely have been expected, are still indicative of the advanced evolutionary stage of IndE considering the fact that it unearthed at least one variable to which IndE seems to be more sensitive than other varieties have been previously shown to be, namely, VOICE.

6 | CONCLUSION AND OUTLOOK

The paper at hand is, to my knowledge, the first multifactorial analysis of the subjunctive alternation focusing singularly on IndE. The results suggest two key takeaways. First, IndE, at least in recent newspaper language, seems to be more sensitive to specific factors like VOICE as compared to other varieties analysed in previous studies. However, the directions of any effects in the data match previous research relatively well, overall. Second, while there are diachronic adjustments brought to light by the random forest, these seem relatively minor suggesting that the predictor constellation determining the realisation of the subjunctive has been relatively steady over the last 15 or so years, again with the caveat that this can only be said for newspaper language as a highly standardised medium.

This caveat invokes the primary suggestion for future research. With an eye on IndE, studies done on different data are certainly needed, particularly given the existence of the, albeit by now relatively old, spoken Indian component of the *International Corpus of English* (Greenbaum, 1996). That being said, the data at hand are particularly useful given that it belongs to large-scale corpora of South Asia and so the results for this study would be easily comparable to other projects on the subjunctive-alternation in South Asian Englishes. Schmidt and Funke (*accepted*) is such a paper on Pakistani English, but, so far, no work has been done on this topic using the other varieties included in SAVE and SAVE2020. Finally, while the mandative subjunctive generally is far from an exhausted topic, I would suggest that particularly its contrast with the indicative (see, e.g., Berg et al., 2020) would be a worthwhile avenue of research, especially also in Outer Circle varieties of English.

ACKNOWLEDGEMENTS

I would like to express my thanks to Theresa Anthonia Mößmer and Pauline Alwina Schierholz for their assistance in hand-coding a part of the data. I would also like to thank Tobias Bernaisch, Stefan Th. Gries and Nina Funke for their

helpful feedback and support. Naturally, all remaining inadequacies are my own. The data for this research come from an associated research project funded by the Deutsche Forschungsgemeinschaft (DFG BE 5812/3-1 & GR 5308/1-1).

Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The author declares no conflicts of interest.

ORCID

Karola Schmidt  <https://orcid.org/0009-0008-0541-1630>

NOTES

¹ English does, of course, also retain the perhaps more well-known hypothetical subjunctive as in *I would buy a mansion if I were rich* but I will not further discuss these here. In the same vein, I will not go into the structure of formulaic subjunctives as in *Long live the King*.

² LEXICAL DIVERSITY was determined by calculating all measures from the *quanteda.textstats* package in R (Benoit et al., 2018) and applying a principal component analysis to the group of measures using the *stats::princomp* function. Thirteen different measures of lexical diversity were taken into account for this, including the *Type-Token-Ratio* and *Carroll's Corrected Type-Token Ratio*. The principal component analysis indicates which of the measures under consideration capture the most variation in the data.

³ Note that this example does not display the expected *should-subjunctive*, thereby acting as something of an outlier. The choice of *subjunctive* could be due to the influence of VOICE. Thanks to an anonymous reviewer for pointing this out.

REFERENCES

- Aarts, B. (2012). The subjunctive conundrum in English. *Folia Linguistica*, 46, 1–20. <https://doi.org/10.1515/flin.2012.1>
- Anderson, J. M. (2001). Modals, subjunctives, and (non-)finiteness. *English Language and Linguistics*, 5, 159–166. <https://doi.org/10.1017/S136067430100017X>
- Auer, A., & González-Díaz, V. (2005). Eighteenth-century prescriptivism in English: A re-evaluation of its effects on actual language usage. *Multilingua - Journal of Cross-Cultural and Interlanguage Communication*, 24, 317–341. <https://doi.org/10.1515/mult.2005.24.4.317>
- Balasubramanian, C. (2016). How Indian is Indian English? Indian words in registers of Indian English. *Asiatic*, 10, 89–110.
- Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). Quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, 3, 774. <https://doi.org/10.21105/joss.00774>
- Berg, T., Zingler, T., & Lohmann, A. (2020). The range of linguistic units: Distance effects in English mandative subjunctive constructions. *Journal of Linguistics*, 56, 231–268. <https://doi.org/10.1017/S0022226719000124>
- Bernaish, T., & Lange, C. (2012). The typology of focus marking in South Asian Englishes. *Indian Englishes*, 73, 1–18.
- Bernaish, T., Gries, S. Th., & Heller, B. (2022). Theoretical models and statistical modelling of linguistic epicentres. *World Englishes*, 41, 333–346. <https://doi.org/10.1111/weng.12580>
- Bernaish, T., Gries, S. Th., & Mukherjee, J. (2014). The dative alternation in South Asian English(es): Modelling predictors and predicting prototypes. *English World-Wide*, 35, 7–31. <https://doi.org/10.1075/eww.35.1.02ber>
- Bernaish, T., Heller, B., & Mukherjee, J. (2021). *Manual for the 2020-update of the South Asian Varieties of English (SAVE2020) Corpus*. Justus Liebig University, Department of English.
- Bernaish, T., Koch, C., Mukherjee, J., & Schilk, M. (2011). *Manual for the South Asian Varieties of English (SAVE) corpus*. Justus Liebig University, Department of English.
- Bhatia, T. K. (2022). English language policy in multilingual India. In E. L. Low & A. Pakir (Eds.), *English in East and South Asia: Policy, features and language in use* (pp. 61–74). Routledge. <https://doi.org/10.4324/9780429433467-5>
- 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). Wiley Blackwell.
- Chand, V. (2010). Postvocalic (r) in urban Indian English. *English World-Wide*, 31, 1–39. <https://doi.org/10.1075/eww.31.1.01cha>
- Collins, P., Macalinga Borlongan, A., Lim, J.-H., & Yao, X. (2014). The subjunctive mood in Philippine English: A diachronic analysis. In S. E. Pfenninger, O. Timofeeva, A.-C. Gardner, A. Honkapohja, M. Hundt, & D. Schreier (Eds.), *Contact, variation, and the change in the history of English* (pp. 259–280). John Benjamins.
- Crawford, W. J. (2009). The mandative subjunctive. In G. Rohdenburg, & J. Schlüter (Eds.), *One language, two grammars?* (pp. 257–276). Cambridge University Press. <https://doi.org/10.1017/CBO9780511551970.015>

- De Cuypere, L., & Verbeke, S. (2013). Dative alternation in Indian English: A corpus-based analysis. *World Englishes*, 32, 169–184. <https://doi.org/10.1111/weng.12017>
- Degenhardt, J. (2021). Requests in Indian and Sri Lankan English. *World Englishes*, 42, 523–543. <https://doi.org/10.1111/weng.12573>
- Deshors, S. C., & Gries, S. T. (2020). Mandative subjunctive versus *should* in world Englishes: A new take on an old alternation. *International Journal of Applied Linguistics*, 30, 214–231. <https://doi.org/10.1111/ijal.12275>
- Ehret, K., Wolk, C., & Szmrecsanyi, B. (2014). Quirky quadratures: On rhythm and weight as constraints on genitive variation in an unconventional data set. *English Language and Linguistics*, 18, 263–303. <https://doi.org/10.1017/S1360674314000033>
- Evans, S. (2002). Macaulay's minute revisited: Colonial language policy in nineteenth-century India. *Journal of Multilingual and Multicultural Development*, 23, 260–281. <https://doi.org/10.1080/01434630208666469>
- Gargesh, R. (2020). South Asian Englishes. In C. L. Nelson, Z. G. Proshina, & D. R. Davis (Eds.), *The handbook of world Englishes* (2nd ed., pp. 107–134). Wiley Blackwell.
- Gargesh, R., & Sailaja, P. (2017). South Asia. In M. Filppula, J. Klemola, & D. Sharma (Eds.), *The Oxford handbook of world Englishes* (pp. 425–447). Oxford University Press.
- Greenbaum, S. (1996). Introducing ICE. In S. Greenbaum (Ed.), *Comparing English worldwide: The International Corpus of English* (pp. 3–12). Oxford University Press.
- Gries, S. Th., & 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>
- Gries, S. Th., & Deshors, S. C. (2020). There's more to alternations than the main diagonal of a 2x2 confusion matrix: Improvements of MuPDAR and other classificatory alternation studies. *ICAME Journal*, 44, 69–96. <https://doi.org/10.2478/icame-2020-0003>
- Hickey, R. (2009). South Asian Englishes. In R. Hickey (Ed.), *Legacies of colonial English* (pp. 536–558). Cambridge University Press. <https://doi.org/10.1017/CBO9780511486920.022>
- Hoffmann, S. (1997). *Mandative sentences: A study of variation on the basis of the British National Corpus* (MA-Thesis, Universität Zürich).
- Hornoiu, D. (2015). The subjunctive in present-day English: Revival or demise? *Language and Literature Studies*, 26, 67–77.
- Hundt, M. (1998). It is important that this study (*should*) be based on the analysis of parallel corpora: On the use of the mandative subjunctive in four major varieties of English. In H. Lindquist, S. Klintborg, M. Levin, & M. Estling (Eds.), *The major varieties of English* (pp. 159–173). Acta Wexionensia.
- Hundt, M. (2006). The committee has/have decided... On concord patterns with collective nouns in inner- and outer-circle varieties of English. *Journal of English Linguistics*, 12, 206–232. <https://doi.org/10.1177/0075424206293056>
- Hundt, M. (2009). Colonial lag, colonial innovation or simply language change? In G. Rohdenburg & J. Schlüter (Eds.), *One language, two grammars?* (pp. 13–37). Cambridge University Press. <https://doi.org/10.1017/CBO9780511551970.002>
- Hundt, M. (2018). It is time that this (*should*) be studied cross a broader range of Englishes. In S. C. Deshors (Ed.), *Modeling world Englishes* (vol. G61, pp. 217–244). John Benjamins. <https://doi.org/10.1075/veaw.g61.09hun>
- Johansson, S., & Norheim, E. H. (1988). The subjunctive in British and American English. *ICAME Journal*, 12, 27–36.
- Kachru, B. B. (1965). The Indianness in Indian English. *WORD*, 21, 391–410. <https://doi.org/10.1080/00437956.1965.11435436>
- Koch, C., Lange, C., & Leuckert, S. (2016). 'This hair-style called as 'duck tail': the 'intrusive as'-construction in South Asian varieties of English and learner English. *International Journal of Learner Corpus Research*, 2, 151–176. <https://doi.org/10.1075/ijlcr.2.2.02koc>
- Lange, C. (2012). *The syntax of spoken Indian English*. John Benjamins.
- Lange, C. (2017). Indian English or Indian Englishes? Accounting for speakers' multilingual repertoires in corpora of post-colonial Englishes. In A. Nurmi, T. Rütten, & P. Pahta (Eds.), *Challenging the myth of monolingual corpora* (pp. 16–38). BRILL.
- Leech, G., Hundt, M., Mair, C., & Smith, N. (2009). *Change in contemporary English: A grammatical study*. Cambridge University Press.
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2, 18–22. <https://CRAN.R-project.org/doc/Rnews/>
- Moessner, L. (2007). The mandative subjunctive in Middle English. In G. Mazzone (Ed.), *Studies in Middle English: Forms and meanings* (pp. 209–226). Peter Lang.
- 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>
- Posit team. (2023). *RStudio: Integrated Development Environment for R*. Posit Software. www.posit.co
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. Longman.
- Schilk, M. (2011). *Structural nativization in Indian English lexicogrammar*. John Benjamins.

- Schmidt, K., & Funke, N. (accepted). Exploration of the mandative subjunctive in Pakistani English. *World Englishes* [special issue].
- Schneider, E. (2003). The dynamics of New Englishes: From identity construction to dialect birth. *Language*, 79, 233–281.
- Schneider, E. W. (2007). *Postcolonial English: Varieties around the world*. Cambridge University Press.
- Sedlatschek, A. (2009). *Contemporary Indian English: Variation and change*. John Benjamins.
- Sharma, D. (2012). Indian English. In B. Kortmann & K. Lunkenheimer (Eds.), *The Mouton world atlas of variation in English* (pp. 523–530). De Gruyter.
- Sridhar, S. N. (2020). Indian English. In K. Bolton, W. Botha, & A. Kirkpatrick (Eds.), *The handbook of Asian Englishes* (pp. 243–277). Wiley Blackwell.
- Tamaredo, I., Röthlisberger, M., Grafmiller, J., & Heller, B. (2020). Probabilistic indigenization effects at the lexis–syntax interface. *English Language and Linguistics*, 24, 413–440. <https://doi.org/10.1017/S1360674319000133>
- Traugott, E. C. (1992). Syntax. In R. M. Hogg (Ed.), *The Cambridge history of the English language. Vol. I: The beginnings to 1066* (pp. 168–289). Cambridge University Press.

How to cite this article: Schmidt, K. (2024). The subjunctive alternation in Indian English. *World Englishes*, 1–17. <https://doi.org/10.1111/weng.12683>

APPENDIX

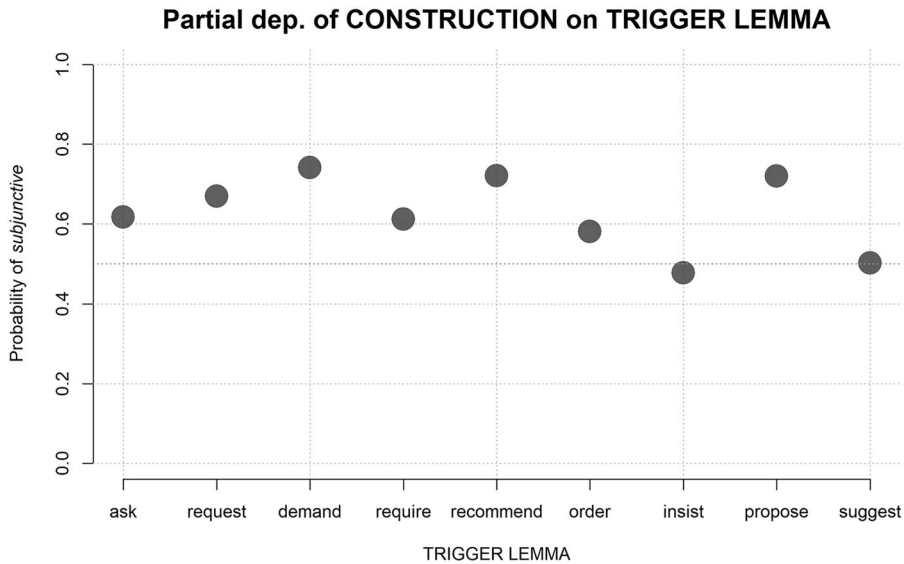


FIGURE A1 Partial dependence plot of TRIGGER LEMMA.