# A Diachronic Study of Modals and Semi-modals in Indian English Newspapers

## Catherine Laliberté[1] iD

## Abstract

Although Indian English is the best-documented South Asian English, its diachronic development has not been described to a great extent. The present study begins to address this gap by offering a real-time perspective on the evolution of modals and semi-modals in Indian English. It sketches the changes in the frequency of modals and semi-modals in three corpora of Indian newspaper texts from 1939, 1968, and the early 2000s. Changes in the frequency of eleven modals and eleven semi-modals are found to be similar to the trends previously observed for written American and British English: semi-modals, as well as the modals *can*, *could*, and *would*, rise greatly in frequency. An analysis of the types of modality expressed by individual modal verbs provides in-depth insights into shifts in Indian English during the period. The study's findings raise methodological and theoretical considerations for the diachronic study of modality in corpora in English generally: the increasing amount of direct quotation in news reportage partly accounts for a rise in modal frequency in this subgenre, which constitutes a confounding aspect seldom articulated in the study of newspaper language. Individual modal verbs exhibit different directions and speeds of change that are not reflected in the trajectory of modals as a category, demonstrating that an aggregate measure is not a suitable point of comparison between varieties to determine their degree of similarity or difference in terms of modal verb usage.

[1]Ludwig Maximilian University of Munich, Germany

**Corresponding Author:**
Catherine Laliberté, Ludwig Maximilian University of Munich, Schellingstr. 3, Munich 80799, Germany.
Email: catherine.laliberte@anglistik.uni-muenchen.de

## 1. Introduction

Indian English (IndE) is considered "the oldest 'non-native' variety of the English language" (Sridhar 2020:243) and is a major variety both in terms of number of speakers and its worldwide influence. IndE has even been described as a norm-providing variety in South Asia (Gries & Bernaisch 2016). Despite its longevity and the appreciable body of research dedicated to it, historical descriptions of IndE have remained scarce. The need for diachronic research on IndE has been observed since at least 2006 (Mukherjee & Hoffmann 2006:166), but, because of lack of available data, such studies have started to appear only recently, notably with Fuchs (2020). In the meantime, much work on language change in the variety has drawn on synchronic data. This approach is premised on the fact that different contemporary genres, varieties, or styles correspond to different stages of change in progress. For example, Collins and Yao (2012) relied on comparisons between present-day spoken and written data to infer direction and speed of change, on the basis that speech universally changes more rapidly than written norms. Others (e.g., Mukherjee & Hoffmann 2006; Mukherjee & Gries 2009) have compared present-day data for IndE and its colonizer parent variety, British English, to pinpoint their degree of difference and deduce IndE's normative orientation. It has been illustrated, however, that such methods may be imperfect substitutes for diachronic work (Gries, Bernaisch & Heller 2018). Genuine diachronic datasets minimize the number of assumptions necessary to uncover trends.

Generally, for Englishes labeled "postcolonial" in the literature, historical data offer the added benefit of relying less on cross-variety comparisons, in particular with American (AmE) and British English (BrE), to sketch language change.[1] Diachronic studies are necessary not only because they work towards describing the linguistic history of Englishes which have been far less scrutinized than British or settler colonial varieties, but also because they can propel forward the theoretical models which have been central to the study of variation in English worldwide (Mukherjee & Bernaisch 2020:756). Most notably, the Dynamic Model of postcolonial Englishes (Schneider 2007) proposes that postcolonial Englishes (including settler colonial Englishes) are the results of specific sociopolitical and linguistic processes that occur in five phases in a temporally linear fashion.[2] Historical linguistic descriptions of these varieties can therefore provide empirical evidence of stages of development (Noël, Van Rooy & Van Der Auwera 2014:3-4) or contribute to the (re)examination of the model's assumptions, for example, by positing the existence of "steady states" not foreseen by the model (Mukherjee 2007).

The present study contributes to the description of IndE from a diachronic perspective. It introduces new historical data in the form of newspaper corpora spanning three sampling points (1939, 1968, and the early 2000s), which are used to gain insights into the recent history of written IndE. The variables investigated here, modal (1) and semi-modal (2) verbs, are expected to exhibit dynamic shifts in frequency and uses, as they have in the history of English more generally.

(1)   Mr M K Jalan, who presented a memorandum to Mr Poonach, said that the operational restrictions on the movement of textile *should* be removed on all railways. (1968_Ln_247)
(2)   This recommendation resulted from the mission's finding that the existing gaps in training were indeed formidable and *needed to* be removed without delay. (1968_Ed_24)

Synchronic evidence has suggested that IndE modals may be especially conservative, a tendency that contrasts with the considerable changes which modals in other varieties have undergone in the twentieth century and beyond. This study therefore seeks to determine whether historical data confirm the conservativeness identified previously with synchronic methods. The hypothesis is that the frequency of modals as a category should remain stable across the period studied, and that verbs which were found to be declining elsewhere may not be in IndE. Stability, or a rise, would provide more evidence regarding the idea that IndE has adopted norms distinct from those of its former colonizer, BrE, in which modals are declining. This research adds to the copious body of work concerned with the study of IndE grammar, providing observations that are relevant to both our understanding of modality in English and language change in this postcolonial variety.

Section 2 summarizes theoretical considerations pertaining to the description of IndE and of modals and semi-modals in AmE and BrE. Section 3 presents the newspaper corpora and lists the variables of interest and their classification. The results, presented in section 4, give a detailed view of modal use in historical Indian newspapers by demonstrating variations in frequency over time, genre, and semantics. These results are discussed in section 5 pointing to their methodological and theoretical implications for the diachronic study of modality as well as language change in IndE.

## 2. Background

### 2.1. Indian English

A large amount of research on IndE has focused on verb-related phenomena such as verb complementation (Mukherjee & Hoffmann 2006), collostructions (Mukherjee & Gries 2009), the subjunctive (Hundt, Hoffmann & Mukherjee 2012), the dative (Bernaisch, Gries & Mukherjee 2014; Gries & Bernaisch 2016), the progressive (Van Rooy 2014; Fuchs 2020), light verb constructions (Hoffmann, Hundt & Mukherjee 2011), and modals and semi-modals (Wilson 2005; Collins & Yao 2012; Deuber et al. 2012; Van Der Auwera, Noël & De Wit 2012; Loureiro-Porto 2016, 2019). Other research on grammar includes lexicosyntax and morphosyntax (Sedlatschek 2009) and syntactic features of spoken IndE (Lange 2012). Some of these studies have related their findings to IndE's current phase in the Dynamic Model, so a consensus has emerged that the variety has reached the endonormative stabilization stage; in other words, it presents tendencies diverging from those of its former colonizer variety, BrE. IndE is said to follow an independent path according to its self-determined norms and even to be a norm-provider for other Englishes in the region (Gries & Bernaisch 2016). For example,

Hoffmann, Hundt, and Mukherjee (2011) observe that present-day IndE exhibits much higher frequencies of light verb constructions (e.g., *take a look*, *give boost*) than BrE and other South Asian Englishes, indicating distinct norms. Other research has exposed tendencies of IndE to be especially conservative, which is also interpreted as evidence of endonormative orientation. This can be seen, for instance, in the comparatively low level of grammaticalization of semi-modals of necessity (Loureiro-Porto 2016, 2019) or the maintenance of the modals *must* (Mair 2015) and *need* (Van Der Auwera, Noël & DeWitt 2012), which have declined in use in BrE. The only (to my knowledge) detailed diachronic investigation of IndE grammar, Fuchs (2020), has brought to light converging trends in the use of the progressive in British and Indian newspaper writing, showing mostly similarity rather than divergence between the two. The emerging picture is that IndE shows, simultaneously, marked innovation (the light verb constructions), preservation of certain forms (the modals), and concord with BrE (the progressive). Hence, Fuchs (2020:13) takes up Mukherjee's (2007) idea that IndE is in a state of equilibrium between endonormative and exonormative forces. These divergent findings may be manifestations of this state, whereby different domains of grammar are subject to different forces. Much of this research differs in the type of data used (various newspaper corpora, written ICE-India and the Kolhapur corpus, spoken ICE-India), which makes generalization problematic. Also, with the exception of Fuchs (2020) and Mair (2015), they rely on synchronic methods.[3] The exclusive use of present-day data is less suitable when it comes to drawing conclusive and detailed insights on temporal trends in IndE. Claims of endonormative stabilization and equilibrium should ideally balance synchronic insights with diachronic ones. The case of modals in particular should benefit from further investigation to determine whether their alleged conservativeness is corroborated by diachronic data.

## 2.2. Modals, Diachronically

In English, modals and semi-modals are the principal means of grammatically encoding modality, i.e., of distinguishing theoretical and hypothetical meaning from factual meaning (Leech [1971] 2004). They differ from each other in that core or central modals (in this paper, simply modals) are fully grammaticalized auxiliaries (e.g., *will*, *must*, *could*) while semi-modals (also called quasi-modals, semi-auxiliaries, or emerging modals) are semantically similar lexical verbs which have acquired, and continue acquiring, auxiliary properties (e.g., BE *going to*, HAVE *to*). As a result, the two types of auxiliaries are conventionally studied alongside each other. Modality has been the object of debates, especially as regards whether modals are declining in frequency in BrE and AmE (Millar 2009; Leech 2011, 2013). A decline has been documented by Leech (2011, 2013) and Mair (2015) in the extended Brown family of corpora (1930s-1990s) for written BrE and AmE, but is not found in the TIME corpus (Millar 2009), in which modals are found to increase greatly.[4] The discussion that ensued (Leech 2011, 2013; Biber & Gray 2013) identified genre as a confounding variable, which highlights the difficulty of determining which type of data is most suitable for generalization

across a variety, and whether generalization is something we should strive for in the first place, as addressed by Biber and Gray (2013).

While change in the frequency of modals is a point of contention, semi-modals are not controversial: the Brown family of corpora as well as the TIME corpus exhibit a clear and important increase in frequency across the twentieth century. A number of hypotheses have been put forward to explain these changes in writing. The rise in frequency of semi-modals is specifically attributed to grammaticalization, as these semi-modals continue acquiring new grammatical uses distinct from their original lexical equivalents (e.g., BE *going to* as a future marker). Leech (2013) proposes that, since grammaticalization is realized above all in spoken language, the rise of semi-modals in writing occurs in conjunction with colloquialization, whereby features of informal speech gradually enter written registers. Relatedly, as the rise of semi-modals is more strongly established in spoken AmE (Leech, Hundt, Mair, and Smith 2009: 104), the same tendency in other varieties has been attributed to Americanization. Collins and Yao (2012), for example, invoke this explanation for their findings on modality in twelve ICE corpora. Here, Americanization can be understood to occur in conjunction with colloquialization: it is a typically American penchant for informality (i.e., speech-like writing) that is adopted into other varieties, rather than American forms themselves, given that no modal or semi-modal can strictly be described as an Americanism. The emergence of semi-modals in particular is a phenomenon which predates the establishment of many postcolonial Englishes, and so can be described, in World Englishes, as the continuation of colonizer input, or a founder effect (Mufwene 1996).

## 3. Method

### 3.1. Newspaper Corpora

There is widespread agreement that corpus-supported observations of diachronic change, especially as regards modal verbs, must take genre differences into account (Biber & Gray 2013; Leech 2013; Mair 2015). This finding can be seen as an argument favoring carefully controlled, aggregate, multi-genre data in order to come close to the idea of representative corpora (the approach championed in Leech [2011]). However, it also highlights the value of corpora or studies that are confined to a single genre (or single publications), insofar as genre is minimized as a confounding parameter. On the basis of differing diachronic tendencies in frequencies exhibited by modals in otherwise minimally different subgenres, Biber and Gray (2013:106) have argued that "historical change should be studied relative to particular registers, rather than attempting a kind of average for English." Consequently, this study is intended as a piece of the puzzle in the historical description of modals in IndE in one genre rather than an attempt to draw conclusions for the variety as a whole.

In India, the press is a major vehicle through which users of English engage with the language. Sridhar's (2020:244) account that "English is more read and heard than it is

written, and more written than spoken" points to the significance and potential impact of printed material on English-using Indian society and IndE itself. English-language newspapers have a particularly potent reach because of their country-wide—and politically powerful—readership (Sridhar 2020:252). As such, they are a particularly relevant standardizing force and vector of norms, representing "acrolectal written English usage" (Bernaisch, Gries & Mukherjee 2014:11). Accordingly, the effects of endonormative stabilization is likely to be observable in the press.

My study uses three corpora of Indian newspapers each containing 180,000 words, of which 120,000 words are from local and national news and 60,000 words come from editorials. The composition of each corpus is detailed in Table 1. The first corpus is made up of selected texts from the IndE component of the South Asian Varieties of English (SAVE) corpus (Bernaisch, Koch, Mukherjee & Schilk 2011), which contains texts from Indian newspapers published between 2002 and 2005. The corpus comprises 90,000 words of news reports from *The Times of India* and 30,000 words of news and 60,000 words of editorials from *The Statesman*. This corpus is the only one of the three used here that includes texts from two different newspapers; the other two are made of texts from *The Times of India* only. These newspapers are both English-language daily newspapers founded in the nineteenth century. *The Times of India* was founded, owned, and edited by British settlers until the country gained its independence in 1947, when its British owners left India. *The Statesman* remained under British ownership until the 1960s. *The Times of India* is currently the leading English-language newspaper in India, with a readership of around 7.6 million in 2013 (Sridhar 2020:252).

**Table 1.** Composition and Size of the Newspaper Corpora

| Corpus (period) | Source | Components | N of texts | N of words | Total Size |
|---|---|---|---|---|---|
| SAVE India 2002-2005 | *The Times of India* | News reports | 342 | 89,946 | 180,447 |
| | *The Statesman* | Editorials | 76 | 60,403 | |
| | | News reports | 83 | 30,098 | |
| 1968 | *The Times of India* | Editorials | 41 | 60,561 | 180,634 |
| | | News reports | 252 | 120,073 | |
| 1939 | *The Times of India* | Editorials | 108 | 60,132 | 180,346 |
| | | News reports | 409 | 120,214 | |

The other two corpora contain texts from *The Times of India* for the years 1939 and 1968. These corpora are not yet publicly available but may be accessed by contacting the author. The material comes from high-quality digitized microfilms and was processed with an Optical Character Recognition (OCR) software, ABBYY FineReader. Individual texts were manually extracted and proofread for OCR errors before they were included in the corpora. Special care was taken to exclude texts from news agencies. The text types and size of these materials match those of the SAVE subcorpus: approximately 120,000 words of local and national news and 60,000 of editorials. In 1939, India's territory extended to what is today Pakistan and Bangladesh. Because of

the need to match the other two corpora as closely as possible, texts that hail from present-day Pakistan and Bangladesh were excluded from the 1939 corpus.

Together, these three corpora allow a time depth of about sixty-seven years and are distributed at roughly thirty-year intervals, i.e., the span of a generation. As shown in Table 1, the corpora and their components are nearly identical in size but differ in the number of individual texts they contain. The 1968 material stands out as containing fewer (and therefore longer) articles than the other periods. The implication of that, and of the relatively small size of the corpora, is that special attention must be paid to individual texts and authors as potential confounding factors, especially as regards editorials. The reporting of the results in section 4 takes those factors into account. The fact that the articles in 1939 and 1968 were generally not signed restricts any further analysis of potentially idiolectal traits. The texts offer no information on the proportion of Indians and colonists employed by the newspaper.

It is noteworthy that the two older corpora match the chronology of B-Brown/ B-LOB and Brown/LOB, as well as newspaper corpora of other postcolonial Englishes (for the 1930s and 1960s) that are currently being compiled. In the absence of much balanced historical IndE data, the newspaper corpora therefore allow for at least tentative comparisons with some findings from Leech, Hundt, Mair, and Smith (2009), Leech (2011, 2013), and Mair (2015) for the Brown family.

## 3.2. Variables and Classification

The modals investigated are the nine core modals and two marginal modals considered in previous research (Leech, Hundt, Mair & Smith 2009; Leech 2013; Mair 2015): *would*, *will*, *can*, *could*, *may*, *should*, *must*, *might*, *shall*, *ought* (*to*), and *need*. An equal number of semi-modals, those researched by Collins (2009a), are examined: HAVE *to*, HAVE *got to*, NEED *to*, *had better*, BE *supposed to*, BE *to*, BE *bound to*, BE *able to*, BE *going to*, WANT *to*, and BE *about to*. This list differs from Leech (2013) and Mair (2015) in that it contains BE *bound to* and BE *about to*, which are considered "semi-auxiliaries" alongside HAVE *to*, BE *going to*, and BE *able to* (Quirk, Greenbaum, Leech & Svartvik 1985:137). Although WANT *to* is not listed as a semi-modal in some grammars, Krug (2000) makes a strong case for its semi-modal status. I follow Collins (2009a), Leech (2013), and many more in including it here.

With the help of AntConc (Anthony 2019), all possible forms of these verbs and constructions were considered, including inflections (e.g., past forms), clitics (e.g., *'ll*, *n't*, *'d*), forms resulting from coalescence (e.g., *gotta*, *wanna*), and negation (*not*) between semi-modal elements. For the semi-modals, all contiguous tokens were retrieved, as well as non-contiguous ones that are split by one element, generally an adverb, as in (3). Such instances account for forty-one of the 1095 semi-modals analyzed.

(3)   Some of these services provided connections to international services and passengers *had* thus *to* walk a considerable distance. (1968_Ln_196)

Tokens were manually analyzed making sure to exclude nouns (e.g., *need*, *May*), non-modal uses, and fixed expressions (e.g., *devil-may-care*). Tokens found in quoted material were included unless it was obvious from the context that the person cited was not Indian, such as Franklin D. Roosevelt or George Harrison.

These procedures resulted in a total of 8255 tokens, each of which was classified according to its meaning. This classification was guided by the terminology and examples in works on modality, such as Palmer ([1979] 1990), Huddleston and Pullum (2002), and Collins (2009a). Any use diverging from those was also recorded. Three main types of modality were distinguished: deontic, epistemic, and dynamic. Deontic modality expresses necessity, obligation, or permission, which come from a deontic source (the speaker, rules, social conventions, etc.). Epistemic modality expresses the likelihood that something is possible, based on logical reasoning. Dynamic modality expresses ability or theoretical possibility which is intrinsic to the subject or forced by circumstances. In (4), for example, contextual elements (*only*) signal that *can* expresses strong likelihood based on logical reasoning, i.e., epistemic necessity. *Can* in (5) refers to an intrinsic (in) ability of the subject to tell the future and is hence classified as dynamic, whereas in (6) the writer (a deontic source) aims to proscribe an idea or behavior. Epistemic (*it is impossible that India regards herself*) or dynamic (*India does not have the ability to regard herself*) readings make little sense here given a premise which sets up a warning of what awaits India if the writer's recommendation is not heeded.

(4)   Further totalitarian aggression, whether by war or threat of war, *can* only have one meaning, namely, aggrandisement at the expense of other States and the reduction of the democracies to helplessness. (1939_Ed_72)

(5)   They have it in their power to plunge Europe and the world into war, and no man *can* forecast what they will do. (1939_Ed_72)

(6)   We are witnessing in Spain the sad end of a democracy which fell into Marxist hands. India *cannot* regard herself today apart from world affairs. When we look outside, our own troubles and divisions seem puny compared with the external danger to our national existence and freedom to which we are ex-posed. (1939_Ed_63)

However, not all modals express these three meanings, and these three meanings are not the only ones which certain modals express. *Would*, *could*, and *might* have a past-tense use (of *will*, *can*, and *may*, respectively) and a hypothetical use, as in (7), in which all three verbs can express speculation. Each past and hypothetical use can express deontic, epistemic, or dynamic modality. *Should* is also a special case, with subjunctive/mandative, adversative, purposive, emotive, and conditional meanings in addition to its more common deontic and epistemic meanings (Huddleston & Pullum 2002:187). Examples of these uses are given in section 4.4. BE *to* also commonly has a conditional sense (*if I were to*) in addition to the three main meanings.

(7)   But it is a grim thought how different it *could* all have been if there had then been a Congress government in West Bengal (1968_Ed_31)

Since context is often unhelpful or ambiguous (in cases of reported speech, for example), the "indeterminate meaning" category is comprised of tokens whose meaning cannot be confidently discerned.

## 4. Results

### 4.1. Overall Frequencies

Table 2 provides the results for the modals in terms of raw frequency. Log-likelihood testing was applied to determine their significance, using the exact corpus sizes.[5] As a category, modals slightly yet significantly increase between 1939 and 2005. *Would*, *can*, *could*, and *ought* (*to*) have increased in frequency during this period and are evidently responsible for the rise in the overall frequency of the category. *Should*, *must*, *might*, and *shall*, by contrast, show significant losses in frequency. The decline of *must* in particular contrasts with that found by Mair's (2015) comparison of the Kolhapur (1978) and ICE-India (1990s) corpora, where *must* exhibits stability. *Will* and *may* cannot be said to have either increased or decreased in frequency; interpretation of the results for *need* are hindered by its modest frequency.

**Table 2.** Modals in Three Indian English Newspaper Corpora (Raw Frequencies)

|            | 1939 | 1968 | % change 1939-1968 | 2000s | % change 1968-2000s | % change 1939-2000s |
|------------|------|------|--------------------|-------|---------------------|---------------------|
| *Would*    | 486  | 561  | +15.4*             | 682   | +21.6***            | +40.3***            |
| *Will*     | 722  | 750  | +3.9               | 704   | -6.1                | -2.5                |
| *Can*      | 217  | 267  | +23.0*             | 315   | +18.0*              | +45.2***            |
| *Could*    | 134  | 203  | +51.5***           | 208   | +2.5                | +55.2***            |
| *May*      | 168  | 112  | -33.3***           | 152   | +35.7*              | -9.5                |
| *Should*   | 311  | 296  | -4.8               | 254   | -14.2               | -18.3*              |
| *Must*     | 152  | 80   | -47.4***           | 94    | +17.5               | -38.2***            |
| *Might*    | 87   | 72   | -17.2              | 42    | -71.4**             | -51.7***            |
| *Shall*    | 24   | 15   | -37.5              | 9     | -40.0               | -62.5**             |
| *Ought (to)* | 5  | 6    | +20.0              | 15    | +150.0*             | +200.0*             |
| *Need*     | 7    | 7    | 0.0                | 3     | -57.1               | -57.1               |
| Total      | 2313 | 2369 | +2.4               | 2478  | +4.6                | +7.1*               |

Note: Significance levels: * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

Looking into the 1939-1968 and 1968-2005 intervals, noteworthy contrasts appear. *May* shows a drop in frequency between 1939 and 1968, followed by a rise of equal

proportion between 1968 and the 2000s. These developments are individually significant, but the overall change (1939-2000s) is not. Conversely, *should* and *shall* show no significant change in frequency during those segments, but the difference between 1939 and 2005 shows significant loss of frequency, although it should be noted that ten of the twenty-four instances of *shall* in 1939 originate from a single text reporting new legislation.

Table 3 provides the raw frequencies for the semi-modals in the three corpora. According to the log-likelihood values, HAVE *to*, NEED *to*, BE *supposed to*, and WANT *to* significantly increase in frequency between 1939 and 2005, whereas BE *to* decreases. *Had better* is unattested in these materials in all periods, and HAVE *got to* occurs only twice. Frequencies for BE *bound to*, BE *able to*, BE *going to*, and BE *about to* show no significant trends. Semi-modals as a category increase across the period, more so than modals do, proportionally. In the case of semi-modals, the 1939-1968 and 1968-2005 intervals show linear tendencies (unlike the case of *may*), at least where significance applies.

**Table 3.** Semi-modals in Three Indian English Newspaper Corpora (Raw Frequencies)

| | 1939 | 1968 | % change 1939-1968 | 2000s | % change 1968-2000s | % change 1939-2000s |
|---|---|---|---|---|---|---|
| HAVE *to* | 103 | 165 | +60.2*** | 169 | +2.4 | +64.1*** |
| HAVE *got to* | 2 | 0 | -100.0 | 0 | — | -100.0 |
| NEED *to* | 5 | 15 | +200.0* | 63 | +320.0*** | +1160.0*** |
| *Had better* | 0 | 0 | — | 0 | — | — |
| BE *supposed to* | 1 | 1 | 0.0 | 15 | +1400.0*** | +1400.0*** |
| BE *to* | 128 | 80 | -37.5*** | 46 | -42.5** | -64.1*** |
| BE *bound to* | 14 | 7 | -50.0 | 10 | +42.9 | -28.6 |
| BE *able to* | 41 | 59 | +43.9 | 41 | -30.5 | 0.0 |
| BE *going to* | 4 | 7 | +75.0 | 10 | +42.9 | +150.0 |
| WANT *to* | 16 | 29 | +81.3 | 54 | +86.2** | +237.5*** |
| BE *about to* | 3 | 3 | 0.0 | 4 | +33.3 | +33.3 |
| Total | 317 | 366 | +15.5 | 412 | +12.6 | +30.0*** |

Note: Significance levels: * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

The near absence of HAVE *got to* and *had better* is unsurprising given their rarity in Balasubramanian's (2009) study and in the Brown family, which Leech (2013:110) attributed to the informal character of the two constructions. While rising significantly in BrE and AmE (Mair 2015:136-137), BE *going to* is rare here. What Leech (2013:110) termed the "prestige barrier," which prevents a greater occurrence of semi-modals in writing because of their perceived colloquiality, may be at play.

HAVE *to*, NEED *to*, and WANT *to* show constant increases from the 1930s on, whereas BE *to* declines. BE *to* never occurs in its non-finite forms in the corpora. This is perhaps unsurprising since non-finite BE *to* has been extremely restricted, if not obsolete, since

the early 1800s in BrE (Hundt 2014). No tendency either way can be discerned for BE *able to*. Finally, BE *supposed to* rises in frequency over the period, despite small token numbers.

With the exception of BE *to*, semi-modals have non-finite forms, which enables them to be used alongside core modals, as in (8). In the IndE corpora, the most frequent semi-modals, namely HAVE *to*, NEED *to*, BE *able to*, and WANT *to*, often occur following a core modal: for example, in 1939, fifteen of the forty-one instances of BE *able to* co-occur with a modal, as do fifty-three of the 165 instances of HAVE *to* in 1968. Such a pattern indicates that while we treat modals and semi-modals as semantic equivalents, they are not mutually exclusive.

(8)  The direction of the hill ranges is generally from north to south, and the railway surveyor *might have to* resort to much zig-zagging to obtain practicable gradients and curves. (1939_Ed_16)

The increase of both modals and semi-modals implies that IndE newspaper writing increasingly modalizes clauses, which goes counter to the findings of Leech (2013) for written AmE and BrE, in which the decline of modals is not proportionally compensated for by rising use of semi-modals. In other words, this "modality deficit" is not found in the present analysis. Leech (2013) suggested that a loss of verbal modality may be compensated for by an increase in the encoding of modality in lexical items, yet he found that the opposite was true: modal lexical expressions had also declined. Thus, Leech (2013) found that modality as a whole was declining in AmE and BrE. Given the rising tendencies of both types of verbs overall in the present study, an important topic for future research would be to explore whether this rise occurs for lexical items as well.

## 4.2. Register and Genre

The fact that the three newspaper corpora are constituted of two subgenres—i.e., editorials and news reports—makes it possible to determine the distribution of modals and semi-modals in terms of style, which this section explores. Given the importance of genre and sub-genre in our assessments of diachronic trends (Leech 2011; Biber & Gray 2013), distinguishing between them is worthwhile, especially since editorials and news have different objectives. Editorials differ from news reports in that their purpose is to share a viewpoint and construct an argument rather than report events. They are more likely to contain judgments of likelihood, recommendations, orders, wishes, and predictions which are expressed through modality. In (9), for example, the author relies on dynamic and epistemic modals to provide rebuttals. Editorials can also be more casual than news reports, in that authors may have more creative freedom in terms of topic, tone, format, and style. Example (10)—taken from the same 1968 text as (9)—may be considered more casual than most news reports, with its rhetorical question, value judgment, *get*-passive, and repetition. This style may also be reflected in the use of modals and semi-modals.

(9) Prof Weiner's reasoning *may* look dotty. But there is method in it. A democratic political system, he argues, has its own dynamic. Whatever changes in popular attitudes it *may* try to bring about it *cannot* wish away the given social environment. (1968_Ed_12)

(10) How can we get rid of that horrid old bore – the too, too slow rate of growth? (1968_Ed_12)

Editorials are fewer in number but usually greater in length than news reports, which means that they are likely less diversified in terms of authors. Consequently idiosyncratic uses may play a greater role in the editorial data. As mentioned in section 3, editorials from 1939 and 1968 are never signed, and many in the SAVE subcorpus are not either, which renders more precise considerations of individuations impossible.

Tables 4 and 5 give the normalized frequency of each modal in the editorial and news genres respectively. For the following reported results, frequency comparisons between editorials and news reports are tested for significance. Modals are more frequent in editorials than in news reports, which was expected given the nature of editorials just outlined. *Can*, *must*, and *may* are especially more frequent in editorials than in news reports over the period studied.

**Table 4.** Modals in Three Indian English Corpora, Editorial Genre (Normalized per 10,000 Words, with Raw Frequency in Parentheses)

|  | 1939 | 1968 | % change 1939-1968 | 2000s | % change 1968-2000s | % change 1939-2000s |
|---|---|---|---|---|---|---|
| Would | 31.26 (188) | 18.99 (115) | -38.8*** | 34.60 (209) | +81.7*** | +11.2 |
| Will | 40.91 (246) | 37.48 (227) | -7.7 | 42.22 (255) | +12.3 | +3.7 |
| Can | 25.94 (156) | 35.01 (212) | +35.9** | 30.96 (187) | -11.8 | +19.9 |
| Could | 5.49 (33) | 8.42 (51) | +54.6* | 13.74 (83) | +62.8** | +151.5*** |
| May | 16.63 (100) | 12.22 (74) | -26.0* | 12.75 (77) | +4.1 | -23.0 |
| Should | 21.79 (131) | 19.81 (120) | -8.4 | 18.38 (111) | -7.5 | -15.3 |
| Must | 18.63 (112) | 7.27 (44) | -60.7*** | 12.58 (76) | +72.7** | -32.1** |
| Might | 4.82 (29) | 6.11 (37) | +27.6 | 3.15 (19) | -48.7* | -34.5 |
| Shall | 0.83 (5) | 0.83 (5) | 0.0 | 0.50 (3) | -40.0 | -40.0 |
| Ought (to) | 0.33 (2) | 0.99 (6) | +200.0 | 2.48 (15) | +150.0* | +650.0*** |
| Need | 0.83 (5) | 0.50 (3) | -40.0 | 0.33 (2) | -33.3 | -60.0 |
| Total | 167.46 (1007) | 147.62 (894) | -11.2** | 171.68 (1037) | +16.0** | -2.9 |

Note: Significance levels: * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

The news reports appear to be responsible for the rise of *would* and *can* found in section 4.1. Despite this fact, *can* is vastly more frequent in the editorial genre. The rise of *could* and *ought* (*to*) is especially strong in the editorial genre, though the frequency of *ought* (*to*) remains low, and five of the fifteen instances in the editorials from the 2000s come from a single text. The decline of *must* is clear in both genres, which is some evidence of a long-term, robust decrease of frequency for this modal in written IndE. The decline of *should*, *might*, and *shall* is

significant only in the news genre. The frequency of *shall* in 1939 is, however, largely due to a single text containing ten instances out of nineteen. While frequency differences for these modals in editorials also appear to show the same tendency, they are not statistically significant. *Will*, *may*, and *need* do not show significant tendencies in either genre.

**Table 5.** Modals in Three Indian English Corpora, News Genre (Normalized per 10,000 Words, with Raw Frequencies in Parentheses)

| | 1939 | 1968 | % change 1939-1968 | 2000s | % change 1968-2000s | % change 1939-2000s |
|---|---|---|---|---|---|---|
| Would | 24.79 (298) | 37.14 (446) | +49.7*** | 39.40 (473) | +6.1 | +58.7*** |
| Will | 39.60 (476) | 43.56 (523) | +9.9 | 37.40 (449) | -14.2* | -5.7 |
| Can | 5.07 (61) | 4.58 (55) | -9.8 | 10.66 (128) | +132.7*** | +109.8*** |
| Could | 8.40 (101) | 12.66 (152) | +50.5** | 10.41 (125) | -17.8 | +23.8 |
| May | 5.66 (68) | 3.16 (38) | -44.1** | 6.25 (75) | +97.4*** | -10.3 |
| Should | 14.97 (180) | 14.66 (176) | -2.2 | 11.91 (143) | -18.8 | -20.6* |
| Must | 3.33 (40) | 3.00 (36) | -10.0* | 1.50 (18) | -50.0* | -55.0** |
| Might | 4.83 (58) | 2.91 (35) | -39.7 | 1.92 (23) | -34.3 | -60.3*** |
| Shall | 1.58 (19) | 0.83 (10) | -47.4 | 0.50 (6) | -40.0 | -68.4** |
| Ought (to) | 0.25 (3) | 0 | -100.0* | 0 | — | -100.0 |
| Need | 0.17 (2) | 0.33 (4) | +100.0 | 0.80 (1) | -75.0 | -50.0 |
| Total | 108.64 (1306) | 122.84 (1475) | +12.9** | 120.04 (1441) | -2.3 | +10.3** |

Note: Significance levels: * = *p* < 0.05, ** = *p* < 0.01, *** = *p* < 0.001.

That news reports exhibit a significant rise in the use of modals when editorials do not points to a change in style in news reports during the twentieth and twenty-first centuries, one which may involve colloquialization. For these data, it is possible that colloquialization is brought about by a shift in journalistic practices. An indication that practices have changed can be seen in the rise of direct quotes throughout the period: a rudimentary corpus search for quotation marks (" ") returns 114 hits for 1939, 277 for 1968, and 474 for the 2000s. The "growing popularity of direct quotes in news reportage" in Indian newspapers has also been observed by Sedlatschek (2009:315), indicating that this observation is not limited to the corpora used here. News reports, then, are increasingly speech-like over time, in the sense that they contain more speech reporting. This is an observation made by Mair (2006:188) for the American and British press as well, a phenomenon he terms "oralization." Given the pervasiveness of modals in speech (Leech, Hundt, Mair & Smith 2009:77), it is logical that an increased reliance on direct quotes would have an effect on the frequency of modals in the press. This potential correlation between genre and modality is explored in more detail in section 4.3.

Tables 6 and 7 give the frequency of each semi-modal in the editorial and news genres. Semi-modals are more frequent in the editorial genre than in news reports. Editorials, then, contain both more modals and more semi-modals than the news reports

do. Once more, this is not surprising given the fact that editorials are generally more argumentative than news reports and so are more likely to express necessity, hypotheticality, and possibility.

The growth of NEED *to* and WANT *to* (Table 3) is constant across all periods in both genres, showing that the trend is especially robust. With the exception of a minute decrease between the 1968 and 2000s editorials, the same holds for HAVE *to.* The decline of BE *to* is also attested for all periods and genres. BE *supposed to* significantly increases in frequency between 1939 and the 2000s in the editorial genre, but not significantly in news reports. Given its tiny raw frequency, though, the concern with idiosyncratic uses raised before applies here too. The other semi-modals, HAVE *got to*, *had better*, BE *bound to*, BE *able to*, and BE *about to*, do not exhibit significant tendencies over the whole 1939-2000s period, although some ephemeral changes can be observed in some periods (e.g., the rise and fall of BE *able to* in Table 6).

In 1939, HAVE *to* is significantly more frequent in news reports. Because it reports on past events, the news genre makes particularly extensive use of its past tense form compared to editorials. Table 8 shows the frequencies of *had to* compared to those of present-tense *have to* and *has to* for each corpus. The past form *had to* is, unsurprisingly, comparatively overrepresented in the news. Seen in conjunction with the high frequency of *must* in editorials compared to news, it may be that the necessity for past tense forms in news reports fosters the use of HAVE *to* as opposed to *must* (which may be preferred in present-tense contexts) in 1939 and 1968. This is supported by the fact that in Table 8, genre distinctions become insignificant by the 2000s, signaling that the distribution of present and past HAVE *to* does not depend on genre the way it did at the two earlier sampling points.

**Table 6.** Semi-modals in Three Indian English Newspaper Corpora, Editorial Genre (Normalized per 10,000 Words, with Raw Frequencies in Parentheses)

| | 1939 | 1968 | % change 1939-1968 | 2000s | % change 1968-2000s | % change 1939-2000s |
|---|---|---|---|---|---|---|
| HAVE *to* | 4.66 (28) | 10.24 (62) | +121.4*** | 9.60 (58) | -6.5% | +107.1** |
| HAVE *got to* | 0 | 0 | — | 0 | — | — |
| NEED *to* | 0.33 (2) | 1.82 (11) | +450.0** | 5.96 (36) | +227.3*** | +1700*** |
| *Had better* | 0 | 0 | — | 0 | — | — |
| BE *supposed to* | 0 | 0.17 (1) | N/A | 1.32 (8) | +700.0* | N/A*** |
| BE *to* | 12.97 (78) | 5.49 (33) | -57.7*** | 4.30 (26) | -21.2 | -66.7*** |
| BE *bound to* | 1.50 (9) | 0.99 (6) | -33.3 | 1.16 (7) | +16.7 | -22.2 |
| BE *able to* | 2.66 (16) | 6.60 (40) | +150.0** | 2.15 (13) | -67.5*** | -18.8 |
| BE *going to* | 0.17 (1) | 1.16 (7) | +600.0* | 0.66 (4) | -42.9 | +300.0 |
| WANT *to* | 0.83 (5) | 1.32 (8) | +60.0 | 3.64 (22) | +175.0** | +340.0*** |
| BE *about to* | 0 | 0.33 (2) | N/A | 0.17 (1) | -50.0 | N/A |
| Total | 23.12 (139) | 28.07 (170) | +22.3 | 28.97 (175) | +2.9 | +25.9* |

Note: Significance levels: * = *p* < 0.05, ** = *p* < 0.01, *** = *p* < 0.001.

**Table 7.** Semi-modals in Three Indian English Newspaper Corpora, News Genre (Normalized per 10,000 Words, with Raw Frequencies in Parentheses)

| | 1939 | 1968 | % change 1939-1968 | 2000s | % change 1968-2000s | % change 1939-2000s |
|---|---|---|---|---|---|---|
| HAVE to | 6.24 (75) | 8.58 (103) | +37.3* | 9.25 (111) | +7.8 | +48.0** |
| HAVE got to | 0.17 (2) | 0 | -100.0 | 0 | — | -100.0 |
| NEED to | 0.25 (3) | 0.33 (4) | +33.3 | 2.25 (27) | +575.0*** | +800.0*** |
| Had better | 0 | 0 | — | 0 | — | — |
| BE supposed to | 0.08 (1) | 0 | -100.0 | 0.58 (7) | N/A** | +600.0* |
| BE to | 4.16 (50) | 3.91 (47) | -6.0 | 1.67 (20) | -57.4*** | -60.0*** |
| BE bound to | 0.42 (5) | 0.08 (1) | -80.0 | 0.25 (3) | +200.0 | -40.0 |
| BE able to | 2.08 (25) | 1.58 (19) | -24.0 | 2.33 (28) | +47.4 | +12.0 |
| BE going to | 0.25 (3) | 0 | -100.0* | 0.50 (6) | N/A** | +100.00 |
| WANT to | 0.92 (11) | 1.75 (21) | +90.9 | 2.67 (32) | +52.4 | +190.9*** |
| BE about to | 0.25 (3) | 0.08 (1) | -66.7 | 0.25 (3) | +200.0 | 0.0 |
| Total | 14.81 (178) | 16.32 (196) | +10.11 | 19.74 (237) | +20.9* | +33.1** |

Note: Significance levels: * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

**Table 8.** *Had to* versus Present Tense Forms (Normalized per 10,000 Words. Raw Frequencies in Parentheses)

| | 1939 | | 1968 | | 2000s | | Total | |
|---|---|---|---|---|---|---|---|---|
| | Past | Present | Past | Present | Past | Present | Past | Present |
| Editorials | 0.83(5) | 2.00(12) | 1.65(10) | 6.11(37) | 1.16(7) | 5.13(31) | 1.21(22) | 4.42(80) |
| News | 3.66(44) | 0.75(9) | 4.50(54) | 0.92(11) | 2.33(28) | 4.33(52) | 3.5(126) | 2.00(72) |
| p-value | <0.001 | <0.05 | <0.01 | <0.0001 | >0.05 | >0.05 | <0.0001 | <0.0001 |

## 4.3. The Role of Direct Quotes

In section 4.2, I called attention to the effect of direct quotations, which may account for overall increases in modal and semi-modal verbs over time. In order to evaluate the potential contribution of quoted speech to the results shown so far, all modals and semi-modals found in direct quotes, such as exemplified in (11), were collected and examined.

(11)    To this, Mr. Sakerlal Balabhal, interrupting said: "It *may* be so in Bombay." (1939_Ln_284)

There are 581 verbs across all three corpora in direct quotation, of which 513 are modals and sixty-eight are semi-modals. Most ($N = 508$) are found in news reports. Table 9 compares the frequency of modals and semi-modals across all corpora, with and without direct quotations. The tendencies for semi-modals are similar in both cases. For the modal

category, however, the effect of quoted material is relevant: the significant 7.1 percent increase is no longer discernible, resulting in a non-significant 0.9 percent decrease. This shows that an increasing number of direct quotations affects modals overall.

Since the semi-modals are largely unaffected by the presence of direct quotations, and since news reports appear to be responsible for most of the changes observed, I focus on the extent to which quoted material affects modal frequencies in news reports specifically. Of the 513 modals found in quotations in all corpora, 448 occur in news reports. Whereas news reports display a significant increase in modals between 1939 and the 2000s, this trend is reversed to a non-significant decrease when the modals in quotations are excluded, as shown in Table 10. Tendencies for individual verbs remain the same, however, with the exception of a now significant 28 percent decrease for *will*.

These findings provide some reassurance that direct quotes have limited effects on the trajectory of modals and semi-modals outlined earlier. Although the effect is not substantial, my data show that quoted material has the potential to change our overall sense of frequencies for individual modals in newspaper language.

**Table 9.** Modals and semi-modals in Three Indian English Newspaper Corpora (Raw Frequencies), with and without Quoted Material

|  |  | 1939 | 1968 | % change 1939-1968 | 2000s | % change 1968-2000s | % change 1939-2000s |
|---|---|---|---|---|---|---|---|
| Modals | full corpus | 2313 | 2369 | +2.4 | 2478 | +4.6 | +7.1* |
|  | without quotations | 2217 | 2233 | +0.7 | 2197 | -1.6 | -0.9 |
| Semi-modals | full corpus | 317 | 366 | +15.5 | 412 | +12.6 | +30.0*** |
|  | without quotations | 302 | 353 | +16.9* | 372 | +5.4 | +23.2** |

Note: Significance levels: * = $p$ < 0.05, ** = $p$ < 0.01, *** = $p$ < 0.001.

**Table 10.** Modals in Three Indian English Corpora, News Genre (Raw Frequencies), without Quoted Material

|  | 1939 | 1968 | % change 1939-1968 | 2000s | % change 1968-2000s | % change 1939-2000s |
|---|---|---|---|---|---|---|
| *Would* | 293 | 425 | +45.1*** | 429 | +0.9 | +46.4*** |
| *Will* | 458 | 485 | +5.9 | 330 | -32.0*** | -28.0*** |
| *Can* | 52 | 34 | -34.6 | 84 | +147.1*** | **+61.5 |
| *Could* | 99 | 151 | +52.5*** | 117 | -22.5* | +18.2 |
| *May* | 58 | 31 | -46.6** | 68 | +119.4*** | +17.2 |
| *Should* | 171 | 170 | -0.6 | 104 | -38.8*** | -39.2*** |
| *Must* | 32 | 30 | -6.3 | 10 | -66.7** | -68.8*** |
| *Might* | 58 | 32 | -44.8** | 21 | -34.4 | -63.8*** |
| *Shall* | 16 | 4 | -75.0** | 4 | 0.0 | -75.0** |
| *Ought* | 2 | 0 | -100.0 | 0 | — | -100.0 |
| *Need* | 1 | 4 | +300.0 | 1 | -75.0 | 0.00 |
| Total | 1240 | 1366 | +10.2* | 1168 | -14.5*** | -5.8 |

Note: Significance levels: * = $p$ < 0.05, ** = $p$ < 0.01, *** = $p$ < 0.001.

## 4.4. Meanings and Uses

This section provides a deeper dive into the uses of *would*, *should*, *must*, HAVE *to*, and *may*, which show trends that are robust or otherwise relate to structural change in progress.

*4.4.1. Would and Should.* As shown in Table 2 (section 4.1), *would* underwent a significant and large increase in frequency. The case of this modal is notable because it is employed in a novel way in IndE: in addition to its hypothetical and preterite uses, so-called "extended *would*" (Collins 2009a) is used in non-hypothetical and non-past contexts, where *will* would be expected in AmE and BrE. These meanings are exemplified here, with *would* in (12) expressing the hypothetical character of an unrealized event, and in (13), as the past form of *will* to express past futurity. By contrast, the extended use illustrated by (14) is neither past nor strictly hypothetical but takes on a pragmatic function indicating "polite and tactful unassuredness" (Collins 2009b:34). In this case, the speaker is tentatively hopeful regarding the realization of an event. When none of these meanings are distinguishable, the token was deemed "indeterminate," as in (15), where reported speech makes it uncertain whether *would* is intended as the past tense of *will* (*the association will welcome…*) or a hypothetical (*the association would welcome… [given the opportunity]*).

(12)   A sheltered bus station was first proposed at Khodadad Circle itself, but both the municipality and the police disapproved of the idea. The police contended that it *would* add to the traffic congestion at that point […] (1968_Ln_150)

(13)   They were ready to board the plane but were told at the last minute that the flight *would* be delayed by 45 minutes as a special was being carried out. (1968_Ln_82)

(14)   "We will soon send our proposal to the Centre. I am hopeful that it *would* be accepted in the interest of the state," added Bhattacharya. (SAVE_Ln_toi_75)

(15)   In reply to a Question by the Chairman, Mr. Kasturbhal said the association *would* welcome an all-India textile board to consider all common questions concerning the textile industry in India. (1939_Ln_116)

The rising frequency of *would* may involve fluctuations in the frequency of its extended meaning. Table 11 shows that the actual frequency of extended *would* is exceptional and cannot be the cause of the increasing frequency of the modal generally. All uses are increasingly more frequent and contribute to the overall trend, however.

**Table 11.** Uses of *Would* in Three Indian English Newspaper Corpora

|  | 1939 | 1968 | % change 1939-1968 | 2000s | % change 1968-2000s | % change 1939-2000s |
|---|---|---|---|---|---|---|
| Hypothetical | 265 | 206 | -22.3** | 296 | +43.7*** | +11.7 |
| Past | 138 | 244 | +76.8*** | 247 | +1.2 | +79.0*** |
| Extended | 4 | 9 | +125.0 | 19 | +111.1 | +375.0** |
| Indeterminate | 79 | 102 | +29.1 | 120 | +17.7 | +51.9** |
| Total | 486 | 561 | +15.4* | 682 | +21.6*** | +40.3*** |

Note: Significance levels: * = *p* < 0.05, ** = *p* < 0.01, *** = *p* < 0.001.

The rarity of extended *would* may be affected by the large amount of backshift in the corpora, which makes extended *would* indistinguishable from its use as the preterite of *will* (hence the comparatively high amount of indeterminate tokens). Another factor influencing the frequency of extended *would* is, once more, the amount of direct quotes present in the corpora. Of the thirty-two instances found in the three corpora, thirteen are in direct quotes. Of those, eleven are to be found in the SAVE mini corpus, which contains the most recent texts. The apparent increase of extended *would* might simply be due to changes in journalistic style, which, as we have seen, appears to rely more on direct quotes in the twenty-first century than in 1939. Ignoring the quoted material, frequencies of extended *would* fall to four (1939), seven (1968) and eight (2000s). None of these differences is significant.

*Should* is also interesting when its meanings are more closely observed. Table 12 gives the details of its meaning changes. A number of meanings significantly drop in frequency between 1939 and 2000s, starting with the subjunctive/mandative use (e.g., *it was imperative that this should be done*). This is meaningful because it is likely a sign that Indian newspaper English is abandoning this periphrastic means of expressing this type of obligation/necessity and replacing it with the morphological subjunctive or the indicative. Sedlatschek (2009:315), for example, reports that the morphological mandative subjunctive has made "noticeable inroads" in Indian newspaper writing. A pilot search for subjunctive "trigger" verbs (Leech, Hundt, Mair & Smith 2009:53-54), reported in Table 13, reveals that this may be the case here, too, since the mandative is expressed with *should* at decreasing rates over the period covered.[6] Whether this use of *should* is substituted with the indicative or the morphological subjunctive is difficult to tell from this limited study, but the morphological subjunctive appears to be more frequent. Note also that recent research on the mandative subjunctive has established that the occurrence of one variant or another strongly depends on the individual trigger verb, including in IndE. Diachronic trends in the use of the subjunctive should thus be evaluated with regard to this predictor, among others (Hundt 2018; Deshors & Gries 2020).

**Table 12.** Meanings of *Should* in Three Indian English Newspaper Corpora

|  | 1939 | 1968 | % change 1939-1968 | 2000s | % change 1968-2000s | % change 1939-2000s |
|---|---|---|---|---|---|---|
| Deontic | 163 | 198 | +21.5 | 218 | +10.1 | +33.7** |
| Epistemic | 23 | 14 | -39.1 | 12 | -14.3 | -47.8 |
| Mandative/subj | 55 | 45 | -18.2 | 15 | -66.7*** | -72.7*** |
| Adversative | 0 | 1 | N/A | 0 | -100.0 | — |
| Purposive | 0 | 0 | — | 0 | — | — |
| Emotive | 22 | 8 | -63.6** | 1 | -87.5* | -95.1*** |
| Conditional | 6 | 4 | -33.3 | 4 | 0.0 | -33.3 |
| Preterite of *shall* | 27 | 2 | -92.3*** | 0 | -100.0 | -100.0*** |
| For *would* | 6 | 4 | -33.3 | 2 | -50.0 | -66.7 |
| Indeterminate | 9 | 20 | +122.2* | 2 | -90.0*** | -77.8* |
| Total | 311 | 296 | -4.8 | 254 | -14.2 | -18.3* |

Note: Significance levels: * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

**Table 13.** Verb Forms after Selected Trigger Verbs in Three Indian English Newspaper Corpora[7]

|  | % in 1939 | % in 1968 | % in 2000s |
|---|---|---|---|
| *Should* | 31 (75.6) | 31 (73.8) | 13 (41.9) |
| Morphological subjunctive | 8 (19.5) | 10 (23.8) | 11 (35.5) |
| Indicative | 0 | 0 | 3 (9.7) |
| Morphological or indicative (unclear) | 0 | 1 (2.4) | 4 (12.9) |
| *Shall* | 2 (4.9) | 0 | 0 |
| Total | 41 | 42 | 31 |

One instance of adversative (*lest something should happen*) and none of purposive (*in order that something should happen*) uses were found in these corpora, but this is not surprising given that Collins (2009a:48) only finds a small handful of them in much bigger and more diversified corpora of British, American, and Australian English.

The emotive meaning of *should* "occurs primarily with predicative items indicating surprise or evaluation" (Collins 2009a:50), as in (16), in which *should* is part of a clause complementing an adjective conveying subjective evaluation (here, a lack of surprise). This meaning declines between 1939 and the 2000s. It may be hypothesized that this use of *should* is replaced by a type of hypothetical *would.* This option would work in (16).

(16)   Nevertheless it was inevitable that they *should* feel the effects of self-government in the neighbouring territories of British India, with which the Orissa States are closely interlinked. (1939_Ed_20)

Finally, *should* as the preterite of *shall* is also falling out of use during the period studied. Again, given that Collins (2009a) does not find any unambiguous instance of this in contemporary AmE, BrE, and Australian English, this result is predictable.

While the mandative/subjunctive, emotive, and preterite uses of *should* are declining, its deontic use (denoting obligation) is increasing. Interplay with other modals and semi-modals might be responsible for this phenomenon: deontic *should* may be less face-threatening than deontic *must*, for instance, or even deontic HAVE *to*. Its lower face-threatening nature is also the explanation of Leech, Hundt, Mair, and Smith (2009:89) for similar findings in AmE and BrE. This avoidance of strong necessity/obligation in AmE has been attributed by Myhill (1995) to democratization, whereby solidarity, as opposed to hierarchical power relationships, is increasingly valued. The data used here intersect with the independence of India from Britain (1947), the creation of its constitution and the installment of its democracy (1950), which makes such a shift in Indian values plausible, too. However, work by Kranich, Hampel, and Bruns (2020) on the pragmatics of requests in IndE shows that Indian speakers tend to be more direct in their demands (i.e., to use fewer modals) than other speakers, and more so when they assume a position of power, which goes against the behavior typically associated with democratization.

*4.4.2. Must and* HAVE *to.* *Must* and HAVE *to* share semantic ground and are sometimes said to be competitors, with HAVE *to* being increasingly preferred to *must* (Leech, Hundt, Mair & Smith 2009:98-99). As was shown in Tables 2 and 3 (section 4.1), *must* is more frequent in 1939 but is increasingly overtaken by HAVE *to* at the later sampling points. Tables 14 and 15 break down these frequencies according to meaning.

**Table 14.** Meanings of *Must* in Three Indian English Newspaper Corpora

|  | 1939 | 1968 | % change 1939-1968 | 2000s | % change 1968-2000s | % change 1939-2000s |
| --- | --- | --- | --- | --- | --- | --- |
| Deontic | 124 | 69 | -44.4*** | 81 | +17.4 | -34.7** |
| Epistemic | 11 | 6 | -45.5 | 10 | +66.7 | -9.1 |
| Dynamic | 3 | 2 | -33.3 | 2 | 0.0 | -33.3 |
| Indeterminate | 14 | 3 | -78.6** | 1 | -66.7 | -92.9*** |
| Total | 152 | 80 | -47.4*** | 94 | +17.5 | -38.2*** |

Note: Significance levels: * = *p* < 0.05, ** = *p* < 0.01, *** = *p* < 0.001.

The only statistically significant tendency for *must* is the decline of its deontic meaning, denoting strong obligation. Deontic HAVE *to*, on the other hand, increases. It is therefore specifically the deontic sense of HAVE *to* that replaces deontic *must*. This phenomenon parallels that of AmE and BrE (Leech, Hundt, Mair & Smith 2009), as well as Australian English (Collins 2014). As argued for deontic *should*, deontic HAVE *to* (denoting milder obligation than deontic *must*) may be a reason for this shift, in line with the idea of democratization already outlined. In the IndE data, epistemic and

dynamic meanings, expressing likelihood (e.g., *you must be hungry*) and need (e.g., *the baby must eat*) respectively, remain stable. Note that the dynamic meaning of HAVE to is particularly frequent, and epistemic *must* is preferred to epistemic HAVE to in all periods. This result aligns with Loureiro-Porto's (2019:124) description of these verbs in their epistemic use in spoken IndE.

*4.4.3. May.* While *may* exhibits conflicting tendencies overall, with a statistically significant decrease in 1939-1968 but an equal increase in 1968-2000s, the tendencies of its various meanings shown in Table 16 give a clear picture: its two least frequent meanings, deontic (permission) and dynamic (ability), have decreased throughout the period under consideration here. Epistemic *may* (indicating possibility), then accounts for an increasingly greater proportions of uses. Thus, despite the V-shaped trend in total frequency reported in Table 2 (section 4.1), *may* shows linear changes which may be indicative of a greater trend, namely that this verb is increasingly monosemous. This was also observed by Leech (2003) in the Brown corpora. A satisfactory explanation for the dip in *may*'s frequency in the 1968 corpus remains to be found, but it may lie in the makeup of that corpus, e.g., in author-specific preferences. Sampling may be an issue, as language change is evidently not always linear, and the year 1968 (or 1939 or 2000s) may not be representative of the trends of that decade, for example.

**Table 15.** Meanings of HAVE *to* in Three Indian English Newspaper Corpora

|  | 1939 | 1968 | % change 1939-1968 | 2000s | % change 1968-2000s | % change 1939-2000s |
|---|---|---|---|---|---|---|
| Deontic | 39 | 79 | +102.6*** | 91 | +15.2 | +133.3*** |
| Dynamic | 55 | 68 | +23.6 | 59 | -13.2 | +6.8 |
| Epistemic | 0 | 0 | — | 2 | N/A | N/A |
| Indeterminate | 9 | 18 | +100.0 | 17 | -5.6 | +88.9 |
| Total | 103 | 165 | +60.2*** | 169 | +2.4 | +64.1*** |

Note: Significance levels: * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

**Table 16.** Meanings of *May* in Three Indian English Newspaper Corpora

|  | 1939 | 1968 | % change 1939-1968 | 2000s | % change 1968-2000s | % change 1939-2000s |
|---|---|---|---|---|---|---|
| Epistemic | 103 | 88 | -14.6 | 133 | +51.1** | +29.1 |
| Deontic | 21 | 5 | -76.2** | 8 | +60.0 | -61.9* |
| Dynamic | 37 | 13 | -64.9*** | 10 | -23.1 | -73.0*** |
| Indeterminate | 7 | 6 | -14.3 | 1 | -83.3* | -85.7* |
| Total | 168 | 112 | -33.3*** | 152 | +35.7* | -9.5 |

Note: Significance levels: * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

# 5. Discussion and Conclusion

This paper set out to describe variation and change in modals and semi-modals in written IndE during the twentieth century. This aim developed from claims that modals and semi-modals have behaved especially conservatively in IndE compared to written AmE and BrE. The hypothesis was that findings from diachronic data should be consistent with this observation. Specifically, I posited that the frequency of modals as a category should remain stable and that verbs declining elsewhere (*must*, *need*, *shall*) would not lose frequency. This would show divergence from IndE's historical input variety, BrE, whose modals have been declining in written language. I assumed that such a tendency would be a sign of endonormative stabilization, whereby the linguistic norms of a postcolonial variety are self-determined.

The results show that modals as a category have slightly but significantly increased in frequency between 1939 and 2000s in IndE newspapers, but only when all quoted material is preserved. Otherwise, this increase is neutralized. Newspaper language is a combination of writers' own words and those of the people they choose to quote, and the way these dimensions of journalism are employed have changed, which has clear linguistic consequences. The measure excluding quoted material is consistent with the hypothesis submitted at the beginning of this paper: in Indian English newspapers, between 1939 and the 2000s, core modals have not significantly decreased or increased in frequency.

Crucially, however, the tendencies exhibited by most core modals do *not* fundamentally differ from those of AmE and BrE as reported in Leech (2011) and Mair (2015). The least frequent modals (*must*, *might*, and *shall*) are also those which decrease most rapidly in IndE. *Can*, *could*, and *would* increase in frequency in IndE, as is the case in AmE (and BrE, except for *would*). Not only are these three modals responsible for the rise in frequency of the whole category, they also appear to be increasing at a higher rate than they do in AmE overall: *can* increases by only 25.7 percent in AmE against 45.2 percent in IndE, *could* by 24.3 percent against 55.2 percent in IndE, and *would* by 18.9 percent against 40.3 percent in IndE. This remains true when excluding quoted material. In the data from the TIME corpus (Millar 2009:199), modals as a category increase by 22.9 percent between the 1920s and 2000s, a change largely driven by massive increases of *can*, *could*, and, contrary to the other corpora mentioned here, *may. Might* also increases there, unlike in the Brown and IndE newspaper corpora. The tendencies of individual verbs in IndE are closer to those of the Brown corpora (which contain multiple genres) than those of TIME (similar in genre), as is the speed of these changes. The results raise questions regarding the relevance of using modals as a category to determine direction or speed of structural change. Clearly, the fact that the category increases in IndE newspapers and not in aggregate corpora for AmE and BrE is by itself no indication that it is moving in a different direction from the other two varieties. Looking into the individual modals, it appears to be the opposite: IndE newspaper writing does not diverge from written AmE and BrE, and it is even possibly faster in the rise of certain modals. In short, the results are consistent with the expectation that modals as a category have remained stable or increased in frequency, but

they do not indicate endonormative stabilization (i.e., divergence from British tendencies), contrary to the rationale formulated at the outset. Rather, Indian English newspapers match British and American writing when individual verbs are considered.

Semi-modals have evolved as expected: they have increased in frequency overall, and this rise is also clear in both editorials and news reports, at almost identical rates in both, which indicates that this trend is robust. In this regard, IndE press language is on par with tendencies discovered for other varieties and does not come across as especially conservative or innovative. While the study did not specifically set out to explore colloquialization, the fact that semi-modals increase overall is a convincing sign that colloquialization is taking place, as these verbs are more frequent in speech (Balasubramanian 2009:195-196). Colloquialization also appears in the shape of oralization (Mair 2006:188), whereby speech is inserted into written language, as news reports were found to contain increasingly more direct quotes. Oralization, the expansion of semi-modals, and the trends presented for most individual modals are visible starting as early as in the 1939-1967 interval, indicating that they are not new, contrary to the post-1990 liberalization of the Indian economy and any concomitant influence of AmE. The implication is that these phenomena constitute trends parallel to that of American and British media, in continuity with those set in motion via the colonial input, rather than recent Americanization.

These conclusions stand in contrast with most other studies on the grammar of IndE but join Fuchs (2020) in reporting little evidence of endonormative tendencies. Notably, both this study and the findings of Fuchs (2020) on the progressive are based on historical newspaper data. While diachronic evidence should depict actual change more precisely than synchronic approaches (Gries, Bernaisch & Heller 2018), the niche representativity of newspaper data remains a problem. Unfortunately, other types of materials are not yet available for historical studies of IndE, owing to a certain inequality between postcolonial Englishes and other Englishes in terms of the types and amount of corpus data readily accessible. Sridhar (2020:254) notes to this effect that such fragmented accounts describe IndE "like the proverbial elephant by the Hindus, correctly in parts but missing the big picture," referring to the parable in which individuals encountering an elephant in the dark are each only able to touch one of its parts and thus end up misidentifying the creature as various unrelated objects rather than one large animal. Future work will hopefully strive to further diversify types of data for IndE and expand the scope of our enquiries to the variety's full range of registers.

## Acknowledgements

## Declaration of Conflicting Interests

## Funding

## ORCID iD

Catherine Laliberté  https://orcid.org/0000-0002-0342-7873

## Notes

1. I adopt Denis and D'Arcy's (2018) use of "postcolonial" to refer to varieties used in postcolonial situations in which English is used by the indigenous population. This therefore excludes settler colonial varieties such as American English, which stem from (ongoing) settler colonization.
2. The five phases are (1) Foundation, (2) Exonormative stabilization, (3) Nativization, (4) Endonormative stabilization, and (5) Differentiation. Each is characterized by distinct (a) historical and political events, which provoke shifts in (b) identity construction, which in turn determine the modalities of (c) language contact and attitudes, which result in (d) linguistic developments and structural effects. See Schneider (2007) for more detail.
3. The contribution of Mair (2015) to the study of Indian English modals consists in reporting frequencies for *must* and *shall* in the Kolhapur (1978) and ICE-India (1990s) corpora.
4. Mair (2015) employs results from Leech, Hundt, Mair, and Smith (2009) and Leech (2011) but includes data from B-Brown, which was not complete at the time of Leech (2011). Mair (2015) also presents frequencies for each verb, including percentage of change, for the 1930s, 1960s, and 1990s components of the Brown family, which are not (all) reported in Leech (2011). Hence, Mair (2015) is used here as the principal reference for comparisons with the Brown family.
5. I used Paul Rayson's UCREL log-likelihood and effect size calculator (http://ucrel.lancs.ac.uk/llwizard.html).
6. Deuber et al. (forthcoming) find recent IndE newspaper data (2011-2013) to exhibit an even lower proportion of mandative *should* (about 29 percent), which further substantiates the decline seen here.
7. The verbs are ADVISE, ASK, BEG, DEMAND, DESIRE, DIRECT, INSIST, MOVE, ORDER, PROPOSE, RECOMMEND, REQUEST, REQUIRE, STIPULATE, SUGGEST, URGE, and WISH (Leech, Hundt, Mair & Smith 2009:53-54).

## Software

Anthony, Lawrence. 2019. AntConc [Computer program]. Version 3.5.8, retrieved February, 2020 from https://www.laurenceanthony.net/software. Tokyo, Japan: Waseda University.

## References

Balasubramanian, Chandrika. 2009. *Register variation in Indian English*. Amsterdam: John Benjamins.

Bernaisch, Tobias, Stefan Th. Gries & Joybrato Mukherjee. 2014. The dative alternation in South Asian English(es): Modelling predictors and predicting prototypes. *English World-Wide* 35(1). 7-31.

Bernaisch, Tobias, Christopher Koch, Joybrato Mukherjee & Marco Schilk. 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.

Biber, Douglas & Bethany Gray. 2013. Being specific about historical change: The influence of sub-register. *Journal of English Linguistics* 41(2). 104-134.

Collins, Peter. 2009a. *Modals and quasi-modals in English*. Amsterdam: Rodopi.

Collins, Peter. 2009b. Extended uses of *would* in some Asian Englishes. *Asian Englishes* 12(2). 34-46.

Collins, Peter. 2014. Quasi-modals and modals in Australian English fiction 1800-1999, with comparisons across British and American English. *Journal of English Linguistics* 42(1). 7-30.

Collins, Peter & Xinyue Yao. 2012. Modals and quasi-modals in New Englishes. In Marianne Hundt & Ulrike Gut (eds.), *Mapping unity and diversity world-wide*, 49-68. Amsterdam: John Benjamins.

Denis, Derek & Alexandra D'Arcy. 2018. Settler colonial Englishes are distinct from postcolonial Englishes. *American Speech* 93(1). 3-31.

Deshors, Sandra C. & Stefan Th Gries. 2020. Mandative subjunctive versus *should* in world Englishes: A new take on an old alternation. *Corpora* 15(2). 213-241.

Deuber, Dagmar, Carolin Diewer, Stephanie Hackert & Michaela Hilbert. 2012. *Will* and *would* in selected New Englishes. In Marianne Hundt & Ulrike Gut (eds.), *Mapping unity and diversity world-wide*, 91-116. Amsterdam: John Benjamins.

Deuber, Dagmar, Stephanie Hackert, Eva Hänsel, Alexander Laube, Mahyar Hejrani & Catherine Laliberté. Forthcoming. The norm orientation of English in the Caribbean: A comparative study of newspaper writing from ten countries. *American Speech*.

Fuchs, Robert. 2020. The progressive in 19th and 20th century settler and indigenous Indian English. *World Englishes* 39(3). 394-410.

Gries, Stefan Th. & Tobias Bernaisch. 2016. Exploring epicentres empirically: Focus on South Asian Englishes. *English World-Wide* 37(1). 1-25.

Gries, Stefan Th., Tobias Bernaisch & Benedikt Heller. 2018. A corpus-linguistic account of the history of the genitive alternation in Singapore English. In Sandra C. Deshors (ed.), *Modeling World Englishes: Assessing the interplay of emancipation and globalization of ESL varieties*, 245-279. Amsterdam: John Benjamins.

Hoffmann, Sebastian, Marianne Hundt & Joybrato Mukherjee. 2011. Indian English – An emerging epicentre? A pilot study on light verbs in web-derived corpora of South Asian Englishes. *Anglia* 129(3). 258-280.

Huddleston, Rodney & Geoffrey K Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Hundt, Marianne. 2014. The demise of the *being to V* construction. *Transactions of the Philological Society* 112(2). 167-187.

Hundt, Marianne. 2018. It is time that this (*should*) be studied across a broader range of Englishes: A global trip around mandative subjunctives. In Sandra C. Deshors (ed.), *Modeling World Englishes: Assessing the interplay of emancipation and globalization of ESL varieties*, 217-244. Amsterdam: John Benjamins.

Hundt, Marianne, Sebastian Hoffmann & Joybrato Mukherjee. 2012. The hypothetical subjunctive in South Asian Englishes: Local developments in the use of a global construction. *English World-Wide* 33(2). 147-164.

Kranich, Svenja, Elisabeth Hampel & Hanna Bruns. 2020. Changes in the modal domain in different varieties of English as potential effects of democratization. *Language Sciences* 79. 1-15.

Krug, Manfred G. 2000. *Emerging English modals: A corpus-based study of grammaticalization*. Berlin: Mouton de Gruyter.

Lange, Claudia. 2012. *The syntax of spoken Indian English*. Amsterdam: John Benjamins.

Leech, Geoffrey. 2003. Modals on the move: The English modal auxiliaries 1961-1992. In Roberta Facchinetti, Manfred Krug & Frank Palmer (eds.), *Modality in contemporary English*, 223-240. Berlin: Mouton de Gruyter.

Leech, Geoffrey. [1971] 2004. *Meaning and the English verb*. Harlow: Pearson Longman.

Leech, Geoffrey. 2011. The modals ARE declining: Reply to Neil Millar's "Modal verbs in TIME: Frequency changes 1923-2006." *International Journal of Corpus Linguistics* 16(4). 547-564.

Leech, Geoffrey. 2013. Where have all the modals gone? An essay on the declining frequency of core modal auxiliaries in recent standard English. In Juana I. Marín Arrese, Marta Carretero, Jorge Arús Hita & Johan van der Auwera (eds.), *English modality: Core, periphery and evidentiality*, 95-115. Berlin: Mouton de Gruyter.

Leech, Geoffrey, Marianne Hundt, Christian Mair & Nicholas Smith. 2009. *Change in contemporary English: A grammatical study*. Cambridge: Cambridge University Press.

Loureiro-Porto, Lucía. 2016. (Semi-)modals of necessity in Hong Kong and Indian Englishes. In Elena Seoane & Cristina Suárez-Gómez (eds.), *World Englishes: New theoretical and methodological considerations*, 143-172. Amsterdam: John Benjamins.

Loureiro-Porto, Lucía. 2019. Grammaticalization of semi-modals of necessity in Asian Englishes. *English World-Wide* 40(2). 115-142.

Mair, Christian. 2006. *Twentieth-century English: History, variation and standardization*. Cambridge: Cambridge University Press.

Mair, Christian. 2015. Cross-variety diachronic drifts and ephemeral regional contrasts: An analysis of modality in the extended Brown family or corpora and what it can tell us about the New Englishes. In Peter Collins (ed.), *Grammatical change in English world-wide*, 119-146. Amsterdam: John Benjamins.

Millar, Neil. 2009. Modal verbs in TIME: Frequency changes 1923-2006. *International Journal of Corpus Linguistics* 14(2). 191-220.

Mukherjee, Joybrato. 2007. Steady States in the evolution of New Englishes: Present-day Indian English as an equilibrium. *Journal of English Linguistics* 35(2). 157-187.

Mukherjee, Joybrato & Tobias Bernaisch. 2020. Corpus linguistics and Asian Englishes. In Kingsley Bolton, Werner Botha & Andy Kirkpatrick (eds.), *The handbook of Asian Englishes*, 741-761. Hoboken: Wiley-Blackwell.

Mukherjee, Joybrato & Stefan Th. Gries. 2009. Collostructional nativisation in New Englishes verb-construction associations in the International Corpus of English. *English World-Wide* 30(1). 27-51.

Mukherjee, Joybrato & Sebastian Hoffmann. 2006. Describing verb-complementational profiles of New Englishes: A pilot study of Indian English. *English World-Wide* 27(2). 147-173.

Mufwene, Salikoko. 1996. The Founder Principle in creole genesis. *Diachronica* 13(1). 83-134.

Myhill, John. 1995. Change and continuity in the functions of American English modals. *Linguistics* 33(2). 157-212.

Noël, Dirk, Bertus van Rooy & Johan van der Auwera. 2014. Diachronic approaches to modality in World Englishes: Introduction to the special issue. *Journal of English Linguistics* 42(1). 2-6.

Palmer, Frank R. [1979] 1990. *Modality and the English modals*. London: Longman.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.

Schneider, Edgar. 2007. *Postcolonial Englishes: Varieties around the world*. Cambridge: University Press.

Sedlatschek, Andreas. 2009. *Contemporary Indian English: Variation and change*. Amsterdam: John Benjamins.

Sridhar, S. N. 2020. Indian English. In Kingsley Bolton, Werner Botha & Andy Kirkpatrick (eds.), *The handbook of Asian Englishes*, 243-277. Hoboken: Wiley-Blackwell.

Van Rooy, Bertus. 2014. Progressive aspect and stative verbs in Outer Circle varieties. *World Englishes* 33(2). 157-172.

Van Der Auwera, Johan, Dirk Noël & Astrid De Wit. 2012. The diverging *need (to)*'s of Asian Englishes. In Marianne Hundt & Ulrike Gut (eds.), *Mapping unity and diversity world-wide*, 69-90. Amsterdam: John Benjamins.

Wilson, Andrew. 2005. Modal verbs in written Indian English: A quantitative and comparative analysis of the Kolhapur corpus using correspondence analysis. *ICAME Journal* 29. 151-170.

## Author Biography

**Catherine Laliberté** is a PhD candidate in English linguistics at the University of Munich. She is completing her dissertation on the use of English by Panamanians of West Indian descent, touching on morphosyntactic variation as well as language maintenance and shift. Her research interests include World Englishes, pidgins and creoles, corpus linguistics, modality, and language in film.