Language Variation and Change, **13** (2001), 343–373. Printed in the U.S.A.
© 2002 Cambridge University Press 0954-3945/02 $9.50

The pluperfect in native and non-native English: A comparative corpus study

DEVYANI SHARMA
Stanford University

ABSTRACT

One of the challenges in characterizing non-native varieties of English is accounting for variant uses of ostensibly standard English forms. The present corpus study examines both quantitative and qualitative aspects of pluperfect use in Indian English (IndE), British English (BrE), and American English (AmE). IndE is found to differ from native usage by associating *had* + V-*ed* with present perfect and preterite meanings. Licensing of pluperfect contexts by time adverbials is also found to be significantly lower in IndE. AmE shows the lowest overall use of the pluperfect and the highest use of disambiguating adverbials. Thus, AmE and IndE show distinct patterns of divergence from BrE. Variation within IndE exhibits a tendency for greater non-nativeness in regional (vs. national) press and in bureaucratic (vs. press) registers, suggesting a multidimensional distribution of IndE nonstandardness in India. These nonstandard uses are shown to convey new pragmatic meanings deriving from ambiguity in the native system and reinforcement from substrate languages. Finally, these changes are evaluated in relation to the broader tense–modality–aspect system of IndE as well as those of other non-native Englishes which exhibit similar characteristics.

This article examines a case of dialectal variation in a subpart of the tense–modality–aspect (TMA) system of Indian English (IndE). Rather than examining nonstandard forms in the dialect, which are relatively few, I examine new functions associated with an existing form—namely, the use of the pluperfect *had* + V-*ed* construction—and contrast this usage with that found in native British (BrE) and American English (AmE). Research on non-native Englishes and creoles has begun to focus on the phenomenon of variation in the meanings associated with standard English forms. In his survey of existing studies on features of Indian English, Shastri (1992:274) observed that "the identification of variant functions of the same forms in Indian English is the immediate task before us." Bickerton (1980:113), in his discussion of decreolization, noted that the development of new functions for existing forms must be treated as a

I am indebted to John Rickford, Elizabeth Traugott, Tom Wasow, Rob Podesva, and members of the audience at the 29th Conference on New Ways of Analyzing Variation, East Lansing, Michigan, October 2000, for valuable input on earlier versions of this article. Special thanks to two anonymous reviewers for their thoughtful suggestions. Any remaining shortcomings are solely my responsibility.

https://doi.org/10.1017/S095439450113303X Published online by Cambridge University Press

distinct mode of change from the acquisition of new forms for existing meanings or functions.

Due to the complexity of English use in multilingual settings and to a delayed recognition of systematicity in indigenized, post-colonial varieties of English (Kachru, 1992; Williams, 1987), quantitative studies of IndE and other non-native varieties are much less common than descriptive overviews of dialect features (some exceptions include Leitner, 1991; Shastri, 1996; Sridhar, 1991). In the present study, I explore techniques for quantifying functional and pragmatic variants of the pluperfect form in IndE. Perhaps even more than structural innovations, it is this class of variants that constitutes the basis of divergence from native usage in this variety. I examine a corpus of present-day IndE print texts for such a shift in usage, and I carry out both a quantitative and qualitative comparison with corpora of two native varieties of English (British and American) to establish the nature and degree of this change.

The central finding is that the pluperfect in IndE may also be used with either a preterite or a present perfect meaning (see Rickford & Rafal, 1996, for a similar phenomenon in AAVE). In addition, the comparison to native varieties reveals crossdialectal variation in the use of this construction and the explicit disambiguation of its meaning (via time adverbials, for instance) across different varieties of English. In particular, AmE and IndE diverge from BrE usage of pluperfect in distinct ways: AmE exhibits greater restrictiveness along certain dimensions, whereas IndE extends the use of the construction beyond the BrE discourse constraints. The new usage in IndE appears to derive from a nonstandard signaling of pragmatic viewpoint and tense reorientation. While language transfer from native language(s) is commonly cited as a source of non-native variation, I suggest that ambiguity inherent in native English usage can play as important a role.

After a brief overview of the introduction, spread, and current status of English in India, I present a description of the corpus data and the methodological approach. This is followed by quantitative data on dialectal variation among the three dialects as well as intradialectal variation within IndE, based on register and regionality. The qualitative discussion examines the pragmatic meanings that accompany nonstandard uses. Finally, I locate these findings within the broader TMA system of IndE and developments in other nonstandard varieties.

ENGLISH IN INDIA

Kachru (1992:301) and Williams (1987:162) used the phrase "non-native institutionalized variety" for English that is rarely the native language of speakers, but is acquired locally as a second language through educational institutions. In these settings, non-native speakers may ultimately become most proficient in English or may use English in the widest range of functional domains, leading to a continuum that approaches nativeness at one extreme. Thus, although I follow the conventional use of "non-native" here, it is necessary to bear in mind that the

https://doi.org/10.1017/S095439450113303X Published online by Cambridge University Press

native/non-native distinction in linguistic theory tends to reify continuous phenomena into discrete categories.

Mehrotra (1998) identified three broad periods of English in India: 1578–1834, 1835–1947, and 1947–present. During the initial 250-year period, when English was first introduced into the subcontinent, the main domains of its use were trade, the military, and missionary work. The second phase of English in India began with an ordinance issued in 1835, which decreed that English be the medium of all schools and universities in India. The ordinance was based on a recommendation by T. B. Macaulay entitled *Minute on Indian Education*, which proposed the formation of "a class of persons, Indian in blood and colour, but English in taste, in morals and in intellect" (Kachru, 1983:22). These individuals were to be interpreters "between us [the British] and the millions whom we govern." As in many post-colonial regions, the long-term result of this policy was an overarching, class-based politicization of the English language throughout the country, leading to divided linguistic ideologies.[1]

The third period of the establishment of English in India began immediately after independence in 1947. As many Indians ceased to aspire to a British acrolectal standard in their English, the ideological distance from the native variety and the range in bilingual competence expanded. This is the crucial point at which indigenized, non-native varieties of English come to be distinct from regular second language acquisition (Sridhar & Sridhar, 1992). Indian languages and vernaculars were initially favored as mediums of instruction; however, today English has not only remained in India, but also spread further across various domains of society, contrary to the predictions of many at the time of independence. English dominates the mass media and the publishing world, and it is a central, functional language in domains such as higher education, administration, court proceedings, and politics. Nevertheless, English has not displaced many indigenous languages, but rather plays a functional role in a complex, multilingual arena.[2]

This brief sketch gives some indication as to why the language has persisted after independence. The roots of English were set early and deep; many current institutions including universities, administrative organizations, and the media were first established in English. The linguistic politics of the region offers no uncontroversial option as a national substitute language, and the national infrastructure as well as the global economic bias toward English increase the demand for the medium. At the same time, English is a unique linguistic marker of class divisions and power asymmetries in the nation, as is the case in many former British colonies (Dasgupta, 1993; see also Spitulnik, 1998, for a discussion of this phenomenon in Zambia).

Few speakers of IndE still aspire to a native standard as their acrolect, as demonstrated in Sahgal (1991). The range of semi-autonomous domains of English use sketched here suggests the existence of multiple standards for IndE rather than a single scale of greater or lesser nonstandardness.[3] In addition, regional standards with distinct sets of nonstandard features also seem to exist,

https://doi.org/10.1017/S095439450113303X Published online by Cambridge University Press

partly influenced by widespread multilingualism. These two issues (domains of use and regional variation) form the basis for the intradialectal variation investigated in this article.

Early studies of IndE tend to be restricted to prescriptive or pedagogical descriptions based on deficiency- or error-oriented approaches (for a detailed discussion, see Kachru, 1992). In response to these relatively limited investigations, more recently developed functional approaches attempt to contextualize forms and patterns found in IndE within their sociolinguistic context, taking into account their functional role within a multilingual setting and their possibly innovative use rather than assuming a failure to match a target native standard or a fossilized stage of incomplete second language acquisition (Sridhar & Sridhar, 1992).

DATA AND METHODOLOGY

The debate over models for the study of non-phonological variation has taken various forms over the past few decades (cf. Cheshire, 1987; Lavandera, 1978; Romaine, 1984; Sankoff, 1973). Central to the debate is the question of referential meaning, which arises out of any attempt to extend methodologies for the study of phonological variation to syntactic variation. The fundamental problem lies in identifying structural variants with exact semantic equivalence. In this study, rather than isolating two forms, one standard and one variant, I examine the non-equivalence in meanings associated with one particular form.

In his suggestions for future research, Leitner (1991:228) observed that "past perfect might signal the notion of 'remote past' in Indian English." A corpus example is given in (1a). The spoken IndE example in (1b), from personal interview data, has the meaning: *My parents just visited; we just saw them off last week.* The present study aims to capture the attested range of meanings associated with this construction, but is not limited a priori to remote past meaning, despite Leitner's hypothesis.

(1) a. This is the second time that such an object *had been* sighted here. (KC A13 142)
b. My parents just visited. Just last week we *had seen* them off.

For the purposes of gaining an overview of pluperfect use in IndE, I divide the quantitative section of this study into two parts. In the first part, I compare the distribution of uses of *had* + V-*ed* in IndE to its distribution in two native varieties: BrE and AmE. This is a necessary initial step, as some earlier studies of syntactic variation in IndE have assumed an overly strict interpretation of how the variable in question is used in the native standard variety. Three assumptions are inherent in such studies: (1) that the proposed variable does in fact exhibit systematic variation in contrast to native varieties; (2) that the standard is not variable; and (3) that there is a fixed standard equivalent for any given form in the non-native variety. All three assumptions are problematic. An idealized reification of native and non-native varieties can misrepresent both varieties and, more

https://doi.org/10.1017/S095439450113303X Published online by Cambridge University Press

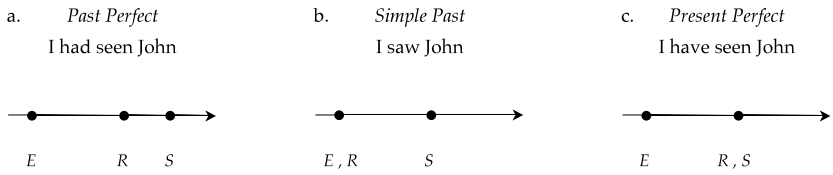


FIGURE 1. Tense and aspect time relations in English (Reichenbach, 1947).

importantly, mask the possible relationships between variation in forms and functions in both (cf. Mufwene, 1996).

With the advent of sufficiently large databases, some of these issues in the study of dialectal syntax are partially resolvable. Shastri (1996:79), in his corpus study of infinitival *to* complementation, showed that variation of the type he examined in IndE also existed to a lesser degree in native English. If non-native uptake and expansion of existing variation in native usage is an important mechanism of change, capturing such native variability is crucial to any insightful analysis of non-native usage.

The second part of the quantitative study examines intradialectal variation in IndE usage. The qualitative discussion that follows turns to the pragmatic basis of the new usage and its relation to other IndE TMA properties as well as other nonstandard TMA systems. Before reporting the results, I provide some background information on the English past perfect, the electronic corpora, and the coding criteria used in the study

The past perfect construction

The perfect construction indicates relational anteriority and relevance to a deictic zero point (Binnick, 1991:161; Bybee, Perkins, & Pagliuca, 1994:55; Comrie, 1976:53). The English past perfect construction is generally used to refer to an event that occurred before a particular point in the past. In one sense, alluded to by Comrie (1976:53), perfect aspect—whether past, present, or future—relates to two time points, whereas tenses such as past or present simply refer to one, with no implication that it bears a necessary relation to others. An intuitive visual representation of such generalizations is provided in Reichenbach (1947:290). In his discussion of systems of time reference, Reichenbach noted that in order to distinguish among types of past and present tenses one must make reference to three, not two, time points. In Figure 1, the pluperfect, preterite, and present perfect are contrasted in terms of the time points R (reference point), E (event point), and S (speech point). The distinguishing feature of the past perfect is that the R point is distinct from both the E point and the S point. This R point may be signaled to varying degrees of explicitness in the clause or the immediate context, which is a crucial issue in the interpretation of the past perfect.

https://doi.org/10.1017/S095439450113303X Published online by Cambridge University Press

In the aspectual system of English, a distinction is made between past and present perfect; as the term "past perfect" suggests, the construction might be seen as a straightforward compound of past tense and perfect aspect. Dahl (1985:144) argued against this view of the pluperfect as a compound, observing that the pluperfect "may develop secondary or extended uses which are not characteristic of the perfect by itself" (e.g., counterfactuals), and that there seem to be "less strict constraints as concerns use with definite time adverbs in the pluperfect than in the Present Perfect." He also noted that the pluperfect occurs in some TMA systems that lack a regular perfect, as in French.

(2) a. Elle *a lu* le livre.
She has read the book = 'She read the book' (simple past)
b. Elle *avait lu* le livre.
She had read the book = 'She had read the book' (pluperfect)

In such languages, the present perfect does not always convey the compound meaning of present tense and perfect aspect. Rather, the present perfect has partly or completely replaced the function of simple past. Comrie (1976) suggested that such shifts in correlation between form and function of TMA marking result from the "gradual relaxation" of such requirements as degree of recentness and presence of adverbials. It is precisely these requirements that are directly identifiable as differing in strictness in the dialects examined here, suggesting that these are some of the mechanisms of change in tense–aspect meaning.

Thus, the innovative use of the pluperfect for marking a more general remote past is a functional extension that has some crosslinguistic support. Given this typological pattern, it is important to consider the possibility of substrate language transfer influencing the direction of this development, in this case from Indo-Aryan languages. One characterization of the Hindi past perfect suggests that it expresses "the completion of an action before a definite moment [or an action] in the past, and also emphasizes the *remoteness* of the action from the present moment, its belonging to a finished segment of time—last year, yesterday, yesterday morning etc." (Katenina, 1960; my emphasis). Some Dravidian languages also show an extension of the perfect construction to past meaning. For instance, Sridhar (1990:229) observed that "the present perfect form is used in Kannada in some of those contexts where the simple past would be appropriate in some languages such as English. This leads to the very common use by Kannada speakers of English of sentences such as 'I have bought the book yesterday' for 'I bought the book yesterday'."

The typological space of possible form–meaning pairings across tense–aspect systems, particularly in substrate languages, constitutes an important backdrop for studying the emergence of nonstandard systems. However, this does not necessarily mean that the shift evident in the data here is exclusively a result of language transfer. The shift of pluperfect use for preterite meaning, to be discussed in detail, can be observed in unrelated dialects of English that have little

https://doi.org/10.1017/S095439450113303X Published online by Cambridge University Press

or no mutual contact and different substrate languages. Furthermore, the qualitative analysis presented in this article identifies some potential sources of IndE usage that derive from native English ambiguity.

The corpora

In order to establish whether the pluperfect in IndE varies systematically from native varieties, the first part of this corpus survey compares the use of *had* + V-*ed* in the press register in IndE to its use in comparable BrE and AmE press registers. Rather than assuming a priori that the use of *had* + V-*ed* deviates from native standards, this comparison is necessary to establish whether there is a difference in usage and whether any comparable variability in usage occurs within and among the native varieties themselves. The study is then extended to bureaucratic texts in the IndE corpus in order to examine register variation within IndE.

The use of written rather than spoken data raises the difficulty of a generally lower frequency of many interesting nonstandard forms. In particular, some structural variants, though very frequent in vernacular speech, are often too infrequent or unsystematic in print media to be quantifiable. However, the most important motivation for selecting a text corpus is that the collected registers represent established domains of English use in India. The analysis of speech data would be rendered problematic by the range of bilingual proficiency levels. This problem of distinguishing between limitations in second language proficiency and stable nativizations of the variety is, to a large degree, avoided by the restriction to print media.

The machine-readable corpora that are examined here are the Kolhapur Corpus of Indian English (texts from 1978), the Brown Corpus of present-day American English (texts from 1961), and the Lancaster–Oslo/Bergen (LOB) Corpus of present-day British English (texts from 1961). All three corpora were obtained from the International Computer Archive of Medieval and Modern English and were specifically selected because they are comparable in terms of design, text selection from various registers, and size.[4] The LOB Corpus was created as a BrE counterpart to the Brown Corpus, and the Kolhapur Corpus was similarly designed for comparative studies of the three varieties. All three corpora consist of roughly one million words, containing 500 texts of approximately 2,000 words each, distributed across 15 text categories.

The press register forms the basis for comparison in the first part of the study, mainly in order to maximize the range of contexts for the use of *had* with various time references but within a comparable register range. Moreover, Kachru referred to the national newspaper register as a "nativized" register (1996:22), implying that English is one of the languages associated with these domains of society for long enough to reflect a nativized version of the language.

For the purposes of interdialectal comparison, only data from the press reportage subsection (category A) of each corpus were extracted. This comprised a set of 44 texts per corpus, yielding three subcorpora of roughly 88,000 words each. Within the IndE corpus analysis, the data from the press reportage section was

https://doi.org/10.1017/S095439450113303X Published online by Cambridge University Press

TABLE 1. *Rate of occurrence of selected contracted forms in the press registers (sections A, B, and C) of three corpora*

Contracted Form	American	British	Indian
will → 'll	5.4%	2.5%	0%
have → 've	4.0%	1.9%	0.1%
not → n't	28.7%	18.2%	0%

contrasted with the bureaucratic register, composed of data taken from the H category of the Kolhapur Corpus (Miscellaneous: Government Documents, Foundation Reports, Industry Reports). This subcorpus consisted of 37 texts, containing a total of 74,000 words. While the restriction to these two registers limited the size of the corpus investigated here to much less than one million words, the number of tokens extracted and examined from each variety was substantial enough to draw robust generalizations.

Much of the recognizable "Indianness" in the Kolhapur Corpus lies in the style and formality features that have often been cited as characteristics of IndE. These include lengthy or wordy prose style, Latinization (e.g., *demise* rather than *death*), and frequent use of archaic forms and idioms and of formal style markers.

An example of stylistic formality is given in Table 1, which shows the degree to which certain forms are contracted in press reporting in the three varieties of English. IndE shows a consistently lower use of contracted forms than the native varieties; Biber (1987) and Hundt (1996) reported a similar difference between BrE and AmE contraction in written registers. Many of the contracted forms in native press reporting occur within direct quotes. IndE appears to paraphrase direct quotes much more than either AmE or BrE, which may account for the lack (or avoidance) of contracted forms. Table 2 shows another commonly noted style marker: IndE uses longer sentences in all registers without exception, even given comparable genre differences. The pattern in Tables 1 and 2, in which AmE and IndE represent two distinct patterns with respect to BrE, is recapitulated in most of the results in the present study.

Although these examples indicate a certain formality in IndE deriving from an earlier high prestige variety of BrE, the corpus also shows clear markers of indigenization. For example, for the same amount of text, the national press data contains more non-English, borrowed lexical forms (170) than the regional press data (159), although the range of lexical items is greater in the latter. Thus, even the relatively standard and formal press register is firmly indigenized in certain respects. In summary, then, the three corpora are very similar in terms of size and types of texts. While the Kolhapur Corpus resembles the native corpora in structure and content, it exhibits a number of distinguishing structural characteristics as well, deriving from a high prestige style as well as from processes of indigenization.

https://doi.org/10.1017/S095439450113303X Published online by Cambridge University Press

TABLE 2. *Average sentence length (in words per sentence) in three corpora*

Register	American	British	Indian
Press (reportage)	7.9	8.7	10.6
Press (editorial)	8.1	9.2	10.8
Skills, trades, hobbies	8.3	9.9	11.1
Scholarly/scientific writing	8.3	9.8	10.3
Detective fiction	8.8	10.1	11.5
Humor	8.7	10.3	11.2

Coding criteria and variables

In examining the use of *had* + V-*ed* across the three corpora, my main interest lay in extracting the range of variability in meanings associated with use of the standard pluperfect construction. In other words, I was interested in the association of nonstandard meanings with a standard form. To this end, I coded for the semantic function of *had* + V-*ed* and also for various internal and external independent variables. In Shastri's (1988:40) terms, this identified "opaque" rather than "transparent" features. Transparent features, such as borrowings or code-mixing in IndE, clearly signal the presence of non-nativeness. By contrast, opaque features are not readily observable; Shastri suggested that for opaque syntactic, semantic, and pragmatic features "it is perhaps not the *form* that is at variance but the function" (1992:274).

In order to delimit the appropriate dataset, all tokens of *had* + V-*ed* were extracted, leaving out all non-pluperfect uses of the form *had*. This includes the following uses of *had*: obligative (*I had to be at the airport on time*), possessive/stative (*They had two seconds remaining on the clock*), causative (*She had the offending item removed*), and counterfactual/subjunctive (*If I had known, I would have left sooner*) uses. Of these, only counterfactuals potentially relate to the use of the pluperfect. For instance, in AmE, the general reduction in the use of the pluperfect construction, as observed in Elsness (1997:229), may be accompanied by an increase in use of the simple past form in the protasis of counterfactuals. However, an exploratory comparison of counterfactuals in the three corpora did not reveal any noteworthy patterns. The incidence of counterfactuals for the press subsections A, B, and C in the three corpora was low and not notably different. An additional complication in including them is that they do not share the temporal frame of the pluperfect; their discourse context, tense environment, and time adverbial requirements—features that are coded for each pluperfect example in the data here—are distinct. While it may be fruitful to conduct a close analysis of counterfactual usage in the three varieties, the discussion of its role in the changing uses of *had* + V-*ed* is restricted here to the qualitative discussion of IndE TMA features.

https://doi.org/10.1017/S095439450113303X Published online by Cambridge University Press

Each remaining instance of *had* + V-*ed* was examined in its textual context to ascertain its intended meaning in that context. This judgment was more subjective than the usual coding of structural variables, which can be objectively classified. However, since the variation in question is in meaning, not form, this approach was necessary (see Harris, 1996:46, for a discussion of methodological issues, such as the need for subjective coding and the infrequency of tokens, in the study of syntactic variation).[5]

Judging the context of each token, regardless of whether it occurred in a native or non-native text, allowed for variability in use within the native varieties as well. As it turned out, very few tokens of the native variety occurred in contexts where a pluperfect meaning was not inferable. The fact that those few do occur in native usage, however, is important to the study and is discussed in the analysis. The three standard English meaning equivalents found with the *had* + V-*ed* construction are listed in (3).

(3) a. pluperfect (standard meaning)
b. present perfect
c. preterite

(3a) is the standard native meaning. The other two variants are other meanings found to be associated with the *had* + V-*ed* construction in the data.[6] The method of establishing the meaning of the *had* + V-*ed* construction in each example involved a close examination of the discourse context in which it occurred for a clear indication of a distinct R point and E point to license the pluperfect meaning. For example, in the standard use in (4), the E point is the taking up of reins by General Rafael. The announcement of the arrests represents an intervening R point between the past E point and the present S point: namely, the report itself. Finally, the R point is embedded within the adverbial clause *a few hours after* . . . , which unambiguously distinguishes the two points (E, R) in past time.

(4) *had* + V-*ed* with standard pluperfect meaning:
The arrests were announced a few hours after Trujillo's son, Gen. Rafael, *had taken* up the reins of power his father held for more than 30 years. (LOB A31 17)

Examples of each of the two types of nonstandard uses of the pluperfect construction found in the Kolhapur Corpus are given in (5).

(5) *had* + V-*ed* with present perfect meaning:
a. Meanwhile there was rain in Chikmagalur since morning today and reports about rain *had been received* from some other areas also. Chikmagalur goes to the polls tomorrow to decide the political future of Mrs. Indira Gandhi. (KC A43 157)
b. This afternoon fresh warnings were flashed to West Delhi areas where the New Friends Colony and Maharani Bagh area lie right in the way of the coming waters. Never before in the Capital's history these colonies *had faced* such a flood threat. The Okhla industrial complex in this sector is deserted. (KC A1 10)

https://doi.org/10.1017/S095439450113303X Published online by Cambridge University Press

had + V-*ed* with past meaning:

c. "India's struggle for freedom was led since 1920 by Mahatma Gandhi who believed in non-violence," said the document. Apart from this single instance Gandhiji did not find a place elsewhere. The capsule *had been silent* about the role of the early pioneers in the freedom struggle … The document only mentions the names of … (KC A34 106)

The main feature of the excerpts in (5) is that there is no evidence of a R point in the past that is distinct from the E point being alluded to as well as from the present S point. For example, in (5b), the clauses surrounding the pluperfect construction describe the approaching floods in present tense, and the past E point (of never before having faced such a threat) only lies anterior to the present R point, not to an intervening past one. Similarly, in (5c), no R point intervenes between the present S point and the writing of the time capsule (in fact, in this particular use, *had* + V-*ed* can even be replaced by the present tense).

In the presentation of the data here, nonstandard uses of *had* + V-*ed* with either present perfect or preterite meanings are generally grouped together as "other" in opposition to "standard" meaning, as it was unclear in a few cases of nonstandard use whether present perfect or preterite was more appropriate in the given context.

Four contextual cues were also coded for each example in order to examine the degree to which each variety explicitly marks the relative anteriority of pluperfect contexts. The few instances of present perfect are classified under "present" in (6b) and (6c). "Preceding" and "following" are interpreted linearly; for instance, an embedded clause in the preceding sentence counts as the closest preceding clause. These internal constraints may collectively convey the pluperfect meaning but are not obligatory and therefore reflect degrees of variability in style and inferable discourse meaning.

(6) a. disambiguating adverbial (present/absent)
b. preceding tense (pluperfect, past, present, future)
c. following tense (pluperfect, past, present, future)
d. reported speech verb (yes/no)

The category of disambiguating adverbials includes time-marking adverbs (*earlier*, *afterwards*, *already*, *later*), as in (7a), and time-marking adverbial clauses, as in (7b).

(7) a. *Five minutes earlier* the London Scottish defence *had* been split asunder. (LOB A23 198)
a′. Compare to: **Five minutes earlier* the London Scottish defence *has* been split asunder.
b. There were also brief discussions on Laos, Berlin and other foreign questions, *after* Mr. Kennedy *had* informed Mr. Macmillan of his discussions with Mr. Krushschov. (LOB A04 20)
b′. Compare to: * … *after* Mr. Kennedy *has* informed …

https://doi.org/10.1017/S095439450113303X Published online by Cambridge University Press

Adverbial phrases able to signal any type of past event, such as *over the past ten years* in (8), were not counted as disambiguating adverbials.

(8) a. He complained that the gap between earnings of manual workers in local government, compared with the average in all industries for men over 21, *had* widened *over the past ten years*. (LOB A04 170)
b. Compare to: . . . *has* widened *over the past ten years* . . .

Since time adverbials only optionally accompany pluperfects, they can act as a measure of explicit signaling of the pluperfect meaning as opposed to inferable discourse meaning, which may be subject to variable interpretation, allowing changes in usage. The tense of both preceding and following clauses also provides contextual, but again not deterministic, cues to R and E points (Comrie, 1985:65; Rickford & Rafal, 1996:228); these were also included as independent, internal variables. Finally, reported or quoted speech seemed to favor the choice of past perfect over present perfect. The examples from the three corpora in (9) show clauses in which the reporting of an event is the main trigger for shifting the R point to the (already past) moment of reporting by the original speaker. In all three examples, the state of affairs described in the pluperfect clause is still in effect at the time of the article.

(9) a. The committee *noted* that the probe into elementary watershed basins *had not so far been* given the importance it deserves. (KC A38 119)
b. The conservation leader *said* other mistakes in spraying *had caused* serious damage in Ohio and Wyoming. (Brown A25 11)
c. Coun. Collinghan *said* there *had been* an improvement on the square, where the buses were now parking at the rear instead of in front of the library. (LOB A43 116)

If reported speech had been found to categorically require *had* + V-*ed* across the corpora, the relevant data would have had to be omitted as reflecting a categorical internal constraint. However, reported speech also occurs with present perfect, as in (10), indicating that it is a constraining but not strictly determining factor. Thus, reported speech was included as another internal constraint.

(10) Mr. Gajendra Prasad Himansu, State Minister for Irrigation, while replying to Mr. Pasupati Singh *said* that the Chief Engineer *has been directed* to examine and to take necessary action on the report. (KC A35 48)

In summary, the coding of the meaning of each *had* + V-*ed* form is based on its textual context and temporal frame, with the internal variables contributing additional information regarding how this context is signaled.[7]

The external variables in (11) were included in order to examine variation within the Kolhapur Corpus; they constitute the second part of the quantitative study, which examines variation within IndE.[8]

https://doi.org/10.1017/S095439450113303X Published online by Cambridge University Press

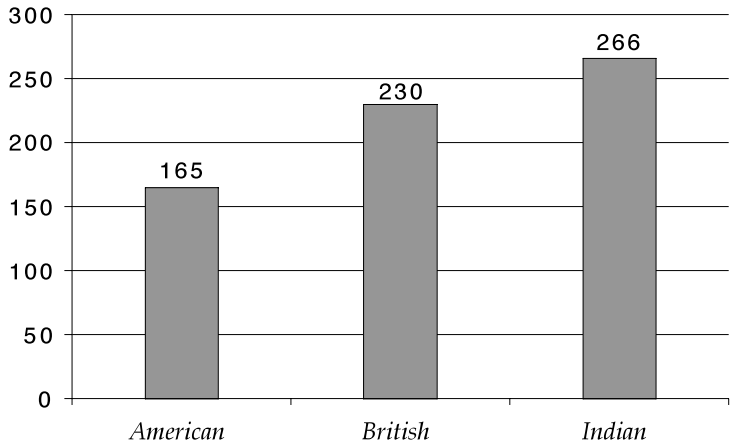


FIGURE 2. Total number of tokens of *had* + V-*ed* in the three corpora.

(11) a. corpus (American, British, Indian)
b. regionality (national, regional)
c. register (press, bureaucracy)

QUANTITATIVE RESULTS

Intervarietal comparison

Figure 2 shows the raw totals of *had* + V-*ed* identified in the three sets of press reportage data, including both standard (pluperfect) and nonstandard uses. As these figures hold for equivalent amounts of text in each corpus (88,000 words per corpus), we can conclude from Figure 2 that the IndE press corpus shows a higher total use of *had* + V-*ed* than the native press corpora, and that the AmE corpus shows the lowest of the three. A lower use of both present and past perfect in AmE in contrast to BrE has been observed in other studies as well (Dusková, 1976:59; Elsness, 1997; Harris, 1996:38).

The graphs in Figures 3 through 6 are in the form of percentages. For the analysis of internal variables, 100 tokens from the two native varieties were coded and almost all tokens in the IndE press corpus were coded (230). More tokens were coded for IndE in order to allow for more detailed analyses within different registers of the IndE corpus for the second part of the quantitative section. In the results that follow, statistical significance in each case was measured non-parametrically by performing a chi-square test on the data in question. The results of the significance tests are included below each figure or table.

Figure 3 reflects the meanings associated with each token; these were coded, as mentioned earlier, based on the E, R, and S points evident in the discourse context. The distribution indicates that the higher proportion of *had* + V-*ed* in the IndE corpus in Figure 2 is distributed over a wider range of meanings. The "other"

https://doi.org/10.1017/S095439450113303X Published online by Cambridge University Press

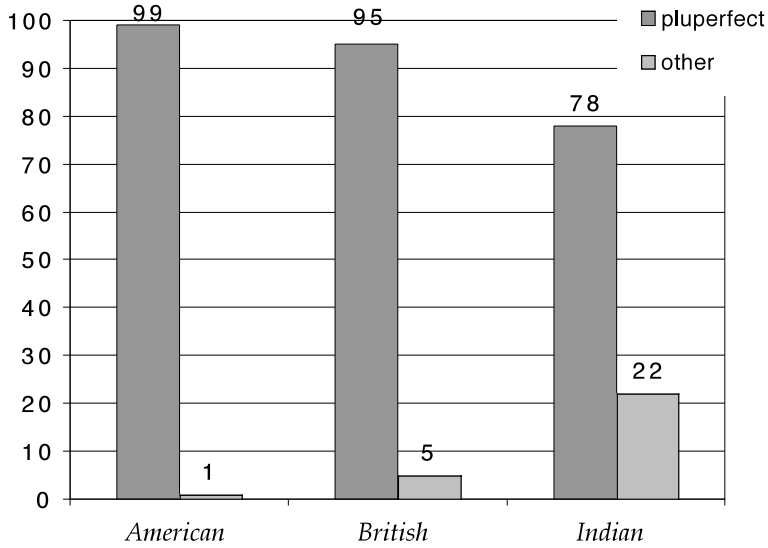


FIGURE 3. Meanings associated with *had* + V-*ed* in percent (IndE vs. AmE/BrE, $p \leq .001$; BrE vs. AmE, not significant).

category for IndE subsumes 51 nonstandard uses, 28 (12%) tokens with present perfect meaning and 23 (10%) with preterite meaning. The difference between the IndE data and each of the native varieties is significant, whereas the slight difference between the native varieties is not. Only 6 tokens in the other corpora appeared in contexts where it was impossible to establish a distinct E point and R point for the context of use. However, even this small number indicates crucially that some ambiguity in pluperfect meaning can arise in native varieties. Two of these examples from the BrE corpus are given in (12).

(12) a. (Report on the newly elected Mayor of Huddersfield) Clr. Brook began his education in two of the local schools. As a boy and a man his interests *had covered* a very wide field … His work on the council over the many years of his service *had been* outstanding in more than one respect. (LOB A30 31)

b. Only a few hours after Mr. Lloyd and his 24-strong delegation landed at Accra this morning, hundreds of shop assistants demonstrated outside the British-owned Kingsway Stores, the largest in town. The stores *had been hit* by the same strike wave that has paralysed the port of Takoradi for the past week. Roots of the discontent: The Austerity Budget, including a compulsory savings scheme … (LOB A2 61)

The underlying factors motivating these instances of ambiguous usage in BrE are examined in the qualitative part of the analysis as a possible source of the non-native usage patterns.

https://doi.org/10.1017/S095439450113303X Published online by Cambridge University Press

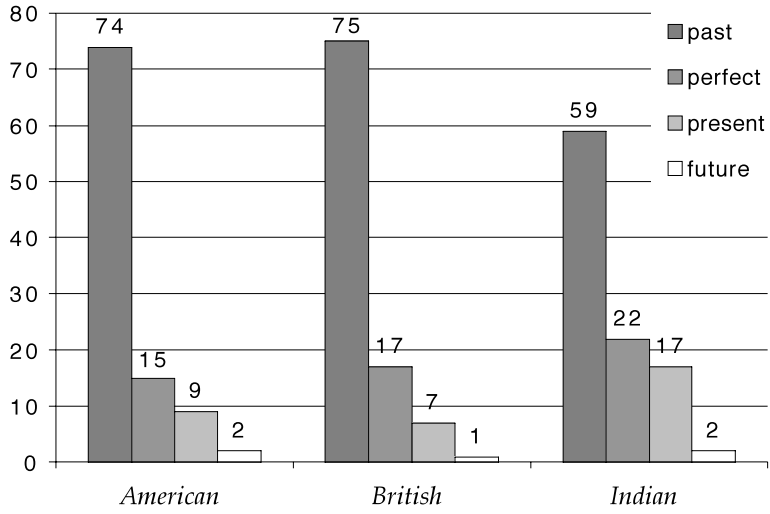


FIGURE 4. Preceding tense with *had* + V-*ed* in percent (IndE vs. AmE/BrE, $p \leq .01$; BrE vs. AmE, not significant).

In order to examine possible factors in the interdialectal differences in Figures 2 and 3, we can examine the contextual factors that were listed in (6). Figures 4 and 5 summarize the frequency with which different tenses precede and follow all coded instances of *had* + V-*ed*. In the native varieties, the past tense context is the most felicitous for a further (past perfect) reference to more specific past tense events. In fact, the nearly identical distribution of tense frequencies in the BrE and AmE corpora in both figures is remarkable. Although past tense is the most common tense environment for *had* + V-*ed* in IndE as well, other tenses are more frequently found adjacent to these clauses. In particular, present tense is more frequent in both preceding and following environments. These tense differences relate to the use of the *had* + V-*ed* construction with alternative meanings, as a preterite or present perfect meaning for *had* + V-*ed* would not require a preset past tense environment as stringently, if at all. Support for this comes from the additional finding that 71% (10/14) of the tokens in IndE that occur with present or future tense in both their preceding and following environments are used with a nonstandard meaning. Some of these tense combinations, in fact, never occur in the BrE or AmE data: for instance, future-pluperfect-future. Again, the contrast between the IndE distribution and the native English distributions of tense environments is statistically significant in both Figures 4 and 5, while the difference between the native varieties is not. In terms of discourse organization and sequence-of-tense phenomena, preceding tense environment (Figure 4) is the more relevant of the two for conditioning pluperfect use.

The presence or absence of explicitly disambiguating adverbs or adverbial clauses, as illustrated in the examples in (7) and (8), was coded for every token to

https://doi.org/10.1017/S095439450113303X Published online by Cambridge University Press

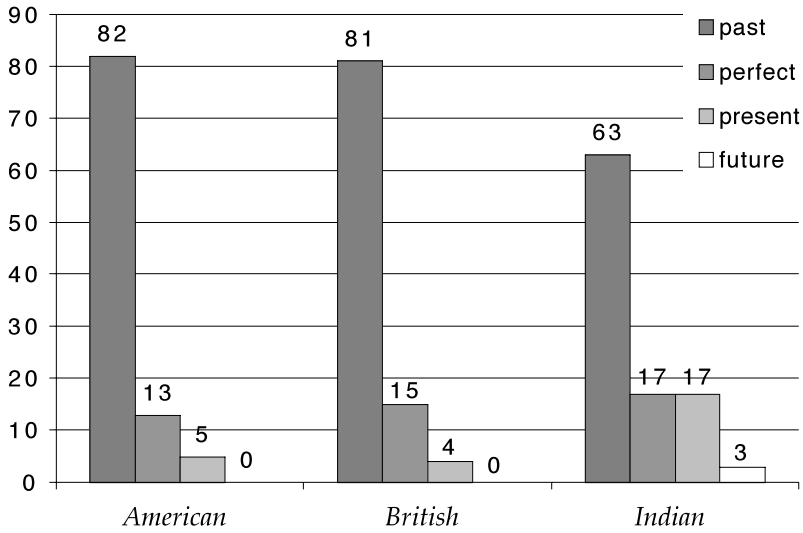


FIGURE 5. Following tense with *had* + V-*ed* in percent (IndE vs. AmE/BrE, $p \leq .01$; BrE vs. AmE, not significant).

TABLE 3. *Presence of disambiguating adverbials in* had + *V*-ed *clauses in three corpora (*N *in parentheses)*

Disambiguating Adverbial	American	British	Indian
Present	59% (59)	36% (36)	30% (70)
Absent	41% (41)	64% (64)	70% (160)

Note: AmE vs. BrE/IndE, $p \leq .001$; IndE vs. BrE, not significant.

quantify the explicit signaling of pluperfect meaning. Disambiguating adverbials were found to be most frequent in the AmE corpus (59%), significantly less common in the BrE corpus (36%), and least common in the IndE corpus (30%), as shown in Table 3.

The importance of time adverbials as well as adjacent tense in the pragmatic interpretation of aspect meanings had been commonly noted (Comrie, 1985:65; Winford, 1993:151), and Crystal (1966:7) suggested that "time relations in English are handled more by the careful use of adverbials … than by any other means." Consequently, the general absence of adverbial marking may facilitate the reinterpretation of pragmatic meanings of aspectual constructions. Given the

https://doi.org/10.1017/S095439450113303X Published online by Cambridge University Press

TABLE 4. *Frequency of occurrence of* had + *V*-ed *in reported speech in three corpora (*N *in parentheses)*

Reported Speech	American	British	Indian
Reported	39% (39)	41% (41)	44% (102)
Not reported	61% (61)	59% (59)	56% (128)

data here, this absence may permit an expansion of the use of the pluperfect in both BrE and IndE. BrE shows a higher rate of pluperfect use, but with only slightly more nonstandard meanings. The significantly greater degree of change in function in IndE may be partly explained by the compounding effect of its status as a non-native language in contact with the more pastlike pluperfect constructions found in Indo-Aryan and Dravidian systems, as described earlier. Thus, stylistic variation such as the use of time adverbials can create environments that allow substrate-based functional reanalysis.

Finally, the occurrence of *had* + V-*ed* in reported speech was also coded, as this seemed to be an important constraint on the use of pluperfect rather than present perfect or past. The systematic divergence of AmE and IndE, already observed, is again present here, but not to a significant degree.

Table 4 shows that IndE, which does not use tense or adverbial signaling as strictly as the other two varieties, has the highest proportion of pluperfect usage in reported speech contexts. This fact is only a tendency in statistical terms. However, the slight difference may relate to a feature mentioned in connection with Table 1: that is, IndE press reporting avoids direct quotations, as compared to AmE and BrE. This leads to more reported speech contexts, allowing for a higher rate of pluperfect use. Also, IndE shows a pattern of maintaining the reported speech pluperfect over several sentences beyond the original reporting verb, as a stylistic device. A reported speech predicate introduces an embedded viewpoint, after which that viewpoint, rather than the point of view of the author, is maintained by the pluperfect.

(13) J. R. Jayawardene told the Sri Lanka Parliament the day the new Constitution came into force earlier this month that it *had enlarged* the people's democratic freedom. It *had extended* their sovereignty to include the process of government, Fundamental Rights and the franchise. It *had ensured* that the President and Parliament did not extend their terms without public approval in a referendum. Prolonged rule through Emergency regulations, as in the past, was out. Parliament must now debate and vote on every Public Security Ordinance and, after three months, it must be passed by a two-thirds majority. Fundamental Rights *had been made* justiciable, there was a provision for an Ombudsman (Parliamentary Commissioner for Administration) and the method of voting by Proportional Representation *had been introduced*. (KC A 22–13)

https://doi.org/10.1017/S095439450113303X Published online by Cambridge University Press

In the IndE example in (13), the quotative voice of the initial reported speaker is maintained over several clauses, even with a break in tense type in between. This usage differs from native usage only in degree, however; as (14) shows, the pluperfect may occur in a sentence subsequent to the quotative sentence in BrE as well.

(14) Mr. Healey denounced the Government for using double standards. Ministers *had rebuked* the unaligned nations for not condemning the new Russian tests, but they themselves had no condemnation for events in Angola or Algeria. (LOB A5 196)

Though intraclausal and crossclausal quotative speech were not distinguished in the quantitative analysis, the latter appears to be relatively rare in AmE. Thus, although the results for reported speech are not statistically significant, stylistic differences do exist between the varieties, which can create or inhibit new contexts for the use of the pluperfect form.

In sum, the pattern that emerges across the internal variables is that AmE and BrE are almost identical in their tense contexts for the pluperfect, but diverge in the use of time adverbials. With regard to time adverbials and, to a lesser degree, reported speech, BrE lies between AmE and IndE, while tense ordering is almost identical for BrE and AmE. More specifically, IndE extends the pluperfect construction to both the shared aspectual meaning of present perfect and the shared tense meaning of simple past. This "leakage" in functions correlates with a less stringent set of restrictions in IndE on contextual cueing of pluperfect meaning via time adverbials. By contrast, AmE seems to have restricted the use of the pluperfect further than BrE, exhibiting stricter requirements for explicit marking and a generally lower overall use of the construction.[9]

The next section presents the second set of quantitative results from the corpus survey, examining variation in pluperfect use within IndE, and follows with a qualitative discussion of the discourse basis of this change and of its relation to other TMA phenomena in IndE and other nonstandard varieties.

Intravarietal variation: Pluperfect use within the IndE corpus

Two external variables were included in the coding of the corpus data in order to allow a quantitative examination of variation within IndE. The distribution of a variable can provide some indication of the relevant dimensions of variation in a speech community, an issue that is of particular interest in emerging varieties of English. The two intravariety factors discussed here are regionality and register.

All the IndE texts, in both press and bureaucratic registers, are classified in the Kolhapur Corpus as either national or regional. Examples of national texts are central government documents or major city newspapers with a national distribution. Regional texts are, for example, local newspapers or state government documents. The data in Table 5 show a statistically significant difference between national and regional usage in IndE. This difference is noteworthy, as it reflects a lectal range that seems to be sensitive to geographic location relative to centralized, urban standards. Given the multilingual landscape and high register of

https://doi.org/10.1017/S095439450113303X Published online by Cambridge University Press

TABLE 5. *Usage of* had+*V*-ed *in IndE according to regionality (*N *in parentheses)*

Geographic Range	Standard Meaning	Other Meaning
National texts	83% (100)	17% (20)
Regional texts	72% (79)	26% (31)

Note: $p \leq .05$.

English in this spectrum, these results support the view that IndE standardness is higher in urban areas. All the same, the occurrence of nonstandard *had* + V-*ed* in the national (i.e., centralized, urban) text sources (press and government) indicates that the feature is not entirely absent in mainstream usage.

The second external variable that was contrasted was register. I compared the press register to the register termed "miscellaneous" in the corpus, which I classify here as bureaucratic documents. This register consists of texts from the H category of the Kolhapur Corpus, which includes central and state government documents, foundation reports, and industry reports.

The selection of these two registers to explore intradialectal variation is based on the historical establishment of domains of English use in India, discussed earlier. Kachru (1983:26) cited the civil services as possibly playing a significant role in the establishment of a native standard in India, as administration is one of the oldest and most widespread institutions of English use both before and, even more expansively, after independence. Unlike the national press, the national administrative or civil services extend to all corners of the country and across classes, thereby integrating a very disparate population into a single institutional language code. For this reason, while government documents in native English-speaking nations may represent a high level of formality, the Indian bureaucratic register in fact reflects a potentially more indigenized standard than, say, that of the national English language media. By contrast, the national English press in India tends to have closer professional ties with international English media and may also derive its standard from native varieties of journalistic prose.

Given these differences in the function of English in the two registers, we would expect print media to show fewer nonstandard features than government and other bureaucratic registers. All the same, English plays a central role in both government and press in India, and these can be considered two of the most common domains of regular IndE usage. By contrast, the national/regional distinction in Table 5 corresponds to greater and lesser degrees of overall English use, and therefore the discrepancy there might be expected to be higher.

Table 6 shows the distribution of *had* + V-*ed* usage across the press and bureaucratic registers; Figure 6 shows a more detailed cross-classification of both the external variables.[10] The difference in percentage of nonstandard uses of *had* +

https://doi.org/10.1017/S095439450113303X Published online by Cambridge University Press

TABLE 6. *Usage of* had+*V*-ed *in IndE according to register (*N *in parentheses)*

Register	Standard Meaning	Other Meaning
Press texts	79% (154)	19% (42)
Bureaucratic texts	73% (25)	27% (9)

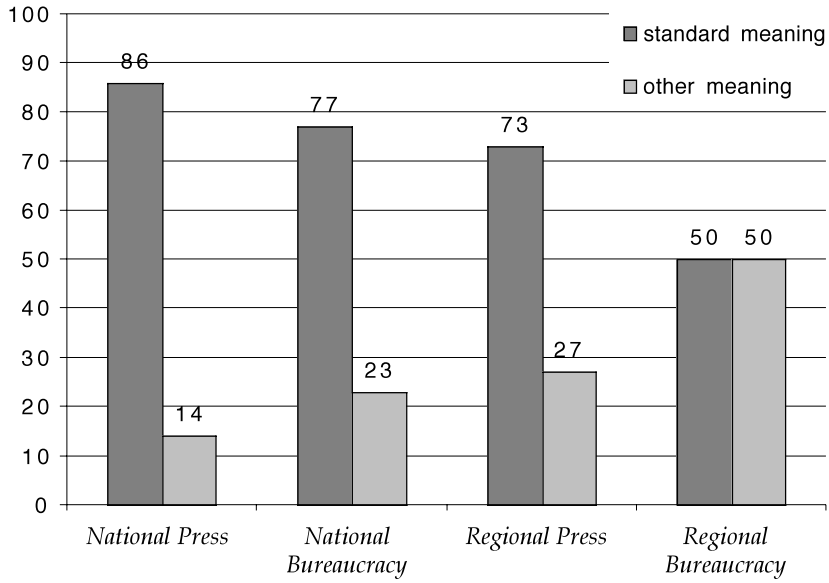


FIGURE 6. Cross-classification of IndE external variables in percent.

V-*ed* in Table 6 supports the prediction that administrative bureaucracy might show greater nonstandardness due to its extension across geographic and socioeconomic divisions. In the native English registers, on the other hand, bureaucratic texts fall high on a scale of standardness and formality, perhaps even more so than press. This difference suggests that the relation of standardness and formality in text registers cannot be assumed to be universal, but may instead be subject to considerable sociocultural determination. In other words, degrees of nonstandardness may not always correlate directly with a universal stylistic cline of formality, as is assumed in many studies of social and stylistic variation (cf. Finegan & Biber, 1994:317).

https://doi.org/10.1017/S095439450113303X Published online by Cambridge University Press

QUALITATIVE DISCUSSION

Pragmatics of had + *V*-ed *usage in IndE*

The actual tense and aspect contexts in which the nonstandard uses of the pluperfect construction occur deserve closer examination. Is a new, specialized semantic meaning associated with these uses, as was found for Trinidadian English in Winford (1993), or is the pluperfect being licensed more freely in the IndE pragmatic or discourse framework? As I show in this section, the semantics and pragmatics of the nonstandard uses supports the latter hypothesis.

At first glance, the 51 nonstandard uses of the *had* + V-*ed* construction that were found in this study seem to fall into a number of groupings, listed in (15). The examples are given without context and are intended only to illustrate the types of clauses that were classified into different groups. For all groups, an intervening R point is not distinct in the discourse context.

(15) a. Sustained state of affairs: these uses often occur with explicit extended time adverbials more commonly associated with present perfect. In all cases, the state of affairs is still in effect.
   i. Similar concessions *had been* in force for years in the southern States … (KC A13 142)
   ii. Politics in Bihar, for decades, *had been* caste-ridden … (KC A1 35)

b. Covertly reported speech: these tokens occur without a quoted speaker but with a remotely inferable reported speech context.
   i. The agitators … *had also disrupted* road traffic. (KC A34 14)
   ii. Good care *had been taken* to use the leverage of canalisation to promote growth. (KC A36 152)

c. Remote past: the E point in these pluperfect uses is located in the remote past. Many of these examples refer explicitly to a remote time point.
   i. In the past, there *had been* criticism in the J.P.P. Executive Committee over the issue … (KC A34 67)
   ii. The National Cooperative Development Corporation *had contributed* Rs.35.45 lakhs during 1976–7. (KC H4 108)

d. Relative to immediate prior description: a previously described event becomes the R point for a subsequent pluperfect clause. Many examples include a reference to the prior description (e.g., *such an object*).
   i. This is the second time that such an object *had been sighted* here. (KC A13 144)
   ii. Rarely *had* a Ranji final *taken* such a course on the opening day itself. (KC A32 16)
   iii. Never before in the Capital's history these colonies *had faced* such a flood threat. (KC A37 177)

Unlike the data in Winford's (1993) study of *have* in Trinidadian English, the IndE data cannot be explained in terms of the four commonly noted meanings of the perfect: result/stative, experiential/existential, persistent situation, and recent past/hot news (Comrie, 1976:57). No single, semantic meaning can be attributed to all the uses in (15) without distorting the actual range of contexts in

https://doi.org/10.1017/S095439450113303X Published online by Cambridge University Press

which the construction appears. In connection with (15c), Leitner's (1991:228) suggestion, mentioned earlier, that pluperfect is used for remote past in IndE turns out to be partially true, but applies only to a subset of all the uses in the data.

Rather than composing a heterogeneous set of unclear uses or failures to match a native standard, I propose that these nonstandard uses found in IndE can be unified into a regular pattern of pragmatic viewpoint marking. Two examples are discussed in (16) and (17).

(16) One of the scientists, Dr P. S. Sehra, said it was a strange luminous object moving from north-west to south-east at 9.12 p.m. The object, sighted from the P.R.L. campus, had a glowing double head and a long, narrow tail. It was seen for about 10 to 15 seconds;
  a. the object *was* "bluish yellow" in colour and the elevation was about 70,
  b. he *added*.
  c. This *is* the second time
  d. that such an object *had been* sighted here.
  e. The first *was* sighted on April 3. Professor D. Lal, director of the Physical Research Laboratory, and his colleagues, Dr J. N. Desai and Dr J. N. Goswami, have came to the conclusion that . . . (KC A13 138–144)

The two past tense clauses in (16a) and (16b) refer to two different E points, one at which a peculiar object was sighted and one at which this sighting was reported by Dr. Sehra to the current speaker/reporter. In (16c), present tense relates the sighting of the object to the present time (S). However, in (16d), rather than continuing to relate the first sighting to the present time (S) knowledge, it is shifted relative to Dr. Sehra's second sighting. This pseudo-R point is the main focus of the report and disrupts the maintenance of present relevance in the reporting. This shift to an apparent R point, instead of maintaining a fixed relevance to the present time (S), is comparable to the pattern in (17).

(17) a. This Bill *is* very timely.
  b. I hope the cultivators *will be benefitted* after passing this Bill to a greater extent.
  c. Other Bills like the Manipur Town and Country Planning Bill *had also been passed* for raising the standard of living of the people in the State.
  d. I *hope* the Bills which are under consideration in the House today *will be passed* for the benefit of the people. (KC M31 39–44)

This example shows an even more radical shift: that is, from future to pluperfect and back to future (with a present quotative verb). Not a single instance of this tense ordering appears in the BrE or AmE corpus data. In (17a), the debate over a bill is the E point or state of affairs being described. In (17b), a future tense reference is made to projected benefits. When a remote past reference is introduced in (17c), the original E point and the focus of the narrative (debating the bill) becomes a pseudo-R point. Note that the passing of earlier bills is still in effect at the present time (S) and would usually be referred to with present relevance (i.e., with present perfect). Thus, as in (16d), it is the anterior E in (17c)

https://doi.org/10.1017/S095439450113303X Published online by Cambridge University Press

that is introduced as being relevant to the previous, focal E rather than the current S, as might be expected in standard English reportage.

The primary effect of the *had* + V-*ed* construction in these and most of the other nonstandard examples is to distance a completed, past event from the narrative focus. This permits a wider range of discourse uses than the standard, native English pluperfect. English conforms to a general principle of permanence of the R point, which eliminates such tense combinations as (18).

(18) *I had mailed the letter when John has come

This principle very broadly states that, in a coherent segment of discourse, R is maintained and E and S may shift relative to it (Reichenbach, 1947:293). However, shifting of the R point does occur under certain circumstances in native standard varieties as well, as in the case of indirect speech in subordinate clauses. Thus, (19b) can be used to express the meaning of (19a).

(19) a. Sue thought, "Max has left."
b. Sue thought that Max had left.

For this reason, the simple principle that R must be maintained is too strong to allow for the range of observable sequence-of-tense phenomena. In fact, a range of contextual factors (e.g., temporal adverbs, telicity of the verb, stativity, and result entailments) determines this phenomena and other characteristics of the perfect (cf. the XN 'extended now' theory of McCoard, 1978:151). The quantitative data in this study have already shown that sequence-of-tense considerations and adverbs play some part in the use of the pluperfect form. Thus, R-shifting is an inherent component of native English but may be subject to variable usage. The BrE example from (12b), repeated in (20), demonstrates this variability in native usage. Here, the pluperfect form may be associated with present perfect meaning.

(20) Only a few hours after Mr. Lloyd and his 24-strong delegation landed at Accra this morning, hundreds of shop assistants demonstrated outside the British-owned Kingsway Stores, the largest in town. The stores *had been hit* by the same strike wave that has paralysed the port of Takoradi for the past week. Roots of the discontent: The Austerity Budget, including a compulsory savings scheme … (LOB A2 61)

The use of the pluperfect in (20) is ambiguous between present perfect and past perfect meanings. Either the demonstration of shop assistants or the arrival of Mr. Lloyd introduces a new R point, licensing the pluperfect shift in tense viewpoint so that the expression *the same strike wave* refers to the time period prior to that. However, the strike wave is not clearly restricted to that time period and continues to have present relevance, permitting a present perfect reading. This ambiguity in the meaning of the pluperfect has been previ-

https://doi.org/10.1017/S095439450113303X Published online by Cambridge University Press

ously noted in Tichy (1980:361); it can either be equivalent to a past-in-the-past or a present-perfect-in-the-past.

Now, rather than seeing the IndE examples as entirely novel, we can consider their salient R-shifted characteristic as extending out of an ambiguous usage already present in native standard English. This extension may be further reinforced by substrate language systems, such as Hindi and Kannada, as mentioned earlier. If this view is correct, it demonstrates the importance of taking into account variability and ambiguity in the lexifier or native variety itself when considering possible sources of new form–function relationships in a new variety (Mufwene, 1996).

Related changes within the TMA system of IndE

The functional reallocation of the *had* + V-*ed* construction must be considered within the context of other forms and meanings in the IndE TMA system. Harris (1996:32) made this argument in his study of the Hiberno English TMA system:

> Of necessity the [sociolinguistic variable] model encourages an atomistic view of variation, whereby pairs of apparently alternating standard and nonstandard forms are studied in isolation from other forms in the relevant grammatical subsystem. This 'worm's eye' approach encourages the impression that differences between the standard and a particular vernacular are merely superficial and tends to obscure whatever deep-seated divergences there might exist between the two varieties. In contrast, the likelihood of radical structural differences coming to light is increased as the angle of observation is increased to encompass the wider grammatical subsystems in which the apparently alternating forms are embedded.

Winford (1993) also endorsed a systematic approach to TMA analysis along these lines, adding that functional equivalences and pragmatic categories across two dialects are crucial for a clear understanding of variation in TMA marking.

Some IndE TMA features have been investigated to some degree in the Kolhapur Corpus and in other studies as well (Katikar, 1984; Leitner, 1991; Williams, 1987). Three additional TMA features of IndE that may be reconsidered in light of the pluperfect usage discussed here are the use of the modals *could* and *would*, the present perfect construction, and the progressive.

In his study of IndE modals, Leitner (1991) primarily restricted his discussion to frequencies of forms and did not identify a noticeable development in the functions of *would* and *could* in IndE: these modals can be used with the standard meanings of *will* and *can*, respectively. The shared extension in both cases is the use of these two modals with non-conditional meanings, which may derive partly from an extension of the polite use of the subjunctive to other modal contexts in order to index politeness or formality.

The two examples in (21), personal communications from different IndE speakers, show instances of *would* being used with a simple future meaning. These modal uses are another instance of a new function being associated with a standard form.

https://doi.org/10.1017/S095439450113303X Published online by Cambridge University Press

(21) a. He *would be coming* to receive you on 15th and he *would be bringing* the other ticket too. This is because M. would have gone to Delhi by then and we just want to ensure that the ticket is not misplaced. He *would be seeing* you off too.
b. I *would be visiting* your place tomorrow . . . I *would be reaching* around 9:30am. I completed my work just now and *would be free* tomorrow.

The modals *could* and *would* in Standard English parallel uses of the pluperfect construction in certain contexts. They are used as the past equivalents of *can* and *will*, as in the reported speech examples in (22a); the equivalent alternation between simple past and pluperfect is shown in (22b).

(22) Reported speech
a. I can/will bring the book.
a′. I told her I *could*/*would* bring the book.
b. I brought the book.
b′. I told her I *had brought* the book.

Furthermore, both modals *could* and *would* are used generally in counterfactual or conditional contexts, as is *had*. This is shown in the examples in (23).

(23) Counterfactual
a. I can/will bring the book.
a′. (If) I *could*/*would* have brought the book . . .
b. I brought the book.
b′. If I *had brought* the book . . .

The IndE uses of *could* and *would* in non-irrealis contexts parallel the IndE use of *had*, which occur similarly in non-pluperfect contexts. Together, the use of *could*, *would*, and *had* in new environments forms an overlapping set of extensions in IndE that may be mutually reinforcing.

Another potentially related TMA change is a past punctual use of the present perfect. The examples in (24) show that, in IndE, the present perfect can indicate a completed past action rather than a past action with ongoing relevance in the present. (24a) is from an online discussion and (24b) is from personal interview data.

(24) a. We *have known*, four years ago, that weightlifting was going to be an Olympic sport. We *have known*, four years ago, that Malleshwari was a medal prospect–heck, four years ago, she was the world champion.[11]
b. Actually, I *have come* here about six months back.

Both examples include punctual adverbial phrases (*four years ago*, *six months back*). In general, the present perfect in Standard English cannot co-occur with punctual time adverbials (Comrie, 1976:54). Williams (1987:182) attributed this non-native use to a "lexicalization of completion," whereby the perfect form indicates completion of the activity rather than the continuing impact of a past event on the current situation. If the present perfect has come to represent lex-

https://doi.org/10.1017/S095439450113303X Published online by Cambridge University Press

icalized completion and has lost the connotation of present relevance, it begins to resemble perfective (rather than perfect) marking, which Dahl (1985:78) described as the denotation of "a single event with a well-defined result or end-state, located in the past." The perfective represents an action as a whole (sometimes considered a type of *Aktionsart*), whereas the perfect indicates relational anteriority and relevance to a deictic zero point (Binnick 1991:161; Bybee et al. 1994:55). This depiction of IndE present perfect use mirrors the innovative uses of the past perfect discussed here, as these new uses have been shown to contribute information such as remoteness and prior completion. Together, past and present perfect use in IndE begin to resemble a generalized perfective viewpoint aspect.

If this is the case, the progressive form in IndE may additionally fit into this picture, as the progressive in IndE appears often to signal imperfectivity. The use of progressive aspect with stative predicates to signal a current, ongoing state is one of the more salient and frequently noted features of IndE (Kachru, 1992; Leitner, 1991; Richter, 2000; Williams, 1987:172). An example from the Kolhapur Corpus is given in (25).

(25) [The ministers] are fully corrupted. They *are having* vindictive motives. (KC H31 1630)

Williams (1987) noted that the use of progressives with stative verbs occurs in West African varieties of English, and Harris (1984:56) cited certain Celtic dialects as also exhibiting this usage. Though it is not particularly common in the Kolhapur Corpus, as noted in Richter (2000), it can also be observed in spoken IndE usage (e.g., *We are knowing each other*, from personal interview data). The consistent use of the progressive form with the future (a nonperfective category) in (21) is also noteworthy in this respect.

Michaelis (1998:59) argued that, although the progressive is often associated with imperfectivity crosslinguistically, the standard English progressive is not a marker of imperfective viewpoint aspect but rather phasal aspect, which accounts for the ungrammaticality of stative progressives. According to her, "the viewpoint aspects perfective and imperfective are covert categories of English; that is, they are not directly encoded morphologically" (Michaelis, 1998:59); the patterns observed in IndE point to the possibility of an emergent viewpoint aspect distinction. Pending a more detailed analysis, the preliminary discussion in this section suggests that it may be possible to conceive of the variation in the pluperfect case as a part of a more systematic tendency, which privileges, among other factors, a perfective/nonperfective distinction in the new system. Naturally, these speculations regarding connections to other IndE TMA features are preliminary at this point and await a more complete investigation.

Similarities in other nonstandard TMA systems

A final consideration in the data in this study of IndE TMA forms is the existence of comparable phenomena in unrelated non-native or nonstandard varieties. In

https://doi.org/10.1017/S095439450113303X Published online by Cambridge University Press

their study of *had* + V-*ed* in narratives of African American adolescents, Rickford and Rafal (1996) identified a use of the construction that marks preterite rather than pluperfect. They cited Cukor-Avila and Bailey (1995) as referring to this construction as "innovative *had* + past," a description they extended in order to explain the usage of AAVE narratives. Rickford and Rafal depicted the innovative use as conveying narrative reorientation or signaling discourse peaks by foreshadowing complicating action or evaluative points. In comparison to the data here, which indicates a local, anteriority-marking function and a possible lexicalization of completion, the innovative functions of preterite *had* in AAVE narratives are slightly different. However, the absence of a canonical, intermediate R point and the extension to preterite meaning is comparable. The anterior, completive aspect marking of perfect constructions discussed for present and past perfect in IndE has also been observed in Philippine English and Singapore English (Platt, Weber, & Ho, 1984).

Another interesting parallel occurs in Trinidadian English.[12] Consider the similarities between the preterite use of *had* in IndE, with the possibly related shifts found in *would* and *could*, and the summary of features presented by Solomon (1972:4) in the following excerpt, which I quote in full to include the theoretical issue that he raises.

> What makes [an analysis of dialect mixture] difficult is the tacit assumption that the changes occasioned by the functional shift are changes, quantitative or qualitative, in the incidence of forms or "sets" of forms, without consideration of the cognitive value of the forms in the different dialects. To put it another way, although there is interference there is no linguistic variable that can describe the shift from English "comes" to French "vient." The shift is from a system where a certain conceptual distinction is absent to one where it is present, and the formal correlates of the concept are not amenable to meaningful manipulation …
>
> Examples are in fact abundant: in Trinidadian "English," "could" is frequently used with the meaning, approximately, of SE "can" or French Creole "pe"; "could have" is used equivalently to French Creole "te pe," roughly translatable as "could" or "was able." "Had gone" is used with the meaning of French Creole "te ale," which is not directly translatable into English but means that the event happened in the past but without any implication of recentness or relevance to the moment of speaking.

Winford (1993:172) also noted precisely these alternate meanings of *could* and *would* in Trinidadian English. Foley et al. (1998:142) observed the same extension of *could* and *would* in Singapore English, providing the example *I am sorry to tell you that I would have to turn down your application.*

Although some first language influence from Indian languages is certainly present in various structural deviations of IndE, the crossdialectal similarities presented in this section suggest that it may not be the only explanation for the TMA changes discussed here. First, TMA systems are rather diverse, particularly between Dravidian and Indo-Aryan languages, and thus may at best only reinforce or inhibit certain ongoing shifts through structural similarities. More importantly, an equivalence appears to exist between several TMA changes in IndE

https://doi.org/10.1017/S095439450113303X Published online by Cambridge University Press

and those in unrelated nonstandard varieties worldwide. Based on these observations and the discussion of native English variability, a more accurate representation of the semantic extension in pluperfect use is that universal principles of reanalysis and generalization begin to operate on inherent ambiguities in certain constructions in the native/lexifier language, accompanied by social mechanisms of spread and possibly secondary native language reinforcement.

CONCLUSIONS

This study has examined several aspects of pluperfect usage in IndE and native varieties. The non-native corpus data additionally associate present perfect and preterite meanings with the *had* + V-*ed* construction, showing significant differences from native usage. Furthermore, the frequency of cases in which pluperfect contexts are explicitly licensed by time adverbials is lower in IndE than in native varieties. Differences between AmE and BrE also emerge from the analysis of the three corpora: in particular, AmE shows a lower overall use of *had* + V-*ed* and also stricter constraints on signaling such contexts.

Variation within IndE shows a tendency for greater non-nativeness in regional texts and in the bureaucratic register. However, both registers examined and both types of press texts exhibit some degree of nonstandardness in pluperfect use, indicating that, while variation may exist within the variety, this usage is fairly widespread. The unifying characteristic among the nonstandard uses is a pragmatic distancing from a given narrative focus, resulting in a generalized marking of remoteness and completion. A comparison of the use of the pluperfect construction with other non-native characteristics of the TMA system of IndE suggests a systematic semantic shift that may cover other variants in meaning associated with *would* and *could* and present perfect *have* as well. This shift may be attributed to an emergent anterior, completive (perfective, rather than perfect) marking and a concomitant loss of the necessity for a distinct reference point in the context, allowing both present perfect and preterite meanings to be associated with all three constructions. While language transfer may play a role in reinforcing this development, ambiguity in native English usage represents an important trigger for variability and change. Finally, the brief comparison of these patterns to the TMA systems of other non-native and nonstandard varieties of English indicates a number of shared characteristics, suggesting further directions for research on possible causes of change in form–function relationships across varieties of English.

NOTES

1. The debate out of which Macaulay's *Minute* emerged has been reinterpreted over the decades, but still persists today in various aspects of language planning and education. The view that English is an inevitably indigenized part of the Indian linguistic landscape stands in contrast to one that emphasizes the need for linguistic self-determination distinct from the "colonial" language. An article posted as recently as July 1998 at http://www.indolink.com/Analysis/politics24.html refers to conflict between "hybrid Macaulyite" (implying a sort of betrayal of national interests) and "swadeshi" (approximately "self-made"; a term that was associated with the Indian independence movement)

https://doi.org/10.1017/S095439450113303X Published online by Cambridge University Press

persuasions in the country. The use of the term "Macaulyite" reflects the persistence of this event in present-day Indian consciousness.

2. The three-language formula for education has been interpreted variably due to regional differences as well as to a North–South divide in attitudes towards Hindi (a Northern, Indo-Aryan language), which predates English as a lingua franca and alternate national language. An example is Tamil Nadu, a state which witnessed language riots in the 1960s in which 70 people were killed and where Hindi and English were both temporarily banished (Kachru, 1983:90). The three-language formula is reduced to a two- or even one-language formula in such regions, for political and practical reasons.

3. See also Rickford (1987:27) on multidimensionality in Guyanese Creole.

4. Abbreviations in corpus references: Kolhapur Corpus of Indian English (KC), Lancaster–Oslo/Bergen Corpus of British English (LOB), Brown Corpus of American English (Brown). Corpus references provide the following information: (corpus name, file name, line number). For further details of these corpora, see http://www.hit.uib.no/corpora.html.

5. In order to control for subjective bias, a subset of the data was independently recoded following the same criteria by a second coder. The two sets of results matched with 94% accuracy.

6. According to Labov's original formulation of the Principle of Accountability, variants such as (3b) and (3c) should ideally be contrasted with all occurrences of regular present perfect and preterite forms as well: that is, tokens of these forms ought to be included. However, Labov later noted that this stipulation is too strict to apply to all studies of variation, as the set of possible variants for certain syntactic features cannot be clearly restricted, citing the study of invariant *be* in AAVE as one such case (1982:30). In this study as well, inclusion of all other past uses is both theoretically and methodologically intractable. I restrict the study to comparisons of sets of occurrences, listed in (3), across dialects.

7. The potential for circularity arises here, because unlike structural variables, which are more commonly studied, semantic variables may derive part of their actual meaning from some of these internal constraints. However, I include the internal variables in (6), as none of them constitutes an explicit coding criterion, and as there appears to be important variation in the use of (optional) contextual cues in different varieties. The basic semantic coding relies primarily on the greater discourse context, not on the internal variables.

8. As this study exclusively uses corpus material, the external variables are regrettably restricted to these broad categories, as detailed social correlates for tokens are irretrievable from the corpus texts.

9. In spite of the interesting patterns of deviation among varieties, it must be noted that IndE shows fairly high levels of standard usage in general. Other researchers on IndE and other indigenized Englishes have, in fact, observed that the analysis of such varieties can be quantitatively intractable due to inconsistency in levels of nonstandardness and the high degree of standardness in text registers (Leitner, 1991; Solomon, 1972). However, this tendency to standardness must also be seen as a real indication that differences between non-native varieties and native varieties often occur at the same low levels as differences among native varieties.

10. The corpus subsection H contained only a few state government documents, which is why the numbers for regional bureaucracy are rather low in Figure 6.

11. From http://www.rediff.com/sports/olyday.htm, accessed September 19, 2000.

12. I follow Solomon (1972) in using the term Trinidadian English. Winford (1993) treated the entire continuum as Trinidadian English and the creole vernacular as Trinidadian Creole.

REFERENCE

Biber, Douglas. (1987). A textual comparison of British and American writing. *American Speech* 62:99–119.

Bickerton, Derek. (1980). Decreolisation and the creole continuum. In A. Valdman & A. Highfield (Eds.), *Theoretical orientations in creole studies*. New York: Academic. 109–129.

Binnick, Robert. (1991). *Time and the verb.* Oxford: Oxford University Press.

Bybee, Joan, Perkins, R., & Pagliuca, W. (Eds.). (1994). *The evolution of grammar: Tense, aspect, and modality in the languages of the world.* Chicago: University of Chicago Press.

Cheshire, Jenny. (1987). Syntactic variation, the linguistic variable, and sociolinguistic theory. *Linguistics* 25:257–282.

Comrie, Bernard. (1976). *Aspect.* Cambridge: Cambridge University Press.

_____ (1985). *Tense.* Cambridge: Cambridge University Press.

https://doi.org/10.1017/S095439450113303X Published online by Cambridge University Press

Cukor-Avila, P., & Bailey, G. (1995). Grammaticalization in AAVE. In *Proceedings of the 21st Annual Meeting of the Berkeley Linguistics Society (BLS 21)*. Berkeley: Department of Linguistics, University of California. 401–413.
Crystal, David. (1966). Specification of English tenses. *Journal of Linguistics* 2:1–34.
Dahl, Östen. (1985). *Tense and aspect systems.* Oxford: Blackwell.
Dasgupta, Probal. (1993). *The otherness of English: India's auntie tongue syndrome.* New Delhi: Sage Publications.
Dusková, L. (1976). On some differences in the use of the perfect and preterite between British and American English. *Prague Studies in Mathematical Linguistics* 5:53–68.
Elsness, Johan. (1997). *The perfect and the preterite in contemporary and earlier English.* Berlin: Mouton de Guyter.
Finegan, Edward, & Biber, Douglas. (1994). Register and social dialect variation: An integrated approach. In D. Biber & E. Finegan (Eds.), *Sociolinguistic perspectives on register*. Oxford: Oxford University Press. 315–347.
Foley, J. A., et al. (1998). *English in new cultural contexts: Reflections from Singapore.* Oxford: Oxford University Press.
Harris, John. (1996). Syntactic variation and dialect divergence. In R. Singh (Ed.), *Towards a critical sociolinguistics*. Amsterdam: Benjamins. 31–57.
Hundt, Marianne. (1996). Has BrE been catching up with AmE over the past thirty years? In M. Ljung (Ed.), *Corpus-based studies in English. Papers from the 17th International Conference on English Language Research on Computerized Corpora (ICAME 17)*. Amsterdam: Rodolpi. 135–151.
Kachru, Braj. (1983). *The Indianization of English: The English language in India.* Oxford: Oxford University Press.
_____ (1992). *The other tongue: English across cultures.* Urbana: University of Illinois Press.
_____ (1996). South Asian English: Toward an identity in diaspora. In R. J. Baumgardner (Ed.), *South Asian English: Structure, uses and users*. Urbana: University of Illinois Press.
Katenina, T. E. (1960). *Jazyk Chindi.* Moscow: Nauka.
Katikar, P. (1984). *The meanings of modals in Indian English.* Doctoral dissertation, Shivaji University, Kolhapur.
Labov, William. (1982). Building on empirical foundations. In W. P. Lehmann & Y. Malkiel (Eds.), *Perspectives on historical linguistics*. Amsterdam: Benjamins. 79–92.
Lavandera, Beatriz R. (1978). Where does the sociolinguistic variable stop? *Language in Society* 7:171–182.
Leitner, Gerhard. (1991). The Kolhapur Corpus of Indian English–intravarietal description and/or intervarietal comparison. In S. Johansson & A-B. Stenström (Eds.), *English computer corpora: Selected papers and research guide*. New York: Mouton de Gruyter. 215–232.
Mehrotra, Raja Ram. (1998). *Indian English.* Amsterdam: Benjamins.
McCoard, Robert. (1978). *The English perfect: Tense-choice and pragmatic inferences*. Amsterdam: North-Holland.
Michaelis, Laura. (1998). *Aspectual grammar and past-time reference.* New York: Routledge.
Mufwene, Salikoko. (1996). Creolization and grammaticization: What creolistics could contribute to research on grammaticization. In P. Baker & A. Syea (Eds.), *Changing meanings, changing functions*. London: University of Westminster Press. 3–28.
Platt, John, Weber, Heidi, & Ho, Mian Lian. (1984). *The new Englishes.* London: Routledge.
Reichenbach, Hans. (1947). *Elements of symbolic logic.* New York: Macmillan.
Richter, Claudia. (2000). The progressive form in Indian, British, and American English. Master's thesis, Zurich University.
Rickford, John. (1987). *Dimensions of a creole continuum.* Stanford, CA: Stanford University Press.
Rickford, John, & Rafal, Christine Théberge. (1996). Preterite *had* + V-*ed* in the narratives of African-American adolescents. *American Speech* 71:227–254.
Romaine, Suzanne. (1984). On the problem of syntactic variation and pragmatic meaning in sociolinguistic theory. *Folia Linguistica* 18:409–437.
Sahgal, Anju. (1991). Patterns of language use in a bilingual setting in India. In J. Cheshire (Ed.), *English around the world*. Cambridge: Cambridge University Press. 299–307.
Sankoff, Gillian. (1973). Above and beyond phonology in variable rules. In C. J. Bailey & R. Shuy (Eds.), *New ways of analyzing variation in English*. Washington, DC: Georgetown University Press. 44–61.
Shastri, S. V. (1988). The Kolhapur Corpus of Indian English and work done on its basis so far. *ICAME Journal* 12:15–16.

https://doi.org/10.1017/S095439450113303X Published online by Cambridge University Press

———— (1992). Opaque and transparent features of Indian English. In G. Leitner (Ed.), *New directions in English language corpora: Methodology, results, software developments*. New York: Mouton de Gruyter. 263–275.

———— (1996). Using computer corpora in the description of language with special reference to complementation in Indian English. In R. J. Baumgardner (Ed.), *South Asian English: Structure, uses, and users*. Urbana: University of Illinois Press.

Solomon, Dennis. (1972). *Form, content, and the post-creole continuum*. Paper presented at the Conference on Creole Languages and Educational Development, University of the West Indies, Trinidad.

Spitulnik, Debra. (1998). Mediating unity and diversity: The production of language ideologies in Zambian broadcasting. In B. Schiefflin, K. Woolard, & P. Kroskrity (Eds.), *Language ideologies: Practice and theory*. Oxford: Oxford University Press. 163–188.

Sridhar, S. N. (1990). *Kannada*. New York: Routledge.

Sridhar, Kamal. (1991). Speech acts in an indigenised variety: Sociocultural values and language variation. In J. Cheshire (Ed.), *English around the world*. Cambridge: Cambridge University Press. 308–318.

Sridhar, Kamal, & Sridhar, S. N. (1992). Bridging the paradigm gap: Second-language acquisition theory and indigenized varieties of English. In B. Kachru (Ed.), *The other tongue: English across cultures*. Urbana: University of Illinois Press.

Tichy, P. (1980). The logic of temporal discourse. *Linguistics and Philosophy* 3:343–369.

Williams, Jessica. (1987). Non-native varieties of English: A special case of language acquisition. *English World-Wide* 8:161–199.

Winford, Donald. (1993). Variability in the use of perfect *have* in Trinidadian English: A problem of categorial and semantic mismatch. *Language Variation and Change* 5:141–187.

https://doi.org/10.1017/S095439450113303X Published online by Cambridge University Press