



# Past inflection around the world: A cross-variety analysis of New Englishes

Stephanie Hackert<sup>\*</sup>, Catherine Laliberté, Diana Wengler

*University of Munich (LMU), Germany*

Received 22 December 2023; revised 4 June 2024; accepted in revised form 20 June 2024; available online 6 July 2024

## Abstract

In this paper, we investigate variable past inflection in four New Englishes. Our data are drawn from the conversational parts of the Hong Kong, India, Jamaica, and Philippine subcomponents of the International Corpus of English. We investigate the entire range of language-internal factors that have been found to influence non-obligatory past marking in varieties of English. This includes morpho-phonological verb class, lexical aspect, grammatical aspect, marker persistence, the presence or absence of a temporal adverbial, and, for consonant-final regular verbs, preceding and following phonological environment. We also consider verb frequency, which has received only scant attention in past inflection research so far. Employing both mixed-effects regression and random forests, we argue that, despite inter-variety differences, there is a core grammar of past inflection, which is constrained by general structural and cognitive phenomena such as grammatical aspect and marker persistence, with frequency also exerting an important and consistent effect. This has implications for debates about universals vs. substrate influence or creole effects in morphosyntactic variation in English.

© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

**Keywords:** Past inflection; International Corpus of English; New Englishes; Creole; Second-language acquisition; Aspect Hypothesis

## 1. INTRODUCTION

Morphosyntactic differences across varieties of English have been a popular research topic of late, fueled, on the one hand, by (partly) competing models of the historical development and current status and functions of English around the world (e.g., [Kachru, 1982](#); [Schneider, 2003, 2007](#); [Buschfeld and Kautzsch, 2017](#)) and, on the other, by handbooks and atlas projects providing a wealth of comparable data on the structural features of Englishes (e.g., [Kortmann et al., 2004](#); [Kortmann et al., 2020](#)). While studies testing models have often focused on usage frequencies in only one or a few varieties at a time, typological studies using handbook and atlas data generally employ a macro approach, aggregating a large number of features in a large number of varieties but making categorical distinctions in terms of these

<sup>\*</sup> Corresponding author.

E-mail address: [st.hackert@lmu.de](mailto:st.hackert@lmu.de) (S. Hackert).

features' presence or absence. A complementary perspective is provided by micro studies zooming in on specific features in carefully annotated corpora, which are analyzed by means of rigorous quantitative methods in order to uncover the abstract grammatical patterning underlying the variation observed. In this paper, we contribute to this scholarship by looking into patterns of variable past inflection on lexical verbs in four so-called "New Englishes," i.e., the increasingly autonomous, indigenized varieties of English used as second languages or dialects in former colonies, where they now enjoy special prestige and status in politics and law, education and business, and the media.

The variation between inflected and unmarked past-reference lexical verbs is one of the best-researched morphosyntactic variables in English, particularly so in African American varieties and Caribbean English-lexifier creoles (CECs). This is partly due to the fact that the variable has played an important role in the discussion around the possible creole origins of African American Vernacular English (AAVE). The absence of overt past-tense marking on consonant-final regular verbs (e.g., *work*) was first interpreted as the result of "TD-deletion" (e.g., Labov et al., 1968) and thus as a purely phonological phenomenon common to all varieties of spoken English. By contrast, Bickerton (1975: 159) argued that variable past inflection in AAVE was constrained by grammatical parameters believed to condition verb marking in creole languages, such as stativity. This controversy already highlights one of the major problems involved in analyses of non-obligatory past inflection in English, namely the variable's location "at the intersection" of different structural processes (Patrick, 1991: 171), which is compounded by extralinguistic factors like speaker age, gender, or social background.

The present study is concerned not with vernacular Englishes but with varieties of "educated" English used in different parts of Asia and the Caribbean. Specifically, we include Hong Kong English as a largely monosubstratal variety, Indian English, which is one of the oldest New Englishes, Jamaican English as a creole-influenced dialect, and Philippine English, which is unique among postcolonial Englishes in having an American superstrate. Our data are drawn from the conversational parts of the respective components of the International Corpus of English. All components show variation in the application of the standard English past inflection, despite being part of a set of corpora representing precisely "standard" English (cf. Greenbaum, 1990). We test all language-internal constraints that have been described in the literature as influencing non-obligatory past marking: morpho-phonological verb class, lexical aspect, grammatical aspect, marker persistence, the presence or absence of a temporal adverbial, and, for consonant-final regular verbs, preceding and following phonological environment. We also consider verb frequency as a separate and independent variable and include individual text file and lexical item, as such random factors have repeatedly been shown to massively affect variable linguistic processes of the kind investigated here. The patterns found not only align in important ways across the varieties under study but also closely resemble those described in other research on non-obligatory past marking in CECs, other New Englishes, and first-language (L1) varieties in North America and elsewhere. They also link variable past inflection with other non-categorical morphosyntactic phenomena, such as the genitive or dative alternation. In sum, there exists a core grammar of variable past inflection in English, which is to a large extent determined by structural factors such as grammatical aspect and cognitive ones such as marker persistence and verb frequency but also susceptible to what has been termed "probabilistic indigenization" (Szmrecsanyi et al., 2016: 133), i.e., the emergence of variety-specific grammars in post-colonial speech communities that differ from the grammars of other speech communities in terms of "stochastic patterns of internal linguistic variation [that] are reshaped by shifting usage frequencies in speakers of post-colonial varieties."

In what follows, we first outline the research context in which our study is embedded (Section 2) and then give details concerning our data and method (Section 3). Section 4 presents the results of our statistical analysis. Section 5 discusses these results in light of previous research on the topic. Section 6 offers concluding remarks.

## 2. RESEARCH CONTEXT

### 2.1. Morphosyntactic variation in English, probabilistic grammar, and comparative sociolinguistics

The study of morphosyntactic variation in English has been a popular research topic of late, employing different sources of data, following distinct methodologies, and pursuing contrasting goals. One strand of research has involved the cross-variety comparison of individual linguistic features in corpora such as the International Corpus of English (ICE) or the Corpus of Global Web-based English (GloWbE). Much of this research has focused on national varieties and has had as its explicit or implicit aim the testing of models of World Englishes such as Kachru's Circle Model (1982), Schneider's Dynamic Model (2003, 2007), or Buschfeld and Kautzsch's Extra- and Intraterritorial Forces Model (2017). While more recent work of this kind has employed sophisticated statistical modeling (e.g., Gries et al., 2018), much of it has been primarily descriptive in orientation, reporting usage frequencies of particular "indigenized" or "nativized" features across varieties.

Another strand of research into morphosyntactic variation is typological, taking a bird's-eye view and employing handbook (e.g., [Kortmann et al., 2004](#)) and atlas (e.g., [Kortmann et al., 2020](#)) data to determine the distribution of sets of features and establish meaningful patterns in terms of geography and variety type. This has involved not only the search for diagnostic features, discussions of linguistic simplicity vs. complexity, and the postulation of various kinds of “-versals” (“angloversals,” “areoversals,” “varioversals,” or “vernacular universals”), i.e., frequent or even universal non-standard features of English (e.g., [Szmrecsanyi and Kortmann, 2009](#)), but also the importation of statistical techniques from other fields of research, such as phylogenetic trees.

Still more recently, a number of empirical studies have sought to move beyond “surfacy” phenomena and explore the “hidden” constraints on morphosyntactic variation within and across varieties of English. Studies along these lines have investigated variable grammatical phenomena such as the genitive or dative alternation ([Szmrecsanyi and Grafmiller, 2023](#)), often in written language and drawing on what is described as “probabilistic grammar,” i.e., the assumption that language involves variation governed by constraints that are tied to “more or less subtle stochastic generalization[s] about usage, which language users implicitly know about” ([Szmrecsanyi et al., 2016: 111](#)). Linguistic variation, in this framework, is assumed to be constrained by multiple, sometimes conflicting factors, which can be semantic, syntactic, pragmatic, or cognitive, i.e., related to language production and processing. Despite their being purely language-internal, such factors are assumed to be in part shaped by any speech community's sociolinguistic makeup and historical development; different patterns in usage thus reflect different experiences with language across different speech communities, or “probabilistic indigenization” (2016: 133).

While this approach brings a new perspective to the study of World Englishes, its basic theoretical assumptions and methodological underpinnings closely resemble those of variationist linguistics, which began in the 1960s with the study of urban varieties like AAVE and phonological variables such as  $\text{TD}$ -deletion. Scholars working in this tradition have always been interested in “alternate ways of saying ‘the same’ thing” ([Labov, 1972: 188](#)) and their language-internal and social conditioning, which have been researched in carefully annotated, often self-compiled corpora in a fully accountable manner ([Labov, 1969: 738](#)) and by means of quantitative methods including inferential statistics. Interest in the nature of earlier AAVE (as either English- or creole-like) led to the development of comparative sociolinguistics, which married principles of comparative historical linguistics to the rigorous analytical framework of variationist linguistics. Specifically, comparative sociolinguistics aims at detecting whether datasets “share an underlying grammar, and to what extent” ([Tagliamonte, 2013: 161](#)). This underlying grammar is visible neither in the sheer presence or absence of features nor in corpus frequencies but only in the abstract patterning of variation as evidenced in the direction, strength, and importance of constraints operating on it. Despite differing terminologies, research interests, and types of data employed, thus, probabilistic grammar and comparative sociolinguistics share crucial theoretical and methodological premises (cf. [Szmrecsanyi, 2017](#)), which the following analysis builds on to uncover similarities and differences in the grammars of variable past inflection in Hong Kong, Indian, Jamaican, and Philippine English.

## 2.2. Previous research on variable past inflection in varieties of English

In contrast to standard English, for which the “prescriptive enterprise implies that verbs are always inflected for tense, this is far from the case in spoken vernaculars” ([Poplack and Tagliamonte, 2001: 5](#)). As documented in the *Electronic World Atlas of Varieties of English* ([Kortmann et al., 2020](#)), about two thirds of non-standard Englishes have unmarked past-tense or participle forms, particularly for high-frequency irregular verbs such as GIVE or RUN (feature 129), and 60% of them show zero past forms of regular verbs (feature 132).

Variable past inflection has received extensive sociolinguistic attention and sophisticated statistical treatment, beginning with the earliest quantitative studies of AAVE ([Labov et al., 1968](#); [Wolfram, 1969](#); [Fasold, 1972](#)). United in their rejection of the verbal “deficit” theory (e.g., [Bernstein, 1971](#)), these studies aimed at demonstrating the variety's systematicity and status as a legitimate dialect of English by uncovering the patterned constraints on variable features such as the copula or past inflection. This included an emphasis on the identity of structural constraints across stigmatized and mainstream varieties of the language. Accordingly, the frequent occurrence of unmarked consonant-final regular verbs such as WALK was construed primarily in terms of the phonological process of syllable-final consonant cluster reduction, which affects all varieties of spoken English, and leads to the deletion of /t/ or /d/ if these follow another coda consonant. The earliest quantitative studies of non-obligatory past marking in CECs, by contrast, emphasized their uniqueness rather than their similarities with other Englishes, in line with the emergence of “a general post-colonial consciousness” ([Patrick, 1999: 4](#)) in the region. [Bickerton \(1975: 159\)](#) proposed that past marking in Guyanese Creole followed the dimensions of anteriority, punctuality, and stativity and thus differed fundamentally from the use of past inflection in English. Subsequent studies of variable past marking in AAVE and CECs have often weighed the contribution of “English” vs. creole constraints on the variation, but generally without questioning the terminological and conceptual problems inherent in [Bickerton's \(1981: 58\)](#) description of the “typical” creole tense-aspect system (cf. [Hackert, 2004: 13-16](#)).

Arguing from a comparative sociolinguistic perspective, [Poplack and Tagliamonte \(2001\)](#) present a number of multivariate analyses of variable past inflection in earlier African American English. Phonological conditioning accounts for most of the variation among regular verbs. The behavior of individual lexical items also has a strong effect, particularly among irregular verbs. Further predictors include “morphological priming” (2001: 129) and “discourse preferences” (2001: 141). The Bickertonian factors of anteriority and stativity are inconclusive at best. Poplack and Tagliamonte conclude from these findings that “the grammar of AAVE originated largely from the regional and nonstandard Englishes to which [...] early African Americans were exposed, and not from any widely-spoken creole” ([Poplack, 2000: 2](#)).

In a comparative study of AAVE and Trinidadian Creole, [Winford \(1992: 335\)](#) demonstrates that, with regard to verbal aspect, “[t]he proper distinction is between verbs that refer to specific past situations and those that refer to habitual/characteristic situations” rather than between punctual and non-punctual ones. In a detailed examination of over 8,000 past-reference verbs from urban Bahamian Creole, [Hackert \(2004: 161-166\)](#) finds that the marking propensities of individual lexical items explain not just the behavior of particular morpho-phonological verb classes but also the apparent stativity effect, which largely disappears when a variably stative and dynamic verb, i.e., HAVE, is investigated in isolation. The effect that remains can be ascribed to the fact that statives often have backgrounding function and therefore tend to carry an overt mark of past temporal reference. This discourse-pragmatic principle has also been described for the historical present tense in English narratives ([Schiffrin, 1981](#)); hence, what appeared as a characteristically creole constraint in earlier work ([Bickerton, 1975, 1981](#)) actually aligns with more general principles of language use. This is also true of the disfavoring effect that habituality has in AAVE and mesolectal creoles: habituals describe situations that are “characteristic of an extended period of time” ([Comrie, 1976: 27-28](#)) and often involve “induction and generalization” ([Patrick, 1999: 187](#)) – two dimensions that are epitomized in generic sentences, i.e., sentences expressing regularities, timeless truths, or species-specific characteristics (e.g., *Birds fly*, *Water boils at 100 degrees centigrade*), which, in the world’s languages, often occur in the maximally unmarked tense forms (cf. [Dahl, 1995](#)).

Other relevant studies of variable past marking in AAVE and CEC include [Weldon’s \(1996\)](#) study of Gullah and [Patrick’s \(1991, 1999\)](#) investigation of mesolectal Jamaican Creole, which explains elevated rates of  $\text{TD}$ -deletion on bimorphemic clusters, i.e., past-tense or participle forms (e.g., *passed*), as opposed to monomorphemic ones (e.g., *past*), as being compounded by tense non-marking. [Rickford \(1999\)](#) links variable past inflection in AAVE with the question of the divergence of Black and white vernaculars in the United States, while [Hackert \(2019\)](#), in a study of personal letters by former Panama Canal workers of West Indian descent, finds, that the constraints operative on variable past inflection in this written dataset closely mirror those found in other studies of CECs, including strictly phonological factors. [Deuber \(2014\)](#), finally, analyzes the first forty conversations of ICE Jamaica with a focus on morpho-phonological verb category but finds “little sensitivity” to this factor in her data (2014: 95).

Research on varieties from outside the Atlantic area is less abundant, but some of it has uncovered similar patterns of variation. An early elaborate analysis is presented by [Ho and Platt \(1993\)](#), who analyze over 8,000 past-reference verbs from a hundred Chinese-Singaporean speakers of English. They discuss lexical, morpho-phonological, and social constraints on the variable as well as the behavior of specific high-frequency verbs and generally confirm the direction of effects found in the North American and Caribbean context: consonant-final regular verbs are subject to phonological reduction, and punctual verb situations are more frequently marked than non-punctual ones (1993: 92). In a later study of educated Singaporean English, [Gut \(2009\)](#) focuses on the phonological conditioning of the variable but includes a range of other factors such as verb semantics and the presence or absence of a temporal adverbial and/or marker on the preceding verb. Overall, irregular verbs employing suppletion or vowel change are past-inflected “far more often” than regular ones employing affixation only (2009: 272). Gut also looks into lexical frequency, finding “a weak but not statistically significant [...] relationship” between this factor and rates of past marking (2009: 271).

Past-tense marking, finally, is among the variables analyzed by [Sharma \(2009\)](#) in a comparative study of Indian and Singaporean English. Carefully disentangling the effects of lexical and grammatical aspect, Sharma notes that, in both varieties, perfective verb situations are substantially more frequently past-marked than stative situations, which in turn show more marking than habitual or progressive ones (2009: 178–179). While she does not dismiss the crossvarietal frequency of this pattern, she also notes that the Indo-Aryan and Chinese L1 grammars of all speakers in her sample are aspect-prominent, featuring an overt marker with perfective reference (2009: 176). Hence, variable past marking in Indian and Singaporean English appears to be straightforwardly describable as “a direct replication [...] of perfectivity marking in the substrate systems,” which, unless investigated specifically for varieties of English with non-perfectivity-marking substrates, cannot be claimed to be a universal effect (2009: 179).

[Biewer’s \(2015\)](#) study of variable past inflection in South Pacific Englishes is closely modeled on [Poplack and Tagliamonte \(2001\)](#) and [Hackert \(2004\)](#) and therefore offers a particularly interesting comparative perspective. Importantly, it demonstrates in an empirically sound and statistically robust fashion that the depressing effect on past marking exerted by habituality is not restricted to creoles and related varieties. “[T]ime adverbials, verb morphology and phonology show less of an effect,” which Biewer explains at least partly with “expectations misguided by frequent listings in

secondary literature” (2015: 259). Biewer also does not find a significant effect for stativity, substantiating further the idea that this factor’s effect may be an artifact of lexical and discourse-pragmatic constraints on past marking and not at all unique to creole varieties.

Bohmann and Babalola (2023) study variable past inflection in the spontaneous conversation section of ICE Nigeria and two additional sets of sociolinguistic interviews with speaker of Nigerian English. Their multivariate analysis reveals that morpho-phonological conditioning operates in the ICE data but is not statistically significant, while stativity actually disfavors marking, which they note “requires further exploration in future research” (2023: 31). Discourse-pragmatic constraints, such as the marking of anterior verb situations, and the cognitive factor of marker persistence show effects in the expected direction. Finally, Hackert et al. (fc.) investigate non-obligatory past marking in Australian Aboriginal English as spoken on Croker Island, Northern Territory, a high-contact L1 variety characterized by extreme and persistent variability (Mailhammer, 2021). As in Caribbean English-lexifier creoles, past marking in English on Croker Island involves not only the variable inflection of lexical verbs but also the preverbal marker *bin*. The constraints governing *bin* relate in systematic ways to those governing inflection, which, in turn, evidence patterns known from both creole and non-creole varieties of English: morpho-phonological verb class is highly influential, stativity favors marking but is lexically based, habituality disfavors marking, and temporal disambiguation by means of adverbials makes no difference overall, in contrast to marker persistence, which is the only factor that emerges as highly significant statistically. Hackert et al. (fc.) conclude that the consistent appearance of phenomena such as morphological priming and the root-ness of such purportedly “typical” creole factors as stativity in lexical effects across all kinds of varieties makes it likely that pan-English or even crosslinguistically valid constraints govern variable past marking. The present study puts this conclusion to the test by examining conversational data from a range of New Englishes of wide geographical distribution, linguistic inputs, and sociolinguistic and historical contexts.

### 3. DATA AND METHOD

#### 3.1. Varieties selected

The four varieties selected for this study are postcolonial Englishes, yet they have somewhat different colonial histories and, accordingly, different sociolinguistic make-ups today. As part of a set of case studies for his Dynamic Model, Schneider (2007) gives detailed and comparable accounts of all of them. We therefore refrain from presenting comprehensive sociohistorical and sociolinguistic descriptions here, noting only a few relevant facts.

Hong Kong English (HKE) and Indian English (IndE) are often listed as examples of New Englishes. In both countries, British English was introduced during the colonial period via the education of a small local elite. Even though IndE is now one of the major varieties of English worldwide in terms of speaker numbers, its local status and functions are still limited. Of over 100 million speakers, only some 250,000 report English as their native language, this number having been largely constant since independence in 1947 (Sharma, 2012: 523). As in many other postcolonial contexts, English constitutes a lingua franca in a highly multilingual, ethnically and religiously diverse nation, and, in fact, the label “IndE” encompasses a “dramatic range of variation determined by L1s, region, socio-economic position, mode of acquisition, register of use, and attitude” (2012: 523). IndE “shows strong signs of structural nativization, involving all levels of language” (Schneider, 2007: 169). At the same time, proficiency in the language is still a marker of higher education and social status, and, unlike in most other postcolonial settings, “a local form of English has not adopted the function of an identity-carrier” (2007: 167). Among the growing urban population, the language has recently been extending its functions and forms of use and acquiring more native speakers (Sharma, 2012: 523), but these developments largely post-date the creation of the corpus we use in the present study, ICE India.

What distinguishes Hong Kong from other postcolonial countries is both the historically monoethnic makeup of its population, which was almost exclusively Cantonese-speaking Chinese, and the longevity of its colonial status, which lasted until the official “handover” from Britain to China in 1997. As in India, English was first introduced to southern China by way of the activities of British traders in the seventeenth century, but its current form and functions there owe much to “the economic transformation of Hong Kong from a relatively poor refugee community to a wealthy commercial and entrepreneurial powerhouse” between the 1960s and 1990s (Bolton, 2000: 268). The language was spread initially by missionaries, who taught it alongside Chinese language and literature in so-called “Anglo-Chinese” schools, which, following a series of educational reforms in the 1970s and 1980s, developed into accessible English-medium schools paving the way to the transition from elite bilingualism to mass bilingualism and the growth of a “new middle class” (2000: 269). With “Hong Kong’s economic change and self-projection as a ‘global city,’ associated with internationalization” came positive attitudes toward the language, which boosted its further spread (Schneider, 2007: 136) as well as a diversification in terms of the ethnic background of its speakers (Bolton, 2000: 275). Language policies during the late colonial phase and after the handover strengthened the role of Chinese. Today, the linguistic conflict in the

country involves not so much English and Chinese as the locally spoken Cantonese and mainland Putonghua, the national language of the People's Republic of China. In this situation, English serves as a lingua franca between speakers of the different Chinese varieties but has also become "the marker of a general middle-class (new middle-class) identity for Hong Kong Chinese" (Bolton, 2003: 115).

Jamaica represents the typical Caribbean postcolonial situation, with over ninety per cent of its population being of West African descent. As elsewhere in the region, local forms of English have their basis in creole formation, which involved face-to-face interaction on large plantations between comparatively small numbers of British planters and indentured servants likely using non-standard dialects and speakers of various West African languages. More standard varieties of the language spread only from the nineteenth century onward, when primary education was made available to the Black population majority (cf. Devonish and Thomas, 2012: 181). Today, Jamaican English (JamE) coexists with its lexically related creole in a continuum characterized by vast vocabulary overlap and gradual but patterned grammatical transitions. Attitudes toward Caribbean English-lexifier creoles have substantially changed over the past decades. Whereas the creoles were formerly simply and uniformly designated as "bad" or "broken English," they are now more positively valued. Still, perceptions of a functional division between them and English remain strong (cf. Deuber, 2014: 30-37). Importantly, unlike the other three varieties studied here, JamE is not an L2 variety but English as a second dialect, as Jamaican Creole is the vernacular of the vast majority of the population (cf. Görlach, 2002: 54).

Unlike HKE, IndE, and JamE, Philippine English (PhilE) emerged as a result of American colonization. The United States obtained authority over the Philippines from Spain in 1898 and began anglicizing the country. English was declared the official language as well as the language of education. Its initial establishment was effected by the "Thomasites," a group of some 500 American teachers sent to the Philippines on board the USAT *Thomas* in 1901. The spread of the language was massive and initially unhampered, even during the early postcolonial phase of the country, aided by a bilingual education policy whose local component, Tagalog, or Filipino, proved somewhat difficult to promote in actual practice. By the early 1990s, over half of all Filipinos read, wrote, and spoke English (Schneider, 2007: 141). Following the demise of dictator Ferdinand Marcos in 1986, however, anti-Western and anti-elitist sentiments led to a "backlash" to the language. English is still firmly entrenched "in higher classes, in the professions, and in discussions of technical subjects, and even as a home language in some families, especially among the 'economic elite'" (Schneider, 2007: 143) but resented as a symbol of colonialism and "undemocratic elitism" (2007: 141) by many others. Structurally, Philippine English has seen not only extensive nativization, well-described in the literature, but also internal diversification, including the development of a mixed code, Taglish, or "mix-mix," which is particularly frequent among young, urban, educated Filipinos but has also acquired symbolic value "as a reaction to the hegemonizing tendencies of Philippine society and modern life" (Bautista and Lourdes, 2004b: 226).

### 3.2. Corpora and samples

Our data are drawn from the conversational parts of ICE Hong Kong (ICE-HK), ICE India (ICE-IND), ICE Jamaica (ICE-JAM), and ICE Philippines (ICE-PHI). The ICE corpus project was initiated in the late 1980s with the aim of providing a data base for comparative studies of national varieties of English around the world. Each ICE corpus comprises one million words, distributed across 500 texts of 2,000 words each. More than half of each corpus (600,000 words) consists of spoken data. The range of text types represented in ICE comprises, among others, face-to-face and telephone conversations, classroom lessons, broadcast interviews and discussions, parliamentary debates, legal cross-examinations and presentations, student essays, social and business letters, academic writing, press news reports and editorials, and novels and short stories. Speakers are included on account of their social characteristics: they must be at least eighteen years old and have completed English-medium secondary school in the country in question. The ICE components analyzed here all belong to the first generation of ICE corpora, i.e., data collection was begun (albeit not always finished) in the 1990s, and all corpora have been completed and released.

We restricted our analysis to the category of face-to-face conversations, which are coded as texts S1A-001 to S1A-090 in ICE. Arguably, this text category is among the more heterogeneous ICE categories, as conversations can be anything from casual to careful: they can involve friends engaged in a private exchange, colleagues discussing a controversial work-related topic, or strangers involved in an interview-like situation. We opted for this text type for two reasons. First, (more or less) spontaneous conversation is exactly the text type elicited in the sociolinguistic interviews that constitute the data base of most previous work on non-obligatory past inflection in varieties of English. The face-to-face conversations provide the closest possible ICE equivalent. Second, conversations often contain personal narratives,

which led us to expect an especially high number of past-reference verbs and sufficient variation in marking patterns specifically in this text type.<sup>1</sup> This was borne out by initial data inspection of the first 15 texts of each corpus, which resulted in a projected number of ca. 12,000 tokens overall.

To keep data processing manageable, we selected the second half of each corpus text and extracted and coded all tokens of variably inflected past-reference verbs occurring in this text portion up to the 50th token.<sup>2</sup> This ensured that all texts were adequately represented in our analysis while also preventing undue influence of any individual text containing a disproportionate number of past-reference verbs (cf. [Bohmann and Babalola, 2023: 23](#)). The second rather than first half of each text was chosen under the assumption that, in any individual conversation, speakers would have become more comfortable with the recording situation by that time and would have produced more relaxed, “natural” speech.<sup>3</sup>

### 3.3. Circumscribing the envelope of variation

As pointed out by [Poplack and Tagliamonte \(2001: 114\)](#), accurate variable extraction and coding is “particularly important in the past temporal reference sector,” as “there is little isomorphy between form and function in this sector in [ . . . ] English.” Variably marked lexical verbs account for the bulk of past-reference verb situations, but numerous other construction types also occur, as illustrated in the following passage from ICE Jamaica.

<\$B> <ICE-JA:S1A-009#X180:1:B>No I don't know but <> <-> my</-> <=> my</=></>what **won** me over he **missed** the class because he **went** home and he **got** the times wrong  
 <ICE-JA:S1A-009#X181:1:B>So he **came** back and he is apologizing to us in this <{> <[> <unclear>word</unclear> ar></>  
 <\$A> <ICE-JA:S1A-009#X182:1:A> <[> Profusely</></>  
 <\$B> <ICE-JA:S1A-009#X183:1:B>Like it's okay you know <?> we all</?> been there  
 <ICE-JA:S1A-009#X184:1:B> <?> Like</?> I **went** home to get some fruit and have a snack and I completely **forgot** the time and he's there with this<unclear>word</unclear>grapes and <{1> <[1>an apple a banana right</[1><ICE-JA:S1A-009#>You just wanna hug him and say it's <{2> <[2>okay</[2>  
 <ICE-JA:S1A-009#X185:1:B>And he was really sorry  
 <ICE-JA:S1A-009#X186:1:B>He was really sorry and he **said**<quote>I just **wanted** to get some fruit</quote>  
 <ICE-JA:S1A-009#X187:1:B> <> <-> And he **came**</-> <=> he **came**</=></><{3><[3><,> and</[3>  
 <\$A> <ICE-JA:S1A-009#X188:1:A> <[1>Ah</[1></[1>  
 <\$A> <ICE-JA:S1A-009#X189:1:A> <[2>Yeah right<O>laughs</O></[2></[2>  
 <\$A> <ICE-JA:S1A-009#X190:1:A> <[3>He didn't actually  
 <ICE-JA:S1A-009#X191:1:A>Oh okay</[3></[3>

Like most studies of variable past marking in AAVE, CECs, and World Englishes, we took form as our point of departure, focusing on the alternation between unmarked and past-inflected lexical verbs with unambiguous past-time reference. This excluded all tokens with non-temporal semantics, such as counterfactuals and conditionals, as well as contexts that permitted both a past and a present-time interpretation. The term *inflection* refers to three different morphological processes: suffixation, as in *I just wanted to get some fruit* (ICE-JA:S1A-009#X186:1:B), vowel change, as in *he got the times wrong* (ICE-JA:S1A-009#X180:1:B), and suppletion, as in *he went home* (ICE-JA:S1A-009#X180:1:B). We counted only tokens of the standard English past-tense suffix on lexical verbs but additionally included two of the “primary” ones ([Quirk et al., 1985: 96](#)), i.e., HAVE, both as a main verb and as the “semi-auxiliary” HAVE TO (1985: 137), and main-verb DO. Both are subject to the same inflectional processes as lexical verbs proper. Auxiliary DO, as in *He didn't actually* (ICE-JA:S1A-009#X190:1:A), by contrast, was discounted. Copula and auxiliary BE structures, as in *He was really sorry* (ICE-JA:S1A-009#X185:1:B), were also omitted, as they undergo entirely different variable processes, i.e., deletion and leveling. We further excluded non-standard forms of inflection, as in *You just*

<sup>1</sup> In fact, the elicitation of such narratives as part of so-called “conversational module[s]” ([Labov, 1984: 33](#)) constitutes a staple of classic sociolinguistic field methodology, aimed at overcoming the “observer’s paradox” in the elicitation of vernacular speech (cf. [Hackert, 2004: 187–192](#) on variable past marking in narratives of personal experience in urban Bahamian Creole).

<sup>2</sup> More than 50 tokens were found only in six out of the 360 text portions we employed (mean = 14.36, median = 10.50).

<sup>3</sup> Some texts (e.g., ICE Jamaica S1A-066) are made up of more than a single conversation, in which case this consideration does not apply.

*wanna hug him* (ICE-JA:S1A-009#), historical present forms marked by means of -s, ambiguous forms such as *cut*, *quit*, or *put*, tokens occurring in unclear contexts, and consonant-final regular verbs before /t/, /d/, /θ/, /l/, /tʃ/, and /dʒ/ (Hackert, 2019: 269). Semantically, the envelope of variation is restricted to verb situations with absolute past reference, which excludes verb situations with perfect meaning (2019: 270). For more details on “count” and “don’t count” cases in quantitative analyses of variable past inflection and reasons for the inclusion or exclusion of particular structures, cf. Hackert (2008). Our final database consisted of 4,787 tokens.

### 3.4. Constraints and annotation

As outlined in Sections 1 and 2.2, a frequent explanation for non-obligatory past inflection in varieties of English is phonological. It maintains that unmarked past-reference verbs represent not an underlying zero form but the result of TD-deletion. Under this hypothesis, consonant-final regular verbs (e.g., WALK) should be past-marked much less frequently than other regular verbs and irregular ones. We therefore coded the predictor of morpho-phonological VERB CLASS as follows:

- regular verbs ending in a consonant other than /t/ or /d/, hence forming a coda cluster in the past tense;
- other regular verbs. This comprises both syllabic regular verbs, whose stem ends in /t/ or /d/, leading to /ɪd/ as the past-tense suffix (e.g., WANT), and regular verbs ending in a vowel and taking /d/ as the dental suffix (e.g., PLAY);
- irregular verbs. This includes all historically strong verbs (e.g., COME) as well as weak verbs that now feature stem vowel alternations because of historical sound changes (e.g., TELL, BUY). It also comprises the smaller irregular class of “devoicing” verbs, i.e., verbs without vowel change whose base forms ends in a liquid or nasal + /d/ (e.g., SEND, BUILD; cf. Quirk et al., 1985: 100, 106), as well as the high-frequency, morphologically idiosyncratic verbs DO, GO, HAVE, MAKE, and SAY.

Irregular past forms often show a high token frequency, which endows them with a “lexical strength” that makes them easy to access in the mental lexicon and resistant to change historically (Bybee, 1995: 428), and, in fact, research on variable past inflection has often revealed the highest marking rates precisely among high-frequency irregular verbs such as DO, GET, or GO, including an elevating effect on the marking rates of particular morpho-phonological verb classes (Hackert, 2004: 145). However, high token frequencies are not exclusive to irregular verbs; in fact, some regular verbs such as WANT, WALK, TALK, START, LOOK, or ASK are among the most frequent in English (Biber et al., 1999: 367-369). Frequency thus cuts across morpho-phonological verb classes, and, in fact, there is research that suggests that “frequent regular verbs may be stored in memory with their inflection” just like frequent irregular ones (Jiráňková and Cilibrasi, 2021: 189). Still, as noted in Section 2.2, lexical frequency has so far received very limited attention in connection with variable past marking.

That said, the phenomenon has recently come to the fore in studies of TD-deletion. While its actual effect is still debated, methodological discussion has involved not only data-base issues, i.e., whether to count the frequency of particular lexical items within the dataset itself or with reference to an external corpus (cf. Walker, 2012: 403; Baranowski and Turton, 2020: 7-8), but also the issue of what frequency means and how it is best measured: by way of the abstract unit of lexical representation, i.e., the lemma, or by considering the word form in question, e.g., past tense, or by some ratio of the two (Purse et al., 2022: 3)? With regard to the text base of frequency counts, individual corpora are often small and sometimes biased in one or the other direction. External reference corpora such as Kučera and Francis (1967), CELEX (Baayen et al., 1993), SUBTLEX (van Heuven et al., 2014), or GloWbE (Davies and Fuchs, 2015), by contrast, may cover different time periods, varieties, and/or text types than the ones represented in the data under investigation; they may therefore feature different verb types altogether or at vastly different token numbers (cf. Biber et al., 1999: 367-369). We opted for a compromise that rests on the corpora under study but extends to the entire range of spontaneous spoken texts, i.e., all private and public dialogues (ICE categories S1A and S1B) as well as unscripted monologues (S2A) of ICE-HK, ICE-IND, ICE-JAM, and ICE-PHI. These texts comprise not only the face-to-face conversations analyzed in this study but also phone calls, classroom lessons, broadcast discussions and interviews, parliamentary debates, business transactions, unscripted speeches, and a few others. Together, they amount to exactly half of each ICE corpus, i.e., 500,000 words. We thus measured the “corpus frequency” of each verb in an expanded database that is identical to the data under investigation in terms of time period and variety and closely mirrors them in terms of text type characteristics. As for measuring frequency, word-form and lemma frequency are not actually independent, and while a ratio measure may be most accurate, the two measures have been found to be useable “more or less interchangeably” (Purse et al., 2022: 14). We chose lemma frequency and used logarithmically transformed values to counteract the effect of extreme outliers.



To return to the phonology of  $\tau$ D-deletion, the occurrence of coda consonant clusters and hence past inflection in consonant-final regular verbs has been found to be strongly influenced by preceding and following segment. We therefore also tested for phonological environment for this class. With regard to PRECEDING ENVIRONMENT, Labov (1989) found the following hierarchy from most to least deletion: /s/ > stops > nasals > other fricatives > liquids. Some subsequent studies (e.g., Patrick, 1991; Tagliamonte and Temple, 2005) have appealed to the sonority hierarchy as the determining factor, while yet others (e.g., Guy and Boberg, 1997) quote the Obligatory Contour Principle as underlying the effect of preceding segment. Unfortunately, the “agreement between theoretical predictions and empirical findings” is “quite poor for the preceding environment” (Patrick, 1999: 131). To operationalize the factor, we followed Patrick’s summary of empirical results (1999: 131), which suggests that preceding sibilants, stops, and nasals promote  $\tau$ D-deletion, while non-sibilant fricatives and laterals inhibit it. FOLLOWING ENVIRONMENT refers to whether the sound following the verb is a vowel, consonant, or pause (as indicated by corresponding markup tags in the corpora, or the end of a speaker turn).

We next coded for aspectual semantics. As noted in Section 2.2, stativity and habituality have long figured among the most frequently discussed constraints on non-obligatory past marking in creoles and related varieties, but their definition and operationalization have not always been made explicit nor based on insights from tense-aspect semantics or typology, which has greatly hampered cross-variety comparisons and made comparisons with non-creole languages almost impossible. The term *aspect* refers to two different yet closely intertwined dimensions of temporal information (cf. Comrie, 1976; Smith, 1997): first, the classification of events according to the properties of stative/dynamic, punctual/durative, and telic/atelic, i.e., lexical aspect or *aktionsart*, and second, the internal temporal viewpoint taken on them as either bounded or unbounded, i.e., grammatical aspect. Lexical aspect, and grammatical aspect are independent theoretical notions, but there are important interactions between them. In English, for example, stative verbs are usually incompatible with the progressive aspect; when the latter is applied to such a verb, a special effect is created. We coded for both dimensions separately and retained them as separate predictors in the statistical analysis.

To code for LEXICAL ASPECT, we first distinguished between stative verb situations, as in *I have a brother*, and dynamic ones, as in *He wrote a letter*. The stative/dynamic distinction has received a lot of attention in studies of variable past inflection in AAVE and CECs, which is clearly owed to Bickerton’s (1981: 58) idea of a “prototypical” creole tense-aspect system, in which an interaction exists between stativity and past temporal reference, in that unmarked statives are interpreted as having non-past reference, whereas the default interpretation of unmarked dynamics is past. If a past marker occurs, statives receive a past interpretation, dynamics a past-before-past one. Numerous studies have attempted to reproduce this pattern, and, in fact, statives generally favor past inflection in both AAVE and CECs (Hackert 2019: 274). To code for stativity, we followed Smith (1997: 17-18), who views the phenomenon as a property not of lexical items but of verb situations, which consist of verbs and their arguments as well as adverbials and other contextual information. Thus, individual verbs may assume different stativity values, as illustrated by *I know* (stative) vs. **Suddenly he knew** (dynamic) or *I have a brother* (stative) vs. *I’m having lunch* (dynamic). We coded for stativity manually, employing established tests (cf. Filip, 2012: 728-730) where necessary.

Apart from the stative-dynamic distinction, lexical aspect is not usually investigated in analyses of past marking in varieties of English. It is in research on L2 acquisition, though, and we took inspiration from this field to take a closer look at dynamic situations. According to the Aspect Hypothesis,

in the initial stages of the acquisition of tense-aspect morphology by adults, the acquisition of past morphology will be influenced by lexical aspectual categories. Namely, verbal morphology will be attracted to and will occur with predicates with similar semantics. Perfective past will occur with telic predicates (Bardovi-Harlig and Comajoan-Colomé, 2020: 1139).

Again, we coded the different dynamic tokens as verb situations rather than based on lexical identity alone and, following Vendler (1957), subdivided dynamic tokens into accomplishments (durative, telic), achievements (non-durative, telic), and activities (durative, atelic). Later accounts of lexical aspect have also identified semelfactives (non-durative, atelic), which represent “single-stage events with no result or outcome” (e.g., COUGH; KNOCK; Smith, 1997: 29). Initial data inspection revealed only between two and four semelfactives per variety. As their inclusion would have added a disproportionately small predictor level group to the statistical analysis, we eventually left out this situation type ( $N = 13$ ) altogether. Statives are by definition atelic, and, in fact, according to the Aspect Hypothesis, they should be treated similarly to activities by learners. “Learners first use (perfective) past marking on achievements and accomplishments, eventually extending use to activities and statives” (Shirai, 2007: 53). Note that, in this contention, the Aspect Hypothesis runs counter to Bickerton’s (1981: 58) “prototypical” creole tense-aspect system, whereby statives but not dynamics must receive an overt mark of past temporal reference.

Next in line is GRAMMATICAL ASPECT, where we distinguished between perfective and habitual verb situations. While a perfective situation is conceptually bounded on both ends (Smith, 1997: 66), habituais are open-ended. They constitute

one type of imperfective verb situation, progressives being another (Comrie, 1976: 25). Both present the view of an event “from within,” i.e., as having internal structure. A perfective, by contrast, presents an event as a whole, “from the outside.” The idea of perfectives as bounded wholes must not be confused with the lexical dimension of punctuality; in fact, a perfective event may well extend in time, as in *Rome was not built in a day*. Perfective aspect is also independent of any particular tense, but it most readily combines with past temporal reference, as the view of an event as having a final bound implies that it has come to an end. English does not have a grammaticalized perfective aspect; in fact, the simple past can express both perfective events, as in *Queen Elizabeth II died in 2022*, and habitual ones, as in *I always walked to school as a kid*. Both are variably inflected in varieties featuring non-obligatory past marking. As noted in Section 2.2, habituales have repeatedly been found to strongly disfavor inflection in both AAVE and CECs.

The factor of “PERSISTENCE” (Szmrecsanyi, 2006) or “morphological priming” (Poplack and Tagliamonte, 2001: 129) codes for whether the verb immediately preceding the one under investigation features the same marker, i.e., inflection for inflected verbs and zero for unmarked verbs. We applied this coding (automatically, by means of an R script) regardless of whether speaker turns changed in between adjacent tokens, as we interpret persistence to follow from the cognitive principle that language users tend to recycle material they have heard or used before (cf. MacDonald, 2013: 4). Since we selected only the second half of each conversation for data extraction and coding, we always had to mark the first token occurring in this text portion as “NA,” as the preceding token was not part of our dataset.

The effect of temporal disambiguation on past marking in African American varieties, CECs and other postcolonial Englishes is elusive. It is widely assumed that, in creoles, the interpretation of a verb as referring to the past or not may follow from contextual clues such as the presence of conjunctions and adverbials (cf. Bickerton, 1975: 150, 160) – this is, in fact, also attested in dialectal English (Tagliamonte, 1991: 231). However, neither Tagliamonte and Poplack (1993: 189-190) nor Hackert (2004: 174, 178), Biewer (2015: 259), or Bohmann and Babalola (2023: 31) find significant effects if the predictor of TEMPORAL ADVERBIAL is considered at large. If semantics (cf. Quirk et al., 1985: 481-482) is taken into account, Hackert (2004: 178-181) reports a compounding effect: durative adverbials, which explicitly bound the verb situations they modify, boost past inflection among perfectives in urban Bahamian Creole, whereas adverbials indicating frequency further dampen the marking rates of habituales. Our operationalization of this factor builds on Quirk et al. (1985: 481-482) and distinguishes the following levels:

- adverbials indicating a point of time (e.g., *yesterday, in 1960, after we had had lunch*),
- adverbials indicating duration (e.g., *for three years, until they left school*), and
- adverbials indicating frequency (e.g., *never, every summer*).

We did not set up a separate class of relationship adverbials (1985: 482), as initial data inspection revealed that this class was small overall ( $N = 52$ ) and consisted almost exclusively of tokens of *again*, which can plausibly be regarded as a point-of-time adverbial meaning ‘on a subsequent occasion,’ too (1985: 530).

With regard to extralinguistic factors, we coded for VARIETY (HKE, IndE, JamE, and PhilE). We also included the two random factors of individual TEXT and LEXICAL ITEM.

### 3.5. Data analysis

The basic tenet of comparative sociolinguistics and probabilistic grammar is that variable linguistic systems can be said to overlap to the extent that the workings of the language-internal constraints operating on them converge in the direction of effects and the constraint hierarchy. To capture these dimensions, we employed two different types of multivariate statistical tool, which have been found to complement each other well (cf. Tagliamonte and Baayen, 2012: 33): mixed-effects logistic regression and random forests. We performed separate analyses for each variety’s dataset, which involved two steps: (1) determining a predictor’s direction of effects by means of a mixed model and (2) assessing each predictor’s overall impact on the variation by means of random forest analysis.

Mixed-effects regressions are superior to fixed-effects models in their ability to handle datasets with non-independent, i.e., grouped, data points. Such grouped data structures occur, for instance, if lexical items are used more than once in the dataset. By adding grouping factors as random effects, fixed effects (like GRAMMATICAL ASPECT or PERSISTENCE) are held constant across lexical idiosyncrasies. We used the R-package “lme4” (Bates et al., 2015) to fit a mixed-effects model with GRAMMATICAL ASPECT, LEXICAL ASPECT, PERSISTENCE, FREQUENCY, TEMPORAL ADVERBIAL, and MORPHOLOGICAL VERB CLASS as fixed effects; we added varying intercepts for individual LEXICAL ITEM and TEXT.<sup>4</sup> Variable

<sup>4</sup> For each variety, we employed the following model formula:  $VARIABLE \sim GRAMMATICAL\_ASPECT + LEXICAL\_ASPECT + PERSISTENCE + \log(FREQUENCY) + TEMPORAL\_ADVERBIAL + VERB\_CLASS + (1|LEXICAL\_ITEM) + (1|TEXT)$ .

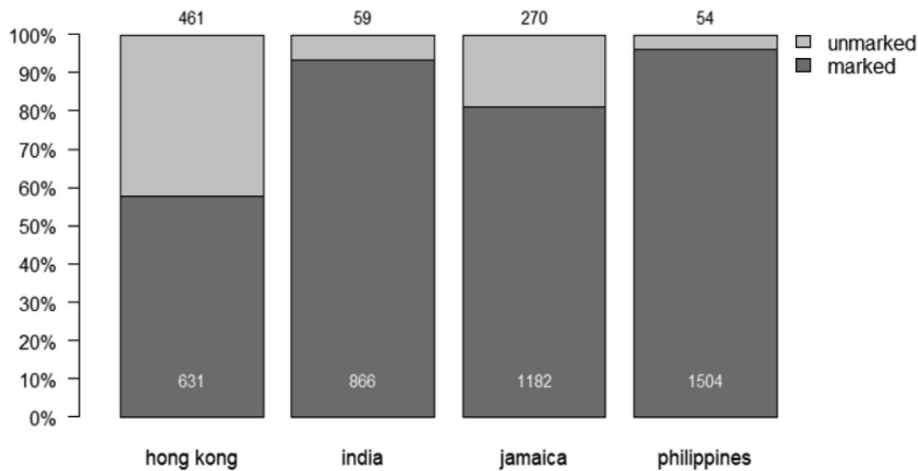


Fig. 1. Absolute frequencies and percentages of past inflection in ICE-HK, ICE-IND, ICE-JAM, and ICE-PHI.

inflation scores, which indicate if a model's predictors are too collinear, were computed via “performance” (Lüdtke et al., 2021). For all fixed effects, these scores were low. The overall significance of each predictor was computed by means of a type II ANOVA in “car” (Fox and Weisberg, 2019). Post-hoc tests assessing the differences between group means were calculated via “multcomp” (Hothorn et al., 2008). Predicted probabilities were obtained via “ggeffects” (Lüdtke, 2018) in order to visualize each predictor's direction of effects.

Random forests, as implemented by the “party” (Hothorn et al., 2006) package in R, constitute a non-parametric regression approach based on recursive binary partitioning of the dataset (Tomaschek et al., 2018: 250). A convenient feature that comes with random forest modeling is the possibility to determine each predictor's impact on the alternation under investigation by means of importance estimations, which render a linguistic variable's constraint hierarchy. We employed the estimation procedure developed by Strobl et al. (2008) in order to counteract the potential inflation of correlated predictors. For each variety, we fit a separate random forest analysis, each consisting of an ensemble of 1,000 trees. Random forests are more robust than individual decision trees because they randomly sample a subset of variables at each split. In our models, we configured this sampling process to consider three variables at each split.<sup>5</sup>

#### 4. RESULTS

Fig. 1 displays absolute frequencies and percentages of past inflection by variety. Marking rates vary between 60% in ICE-HK and near-categorical in ICE-IND (95%) and ICE-PHI (96%). At 82% past inflection, ICE-JAM ranges in between. Total frequencies of past-reference verbs also vary substantially between corpora, which is likely owed to differences in topic. Particularly ICE-JAM contains a high amount of animated talk between interlocutors familiar with each other, including narratives of personal experience; such narratives, as indicated in Section 3.2, are especially conducive to the occurrence of past-reference lexical verbs.

As outlined in Section 3.5, we fit four different models, one for each variety. All four models achieved very good goodness-of-fit values (ICE-HK:  $C = 0.89$ ,  $D_{xy} = 0.79$ , accuracy = 82%, baseline = 60%; ICE-IND:  $C = 0.86$ ,  $D_{xy} = 0.73$ , accuracy = 96%, baseline = 94%; ICE-JAM:  $C = 0.89$ ,  $D_{xy} = 0.78$ , accuracy = 87%, baseline = 81%; ICE-PHI:  $C = 0.95$ ,  $D_{xy} = 0.89$ , accuracy = 97%, baseline = 96%). Table 1 displays the significance of language-internal predictors of past inflection by variety, based on a type II ANOVA employing Wald chi square tests. The significance codes are as follows: 0 ‘\*\*\*’, 0.001 ‘\*\*’, 0.01 ‘\*’, 0.05 ‘.’, and 0.1 ‘.’.

As seen in Table 1, the only predictor that emerges as statistically significant below the 0.05 threshold in all four models is GRAMMATICAL ASPECT. LEXICAL ASPECT is marginally significant in three of them, i.e., ICE-HK, ICE-IND, and ICE-PHI. Two constraints reach statistical significance in two models each: temporal adverbial in ICE-HK and ICE-PHI, frequency in ICE-IND and ICE-JAM. Persistence is highly significant in ICE-JAM but only marginally so in ICE-PHI. Morpho-phonological verb class turns out to be significant in ICE-IND only.

<sup>5</sup> For each variety, we employed the following model formula: VARIANT ~ GRAMMATICAL\_ASPECT + LEXICAL\_ASPECT + PERSISTENCE + log(FREQUENCY) + TEMPORAL\_ADVERBIAL + VERB\_CLASS.

Table 1

Significance of language-internal predictors of past inflection in ICE-HK, ICE-IND, ICE-JAM, and ICE-PHI.

	chisq	df	p	signif
<b>1. ICE-HK</b>				
LEX_ASPECT	6.8920	3	0.075421	.
GRAMM_ASPECT	9.3939	1	0.002177	**
PERSISTENCE	0.3161	2	0.853814	
FREQ_LOG	0.4340	1	0.510038	
TEMP_ADV	8.0586	3	0.044817	*
VERB_CLASS	2.6982	2	0.259477	
<b>2. ICE-IND</b>				
LEX_ASPECT	6.8816	3	0.07577	.
GRAMM_ASPECT	24.8936	1	6.058e-07	***
PERSISTENCE	2.8004	2	0.24654	
FREQ_LOG	3.1585	1	0.07553	.
TEMP_ADV	0.2210	3	0.97414	
VERB_CLASS	10.1696	2	0.00619	**
<b>3. ICE-JAM</b>				
LEX_ASPECT	3.0417	3	0.3852332	
GRAMM_ASPECT	4.4439	1	0.0350252	*
PERSISTENCE	17.7128	2	0.0001425	***
FREQ_LOG	2.8788	1	0.0897536	.
TEMP_ADV	2.4307	3	0.4879499	
VERB_CLASS	4.2115	2	0.1217541	
<b>4. ICE-PHI</b>				
LEX_ASPECT	6.9753	3	0.072690	.
GRAMM_ASPECT	4.4984	1	0.033927	*
PERSISTENCE	5.3363	2	0.069381	.
FREQ_LOG	0.5597	1	0.454392	
TEMP_ADV	13.1835	3	0.004256	**
VERB_CLASS	0.4325	2	0.805553	

Fig. 2 contains partial effects plots for each predictor and variety, i.e., visual representations of the predicted values of each variable with all other variables held constant (Fox 2003: 6).<sup>6</sup> Overall, effect directions are remarkably similar from one variety to the next, except for LEXICAL ASPECT (Fig. 2.1). Whereas HKE clearly follows the Aspect Hypothesis for dynamic verb situations but not for stative ones, PhiE shows the entire pattern posited by the Hypothesis, though weakly so. IndE neatly distinguishes telic from atelic verb situations, including stative ones, but exactly reverses the predictions of the Hypothesis. JamE, finally, makes no clear distinction between telic and atelic dynamics and shows only a slightly elevated probability of past inflection for statives, which supports neither the Aspect Hypothesis nor a creole tense-aspect account.

A highly consistent pattern is found for GRAMMATICAL ASPECT (Fig. 2.2), with perfectives past-inflected at higher probabilities than habituales in all four corpora. The difference in marking propensities is massive in ICE-IND, which correlates with the overall significance of this predictor in the ANOVA results displayed in Table 1.2. PERSISTENCE (Fig. 2.3) and FREQUENCY (Fig. 2.4) also show unambiguous effects. A preceding verb with identical marking value increases the odds of a subsequent past-inflected or unmarked verb in all varieties. The more often a particular verb type occurs, the more likely it is to be marked for past reference. The effect of TEMPORAL ADVERBIAL (Fig. 2.5) is somewhat inconsistent. Still, in all varieties, frequency adverbials tend to correlate with lower probabilities of past inflection, a likely explanation being the association between frequency adverbials and the grammatical aspect of habituality, which has been found to dampen rates of past marking crossvarietally (cf. Section 3.4). Morpho-phonological VERB CLASS (Fig. 2.6), finally, shows the predicted pattern in three out of the four varieties: consonant-final regular verbs are less likely to be past-inflected than both other regular verbs and irregular ones in ICE-HK, ICE-IND, and ICE-JAM, even though in ICE-JAM this effect is weak.

<sup>6</sup> The numbers underlying the statistics presented in this section may be found in the Appendix. Table 2 presents absolute frequencies and percentages of past inflection by predictor, predictor level, and variety. Model summaries may be found in Table 3.

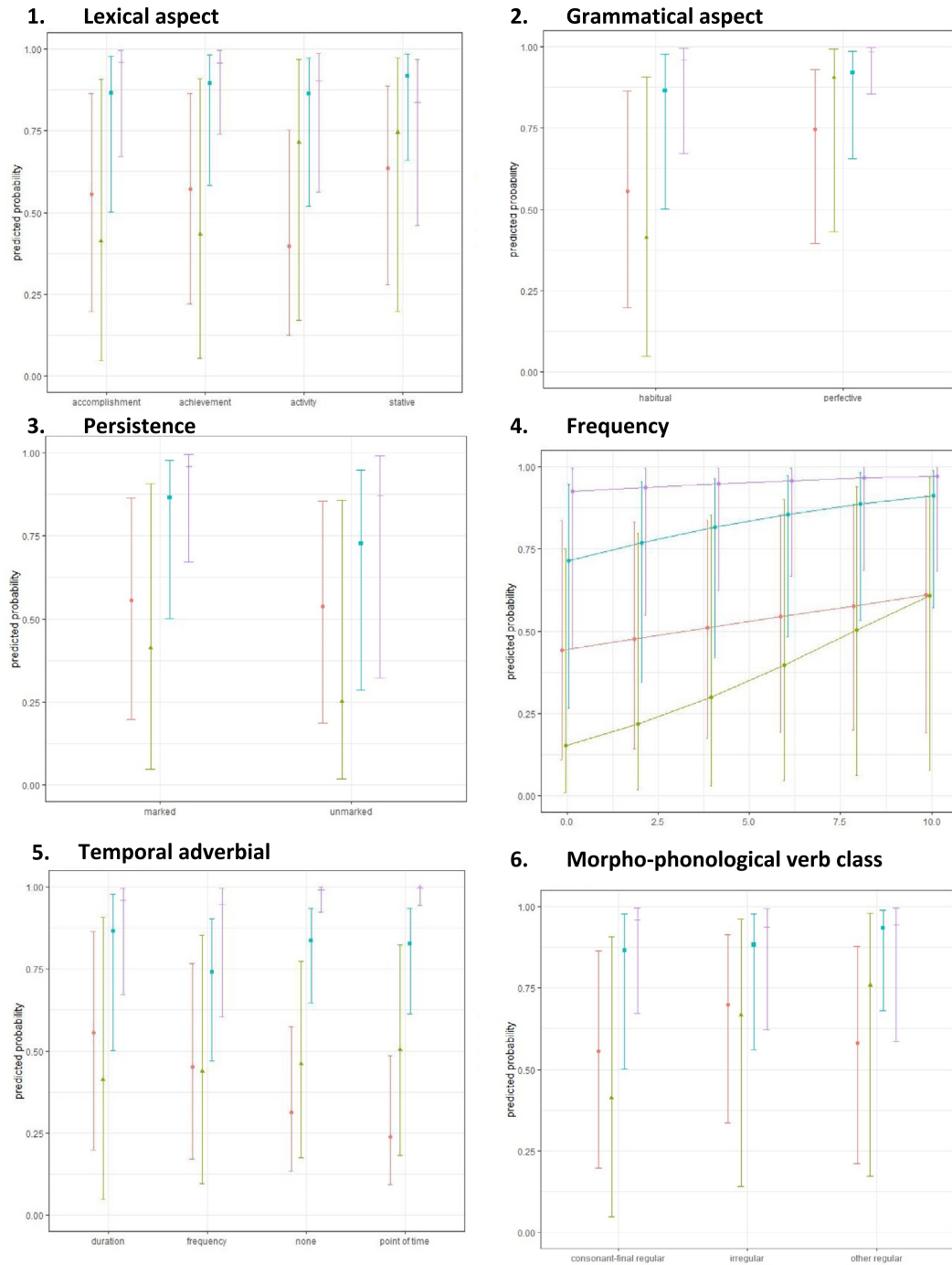


Fig. 2. Language-internal predictors' effects on past inflection in ICE-HK (red), ICE-IND (green), ICE-JAM (turquoise), and ICE-PHI (purple) (predictions for marking). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

ICE-PHI appears to contradict the pattern, with consonant-final regular verbs most likely to show past marking, but overall the distinctions between verb classes are minimal.

As outlined in Section 3.4,  $\tau$ D-deletion, i.e., the reduction of word-final consonant clusters on regular verbs, has been found to be subject to a number of phonological constraints, including preceding and following segment. These

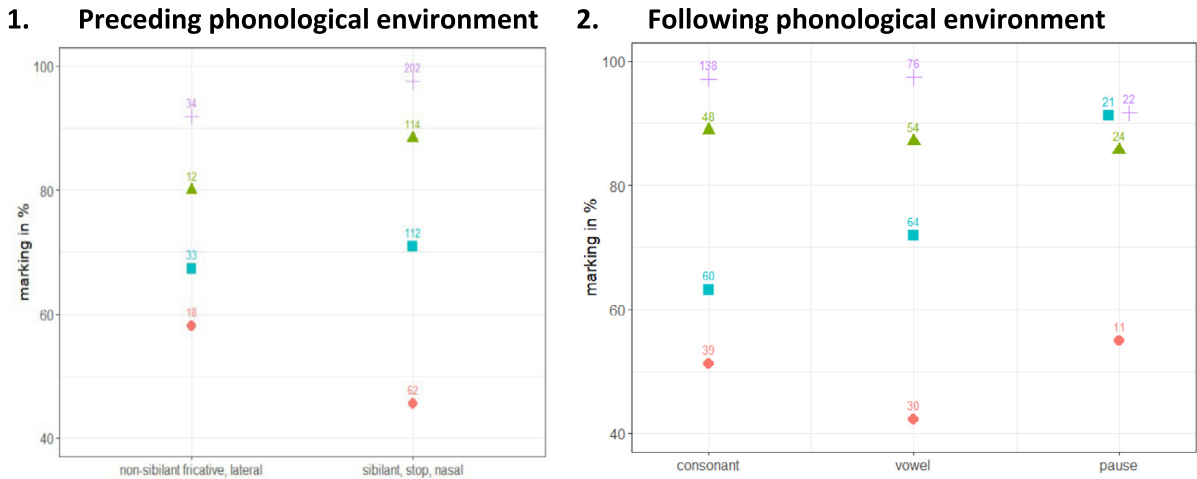


Fig. 3. Past inflection of consonant-final regular verbs by preceding and following phonological environment in ICE-HK (red), ICE-IND (green), ICE-JAM (turquoise), and ICE-PHI (purple). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

constraints also operate in our data but do so rather weakly and partly inconsistently. Figs. 3.1 and 3.2 display absolute frequencies and percentages for both phonological environments by corpus.<sup>7</sup> ICE-HK is the odd one out in both cases, showing less marking both after preceding sibilants, stops, and nasals and before vowels. The effect of preceding segment is weak but runs in the expected direction in all other datasets. With regard to following segment, only ICE-JAM evidences the expected pattern of more marking before vowels than before consonants. ICE-IND and ICE-PHI speakers hardly differentiate between these two phonological environments. Pause, finally, patterns inconsistently across varieties, which, however, is a finding that is in line with much previous research (cf. Schreier, 2005: 206–207).

The results of our random forest analyses may be seen in Fig. 4. *C* values vary between 0.78 and 0.87, which indicates good model-data fits (Brezina, 2018: 126).

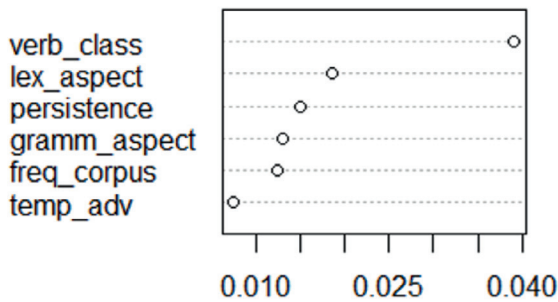
Predictor rankings differ substantially in the four corpora, with each corpus dominated by a different predictor: *VERB CLASS* in ICE-HK, *GRAMMATICAL ASPECT* in ICE-IND, and *PERSISTENCE* in ICE-JAM. In ICE-PHI, we see three highly influential predictors: *TEMPORAL ADVERBIAL*, *GRAMMATICAL ASPECT*, and *LEXICAL ASPECT*. That said, the latter two predictors, i.e., *GRAMMATICAL* and *LEXICAL ASPECT*, play a role in all datasets except ICE-JAM. As seen in Figs. 2.1 and 2.2, however, despite being generally weak, their effects in that variety go in the same general directions as observed for the other varieties and as predicted by the results of previous studies (cf. Section 3.4). The processing-related factors of *PERSISTENCE* and *FREQUENCY* also figure among the more important predictors in all varieties. The discourse-related predictor of *TEMPORAL ADVERBIAL*, by contrast, is entirely unimportant in all corpora except ICE-PHI, where it emerges as on a par with lexical and grammatical aspect.

## 5. DISCUSSION

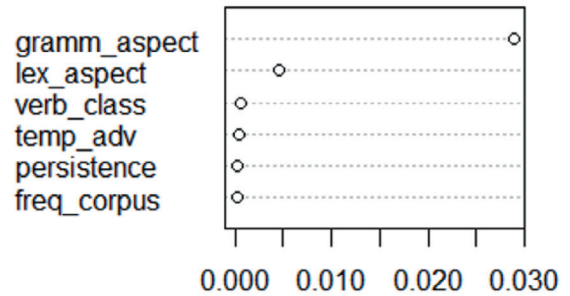
The most significant finding of this study is that, despite differences in terms of overall frequency, effect strength, predictor importance ranking, and – for some predictors – direction of effects, variable past inflection in the four New Englishes investigated here is subject to an identifiable set of structural and cognitive constraints, which are also well-attested in other studies of non-obligatory past marking in varieties of English. Grammatical aspect, persistence, and verb frequency affect the variation observed in all four varieties, and they do so in the same direction. Differences exist in the effects of lexical aspect, morpho-phonological verb class, and temporal adverbial. These differences can be explained by the sociohistorical and sociolinguistic circumstances in which the varieties evolved and are spoken today. More specifically, they relate to substrate influence as well as patterns of acquisition and hence levels of proficiency in English.

<sup>7</sup> Regression analyses were not possible for this subset of our data, as the number of consonant-final regular verbs per corpus was too small ( $N_{ICE-HK} = 167$ ,  $N_{ICE-IND} = 144$ ,  $N_{ICE-JAM} = 207$ ,  $N_{ICE-PHI} = 244$ ).

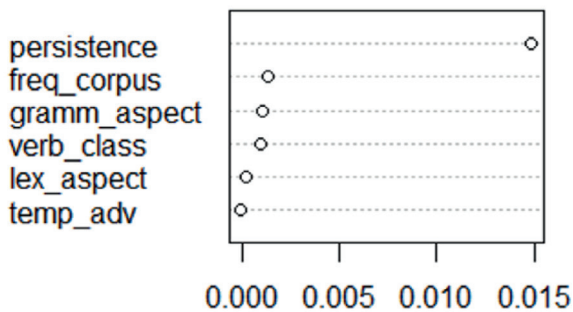
## 1. ICE-HK



## 2. ICE-IND



## 3. ICE-JAM



## 4. ICE-PHI

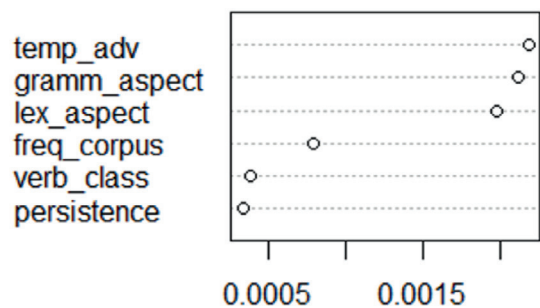


Fig. 4. Predictor importance ranking for past inflection in ICE-HK, ICE-IND, ICE-JAM, and ICE-PHI.

As seen in Fig. 1, overall rates of past inflection in our corpora vary between 60% and over 90%. While all of these rates are substantially higher than what is found in creole data, where marking frequencies generally amount to less than 50%, they align well with what has been observed for other New Englishes and/or acrolectal Caribbean data, where past-inflection frequencies of 75% and more are common (cf., e.g., Deuber, 2014: 94; Gut, 2009: 266; Biewer, 2015: 226; Bohmann and Babalola, 2023: 28). Particularly the IndE and PhilE rates are remarkable but not actually surprising if the sociolinguistic backgrounds of ICE-IND and ICE-PHI are considered. The Indian conversations are largely drawn from recordings made at a refresher course for English teachers held at the Central Institute of Indian Languages in Mysore, Karnataka, with another set of recordings made at or around Shivaji University in Kolhapur, Maharashtra (Lange, 2012: 5). Even though not all speakers conformed to the general ICE requirement of having undergone “formal education through the medium of English to the completion of secondary school” (Greenbaum, 1996: 6), as a group they are clearly overeducated in comparison to, e.g., ICE Great Britain, with most of them having completed tertiary education and holding at least an MA, many a Ph.D. The English teachers among them had additionally received formal instruction in English grammar as part of their job training. As summarized by Lange (2012: 2), “[b]y including those IndE speakers who effectively act as the norm providers within the IndE speech community, the corpus compilers have come to a [...] realistic assessment of [...] standard IndE.” Incidentally, other morphosyntactic features such as zero copula and verbal -s absence also occur at extremely low frequencies in ICE-IND (Lange, p.c., 2023), whereas features at the lexicogrammatical level (cf. Mukherjee and Schilk, 2008), at the purely syntactic level (cf. Lange, 2012), or at the pragmatical level (cf. Revis and Bernaisch, 2020) show a high degree of indigenization. Such a division appears to indicate that the omission of function words or inflectional endings is stigmatized in IndE and that speakers are aware of the non-standard value of such omissions.

ICE-PHI also represents the very standard end of PhilE.

The Filipino speakers [...] represented here come from the highly educated sector, with the least educated among them being college freshmen. A great majority of them are accomplished users of English as a second language, and speak [...] what sociolinguists refer to as an acrolectal variety of English, i.e., a kind of English that approaches the world standard (Bautista and Lourdes, 2004a: 9).

As also noted by [Bautista and Lourdes \(2004a: 12\)](#), many of the Philippine conversations have a somewhat stilted character, having been recorded as part of a class assignment or in other circumstances hampering “natural” speech. This might have prompted speakers to pay specific attention to their speech and aim at “correctness.”

While the Jamaican speakers were also highly educated, many of them being university students or lecturers, the circumstances surrounding the recordings were much less formal: “fieldworkers tried for the most part to record educated Jamaicans in private interactions in which, in spite of their relative informality, a variety considered as English was used at least at the outset, but they were not supposed to intervene if the language changed in the course of the conversation” ([Deuber, 2009: 6-7](#)). Such a change would have inevitably involved Jamaican Creole, and, in fact, the conversations in ICE-JAM do “show a greater range of variation than the initiators of the ICE project may have anticipated” (2009: 7), including morphosyntactic creolisms such as unmarked past-reference verbs, zero copula, and absence of verbal -s.

The Hong Kong data are clearly the most non-standard in terms of past inflection, which concurs with contemporaneous survey results from the early 1990s in which only about a third of respondents claimed they knew English well ([Bolt and Bolton, 1996: 200](#)). Such cautious self-assessments might not entirely reflect the average competence of the speakers sampled for the ICE-HK conversations, with the “vast majority” of them having been between the ages of 20 and 30 when being recorded (1996: 208), but Bolt and Bolton also note (1996: 202) that even though English was generally claimed as the medium of instruction at the time, “Cantonese is in fact quite widely used within secondary and tertiary education, increasingly so in the latter case as the group size becomes smaller and the situation less formal.” Such lesser proficiency in and use of English could explain the comparatively non-standard character of HKE as displayed in [Fig. 1](#).

At first sight, thus, the four datasets appear difficult to compare, seeing as they differ so substantially in terms of speaker competence in English and the register apparently aimed at by these speakers, despite the fact that the vast majority of them had received tertiary education and all produced speech for the same ICE text type, i.e., face-to-face conversations. However, rates of occurrence of a feature are not actually decisive in comparing language samples, as they may vary according to a wide variety of extralinguistic factors. As noted in [Section 2.1](#), the basic principle of the comparative sociolinguistic approach that we have followed here is to look for whether varieties or lects “share an underlying grammar, and to what extent” ([Tagliamonte, 2013: 161](#)), which necessitates attention to more abstract, deeper-level patterns of variation, as evident particularly in the direction of effects operating on the variable in question and the ranking of constraints. This principle is what our comparison with non-standard vernaculars such as AAVE and CECs is crucially premised on; it could also be brought to bear on any cross-variety comparison within the field of World Englishes, whether such comparison involves standard(ized) or non-standard varieties, L1 or L2, or high-contact or low-contact forms of the language.

The effects plot for LEXICAL ASPECT ([Fig. 2.1](#)) suggests substantial heterogeneity between the varieties studied here. Note, though, that the distinctive HKE pattern of past-inflecting telic but not atelic dynamics conforms fully to the earliest formulations of the Aspect Hypothesis, which employed a binary framework based on telicity alone ([Bardovi-Harlig and Comajoan-Colomé, 2020: 1143](#)). Also, recall ([Section 3.4](#)) that the Hypothesis predicts that “second language learners will *initially* be influenced by the inherent semantic aspect of verbs or predicates” ([Andersen and Shirai, 1996: 533](#); emphasis added). As just outlined, of the four groups of speakers represented here, the HKE speakers are the least advanced in terms of English proficiency. Viewed from this angle, their treatment of lexical aspect may be described as a learner strategy, which would fade as proficiency in English increases. This, in fact, is suggested by the comparison with the ICE-PHI and ICE-JAM data, where hardly any distinction is visible in terms of marking probabilities between telic and atelic dynamic verb situations. What exactly is behind the ICE-IND pattern of past inflection by lexical aspect, which, as noted in [Section 4](#), exactly reverses the Aspect Hypothesis, remains unclear at this point, but L1 influence represents a potential explanation.

The behavior of statives with regard to the Aspect Hypothesis has been more equivocal than that of dynamics, with some studies seeing them align with activities but others finding an early emergence of past marking with that situation type, particularly for English, which has been explained with “limited stative vocabularies [. . .], the fact that ‘be’ (the most dominant stative) almost always appears tensed (not in base form), and a lack of competition from other past forms” such as the imperfective in Romance ([Bardovi-Harlig and Comajoan-Colomé, 2020: 1144](#)). Recall, also, that the effect of stativity has been inconsistent across studies of non-obligatory past marking in varieties of English ([Section 3.4](#)). For English-lexifier creoles, it is commonly found that stativity favors past marking (cf. [Hackert, 2004: 164](#)). A careful analysis of individual stative verbs and their discourse contexts in urban Bahamian Creole, however, also revealed that this stativity effect was actually an artifact of the strong marking propensity of individual, high-frequency lexical items such as HAVE, THINK, and WANT, compounded by discourse function, with statives often occurring in backgrounding clauses, where they receive an overt past mark to endow the surrounding discourse with past temporal reference (2004: 161–166). This



aligns with Poplack and Tagliamonte's finding in their earlier AAVE data (2001: 132) of a higher likelihood of past marking for stative verbs expressing anteriority than for other statives. With regard to New Englishes, Biewer (2015: 259) finds no statistically significant differences between stative and dynamic verb situations in her South Pacific data, and Bohmann and Babalola (2023: 31) actually note a depressing effect of the factor on past inflection in Nigerian English, having accounted for verb-specific marking patterns by means of a lexical-item intercept in their mixed model. Clearly, the effect of stativity on verb marking is more complex than either the Aspect First hypothesis or Bickerton's "prototypical" creole tense-aspect model would predict. It depends on individual lexical preferences, which may differ across varieties, as well as on discourse patterns, including the functional organization of narratives (cf. Labov and Waletzky, 1967), which are rarely taken into account in studies of variable past inflection.

The four varieties also differ with respect to the effect of morpho-phonological verb class, which, as seen in Table 1.2, is statistically significant in ICE-IND but has high predictive importance only in ICE-HK (Fig. 4.1). The IndE pattern is clearly owed to the workings of  $\tau$ D-deletion, with consonant-final regular verbs showing a substantially lower likelihood of past inflection than both other verb classes. The HKE pattern, by contrast, pits regular verbs against irregular ones (Fig. 2.6). English is a taught language in all four contexts investigated here, i.e., it is acquired primarily through formal instruction in the school setting, either against the background of other languages, as in Hong Kong, India, or the Philippines, or against an English-lexifier creole, as in Jamaica. Traditionally, rote learning, pattern drill, and explicit attention to formal correctness play an important role in teaching English in all postcolonial communities investigated here (cf., e.g., Oenbring and Fielding, 2014: 29-30; Devonish and Thomas, 2012: 193 for the Caribbean; Majumdar and Mooij, 2012: 226 for India), with regular verbs subject to the "attach *-ed*" rule and irregular ones committed to memory. If rote-learned verbs were subject to qualitatively distinct psycholinguistic processes from rule-based ones, we would expect the two groups to pattern differently with respect to variable past inflection. We see this pattern only in the HKE data, which, as argued above, were sampled from the least advanced speaker group in terms of proficiency in English.

This ties in with what is observable for the phonological conditioning of consonant-final regular verbs. For preceding environment (Fig. 3.1), all corpora except ICE-HK follow the hierarchy of stop, sibilant, nasal > non-sibilant fricative, lateral that has been found to be empirically valid for many L1 dialects; for following environment, ICE-JAM very clearly displays the typical pattern of consonant > vowel, while ICE-HK reverses this pattern, with ICE-IND and ICE-PHI not showing much of an effect (Fig. 3.2). This may simply be a function of the overall extremely high rates of past inflection found in these two corpora. Alternatively, final consonant cluster reduction may not operate as strongly in IndE and PhilE as in other varieties, which is what Bohmann and Babalola (2023: 32), following Gut (2007), claim for their Nigerian ICE data.<sup>8</sup> What always needs to be kept in mind with regard to verb class is that particularly this factor is sensitive to lexical outliers, i.e., individual verb types that occur particularly frequently and show a marking behavior that is not in line with that of their respective class overall (cf. Poplack and Tagliamonte, 2001: 140; Hackert, 2004: 147). This point is further discussed below.

Finally, our corpora show differences in the presence or absence of a temporal adverbial. What is interesting from a creolist perspective is that ICE-JAM replicates what has been found for urban Bahamian Creole (cf. Hackert, 2004: 180), i.e., a depressing effect of frequency adverbials on rates of past inflection (Fig. 2.5). That this is a creole-specific effect appears doubtful, however, as frequency adverbials tend to co-occur with the grammatical aspect of habituality, which has been found to dampen rates of past marking in all varieties for which it has been tested (cf. Section 2.2). A noteworthy pattern is also found in ICE-HK, where adverbials indicating a point of time are associated with particularly low rates of past inflection (Fig. 2.5). It is interesting to note in this context that ICE-HK contains a higher proportion of such adverbials by overall number of verb tokens (23%) than any other corpus (ICE-IND: 18%, ICE-JAM: 16%, ICE-PHI: 20%).<sup>9</sup> This finding once more characterizes the ICE-HK conversations as having been produced by users at a lower level of proficiency in English. Tense marking is one way of expressing temporality, but it interacts with other linguistic devices, including lexical and grammatical aspect, discourse-pragmatic principles, and – most notably – adverbials. Not all languages have grammaticalized tense categories, but all possess a rich inventory of temporal adverbials, and L2 learners fundamentally rely on them in their acquisition of tense-aspect morphology. Particularly important in this respect are point-of-time adverbials, which are not only very frequent but may also express all kinds of temporal relations in early L2 acquisition (Dietrich et al., 1993: 80). In fact, "[w]hen reference to the past is first expressed explicitly, it

<sup>8</sup> Of course, there is always the possibility that difficult-to-detect phonological phenomena such as  $\tau$ D-deletion are not entirely accurately represented in the orthographic transcripts in which the majority of ICE corpora are available to the research community. That consistent and well-attested patterns did emerge from our phonological analysis after all endowed us with some confidence in both our material and this analysis, but of course its results must be taken with a grain of salt.

<sup>9</sup> The differences between ICE-HK and ICE-PHI are not statistically significant ( $p = 0.2716$ ), but the differences between ICE-HK and ICE-IND ( $p = 0.05$ ) and ICE-JAM ( $p < 0.001$ ) are.

is expressed exclusively by adverbial expressions and connectives” (Bardovi-Harlig, 1992: 300). As learner language matures and verbal morphology becomes more stable and systematic, tense marking increases, while the reliance on adverbials decreases (1992: 315). Clearly, of all speaker groups sampled here, the ICE-HK speakers rely most on temporal adverbials and least on past inflection, but this perfectly matches their self-estimation as not-so-competent speakers of English reported above.

All varieties under study here align in three important dimensions. The first is grammatical aspect, with perfective verb situations invariably favoring and habituals disfavoring past inflection (Fig. 2.2). As described in Section 2.2, this effect has been found in all varieties for which the predictor has been tested. Hackert (2004: 170-171) invoked typology to explain this finding for her urban Bahamian Creole data, noting that habituals resemble generics, which tend to be formally unmarked in the world’s languages. This argument is supported by other data from non-creole varieties of English (e.g., Gut, 2009: 269; Biewer, 2015: 240; Bohmann and Babalola, 2023: 31; Hackert et al., *fc.*) and the results of the present analysis. What remains to be explained is the strong effect that grammatical aspect has in IndE and, to a slightly lesser extent, HKE. Recall that Sharma (2009) dismisses a universalist explanation for this pattern in her Indian and Singaporean English data, arguing that both varieties have been shaped by the “perfectivity-marking systems” of their Indo-Aryan and Chinese substrate languages (2009: 177). We cannot directly transfer this argument to the data at hand, as the set of ICE-IND L1s is dominated by Dravidian languages,<sup>10</sup> and Cantonese is the only L1 of the ICE-HK speakers, in contrast to Singaporean English, where Mandarin plays an important role. That said, we find morphological encoding of the perfective aspect, often by means of suffixes, not just in the Indo-Aryan ICE-IND substrates Marathi and Hindi but also in the Dravidian ones Kannada and Tamil (Kulkarni-Joshi, 2017: 174-6), which might be argued to constitute an obvious template for the application of the standard English past inflection by speakers of these languages. Like Mandarin, Cantonese is an aspect-prominent language that possesses an overt perfective marker, *jó*, which follows the verb. Unlike the former, however, it also formally encodes the habitual aspect by means of a postverbal element, *hōi*. Both “are bound forms, behaving essentially as suffixes” (Matthews and Yip, 2011: 228). Hindi, Marathi, and Tamil also mark imperfective aspect(s) formally; Kannada does not (Kulkarni-Joshi, 2017: 174-176). It appears plausible, then, to interpret the strong effect of grammatical aspect on past inflection in IndE and HKE as owed to the presence of a perfective/imperfective distinction in the aspectual systems of all substrate languages and their predilection for marking perfective verb situations with a suffix(-like) element, compounded by the universal preference for non-marking habitual verb situations, which is also evident in JamE and PhilE as well as in numerous other varieties of English.

The varieties are also alike with respect to the effects of persistence and frequency, with both preceding marking and a higher token frequency favoring inflection in all of them (Figs. 2.3 and 2.4). As noted by Szmrecsanyi (2005), “[l]anguage users [...] are creatures of habit,” who tend to recycle material they have heard or used recently, which leads to a variety of phenomena, from morphological persistence in variable patterns of past marking as studied here over lexical priming to persistence in syntactic structures such as the genitive alternation (cf. Szmrecsanyi, 2006: 10; Tammings, 2016: 336). Persistence has been described as a facet of the principle of “Plan Reuse,” according to which speakers minimize the cognitive effort they spend on planning and producing utterances by reactivating “recently executed utterance plans” (MacDonald, 2013: 4). While morphosyntactic persistence is generally described in terms of short-term repetition, Plan Reuse has been shown to extend across both intervening material and time and thus appears to involve not merely short-term memory but also “long-term implicit learning.” In this way, the principle might even underlie the strong cognitive entrenchment of high-frequency past forms to be discussed below. Clearly, it affects language production, linguistic variation, and diachronic development universally and even “appears in many non-linguistic motor behaviors in humans and animals” (2013: 4); it is therefore not surprising that all speakers analyzed here should adhere to it consistently, albeit to different degrees.

It is interesting to note that persistence has the strongest effect in ICE-JAM, where it is not only statistically highly significant (Table 1.3) but also figures as the sole important predictor in the random forest analysis (Fig. 4.3). A potential reason behind this finding is the nature of the Jamaican corpus data. As outlined above, the ICE-JAM conversations consist of private encounters that were relatively informal in nature. A closer inspection of the contents of these conversations revealed that they also contained numerous animated narratives. Variation in tense marking in such narratives, in both English (Schiffrin, 1981) and English-lexifier pidgins and creoles (Tagliamonte, 2000; Hackert, 2004) as well as in other languages (Silva-Corvalán, 1983; Fleischman, 1985), is not haphazard but involves verbs of the same tense clustering together. Narratives thus constitute a textual environment particularly conducive to the application of the persistence principle in tense marking, and it appears plausible to ascribe the disproportionate importance of this predictor

<sup>10</sup> As Lange (2012: 83) outlines, speakers of the Dravidian languages Kannada (ca. 20%), Tamil (13%), Malayalam (8%), and Telugu (7%) account for roughly half the ICE-IND conversations, with the Indo-Aryan languages Marathi (ca. 20%), Punjabi (7%), and Hindi (6%) comprising another third of speaker L1s.

in the Jamaican data vis-à-vis the other datasets to a higher-than-usual share of connected narratives in the ICE-JAM conversations.

While frequency, in the form of variable rule application, has always played an important role in variationist linguistics (cf. Section 2.1), frequency effects in grammar only came to the fore with the emergence of usage-based approaches to language. Today, there is general agreement that “frequency is one of the main determinants for the emergence of linguistic structure and the organization of our grammatical knowledge” (Diessel and Hilpert, 2016), and, in fact, the predictor has the same, if not always statistically significant (Table 1), effect in all our datasets: verb types with higher token numbers are more likely to be past-inflected than those with lower frequencies (Fig. 2.4). This holds true regardless of morpho-phonological verb class, and we find a positive correlation between verb frequency and marking probability for irregular as well as for consonant-final and other regular verbs.<sup>11</sup> For irregular verbs, our finding is easily married to the idea that past forms like *went*, *came*, *told*, *got*, etc. possess great “lexical strength” (Bybee, 1995: 428), which makes them “easy to access” in language production and “resistant to change.” In other words, verbs like *GO*, *COME*, *TELL*, *GET*, etc. occur primarily in inflected form because *went*, *came*, *told*, *got*, etc. are exceptionally frequent, which results in their early acquisition in language learning, non-eligibility to analogical leveling in language change, and preferential selection in a speaker’s choice of variants. What is remarkable about our data is that the positive correlation between token frequency and marking probability also holds for consonant-final regular verbs, for which research on TD-deletion would have predicted otherwise.<sup>12</sup> While a phonological account thus does not explain the frequency effect we find for those verbs, a usage-based one does. It predicts that frequent regular morphological forms are stored in the lexicon just like irregulars (cf. Section 3.4). “This follows from the lexical strength proposal – for regular forms just as for irregular forms, lexical strength varies according to frequency of use” (Bybee, 1995: 450). Hence, frequent verb forms, whether irregular or regular, are strongly entrenched cognitively. They are both acquired first in language learning and activated more readily later in production than infrequent ones, regardless of morpho-phonological class.

In fact, it stands to reason that this predictor, just like stativity, is in large part an artifact of the marking propensity of individual high-frequency lexical items, or, as Walker (2012: 410) puts it, “frequency categories overlap with morphological categories, which themselves consist of sets of individual lexical items that have different preferences” in terms of variable phonological and morphosyntactic processes. If viewed from this angle, the effect of verb frequency on variable past marking appears as a facet of the “Easy First” principle of language production, whereby language users show a predilection for easily planned elements. “‘Easier’ (also termed more accessible or available) words and phrases have been described as more frequent,” among other things (MacDonald, 2013: 3), and while variable past marking is not the kind of sequential syntactic operation in focus in MacDonald (2013) and visible in, e.g., the genitive alternation (cf. Heller et al., 2017; Hackert and Wengler, 2022), it is clear that morphological processes such as past inflection also crucially depend on it. “New plan development in turn relies on retrieval from long term memory, and when this retrieval fails or requires extra time, production is delayed or derailed” (MacDonald, 2013: 3). Together with Plan Reuse, Easy First cognitively organizes language production in a larger framework of statistically-based utterance planning biases, which underlie not only the variable morphological process at stake in the present study but also numerous other distributional and ultimately typological characteristics of language(s).

## 6. CONCLUSION

This study has investigated variable past inflection in parallel corpora for four New Englishes with different postcolonial histories and sociolinguistic backgrounds. The aim was to uncover in how far the Englishes of Hong Kong, India, Jamaica, and the Philippines share a grammar when it comes to this variable, taking complex patterns of constraints rather than simple variant frequency as indicative of underlying rules. Our analysis showed that those patterns are remarkably similar across varieties, with grammatical aspect, marker persistence, and verb frequency exerting effects in the same direction in all of them. Importantly, these constraints have also been found to operate in numerous other varieties of English, including creoles, high-contact L1 dialects, and other New Englishes. This suggests that there is a core grammar of variable past inflection in English, which is governed precisely by these three constraints, which are structural and cognitive in nature and instantiate crosslinguistic principles of grammatical organization. We have suggested that where individual varieties (in particular, IndE and HKE) pattern differently from the others, these differences

<sup>11</sup> Correlation coefficients range between 0.28 and 0.73. The correlations are all statistically significant; in ICE-HK, they show the greatest dispersion of values. They were calculated without categorically marked verbs, as these can be of any frequency and thus constitute a disproportionately large group of outliers, particularly in ICE-IND and ICE-PHI.

<sup>12</sup> The question of whether TD-deletion operates more strongly in more frequent words has been debated controversially. Clearly, high-frequency content words like *just* and function words like *and* are subject to exceptional rates of deletion, which is why they are sometimes left out of analyses entirely. For past-tense forms, the effect has been found to be either bound to “a small class of lexical items” (Walker, 2012: 399) or not statistically significant (Baranowski and Turton, 2020: 17).

can often plausibly be attributed to aspects of second-language acquisition and variety status, a prime example being the effect of lexical aspect in HKE. Of course, substrate influence also needs to be reckoned with, as evidenced in the boost that speakers' perfectivity-marking L1s give to the effect of grammatical aspect in HKE and, even more so, IndE. Stativity, a long-time favorite in analyses of past marking in creoles and related varieties, emerged as surprisingly unstable in our corpora and least significant in JamE, where it would have been expected most.

Our findings further substantiate the potential of the comparative sociolinguistic approach for the exploration of what unites English speakers worldwide as well as how language contact operates on the level of directions of effects and constraint hierarchies. In this way, the study reaffirms one of the initial principles of comparative sociolinguistics that variation encodes linguistic history and relationships between varieties. This perspective may usefully be brought to bear on the traditional corpus-linguistic approach to World Englishes that underlies many of the efforts at modeling the historical spread and current status of varieties of the language. The findings also add to a growing body of evidence that patterns of grammatical variation once thought to be unique to creoles, for example, are in fact unexceptional. They provide a useful line of enquiry for further examinations of variable past marking in other postcolonial Englishes and of other variation phenomena found in varieties worldwide.

### CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

**Stephanie Hackert:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Conceptualization. **Catherine Laliberté:** Writing – review & editing, Data curation. **Diana Wengler:** Visualization, Software, Formal analysis.

### Data availability

Data will be made available on request.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### APPENDIX

Table 2

Absolute frequencies and percentages of past inflection by predictor, predictor level, and corpus.

		ICE-HK		ICE-IND				ICE-JAM				ICE-PHI					
		Unmarked		Inflected		Unmarked		Inflected		Unmarked		Inflected		Unmarked			
Inflected		N	%	N	%	N	%	N	%	N	%	N	%	N	%		
verb class	cons.-final regular	87	52	80	48	18	12	126	88	62	30	145	70	8	3	236	97
	irregular	231	33	462	67	23	4	549	96	157	17	785	83	30	3	921	97
	other regular	99	57	75	43	8	4	172	96	38	16	203	84	16	6	256	94
prec. phon. env.	NA	330	38	537	62	31	4	721	96	195	16	988	84	46	4	1177	96
	non-sibilant fricative, lateral	13	42	18	58	3	20	12	80	16	33	33	67	3	8	34	92
	sibilant, stop, nasal	74	54	62	46	15	12	114	88	46	29	112	71	5	2	202	98
foll. phon. env.	NA	330	38	537	62	31	4	721	96	195	16	988	84	46	4	1177	96
	cons.	37	49	39	51	6	11	48	89	35	37	60	63	4	3	138	97
	vowel	41	58	30	42	8	13	54	87	25	28	64	72	2	3	76	97
	pause	9	45	11	55	4	14	24	86	2	9	21	91	2	8	22	92
lexical aspect	accomp.	67	39	104	61	7	4	160	96	48	18	225	82	3	2	196	98
	achieve-ment	161	32	342	68	26	6	419	94	109	20	435	80	10	1	709	99
	activity	84	58	62	42	5	5	98	95	42	25	127	75	7	4	168	96
	semel-factive	2	67	1	33	0	0	3	100	1	25	3	75	0	0	2	100
	stative	103	49	108	51	11	6	167	94	57	14	343	86	34	9	338	91
gramm. aspect	habitual	110	58	81	42	22	14	138	86	89	21	341	79	33	10	303	90
	pfv.	307	36	536	64	27	4	709	96	168	18	792	82	21	2	1110	98

<b>adv.</b>	<b>duration</b>	6	27	16	73	0	0	20	100	3	13	20	87	4	14	24	86
	<b>frequency</b>	9	35	17	65	3	14	18	86	17	28	44	72	7	19	29	81
	<b>none</b>	289	39	460	61	38	5	653	95	198	18	891	82	37	3	1058	97
	<b>point of time</b>	113	48	124	52	8	5	156	95	39	18	178	82	6	2	302	98
<b>persist-ence</b>	NA	38	43	51	57	3	3	83	97	13	15	72	85	3	3	86	97
	marked	180	32	386	68	38	5	731	95	138	13	924	87	39	3	1288	97
	unmarked	199	53	180	47	8	20	33	80	106	44	137	56	12	24	39	76

Table 3  
Model summaries for the four corpora.

Corpus		Coefficient	SE	z-value	p
<b>ICE-HK</b>	(Intercept)	-0.233	0.955	-0.244	0.807
	lex_aspectachievement	0.068	0.351	0.193	0.847
	lex_aspectactivity	-0.637	0.395	-1.611	0.107
	lex_aspectstative	0.328	0.422	0.778	0.436
	gramm_aspectperfective	0.855	0.279	3.065	0.002
	persistenceNA	-0.155	0.297	-0.522	0.601
	persistenceunmarked	-0.067	0.192	-0.349	0.727
	freq_log	0.068	0.103	0.659	0.510
	temp_advfrequency	-0.420	0.819	-0.513	0.608
	temp_advnone	-1.007	0.633	-1.592	0.111
	temp_advpoint of time	-1.380	0.648	-2.131	0.033
	verb_classirregular	0.621	0.397	1.565	0.117
	verb_classother regular	0.106	0.401	0.265	0.791
<b>ICE-IND</b>	(Intercept)	-1.711	1.432	-1.195	0.232
	lex_aspectachievement	0.092	0.492	0.186	0.852
	lex_aspectactivity	1.275	0.739	1.726	0.084
	lex_aspectstative	1.427	0.673	2.120	0.034
	gramm_aspectperfective	2.612	0.524	4.989	<0.001
	persistenceNA	0.579	0.682	0.849	0.396
	persistenceunmarked	-0.727	0.553	-1.314	0.189
	freq_log	0.215	0.121	1.777	0.076
	temp_advfrequency	0.107	1.410	0.076	0.940
	temp_advnone	0.195	1.159	0.168	0.867
	temp_advpoint of time	0.372	1.210	0.308	0.758
	verb_classirregular	1.051	0.469	2.241	0.025
	verb_classother regular	1.500	0.527	2.847	0.004
<b>ICE-JAM</b>	(Intercept)	0.919	0.987	0.931	0.352
	lex_aspectachievement	0.286	0.298	0.958	0.338
	lex_aspectactivity	-0.021	0.360	-0.057	0.954
	lex_aspectstative	0.546	0.405	1.347	0.178
	gramm_aspectperfective	0.571	0.271	2.108	0.035
	persistenceNA	-0.178	0.371	-0.480	0.632
	persistenceunmarked	-0.879	0.209	-4.203	<0.001
	freq_log	0.142	0.084	1.697	0.090
	temp_advfrequency	-0.809	0.870	-0.931	0.352
	temp_advnone	-0.239	0.798	-0.300	0.764
	temp_advpoint of time	-0.304	0.822	-0.370	0.712
	verb_classirregular	0.150	0.360	0.417	0.676
	verb_classother regular	0.769	0.384	2.005	0.045
<b>ICE-PHI</b>	(Intercept)	2.500	1.383	1.807	0.071
	lex_aspectachievement	-0.045	0.801	-0.056	0.956
	lex_aspectactivity	-0.898	0.849	-1.058	0.290
	lex_aspectstative	-1.500	0.846	-1.772	0.076
	gramm_aspectperfective	0.987	0.466	2.121	0.034
	persistenceNA	-0.424	0.667	-0.635	0.525
	persistenceunmarked	-1.211	0.535	-2.262	0.024
	freq_log	0.101	0.135	0.748	0.454
	temp_advfrequency	-0.289	0.935	-0.309	0.757
	temp_advnone	1.653	0.737	2.242	0.025
	temp_advpoint of time	2.127	0.854	2.492	0.013
	verb_classirregular	-0.440	0.717	-0.613	0.540
	verb_classother regular	-0.337	0.656	-0.513	0.608

## References

- Andersen, R.W., Shirai, Y., 1996. The primacy of aspect in first and second language acquisition: The pidgin-creole connection. In: Ritchie, W.C., Bhatia, T.K. (Eds.), *Handbook of Second Language Acquisition*. Academic Press, San Diego, pp. 527–570.
- Baayen, R. H., Piepenbrock, R., van Rijn, H., 1993. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania. [CELEX lexical database (CD-ROM)].
- Baranowski, M., Turton, D., 2020. TD-deletion in British English: New evidence for the long-lost morphological effect. *Lang. Var. Chang.* 32, 1–23.
- Bardovi-Harlig, K., 1992. The use of adverbials and natural order in the development of temporal expression. *Int. Rev. Appl. Linguist. Lang. Teach.* 30, 299–320.
- Bardovi-Harlig, K., Comajoan-Colomé, L., 2020. The aspect hypothesis and the acquisition of L2 past morphology in the last 20 years: A state-of-the-scholarship review. *Stud. Second. Lang. Acquis.* 42, 1137–1167.
- Bates, D., Maechler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48.
- Bautista, M., Lourdes, S., 2004a. An overview of the Phillipine component of the International Corpus of English (ICE-PHI). *Asian Englishes* 72, 8–26.
- Bautista, M., Lourdes, S., 2004b. Tagalog-English code switching as a mode of discourse. *Asia Pac. Educ. Rev.* 5, 226–233.
- Bernstein, B., 1971. *Class, Codes and Control: Theoretical Studies towards a Sociology of Language*. Routledge & Kegan Paul, London.
- Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., 1999. *Longman Grammar of Spoken and Written English*. Pearson Education Ltd, Harlow.
- Bickerton, D., 1975. *Dynamics of a Creole System*. Cambridge University Press, Cambridge.
- Bickerton, D., 1981. *Roots of Language*. Karoma, Ann Arbor.
- Biewer, C., 2015. *South Pacific Englishes: A Sociolinguistic and Morphosyntactic Profile of Fiji English, Samoan English and Cook Islands English*, Benjamins, Amsterdam/Philadelphia.
- Bohmann, A., Babalola, A., 2023. Verbal past inflection in Nigerian English: A case for sociolinguistic compound vision. In: Wilson, G., Westphal, M. (Eds.), *New Englishes, New Methods*. Benjamins, Amsterdam/Philadelphia, pp. 16–41.
- Bolt, P., Bolton, K., 1996. The international corpus of English in Hong Kong. In: Greenbaum, S. (Ed.), *Comparing English Worldwide: the International Corpus of English*. Clarendon Press, Oxford, pp. 197–214.
- Bolton, K., 2000. The sociolinguistics of Hong Kong and the space for Hong Kong English. *World Englishes* 19, 265–285.
- Bolton, K., 2003. *Chinese Englishes. A Sociolinguistic History*. Cambridge University Press, Cambridge.
- Buschfeld, S., Kautzsch, A., 2017. Towards an integrated approach to postcolonial and non-postcolonial Englishes. *World Englishes* 36, 104–126.
- Bybee, J., 1995. Regular morphology and the lexicon. *Lang. Cognit. Process.* 10, 425–455.
- Comrie, B., 1976. *Aspect: An Introduction to the Study of Verbal Aspect and Related Problems*. Cambridge University Press, Cambridge.
- Dahl, Ö., 1995. Areal tendencies in tense-aspect systems. In: Bertinetto, P.M., Bianchi, V., Dahl, Ö., Squartini, M. (Eds.), *Temporal Reference, Aspect and Actionality*. Rosenberg & Sellier, Torino, pp. 11–28.
- Davies, M., Fuchs, R., 2015. Expanding horizons in the study of World Englishes with the 1.9 billion word Global Web-based English Corpus (GloWbE). *English World-Wide* 36, 1–28.
- Deuber, D., 2009. 'The English we speaking' – Morphological and syntactic variation in educated Jamaican speech. *J. Pidgin Creole Lang.* 24, 1–52.
- Deuber, D., 2014. *English in the Caribbean: Variation, Style and Standards in Jamaica and Trinidad*. Cambridge University Press, Cambridge.
- Devonish, H., Thomas, E.A.C., 2012. Standards of English in the Caribbean. In: Hickey, R. (Ed.), *Standards of English: Codified Varieties around the World*. Cambridge University Press, Cambridge, pp. 179–197.
- Diessel, H., Hilpert, M., 2016. Frequency effects in grammar. In: *Oxford Research Encyclopedia of Linguistics* <https://doi.org/10.1093/acrefore/9780199384655.013.120> (accessed December 15, 2023).
- Dietrich, R., Klein, W., Noyau, C., 1993. The acquisition of temporality. In: Perdue, C. (Ed.), *Adult Language Acquisition: Cross-linguistic Perspectives*, vol. 2 *The Results*. Cambridge University Press, Cambridge, pp. 73–118.
- Fasold, R.W., 1972. *Tense Marking in Black English: A Linguistic and Social Analysis*. Center for Applied Linguistics, Washington D.C..
- Filip, H., 2012. Lexical aspect. In: Binnick, R.I. (Ed.), *The Oxford Handbook of Tense and Aspect*. Oxford University Press, Oxford, pp. 721–751.
- Fleischman, S., 1985. Discourse functions of tense-aspect oppositions in narrative: Toward a theory of grounding. *Linguistics* 23, 851–882.
- Fox, J., 2003. Effect displays in R for generalized linear models. *J. Stat. Softw.* 8 (15), 1–27.
- Fox, J., Weisberg, S., 2019. *An R Companion to Applied Regression*. Sage, Thousand Oaks, CA.
- Görlach, M., 2002. *Still More Englishes*. Benjamins, Amsterdam/Philadelphia.
- Greenbaum, S., 1990. Standard English and the international corpus of English. *World Englishes* 9, 79–83.
- Greenbaum, S. (Ed.), 1996. *Comparing English Worldwide: the International Corpus of English*. Clarendon Press, Oxford.

- Gries, S.T., Bernaisch, T.J., Heller, B., 2018. A corpus-linguistic account of the history of the genitive alternation in Singapore English. In: Deshors, S.C. (Ed.), *Modeling World Englishes: Assessing the Interplay of Emancipation and Globalization of ESL Varieties*. Benjamins, Amsterdam/Philadelphia, pp. 245–279.
- Gut, U., 2007. First language influence and final consonant clusters in the new Englishes of Singapore and Nigeria. *World Englishes* 26, 346–359.
- Gut, U., 2009. Past tense marking in Singapore English verbs. *English World-Wide* 30, 262–277.
- Guy, G.R., Boberg, C., 1997. Inherent variability and the obligatory contour principle. *Lang. Var. Chang.* 9, 149–164.
- Hackert, S., 2004. *Urban Bahamian Creole: System and Variation*. Benjamins, Amsterdam/Philadelphia.
- Hackert, S., 2008. Counting and coding the past: Circumscribing the variable context in quantitative analyses of past inflection. *Lang. Var. Chang.* 20, 127–153.
- Hackert, S., 2019. *Memoirs from Central America: A linguistic analysis of personal recollections of West Indian laborers in the construction of the Panama Canal*. In: Hickey, R. (Ed.), *Keeping in Touch. Emigrant Letters across the English-Speaking World*, Benjamins, Amsterdam/Philadelphia, pp. 261–285.
- Hackert, S., Laliberté, C., Mailhammer, R., Wengler, D., Zeidan, R. Fc. Past marking in Australian Aboriginal English on Croker Island: Local vs. cross-variety patterns and principles.
- Hackert, S., Wengler, D., 2022. Recent grammatical change in postcolonial Englishes: A real-time study of genitive variation in Caribbean and Indian news writing. *Journal of English Linguistics* 50 (1), 3–38. <https://doi.org/10.1177/00754242211052490>.
- Heller, B., Bernaisch, T.J., Gries, S.T., 2017. Empirical perspectives on two potential epicenters: The genitive alternation in Asian Englishes. *ICAME Journal* 41, 111–144.
- Ho, M., Platt, J., 1993. *Dynamics of a Contact Continuum: Singaporean English*. Clarendon Press, Oxford.
- Hothorn, T., Buehlmann, P., Dudoit, S., Molinaro, A., van der Laan, M., 2006. Survival ensembles. *Biostatistics* 7, 355–373.
- Hothorn, T., Bretz, F., Westfall, P., 2008. Simultaneous inference in general parametric models. *Biom. J.* 50, 346–363.
- Jiráňková, L., Cilibrasi, L., 2021. Second-language acquisition of the English past-tense: From rules to analogy. *Linguistica Pragensia* 31, 188–213.
- Kachru, B.B. (Ed.), 1982. *The Other Tongue: English across Cultures*. Pergamon, Oxford.
- Kortmann, B., Burridge, K., Mesthrie, R., Schneider, E.W., Upton, C. (Eds.), 2004. *A Handbook of Varieties of English*, 2 vols. De Gruyter Mouton, Berlin, Boston.
- Kortmann, B., Lunkenheimer, K., Ehret, K. (Eds.), 2020. *The Electronic World Atlas of Varieties of English*. <https://www.ewaveatlas.org> (accessed March 31, 2023).
- Kučera, H., Francis, W.N., 1967. *Computational Analysis of Present-day American English*. Brown University Press, Providence.
- Kulkarni-Joshi, S., 2017. Verbal categories in Marathi (with side glances at Hindi, Kannada and Tamil) 77, 171–190.
- Labov, W., 1969. Contraction, deletion, and inherent variability of the English copula. *Language* 45, 715–762.
- Labov, W., 1972. *Sociolinguistic Patterns*. University of Pennsylvania Press, Philadelphia.
- Labov, W., 1984. Field methods of the project on linguistic change and variation. In: Baugh, J., Sherzer, J. (Eds.), *Language in Use: Readings in Sociolinguistics*. Prentice-Hall, Englewood Cliffs, NJ, pp. 28–53.
- Labov, W., 1989. The child as linguistic historian. *Lang. Var. Chang.* 1, 85–94.
- Labov, W., Waletzky, J., 1967. Narrative analysis: Oral versions of personal experience. In: Helm, J. (Ed.), *Essays on the Verbal and Visual Arts: Proceedings of the 1966 Annual Spring Meeting of the American Ethnological Society*. University of Washington Press, Seattle, WA, pp. 12–44.
- Labov, W., Cohen, P., Robins, C., Lewis, J., 1968. *A Study of the Non-Standard English of Negro and Puerto Rican Speakers in New York City*. Columbia University, New York.
- Lange, C., 2012. *The Syntax of Spoken Indian English*. Benjamins, Amsterdam/Philadelphia.
- Lüdecke, D., 2018. Ggeffects: Tidy data frames of marginal effects from regression models. *J. Open Source Softw.* 3 (26), 772.
- Lüdecke, D., Ben-Shachar, M.S., Patil, I., Waggoner, P., Makowski, D., 2021. Performance: An R package for assessment. *J. Open Source Softw.* 6 (60), 3139.
- MacDonald, M.C., 2013. How language production shapes language form and comprehension. *Front. Psychol.* 4, 1–16.
- Mailhammer, R., 2021. *English on Croker Island: The Synchronic and Diachronic Dynamics of Contact and Variation*. De Gruyter Mouton, Berlin, Boston.
- Majumdar, M., Mooij, J., 2012. The mark's race: India's dominant education regime and new segmentation. In: Sleeter, C., Upadhyay, S.B., Mishra, A.K., Kumar, S. (Eds.), *School Education, Pluralism and Marginality. Comparative Perspectives*, Orient Blackswan, New Delhi, pp. 207–246.
- Matthews, S., Yip, V., 2011. *Cantonese: A Comprehensive Grammar*. Routledge, New York.
- Mukherjee, J., Schilk, M., 2008. Verb-complementational profiles across varieties of English: comparing verb classes in Indian English and British English. In: Nevalainen, T., Taavitsainen, I., Pahta, P., Korhonen, M. (Eds.), *The Dynamics of Linguistic Variation: Corpus Evidence on English past and Present*. Benjamins, Amsterdam/Philadelphia, pp. 163–181.
- Oenbring, R., Fielding, W., 2014. Young adult's attitudes to standard and nonstandard English in an English-creole speaking country: The case of the Bahamas. *Lang. Discourse Soc.* 3, 28–51.
- Patrick, P., 1991. Creoles at the intersection of variable processes: -t, d deletion and past-marking in the Jamaican mesolect. *Lang. Var. Chang.* 3, 171–189.
- Patrick, P., 1999. *Urban Jamaican Creole: Variation in the Mesolect*. Benjamins, Amsterdam/Philadelphia.
- Poplack, S., 2000. Introduction. In: Poplack, S. (Ed.), *The English History of African American English*. Blackwell, Oxford, pp. 1–34.

- Poplack, S., Tagliamonte, S., 2001. *African American English in the Diaspora*. Blackwell, Oxford.
- Purse, R., Fruehwald, J., Tamminga, M., 2022. Frequency and morphological complexity in variation. *Glossa* 7, 1–34.
- Quirk, R., Greenbaum, S., Leech, G., Svartvik, J., 1985. *A Comprehensive Grammar of the English Language*. Longman, London.
- Revis, M., Bernaisch, T., 2020. The pragmatic nativisation of pauses in Asian Englishes. *World Englishes* 39, 135–153.
- Rickford, J.R., 1999. *African American Vernacular English: Features, Evolution. Educational Implications*, Blackwell, Oxford.
- Schiffrin, D., 1981. Tense variation in narrative. *Language* 57, 45–62.
- Schneider, E.W., 2003. The dynamics of New Englishes: From identity construction to dialect birth. *Language* 79, 233–281.
- Schneider, E.W., 2007. *Postcolonial English: Varieties around the World*. Cambridge University Press, Cambridge.
- Sharma, D., 2009. Typological diversity in New Englishes. *English World-Wide* 30, 170–195.
- Sharma, D., 2012. Indian English. In: Kortmann, B., Lunkenheimer, K. (Eds.), *The Mouton World Atlas of Variation in English*. De Gruyter Mouton, Berlin, Boston, pp. 523–530.
- Shirai, Y., 2007. The aspect hypothesis, the comparative fallacy and the validity of obligatory context analysis: A reply to Lardiere. *Second. Lang. Res.* 23, 51–64.
- Silva-Corvalán, C., 1983. Tense and aspect in oral Spanish narrative: Context and meaning. *Language* 59, 760–780.
- Smith, C.S., 1997. *The Parameter of Aspect*. Kluwer, Dordrecht.
- Strobl, C., Boulesteix, A.L., Kneib, T., Augustin, T., Zeileis, A., 2008. Conditional variable importance for random forests (accessed October 31, 2020) *BMC Bioinf.* 9, 307 <http://www.biomedcentral.com/1471-2105/9/307>.
- Szmrecsanyi, B., 2005. Language users as creatures of habit: a corpus-linguistic analysis of persistence in spoken English. *Corpus Linguist. Linguist. Theory* 1, 113–150.
- Szmrecsanyi, B., 2006. *Morphosyntactic Persistence in Spoken English: A Corpus Study at the Intersection of Variationist Sociolinguistics, Psycholinguistics, and Discourse Analysis*. Mouton de Gruyter, Berlin, New York.
- Szmrecsanyi, B., 2017. Variationist sociolinguistics and corpus-based variationist linguistics: Overlap and cross-pollination potential. *Can. J. Linguist./revue Canadienne De Linguistique* 62, 1–17.
- Szmrecsanyi, B., Grafmiller, J., 2023. *Comparative Variation Analysis: Grammatical Alternations in World Englishes (Studies in Language Variation and Change)*. Cambridge University Press, Cambridge New York, NY.
- Szmrecsanyi, B., Kortmann, B., 2009. The morphosyntax of varieties of English worldwide: A quantitative perspective. *Lingua* 119, 1643–1663.
- Szmrecsanyi, B., Grafmiller, J., Heller, B., Röthlisberger, M., 2016. Around the world in three alternations: Modeling syntactic variation in varieties of English. *English World-Wide* 37, 109–137.
- Tagliamonte, S.A., 1991. *A Matter of Time: Past Temporal Reference Verbal Structures in Samaná ´ English and the Ex-Slave Recordings*. University of Ottawa, Ph.D. dissertation.
- Tagliamonte, S., 2000. The story of *kom* in Nigerian Pidgin English. In: McWhorter, J.H. (Ed.), *Language Change and Language Contact in Pidgins and Creoles*. Benjamins, Amsterdam/Philadelphia, pp. 353–382.
- Tagliamonte, S.A., 2013. Comparative sociolinguistics. In: Chambers, J.K., Schilling, N. (Eds.), *The Handbook of Language Variation and Change*. second ed. Wiley-Blackwell, Chichester, pp. 128–156.
- Tagliamonte, S.A., Baayen, R.H., 2012. Models, forests and trees of York English: Was/were variation as a case study for statistical practice. *Lang. Var. Chang.* 24, 135–178.
- Tagliamonte, S.A., Poplack, S., 1993. The zero-marked verb: Testing the creole hypothesis. *J. Pidgin Creole Lang.* 8, 171–206.
- Tagliamonte, S., Temple, R., 2005. New perspectives on an *ol'* variable: (t, d) in British English. *Lang. Var. Chang.* 17, 281–302.
- Tamminga, M., 2016. Persistence in phonological and morphological variation. *Lang. Var. Chang.* 28, 335–356.
- Tomaschek, F., Hendrix, P., Baayen, R.H., 2018. Strategies for addressing collinearity in multivariate linguistic data. *J. Phon.* 71, 249–267.
- van Heuven, W.J.B., Mandera, P., Keuleers, E., Brysbaert, M., 2014. SUBTLEX-UK: A new and improved word frequency database for British English. *Q. J. Exp. Psychol.* 67, 1176–1190.
- Vendler, Z., 1957. Verbs and times. *Philos. Rev.* 66, 143–160.
- Walker, J.A., 2012. Form, function, and frequency in phonological variation. *Lang. Var. Chang.* 24, 397–415.
- Weldon, T.L., 1996. Past marking in Gullah. *UPenn Work. Pap. Linguist.* 3, 63–72.
- Winford, D., 1992. Back to the past: The BEV/creole connection revisited. *Lang. Var. Chang.* 4, 311–357.
- Wolfram, W., 1969. *A Sociolinguistic Description of Detroit Negro speech*. Center for Applied Linguistics, Washington, DC.