

Past Marking in Australian Aboriginal English on Croker Island: Local Versus Cross-Variety Patterns and Principles

Journal of English Linguistics

1–31

© The Author(s) 2025




Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/00754242241298990

journals.sagepub.com/home/eng

Stephanie Hackert¹ , Catherine Laliberté¹ ,
Robert Mailhammer², Diana Wengler¹ ,
and Ronia Zeidan²

Abstract

In this paper, we investigate variable past marking in Australian Aboriginal English as spoken on Croker Island, Northern Territory. Employing data from twenty speakers and both mixed-effects regression and random forests, we show that despite a high degree of individual variability the occurrence or non-occurrence of a past-marked verb is subject to conditioning factors that are known from other varieties of English, most notably lexical and grammatical aspect and marker persistence. Moreover, the constraints governing the preverbal marker *bin* relate in systematic ways to those governing inflection. Our results suggest that the specifics of contact influence may be less relevant to explaining variable linguistic processes such as past marking than more general discourse-pragmatic and cognitive principles of language variation and change. This has implications for the debate about the uniqueness of creole languages, which have often been considered a language type like no other.

Keywords

Australian Aboriginal English, variable past inflection, preverbal *bin*, creole, second-language acquisition, mixed models, random forests, stativity, Lexical Aspect Hypothesis, habituality, persistence, syllable-final consonant cluster reduction

¹LMU Munich, Munich, Germany

²Western Sydney University, Sydney, NSW, Australia

Corresponding Author:

Stephanie Hackert, LMU Munich, Schellingstr. 3, Munich 80799, Germany.

Email: stephanie.hackert@anglistik.uni-muenchen.de

I. Introduction

Past marking is one of the best-studied morphosyntactic variables in English. Studies have examined the use of inflected versus unmarked past-temporal reference verbs in second-language (L2) acquisition (e.g., Wolfram & Hatfield 1984; Bayley 1996; Hawkins & Liszka 2003), English as a lingua franca (e.g., Kirkpatrick & Subhan 2014), New Englishes (e.g., Gut 2009; Biewer 2015; Bohmann & Babalola 2023; Hackert, Laliberté & Wengler 2024), and traditional (e.g., Jankowski & Tagliamonte 2022) and high-contact (e.g., Schreier 2005) first-language (L1) varieties.¹ The focus, however, has been on non-obligatory past inflection in African American Vernacular English (AAVE) and Caribbean English-lexifier creoles (CECs), with at least part of this interest probably owed to the variable's central position in the once heated debate about the origins of AAVE. Whereas the absence of morphological marking on consonant-final regular verbs (e.g., *WORK*) was first interpreted as the result of "TD-deletion" (e.g., Labov, Cohen, Robins & Lewis 1968) and thus as a purely phonological phenomenon common to all varieties of spoken English, Bickerton (1975:159) proposed that, were the AAVE data reanalyzed taking into account grammatical distinctions such as stativity, a creole pattern would emerge.

This controversy already highlights one of the major problems involved in analyses of non-obligatory past inflection in English: the variable's location "at the intersection" of different structural processes (Patrick 1991:171), with extralinguistic factors like age, gender, and social background also entering into the equation. Analyses "have gained in sophistication and breadth" (Poplack & Tagliamonte 2001:103); they now generally make fine distinctions with respect to these factors and processes. However, irrespective of variety, some recurrent constraints have emerged, including structural factors such as morpho-phonological verb class and lexical and grammatical aspect as well as cognitive ones such as priming. There is mounting evidence that, unlike previously assumed, at least some of these factors are not at all specific to creoles and related varieties such as AAVE and thus unconnected to potential creole universals, particular substrates, or the existence of a specific historical contact situation.

In this paper, we analyze patterns of variable past marking in English on Croker Island, Northern Territory, Australia. English on Croker Island is interesting for at least two reasons. First, it cannot be seen as a homogenous variety or even a linearly ordered continuum of features. This situation raises significant questions about functioning communication in the community and the acquisition of English. It also sheds further light on the process of indigenization, that is, the adoption of English as a distinct variety and means to express Indigenous identity. Although English is acquired as an L1 by all speakers on Croker Island, it has not undergone other steps that are usually associated with indigenization, especially the reduction of variation and the adoption of English as the main community language and a means to express Indigenous identity (cf. Schneider 2007). Second, English on Croker Island has been shaped by language contact but in a highly complex way due to its multilingual ecology and longstanding and variable contact situation. Consequently, expected linguistic outcomes of language contact cannot always be diagnosed with sufficient certainty (cf.

Mailhammer 2021:165-169). Our main research question is what constraints govern past marking in English on Croker Island and if these constraints are specific to the variety or resemble those found in other world regions and variety types.

We coded and analyzed 1274 tokens of past marking on lexical verbs using a multivariate approach. We tested the standard constraints described in the literature: morpho-phonological verb class, grammatical aspect, lexical aspect, and temporal disambiguation by means of adverbials. For consonant-final regular verbs, we investigated following phonological environment. We considered age and gender as social factors and linguistic task as a stylistic factor. We also coded for individual speaker and lexical item, as these constraints have repeatedly been shown to strongly affect variable linguistics processes of the kind investigated here. The effects of the factors we investigated align with what has been found in other studies of non-obligatory past marking in varieties of English, including CECs, New Englishes, and L1 varieties of English in North America and elsewhere. This has significant ramifications for our understanding of language contact and variation, especially in the context of postcolonial Englishes and creole studies.

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 4.1 presents the results of our statistical analysis of variable past inflection in the community at large. Section 4.2 turns to preverbal *bin*, a minor yet important past-marking strategy employed primarily by one elderly speaker, and investigates in what way this strategy aligns with the majority pattern. Section 5 discusses our results in the context of previous research; Section 6 offers concluding remarks.

2. Research Context

2.1. English on Croker Island

Croker Island is located off the coast of Northwestern Arnhem Land, Northern Territory, Australia, as shown in Figure 1.

The island measures about 45 km from north to south and at most 15 km from east to west. Minjilang, Croker Island's only settlement, is located at Mission Bay on the east coast. The community facilities include a primary school, a shop, a preschool, a community center, and the local government administration. According to the 2021 census, 88 percent of Minjilang's population of 265 are Aboriginal people. The linguistic situation is complex. Most Indigenous community members speak at least one Indigenous language, mostly Iwaidja, Mawng, and Kuwinjku, or have at least a passive understanding of Iwaidja, which, since the early twentieth century, has been considered Croker Island's primary language. In addition, everyone speaks English as an L1. However, input, domains, and patterns of usage are highly variable. While some community members use English very frequently or exclusively, others only use it in domains that are associated with "white" contexts or in the absence of a common Indigenous language. Indigenous languages occupy almost all community domains except school and government-related communication. For official community

announcements, the default language is Iwaidja, and there is no translation into other languages. For English on Croker Island, this means that the local languages, and especially Iwaidja, Mawng, and Kuwijnjku, functioned not only as substrates historically but are also concurrently spoken languages that exert ongoing influence on English and are influenced by it. Linguistic commonalities between local Englishes and Kriol, an English-lexifier creole spoken in the area (cf. Schultze-Berndt, Meakins & Angelo 2013), suggest that Kriol might have also formed part of the English “feature pool” (Mufwene 2001) available to Indigenous people, even though it was never used as an official medium of instruction.

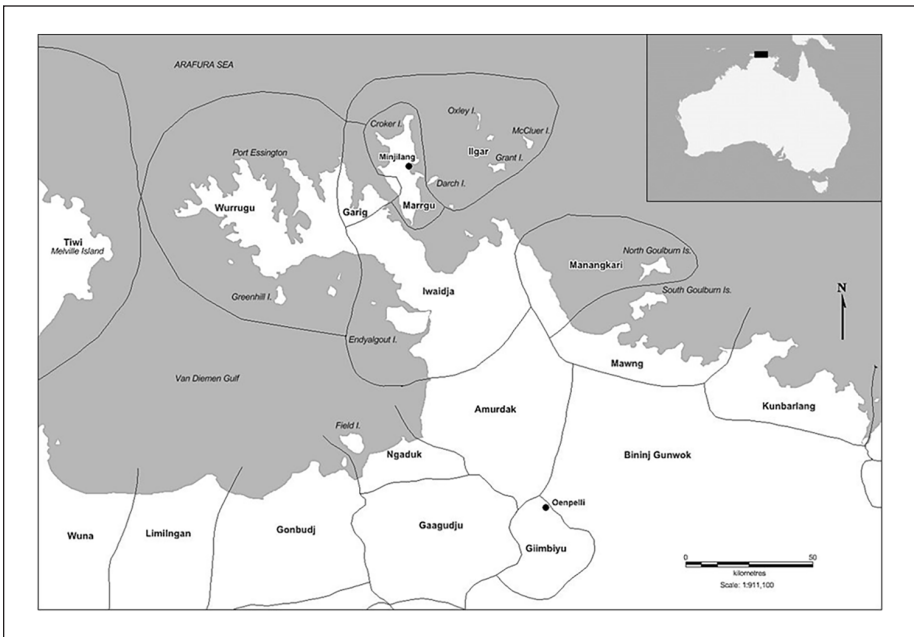


Figure 1. Croker Island, Northern Territory, Australia

This linguistic variability is closely connected to Croker Island’s settlement history. Via British military posts that were established in the early nineteenth century, Aboriginal people in Northwestern Arnhem Land were first exposed to non-standard dialects as well as institutionalized forms of English. They initially used the language as a means to communicate with soldiers, explorers, and then settlers with whom they interacted. But with increasing dominance of the colonizers, English became increasingly forced on them, especially through Christian missions that were set up from the early twentieth century onwards. The mission on Croker Island was established in 1941. It was reserved for children and teenagers of mixed ancestry, who were forcibly brought there from all over northern Australia; almost all Aboriginal residents had to

leave, which led to a break in settlement continuity. From 1958 onwards, Croker Island was re-settled, partly by original inhabitants but also by many newcomers, which increased the Aboriginal community's heterogeneity. While white people resided at the mission community, which was located where Minjilang is today, Indigenous people lived at a beachside camp just outside the mission settlement. As a result, Aboriginal people used English at the mission and Indigenous languages at the camp; children would sometimes speak English among themselves. This situation lasted until the early 1970s, after the end of the mission, when the Indigenous population, for many of whom this was their ancestral land and home, moved into the mission area with the subsequent construction of new housing. It was only then that Minjilang's mixed population formed a community. The persistent heterogeneity of English on the island is probably rooted in differences in acquisition and usage that remain considerable. A more detailed account of the history and current sociolinguistic setting of English on Croker Island is provided by Mailhammer (2021:40-54).

2.2. Previous Research on Variable Past Inflection in English

As noted in Section 1, variable past inflection has received extensive sociolinguistic attention and sophisticated statistical treatment, beginning with the earliest quantitative studies of AAVE (Labov, Cohen, Robins & Lewis 1968; Wolfram 1969). United in their rejection of the verbal "deficit" theory, 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, standard and non-standard, 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 not the similarities of anglophone Caribbean vernaculars with other Englishes but their uniqueness, which was in line with the emergence of "a general post-colonial consciousness" (Patrick 1999:4) in the region. In setting up what later came to be described as a relative tense system, 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 often focused precisely on determining whether the variety under discussion was governed by an "English" or a "creole" system, but generally without questioning the terminological and conceptual problems inherent in Bickerton's (1981:58) description of the "typical" creole tense-aspect system.

Arguing from a comparative sociolinguistic perspective (cf. Tagliamonte 2013), Poplack and Tagliamonte (2001) present 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).

Other studies have confirmed the effects identified by Poplack and Tagliamonte (2001) and added additional pertinent findings. 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 eight thousand past-reference verbs from urban Bahamian Creole, Hackert (2004:161-166) finds that the marking propensities of individual lexical items explain not only the behavior of particular morpho-phonological verb classes but also the apparent stativity effect, which largely disappears when a single, variably stative and non-stative verb, that is, HAVE, is investigated in isolation. The residual stativity effect 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 was considered a characteristic creole constraint in earlier work (Bickerton 1975, 1981) actually is better explained by 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, which are rarely overtly and unequivocally marked in the world’s languages.

More recent research on varieties from outside the Atlantic area and with different sociohistorical origins has uncovered similar patterns of variation. 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 (cf. also Bohmann & Babalola 2023:31). Biewer also fails to find a significant effect for stativity (2015:259), further supporting the idea that the effect of this factor posited in earlier studies may actually be an artifact of lexical and discourse-pragmatic constraints on past marking. Hackert, Laliberté, and Wengler’s (2024) comparative study of conversational data from the Hong Kong, India, Jamaica, and Philippine subcomponents of the International Corpus of English corroborates this view and also highlights the influence of marker persistence and verb frequency.

As noted above, Poplack and Tagliamonte (2001) had concluded that patterns of variation in earlier AAVE were more consistent with an English history than with a

creole background. Extrapolating from this conclusion in the absence of creole as a factor in the context of Croker Island, we propose that the recurring appearance of constraints on variable linguistic processes such as past-tense marking across all varieties, including factors that have been specifically linked to creoles, makes it likely that we are dealing with pan-English or even crosslinguistic constraints. This is the underlying hypothesis of the present study.

Past marking in Aboriginal English (AbE) has been described as generally reduced (Rodriguez Louro & Collard 2021:6). However, the only study that has empirically investigated this is Malcolm (1996), who recorded seven school children between five and ten years old (two female; mean age: 8) in interviews with teachers² in the early 1980s. The two key points to note about this study are, first, that all children employ the base form of the verb about 20 percent of the time and, second, that the range of variants also includes not just inflected and unmarked forms but also the preverbal particle *bin* followed by the base form of the verb, with use of this construction ranging from 2 percent to more than 20 percent among the children (Malcolm 1996:153).

3. Data and Method

3.1. The Sample

The data used in this study were collected between 2013 and 2018 in the community of Minjilang on Croker Island, as part of a larger project investigating the dynamics of language contact between English and Australian Aboriginal languages (Mailhammer 2021). They come from twenty speakers (ten female, 11-78 years, mean age=45.5), who performed a variety of linguistic tasks including semi-structured interviews and elicitation sessions (2021:13-14). Elicitation included the so-called “pear story” stimulus, a short, silent video clip about children stealing fruit commonly used in typological fieldwork (cf. Chafe 1980), as well as another (self-designed) video elicitation task aimed primarily at depicting aspectual contrasts. Other tasks, such as a card game, did not enter the corpus underlying the present study, as they did not contain past-reference verb situations in sufficient numbers. This excluded eight speakers—primarily children—of Mailhammer’s original community sample of 28. As in Mailhammer (2021:62), speakers are referred to by their initials in this paper.³

3.2. 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 either creoles or English: both marked and unmarked forms can encode a number of tense/aspect functions.” This situation is illustrated for Croker Island English in Examples (1) and (2). Forms included in the following statistical analyses are marked in bold, whereas those omitted are underlined.

- (1) you know one year . . . you know small we **went** to a station called Waterloo . . . and my father **do** a job there . . . **build** yards in the bush . . . we **went** ‘n’ . . . seen the manager an’ they **said**: why don’t you go over that way . . . ‘n’ we **went** down under the track . . . then there’s three, three everyday stockman and one white man so my father **spoke** to him – white man – in English, but he didn’t know English . . . so he **had** to speak him in language . . . he **knew** the language, that’s all, yeah they all **grow** up together . . . an’ he didn’t know speak English he **knew** language, so my father **rattle** out the language (AbESL20151109KO.RZ_checked.txt)
- (2) CM I’ve bin working there to that lighthouse
 Researcher 1 yeah
 CM they **bin put** automatic balanda, one balanda mechanic
 CM and that old man at Arrarru we **bin dig** it from
 CM all over and we **bin dig** it and put that wire under
 Researcher 1 right, ok
 CM and we **bin climb** on top of the lighthouse
 CM that wire, he **get** hot **pull**’im, **pull**’m, **pull**’im
 CM for that light
 Researcher 1 what was the wire for? oh right
 Researcher 2 it’s a big light, must have been a big cable
 CM **take** it they **chopped** it up it was a kerosene one.
 CM big light - we bin used to see light here
 Researcher 1 yes, kerosene.
 CM kayirrk you can’t see only a little bit there . . .
 Researcher 2 . . . so the lights over there at Cape Don
 CM that’s all morning time . . . off afternoon it **go** on
 Researcher 1 and how old were you then when you did that job?
 CM bin little bit young
 Researcher 1 what like Justin or –
 CM before that but I bin Darwin first (AbE_20130923_CM_bio_checked.txt)

Like most studies of variable past marking in AAVE, CECs, and World Englishes, we focus on the interplay 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 *chopped it up* (Example 2), vowel change, as in *my father spoke to him* (Example 1), and suppletion, as in *we went down under the track* (Example 1). We counted only tokens of past inflection on lexical verbs but additionally included two of the “primary” ones (Quirk, Greenbaum, Leech & Svartvik 1985:96), that is, HAVE, both as a main verb and as the “semi-auxiliary” HAVE TO (1985:137), as in *he had to speak him in language* (Example 1), and main-verb DO, as both are subject to the same inflectional processes as lexical verbs proper. Auxiliary DO, as in *he didn’t know English* (Example 1), was discounted. Copula and auxiliary BE structures, as in *then there’s three* (Example 1), were also omitted, as they undergo entirely different variable processes. We excluded non-standard forms of inflection ($N=26$), as in *seen the manager* (Example 1), historical present forms, ambiguous

forms such as *cut*, *quit*, or *put*, and tokens occurring in unclear contexts or before the dental/alveolar consonants /t/, /d/, /θ/, /ð/, /tʃ/, and /dʒ/. Semantically, our analysis is restricted to verb situations with absolute past-time reference. In terms of grammatical aspect, both perfective and habitual past-reference situations are included. For more details on “count” and “don’t count” cases and reasons for the inclusion or exclusion of particular structures, cf. Hackert (2008).

In addition to variably inflected past-reference lexical verbs and auxiliary structures, Example (2) additionally features the preverbal particle *bin*. This marker is common not only in CECs but also in the Australian creole Kriol, which is spoken across the Northern Territory (Schultze-Berndt, Meakins & Angelo 2013:246). The speaker in (2) uses *bin* not only before lexical verbs in their base form (*bin V*), as in *we bin dig it*, but also before other tense-aspect markers, as in *we bin used to see light here*, as well as before noun phrases indicating location or direction, as in *I bin Darwin first*, and adjectives, as in *bin little bit young*. Finally, he also employs it in a standard English perfect progressive construction: *I’ve bin working there to that lighthouse*. In this versatility, *bin* in Croker Island English closely resembles what is found in mesolectal CECs such as Bahamian Creole (cf. Hackert 2004:115). Whereas all *bin* constructions involving auxiliaries or non-verbal elements were automatically excluded, closer inspection of *bin V* revealed that it appears to have the same temporal-aspectual semantics as variably inflected lexical verbs in Croker Island English. As only CM used the construction to any significant extent, we treat this speaker’s data separately in Section 4.2.

3.3. Predictor Variables

As outlined in Sections 1 and 2.2, a frequent explanation for variable past inflection in AAVE, CECs, and World Englishes is phonological and maintains that unmarked verbs do not represent an underlying zero form but are the result of TD-deletion. Under this hypothesis, vowel-final (e.g., *PLAY*) and syllabic regular verbs (e.g., *WANT*) as well as irregular ones should be much less subject to non-marking than consonant-final regular verbs (e.g., *WALK*), if at all. We coded the predictor of morpho-phonological *VERB CLASS* as follows:

- irregular verbs;
- syllabic regular verbs, whose stem ends in /t/ or /d/, with /ɪd/ as their past-tense suffix (e.g., *WANT*);
- regular verbs ending in a vowel and taking /d/ as the dental suffix (e.g., *PLAY*);
- regular verbs ending in a consonant other than /t/ or /d/, hence forming a coda cluster in the past tense (e.g., *WALK*).

The class of irregular verbs includes both verbs undergoing only vowel-change (e.g., *GET*, *SIT*, *COME*) and verbs marked by vowel change as well as /t/ or /d/ affixation (e.g., *TELL*, *BUY*), as the occurrence of a vowel change is sufficient to indicate overt past marking (cf. Winford 1992:320). For consonant-final regular verbs, we also coded for following phonological segment, as previous work had identified this as a strong predictor of past inflection for this group of verbs (cf. Patrick 1999:145). We distinguished between consonant, vowel, and pause. We considered *DO*, *GO*, *HAVE*, *MAKE*, and *SAY*

separately, as these verbs show idiosyncratic morphological behavior, are also highly frequent, and often marked. We excluded verbs without vowel change whose base forms ends in liquid or nasal + /d/ and which form the past by way of devoicing (e.g., SEND, BUILD; cf. Quirk, Greenbaum, Leech & Svartvik 1985:106), as our final data set evidenced only seven tokens of such verbs, six of which were unmarked instances of BUILD produced by a single speaker.

We next coded for aspectual semantics. The term *aspect* refers to two different yet closely intertwined dimensions of temporal information. There is, first, the classification of events according to the dimensions of stative/dynamic, punctual/durative, and telic/atelic. Smith (1997) uses the term *situation aspect*, other terms are *aktionsart* and *lexical aspect*, which is what we will use here. Second, there is the internal temporal viewpoint taken on events as either bounded or unbounded, referred to as *viewpoint* or *grammatical aspect* (cf. Comrie 1976; Smith 1997), which is what we use here. Lexical and grammatical aspect are independent theoretical notions, but there are important interactions between them. For example, the use of the progressive aspect with stative verbs is generally marked (Smith 1997:77). We coded for lexical and grammatical aspect 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 non-stative or 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. This is due to Bickerton's (1981:58) idea of a "prototypical" creole tense-mood-aspect system, in which an interaction exists between stativity and past temporal reference. Accordingly, unmarked statives are interpreted as having non-past reference, whereas the default interpretation of unmarked non-statives is past. If a past marker occurs, statives receive a past interpretation, non-statives a pluperfect reading. Numerous studies have attempted to replicate this pattern, and, in fact, stativity has generally been found to favor past inflection in both AAVE and CECs. To code for stativity, we followed Smith (1997:17-18), who considers the phenomenon not merely a lexical property but primarily one 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) versus *Suddenly he knew* (non-stative) or *I have a brother* (stative) versus *I'm having lunch* (non-stative). We coded for stativity manually, employing established tests (cf. Filip 2012:728-730) where necessary.

Apart from the stative-dynamic dimension, lexical aspect is not usually investigated in analyses of variable past marking in AAVE, CECs, or World Englishes. It is, however, in studies of language acquisition, and we took inspiration from this field to take a closer look at the non-stative situation types of accomplishment, achievement, activity, and semelfactive. According to the Aspect Hypothesis, learners "will initially be influenced by the inherent semantic aspect of verbs or predicates in the acquisition of tense and aspect markers associated with or affixed to these verbs" (Andersen & Shirai 1994:133). More specifically, past marking has been found to spread from the telic situation types of achievements and accomplishments to the atelic ones of activities and states. While numerous studies have confirmed this for both L1 and L2

learners, especially among L2 speakers there seems to be considerable variation, which is often explained by confounding factors such as individual variation, transfer effects, and task variation (Bardovi-Harlig & Comajoan-Colomé 2020). Semelfactives are “single-stage events with no result or outcome” (e.g., COUGH; KNOCK; Smith 1997:29). Just as with statives versus non-statives, we coded the different dynamics as verb situations rather than based on lexical identity alone.

FOR GRAMMATICAL ASPECT, 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 one (cf. Comrie 1976:25). Both present the view of an event “from within,” that is, 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*. The perfective aspect is also logically independent of 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*. As noted in Section 2.2, both types of event may occur in marked and unmarked form in varieties of English featuring variable past inflection, but habituais have repeatedly been found to strongly disfavor marking in both AAVE and CECs.

Another factor we included is “PERSISTENCE” (Szmrecsanyi 2006), or “morphological priming” (Poplack & Tagliamonte 2001:129), which indicates whether the verb immediately preceding the one under investigation features the same marker, that is, inflection for inflected verbs, *bin* for *bin*-marked verbs, and zero for unmarked verbs. We coded for persistence (automatically using an R script) regardless of whether speaker turns changed in between immediately adjacent verbs, as we interpret persistence as a consequence of the cognitive principle underlying the fact that language users tend to recycle material they have used or heard before (cf. Szmrecsanyi 2006). “N/A” contexts either initiated conversational exchanges or followed verbs of non-past or unclear reference.

Another controversial constraint on past marking in CECs and related varieties is temporal disambiguation. It is a popular assumption that creole tense-mood-aspect systems rely heavily on the surrounding discourse context, including conjunctions and adverbials, for a verb’s temporal characterization (cf. Bickerton 1975:150-160), but Tagliamonte and Poplack (1993:189-190), Hackert (2004:174, 178), Biewer (2015:259), and Bohmann and Babalola (2023:31) find no significant effects if the category of temporal ADVERBIAL is considered at large. We nevertheless tested this predictor, operationalizing it according to Quirk, Greenbaum, Leech, and Svartvik’s semantic categories (1985:481-482):

- adverbial indicating a point of time (e.g., *yesterday, in 1960, after we had had lunch*);
- adverbial indicating duration (e.g., *for two months, until they left school*);

- adverbial indicating frequency (e.g., *ever*; *every summer*);
- adverbial indicating relationship: *already*, *still*, *yet*;
- *then*.

Even though non-temporal uses of *then* were omitted, we still assigned a separate category to this adverb because of its sheer frequency ($N=110$).

Regarding social factors, we only tested for speaker AGE and GENDER, as other dimensions potentially affecting language use do not sufficiently distinguish the speakers in the sample. All of them speak English as an L1 now, and with few exceptions all acquired the language as children, mostly through formal instruction but also unsupervised. Educational levels are also highly similar, varying between (completed) primary schooling and (incomplete) secondary education in most cases. For some speakers, both dimensions remained partially or entirely unclear. In terms of TASK, we distinguished between interview speech, the pear story, and the aspectual video task (cf. Section 3.1).

3.4. Statistical Analysis

As outlined in Section 3.3, variable past inflection in English is subject to numerous linguistic and social constraints, whose impact can be captured with a multivariate statistical analysis. Our analysis consists of two parts. First, we investigate the alternation between inflected and unmarked past forms at the community level (Section 4.1). Second, we separately consider the data from an individual, CM, whom we excluded from the community analysis due to his idiosyncratic use of an additional third variant, *bin* (Section 4.2). Each analysis consists of two steps: (I) determining each predictor's strength and direction of effects, and (II) assessing each predictor's overall impact on the alternation. Due to the different data structures, step I involves different multivariate techniques: we employ generalized mixed-effects regression modeling for the community analysis and a random forest-based technique suggested by Gries et al. (e.g., Heller, Bernaisch & Gries 2017) for the analysis of CM's data.

Generalized mixed-effects regressions can handle datasets which comprise non-independent, that is, grouped, data points. In sociolinguistic data, such grouped data structures occur if speakers contribute more than a single observation or if lexical items occur more than once. By adding grouping factors as random effects, fixed effects like AGE or VERB CLASS are held constant across individual as well as lexical idiosyncrasies. We used the R package "lme4" (Bates, Maechler, Bolker & Walker 2015) to fit a mixed-effects model. To obtain normally distributed data, age values were logarithmically transformed. Initial data inspection had revealed that the small number of semelfactives in the community sample ($N=17$) would have caused convergence issues; they were therefore left out. We followed the protocol suggested by Zuur, Ieno, Walker, Saveliev, and Smith (2009:101-142) and Gries (2015) for model selection, that is, we started out with a full model from which we successively removed non-significant effects. Our final model consisted of individual SPEAKER and LEXICAL ITEM as random effects, and VERB CLASS, GRAMMATICAL ASPECT, LEXICAL ASPECT, PERSISTENCE, ADVERBIAL, GENDER, as fixed effects. The predictors AGE and TASK had emerged as non-significant and were therefore removed.

As more complex random- and fixed-effects structures caused convergence problems, we only added varying intercepts for the two random effects and did not consider any interaction terms between the fixed effects in the full model. The final model achieved a C-value of 0.86, which indicates a good fit (Tagliamonte & Baayen 2012:156). For all factors, variable inflation scores (VIFs) were computed using the R-package “performance” (Lüdtke, Ben-Shachar, Patil, Waggoner & Makowski 2021). VIFs measure the magnitude of multi-collinearity between factors and were below 3 in our final model (VERB CLASS: 1.20; GRAMMATICAL ASPECT: 1.35; LEXICAL ASPECT: 1.60; PERSISTENCE: 1.05; ADVERBIAL: 1.22; GENDER: 1.10), indicating low correlation (James, Witten, Hastie & Tibshirani 2013). We can thus be confident that the model coefficients are not distorted by other factors. Each predictor’s overall significance and thus strength was computed by means of a type II ANOVA in “car” (Fox & Weisberg 2019). Differences between group means were computed via “multcomp” (Hothorn, Bretz & Westfall 2008). For each factor, predicted probabilities were obtained via “ggeffects” (Lüdtke 2018).

CM’s analysis did not lend itself to parametric regression modeling owing to the “small n large p ” problem (cf. Tagliamonte & Baayen 2012:24). Also facing multicollinearity and empty cells, we employed random-forest (RF) modeling, which constitutes a non-parametric regression approach that is based on recursive binary partitioning of the data such that each split results in increasingly homogeneous subgroups. RFs have been found to be robust against overfitting, resulting in reliable outcomes (e.g., Tomaschek, Hendrix & Baayen 2018:250). The RF was based on an ensemble of 1000 inference trees; the parameter specifying the number of randomly selected predictors at each split (mtry) was set to 3. To visualize the direction of effects for each predictor, we created effects plots based on group averages of the RF’s predicted probabilities (cf. Heller, Bernaisch & Gries 2017:123-124). Even though the resulting effects “do not control for the effects of all other predictors at the same time, [. . . a number of] studies have used this [method] successfully and comparisons of such plots with corresponding effects plots of regressions have been very encouraging” (Heller, Bernaisch & Gries 2017:124). Plotting was done in “ggplot2” (Wickham 2016).

To assess the overall impact of each predictor on the past tense alternation (step II), we also used RF modeling. This means that for the analysis of the community at large, we fitted an RF including the same factors that featured in the final step I model (ntree=1000, mtry=3). For CM’s analysis, we used the RF created during step I. Based on these RFs, all predictors’ impacts were calculated using Strobl, Boulesteix, Kneib, Augustin, and Zeileis’s (2008) unconditional variable importance estimation, an algorithm that has been found to be unaffected by multicollinearity.

4. Results: Patterns of Past Marking in Croker Island English

Figure 2 summarizes the distribution of past markers by speaker in terms of absolute numbers and percentages. Of the overall number of 1274 tokens, 65 percent are inflected. Frequencies of inflection range between 56 percent and 96 percent

(mean = 77%, median = 78%), with the exception of CM, who has only about 12 percent past-inflected verbs. In fact, for this speaker, the major competitor for unmarked verbs is not the standard English past inflection but *bin* V. No other speaker in the sample makes anywhere near as much use of this construction as CM, and it appears as if this speaker's past-tense system operates on principles quite different from those of the other speakers, based as it is on a tripartite instead of on a bipartite structure.

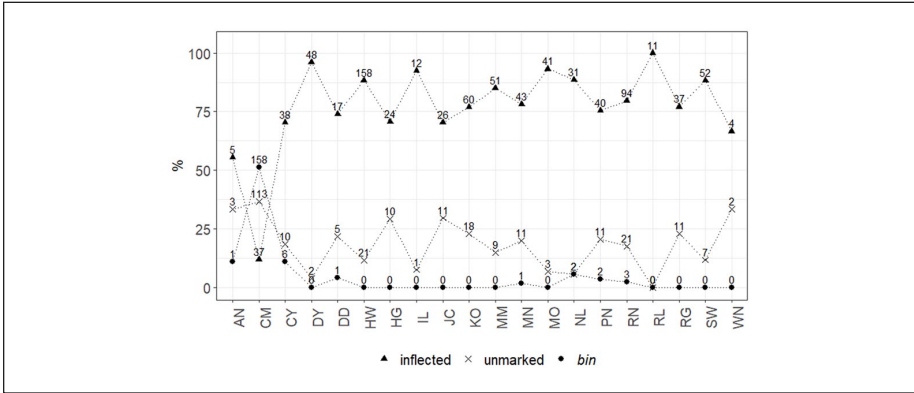


Figure 2. Distribution of Past Markers by Speaker in Terms of Absolute Numbers and Percentages

In brief, CM must be considered a “sociolinguistic outlier,” that is, an individual “whose linguistic behavior for some reason falls well outside that of the wider speech community” (Britain 2003:191). Such outliers, as noted by Britain (2003:191), have attracted the attention of researchers since “the early days of variation studies,” but in almost all cases, “a fuller exploration of the social background of the individual has shed light on their anomalous linguistic behaviour.” This fuller exploration constitutes the subject not just of a separate statistical analysis but also a separate section (4.2) of this paper. What follows in Section 4.1 is restricted to the data from the remaining nineteen speakers and the variation between unmarked and past-inflected verbs ($N=933$). All these speakers’ *bin* constructions ($N=16$) were discarded. Tables 1 and 2 present an overview of tokens and marking rates by predictors and predictor levels for the community sample and CM’s sample.

4.1. The Community at Large

Table 3 shows the results from a type II ANOVA based on our final model. During model selection, AGE and TASK emerged as statistically insignificant, which is why they were excluded from the final model. The summary of the model providing estimate coefficients, standard errors, z -values, and p -values is displayed in Table A1 in the Appendix. In the following, we present effects plots, that is, visual representations of the predicted values of each variable with all other variables held constant.

Table 1. Overview of Predictors and Predictor Levels, Community Sample

Predictor	Predictor level	N	% marking
VERB CLASS	consonant-final regular	164	66.5
	vowel-final regular	33	84.8
	syllabic regular	82	79.3
	Irregular	366	83.9
	DO, GO, HAVE, MAKE, SAY	288	93.4
LEXICAL ASPECT	accomplishment	142	76.1
	achievement	455	85.1
	activity	104	70.2
	stative	232	90.5
GRAMMATICAL ASPECT	habitual	212	79.7
	perfective	721	84.5
PERSISTENCE	NA	34	85.3
	marked	776	86.3
	unmarked	123	64.2
ADVERBIAL	<i>already/still/yet</i>	12	91.7
	duration	19	84.2
	frequency	13	84.6
	point of time	94	85.1
	<i>then</i>	87	67.8
	none	708	84.9
	TASK	interview	812
pear story		95	87.4
video task		26	80.8
GENDER	men	323	79.9
	women	610	85.2

Figure 3 displays the effects of the linguistic predictors.

Figure 3a shows considerable differences between the most and least marked verb classes in English on Croker Island. However, none of these differences are statistically significant, except for those between consonant-final regular verbs (“cons.fin.reg”) and DO, GO, HAVE, MAKE, and SAY (“*d-g-h-m-s*”; $est=2.41, p < .01$) and consonant-final and irregular verbs ($est=1.04, p = .06$). Consonant-final regular verbs show the lowest marking probability by far, but there is less homogeneity among the remaining verb classes than would be predicted by a purely phonological account based on TD-deletion. In fact, vowel-final regular verbs (“v.fin.reg”) display substantially more inflectional marking than any other class apart from DO, GO, HAVE, MAKE, and SAY. The exceptionally high marking rate of these verbs replicates what is generally found not only in other high-contact varieties of English (Hackert 2004:142-145) but also in learner language (Di Biase, Kawaguchi & Yamaguchi 2015:98-99). Their high token frequency endows such verbs’ irregular past forms with considerable “lexical strength,” which makes them easy to access in the mental lexicon and resistant to change historically (Bybee 1995:428).

Table 2. Overview of Predictors and Predictor Levels, CM's Sample

Predictor	Predictor level	N	% inflected	% bin
VERB CLASS	consonant-final regular	63	1.6	55.6
	vowel-final regular	13	0	46.2
	syllabic regular	7	57.1	28.6
	irregular	97	3.1	61.9
	DO, GO, HAVE, MAKE, SAY	62	27.4	56.5
LEXICAL ASPECT	accomplishment	66	12.1	65.2
	achievement	95	13.7	62.1
	activity	47	4.3	51.1
	semelfactive	2	0	0
	stative	32	6.2	37.5
GRAMMATICAL ASPECT	habitual	70	5.7	40.0
	perfective	172	12.2	64.0
PERSISTENCE	<i>bin</i>	137	6.6	62.8
	inflected	25	16.0	48.0
	unmarked	79	13.9	50.6
ADVERBIAL	yes	48	10.4	43.8
	no	194	10.3	60.3

Table 3. Analysis of Deviance for Past Inflection, Community Sample

Analysis of Deviance Table (Type II Wald chi square tests), response: marking			
	chisq	df	p
LEXICAL ASPECT	15.7240	3	.001292**
VERB CLASS	17.8229	4	.001336**
GRAMMATICAL ASPECT	9.7638	1	.001780**
PERSISTENCE	18.4742	2	9.736e-05***
ADVERBIAL	13.6990	5	.017639*
GENDER	7.2881	1	.006941**

Note: Signif. codes: 0 '***' .001 '**' .01 '*' .05.

Figure 4 presents absolute frequencies and percentages of past inflection for consonant-final regular verbs by following phonological environment. We subjected this verb class to a separate analysis to investigate syllable-final consonant cluster reduction, that is, TD-deletion (cf. Sections 1 and 2.2), which is affected by whether the cluster appears before a following consonant, vowel, or pause. A separate regression was not possible due to the small number of tokens ($N=164$). Although the difference between a following consonant and a following vowel is not statistically significant ($\text{chisq}=0.93, p=.33$), the effect of following phonological environment goes in the expected direction: more TD-deletion, and hence less inflection, before consonants than before vowels. While this pattern “shows a resounding consistency across studies” (Baranowski & Turton 2020:3),

the effect of pauses is less clear. In a comparative study of TD-deletion, Schreier (2005:206-207) finds that pauses tend to pattern with vowels in Englishes used by speakers of European descent but behave like consonants in AAVE. In English on Croker Island, this trend is even more pronounced. A potential explanation (Schreier 2005:207) is that in varieties which have a non-European contact history, the pre-pausal environment constitutes a very salient context for the application of the strong crosslinguistic preference for syllables ending in a single consonant or no consonant at all. We discuss the implications of this hypothesis for AbE in Section 5.

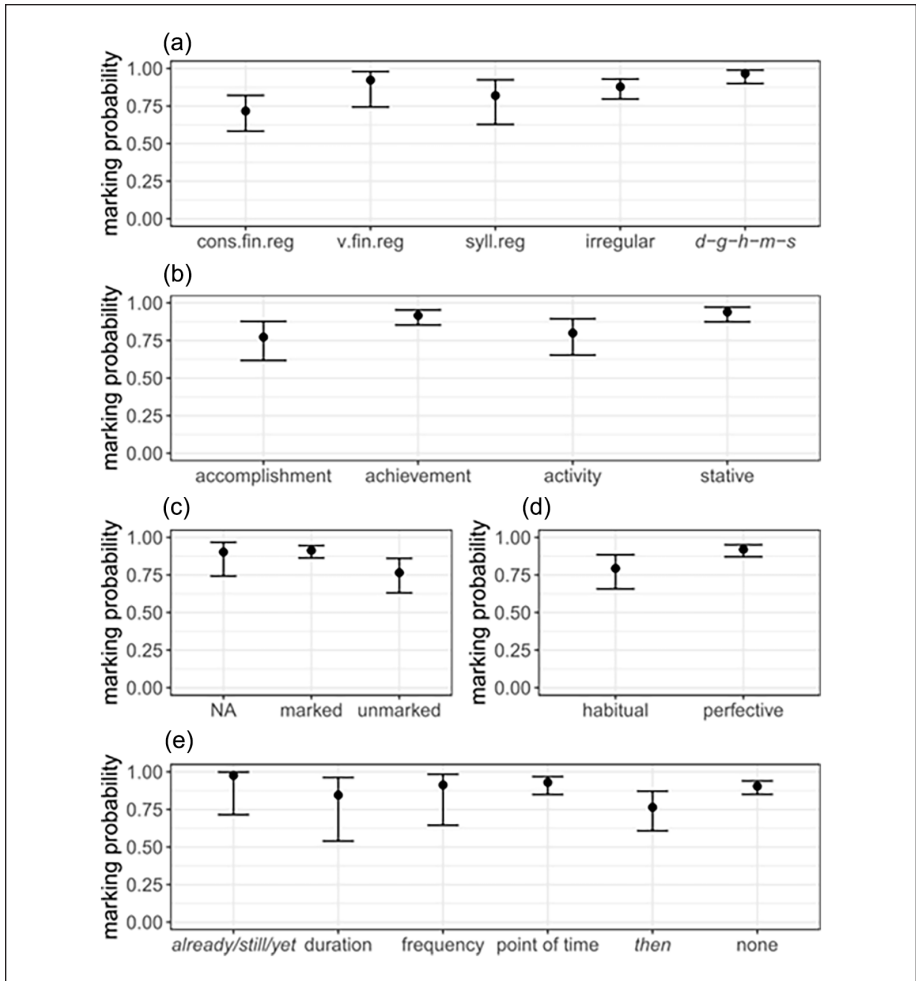


Figure 3. Linguistic Predictors' Effects on Past Inflection, Community Sample (Error Bars for 95% Confidence Intervals): (a) verb class, (b) lexical aspect, (c) persistence, (d) grammatical aspect, and (e) adverbial

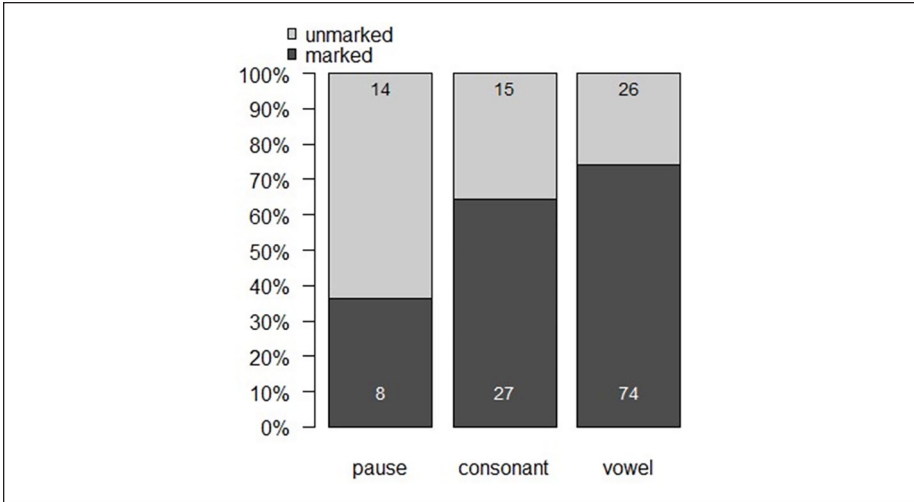


Figure 4. Frequencies and Percentages of Marked and Unmarked Consonant-Final Regular Verbs by Following Phonological Environment

As outlined in Section 2.2, stativity has generally been found to favor past marking in both AAVE and CECs, and Figure 3b shows this effect in English on Croker Island, too. In fact, the difference between statives and other lexical aspect types is statistically significant (accomplishments: $est=1.25$, $p < .05$, activities: $est=1.36$, $p < .05$) except for achievements ($est=0.35$, $p = .88$). A closer inspection of the phenomenon in Bahamian Creole (Hackert 2004:164-166) revealed, however, that the marking propensity of statives is at least in part an artifact of lexical idiosyncrasies. The class of statives is made up to a large extent of the high-frequency verbs *HAVE*, *THINK*, and *WANT*, which very often show past-inflections anyway. To check whether such a lexical effect also underlies the pattern evident in Figure 3b, we separately tallied the stative verb situations in our data ($N=232$), and, in fact, two thirds of them contain one of the five items *HAVE* ($N=96$), *HEAR* ($N=14$), *KNOW* ($N=11$), *THINK* ($N=16$), and *WANT* ($N=18$). Moreover, these verbs are marked between 91 percent and 100 percent of the time—as opposed to 80 percent for all other statives. The difference is statistically significant ($chisq=11.734$, $df=1$, $p < .001$). We also examined the behavior of *HAVE* on its own, which may occur in both stative and non-stative verb situations (Section 2.2). That way stativity can be isolated as a factor, and even though non-stative *HAVE* is rare in our data ($N=9$), it is inflected for past tense at an even higher rate (100 percent) than stative *HAVE* (95 percent inflection), even though the difference is not statistically significant (Fisher's exact test: $p = .63$). Consequently, the favoring effect that stativity has on past inflection in English on Croker Island is as much an artifact of the marking propensity of certain high-frequency verbs as it is in Bahamian Creole.

Our findings concerning the Aspect Hypothesis are somewhat mixed. While achievements fully conform to expectations, being substantially more marked than

both other dynamic situation types (Figure 3b), the overall distinction seems to be between punctual (achievements) and durative (accomplishments, activities) situation types rather than between telic (accomplishments, achievements) and atelic (activities) ones. Only the contrast between achievements and accomplishments is statistically significant ($est=1.17, p < .05$).

Grammatical aspect (Figure 3d) exerts a very significant statistical effect in the expected direction, with perfectives clearly more frequently past-inflected than habituals ($est=1.09, p < .01$). The persistence principle operates in our data, too (Figure 3c): the likelihood of finding an inflected or unmarked verb increases when this verb follows a verb of like marking status, resulting in a highly significant difference between unmarked and marked forms in the preceding slot ($est=-1.17, p < .001$). Thus, speakers of Croker Island English conform to a more general tendency (cf. Szmrecsanyi 2006): they tend to reuse morphosyntactic patterns they have produced or heard before. Interestingly, “N/A” contexts also show high probabilities of marking, which can easily be explained by way of the fact that such verb situations carry the functional load of framing the following stretch of discourse within past temporal reference.

As seen in Figure 3e, in English on Croker Island only a single adverbial type, that is, *then*, noticeably affects rates of past inflection, with statistically significant contrasts found only between *then* and adverbials indicating a point of time ($est=-1.40, p < .05$) on the one hand, and *then* and no adverbial at all ($est=-1.07, p < .05$) on the other. Importantly, as hinted at in Section 3.3, *then* is not only a very frequent adverb but also often functions as a discourse marker indicating progression not through reference time but through discourse time, as in Example (3).

- (3) I didn't talk Mawng, only Iwaidja and like Kunwinjku. And I could go out and we met together, and old man Snowy **then, then** he start teaching me how to talk Mawng. (AbE20130926MMRN_bio.txt)

There are also other non-temporal uses, such as (*but*) *then* in the sense of ‘on the other hand’:

- (4) . . . only the Kunwinjku kids, they speak to each other in Kunwinjku, **but then** they speak to each other in English (AbES12051110RoG.txt)

Even though we did not count such uses, they might well have had a depressing effect on past marking rates in clauses containing *then*.

Figure 5 shows the effects of the only statistically significant non-linguistic predictor in our mixed model, that is, GENDER, with women making more use of the standard English past inflection than men ($est=0.76, p < .01$).

Figure 6 presents the variable importance ranking for past inflection in the community at large. As in the mixed model, we included only the statistically significant predictors. The random forest achieved a *C* value of 0.96.

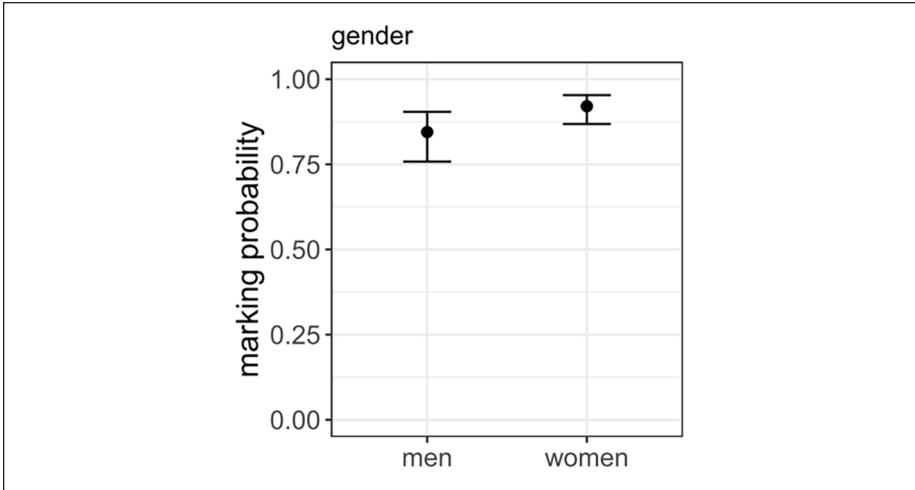


Figure 5. The Effect of GENDER on Past Inflection in Croker Island English

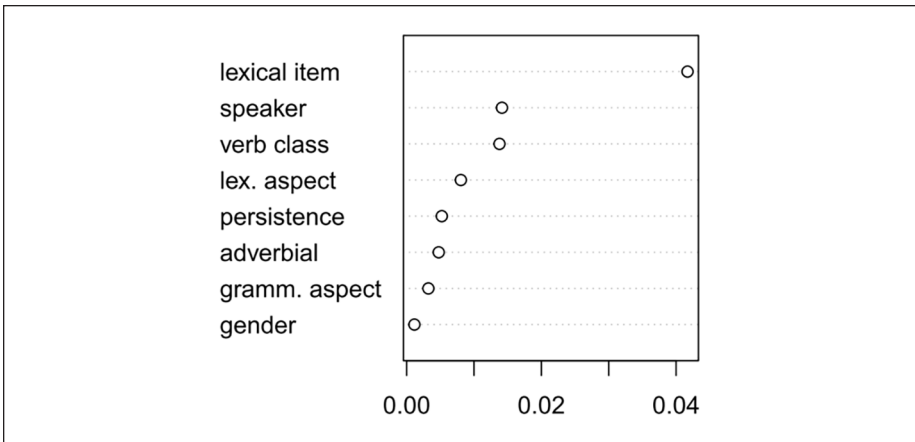


Figure 6. Predictor Importance Ranking for Past Inflection, Community Sample

As seen in Figure 6, by far the largest amount of variation in the past inflection of lexical verbs in Croker Island English is explained by individual LEXICAL ITEM and SPEAKER, which figured as random factors in our mixed model. This is entirely in line with what we see in other analyses of variable linguistic features in English. The categorical factor of VERB CLASS is also important, but this predictor is at least partly collinear with LEXICAL ITEM. Among the remaining categorical factors, LEXICAL ASPECT has the greatest importance, followed by PERSISTENCE, ADVERBIAL, and GRAMMATICAL ASPECT and the only statistically significant social factor of GENDER. PERSISTENCE is a cognitive

factor which has been found to affect variable linguistic patterns more generally, which hints at the fact that what we are dealing with in past inflection in Croker Island English is primarily general, structural-linguistic constraints or even cognitive ones, rather than creole patterns or transfer from substrate languages. Before we discuss this finding further, we take a brief look at the system of past marking in operation in the speech of the community outlier, CM.

4.2. CM’s System of Past Marking

CM’s system of past marking lexical verbs has a tripartite structure, with the preverbal marker *bin* accounting for over half ($N=158$) of this speaker’s tokens ($N=308$; cf. Figure 2). CM produced both interview and video task speech, but the latter contained so few tokens in several important contexts, in particular statives ($N=3$) and habituais ($N=6$), that the following statistical analysis is restricted to his interview speech ($N=242$).

Figure 7 presents the variable importance ranking for CM’s interviews. The random forest predicted 80 percent of all tokens correctly, which is substantially better than chance prediction based on the most frequent of the three alternating forms, that is, *bin* (51 percent). Individual LEXICAL ITEM once more has by far the greatest predictive importance, followed, at a distance, by VERB CLASS, LEXICAL ASPECT, and the remaining linguistic predictors. This pattern is identical with that found for past inflection in the community at large.

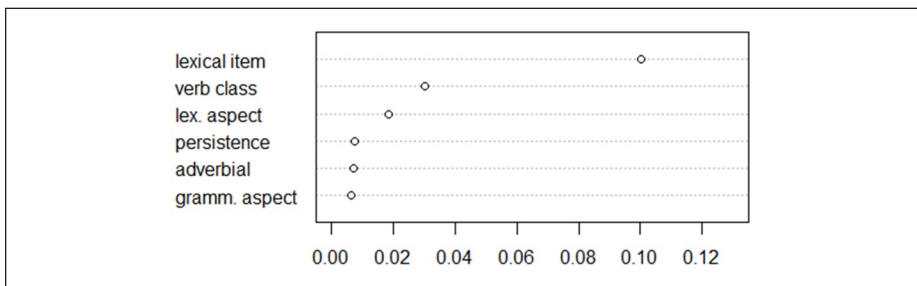


Figure 7. Predictor Importance Ranking for Past Marking, CM’s Sample

Figure 8 zooms in on the effects of the linguistic predictors. The pattern found for morpho-phonological VERB CLASS (Figure 8a) appears surprising at first glance, as this predictor apparently has hardly any effect on the application of *bin*, given that all verb classes are about equally likely to take the marker. The differences between the classes are not statistically significant. The standard English past inflection, by contrast, is strongly disfavored by both consonant-final and vowel-final regular verbs and, to a lesser extent, irregular verbs. Syllabic verbs (“syll.reg”) and the high-frequency ones, that is, DO, GO, HAVE, MAKE, and SAY, are somewhat more likely to be past-inflected in

CM's speech. This pattern appears to be at least in part owed to transfer effects from (untutored) L2 learning. Consonant-final regular verbs feature clusters in the past tense, which the other languages in CM's repertoire do not have. Past-marked vowel-final verbs often have diphthongs or long vowels followed by /d/, which the relevant substrate languages do not have either. The past tenses of DO, HAVE, and SAY are phonotactically unproblematic, since the vowel is short. This may also explain why MAKE has a standard past tense: the vowel is not a diphthong in CM's pronunciation, and so the past tense is unproblematic, too. Finally, syllabic regular verbs ending in [ɪd] also fit into the phonotactic pattern of CM's other languages and may therefore show standard past tenses.

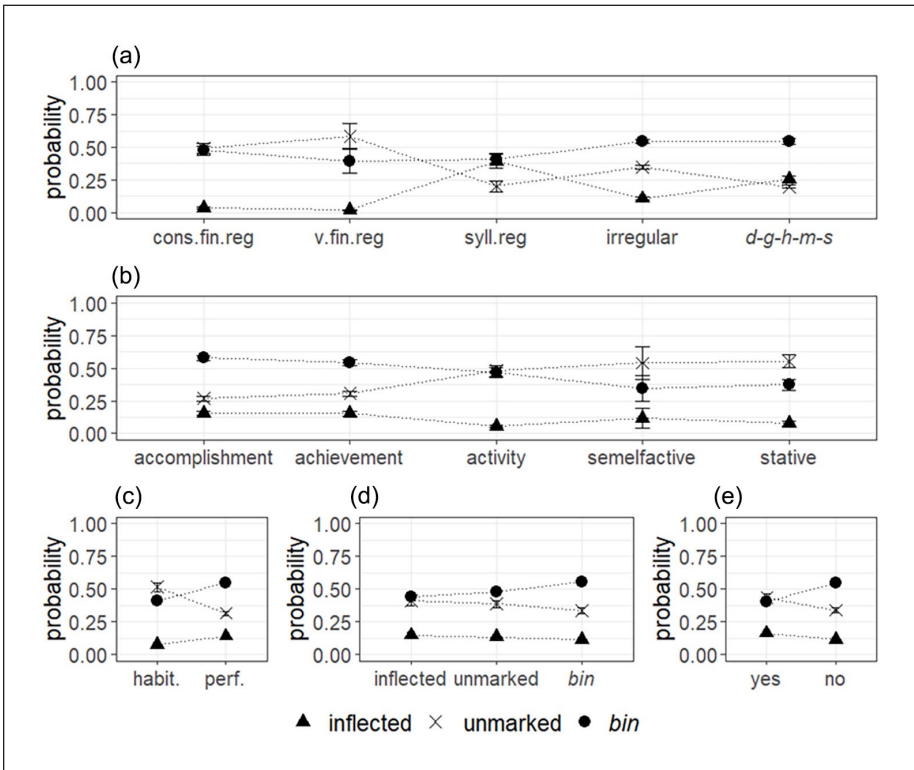


Figure 8. Linguistic Predictors' Effects on Past Marking, CM's Sample: (a) Verb Class, (b) Lexical Aspect, (c) Gramm. Asp., (d) Persistence, and (e) Adverbial

When it comes to LEXICAL ASPECT (Figure 8b), the two past markers show a very similar pattern, which is based on telicity, conforming to the Aspect Hypothesis: achievements and accomplishments are more likely to feature both *bin* and inflection than activities, semelfactives, and statives. GRAMMATICAL ASPECT (Figure 8c) also

shows a parallel pattern for the two markers: both are favored by perfective verb situations. The effect of PERSISTENCE emerges most clearly for *bin* (Figure 8d). As for temporal ADVERBIAL, we lacked sufficient data for the individual categories and therefore collapsed all of them into a single factor (“yes”). The emergent pattern is that *bin* is favored in contexts in which no adverbial is present (“no”). In sum, even though CM’s system of past marking appears quite different at the surface levels of forms and frequencies, the constraints conditioning the variation largely parallel those found in the larger community sample, even if *bin* is factored in.

It is plausible that *bin* in English on Croker Island originally derives from New South Wales Pidgin, where it was the standard way to express past reference (Troy 1994:250). This variety spread throughout Australia in the nineteenth century and gave input to restructured varieties and creoles, including Kriol (Munro & Mushin 2016:83). It is likely that *bin* V was the past tense for many speakers of Aboriginal varieties of English before the start of formal schooling, which is consistent with CM’s data, who acquired English largely unsupervised in the 1940s, when formal schooling had not yet become standard for Aboriginal people in Northern Arnhem Land. Since then, English has formed part of the linguistic input for people growing up in Minjilang, either directly through the local speech community or through contact with Kriol and other varieties of AbE.

5. Discussion

The most significant finding of this study is that past inflection in English on Croker Island is subject to the same recurrent structural constraints as in other varieties of English. This is interesting from at least two perspectives. First, because of the specific history and variability of English on Croker Island, we expected relevant constraints to be difficult to identify. In fact, individual variation is a strong factor, but this situation is by no means unique to Croker Island, and it also does not mean that variability in the speech community cannot be accounted for rigorously. It may well be that further studies on other variables show a similar picture, which would be interesting for the relationship of grammar and variation and for the question of how phenotypically homogenous grammatical systems must be. Second, the fact that variable past inflection is governed by constraints that are found elsewhere, irrespective of variety type and history, for example, high-contact, low-contact, creole, Caribbean, Asian, etc., suggests that exactly those two characteristics—type of variety and history—may not be as relevant to explaining the phenomenon as previously thought and that other factors are much more important. We return to this point below, after we contextualize our findings with respect to English on Croker Island and AbE.

The overall level of past inflection across the sample is comparable to the much smaller sample of Western Australian Aboriginal children from the early 1980s (Malcolm 1996:152). That said, rates of occurrence of a feature are not actually decisive in such comparisons, as they may vary according to a wide variety of extralinguistic factors. What matters is 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 in the direction and strength of

the structural constraints operating on the variable in question. Malcolm's study provides much more coarse-grained data on the types of past tense employed, but some details are worth comparing.

Malcolm's (1996:153) data suggest a trend toward standard marking correlating with age that is not evident in our data. As noted in Section 4.1, AGE did not emerge as a statistically significant predictor in the community sample; in fact, not only is this predictor not statistically significant, but it also has no discernible direction, in the sense that older speakers might consistently behave different than younger ones, and vice versa. This suggests that variable past inflection in Croker Island English is not involved in language change in progress. Interestingly, Malcolm (1996:152) also remarks that *bin* is used "with decreasing frequency across the age range," which could, of course, simply indicate a greater awareness among the older children of his sample that overt non-standard markers such as *bin* are not part of the linguistic behavior expected in school contexts. In the Croker Island sample, the form is disproportionately frequent with the oldest speaker, who is a clear outlier. However, it is not the case that there is a downward trend in the use of this past marker; in fact, the second-highest frequency of *bin* comes from a speaker in her twenties at the time of recording. *Bin* could be either speaker-dependent (CM) or deployed for specific reasons. One such reason could be emotional attachment (Mailhammer 2021:77), but this hypothesis would have to be investigated specifically.

Unfortunately, Malcolm's analysis does not go beyond identifying different types of marking for individual speakers in his sample. Consequently, it is impossible to compare this sample of AbE with that of Harkins (1994), who claims that past marking in her Central Australian sample is pretty much identical with the standard system and that deviations are phonologically conditioned (Malcolm 1996:153). Not unexpectedly, phonological conditioning also plays a role in our sample. Section 4.1 showed that consonant-final regular verbs evidence the lowest inflection rates and that this trend is even more pronounced before consonants and pauses, which is consistent with other high-contact varieties of English. The general explanation for elevated rates of TD-deletion in such varieties consists in constraints against consonant clusters in syllable codas in the substrate languages. The main heritage languages on Croker Island are not particularly tolerant with respect to syllable-final consonant clusters, but they do allow clusters whose first element is a liquid (Iwaidja, cf. Pym & Larrimore 1979:8), a liquid or a glide (Mawng, cf. Capell & Hinch 1970:25), or even a liquid, glide, or nasal (Kunwinjku, cf. Evans 2003:92). This tolerance speaks against a purely substrate-based explanation for the pattern of zero-marking in consonant-final regular verbs in English on Croker Island. However, it seems conceivable that coda clusters, especially clusters without a drop in sonority, are dispreferred in high-contact Englishes due to their origin in group L2 acquisition. Even in highly cluster-tolerant Englishes, such clusters are restricted to past-tense and participle forms belonging to consonant-final regular verbs. Contact varieties with a less rich input of forms may therefore have less support for universally dispreferred coda clusters and consequently show higher levels of TD-deletion, even if they permit certain clusters.

It is remarkable that Croker Island English at first glance follows the presumably creole pattern of favoring past marking on stative verb situations. However, upon

closer inspection, this apparent stativity effect is as much an artifact of lexical idiosyncracies as it is in mesolectal CECs such as Bahamian Creole. Our data partly confirmed the Aspect Hypothesis, which says that telicity favors past marking in language acquisition more generally. At least for Iwaidja speakers, there may also be the supporting factor that Iwaidja has an aspectually underspecified past tense, an anterior, which is contrasted with a past imperfective (Caudal & Mailhammer 2022:5). In the anterior, it is the telicity of the verb that determines its aspectuality: if the verb is telic, the default reading is perfective, if it is atelic, the default reading is imperfective. Consequently, speakers of English on Croker Island who also know Iwaidja—and this is probably a significant number (cf. Mailhammer 2021:143)—would have access to a tense category that would support the use of the Aspect Hypothesis in the acquisition process. As for the other substrate languages, they have been described as having an aspectual contrast in the past tense, that is, normally between a perfective and an imperfective past tense (for Kunwinjku cf. Evans 2003; for Mawng cf. Capell & Hinch 1970). However, it is uncertain what the respective interactions with lexical aspect look like, and whether they would support the Aspect Hypothesis as an acquisition strategy.

Finally, the gender effect we observed is in line with the more general trend in Western societies that women tend to use more formal or prestigious speech forms than men. This is not the norm in postcolonial or non-Western societies (e.g., Bakir 1986), especially if the interaction between gender and other social variables such as social status and education is considered (e.g., Hackert 2004), and even in Western contexts female speakers sometimes show more non-standard speech than male speakers (e.g., Eckert 2000), but the same observation has been made in other postcolonial contexts (e.g., Leimgruber, Lim, Gonzales & Hiramoto 2021). On Croker Island, women have often assumed leadership roles in the community, showing high levels of functionality in Indigenous and Western contexts combined with multilingualism and high levels of education. However, as Indigenous women, their status in the community is likely not reflected in their position in monolingual mainstream society. Hence, it is not unlikely that Croker Island women may move toward prestige norms in mainstream contexts (including interview settings), as has often been theorized for women elsewhere, and for similar reasons related to power and sociolinguistic awareness. Our results indicate that it would be worthwhile to further examine the relationship between variation and community-specific ways of doing gender on Croker Island.

Returning to the issue of how patterns of past marking in high-contact varieties of English are best explained, our results confirm the general picture that has been emerging in the last twenty years or so: the variable is governed largely by the same set of lexical, structural, discourse-pragmatic, and cognitive factors irrespective of whether we are dealing with creoles, high-contact, or L1 or L2 varieties. None of these factors are specific to creoles. Consequently, they cannot be explained as stemming from a “creole prototype.” Similarly, they are not germane to varieties with contact histories and therefore cannot be explained as substrate effects, either. Instead, we think that they constitute general linguistic principles that affect the acquisition and use of past marking in English more generally.

6. Conclusion

In this paper we analyzed past marking in English on Croker Island, Northern Territory, Australia using data from twenty speakers supplying 1274 tokens in total. Our analysis showed that the most predictive factors apart from SPEAKER and individual LEXICAL ITEM were morpho-phonological VERB CLASS and LEXICAL ASPECT, with other linguistic and social predictors such as PERSISTENCE, GRAMMATICAL ASPECT, the presence or absence of a temporal ADVERBIAL, and GENDER also contributing to the variation observed. Our statistical analysis achieved a high predictive accuracy. This means that despite the overall high variability of English on Croker Island it is possible to make robust generalizations about this particular subsystem of grammar. Another interesting result is that, in our data, preverbal *bin* is functionally highly similar to the standard English past inflection. Whether this also holds for other varieties of AbE and/or the Australian English-lexifier creoles warrants further research. Another major conclusion that we draw from our analysis is that the constraints governing the presence or absence of past marking in varieties of English are general linguistic, discourse-pragmatic, or even cognitive rather than creole or contact-specific ones. Consequently, contact processes or substrate languages may be less relevant to explaining patterns of past marking in English than previously thought.

Appendix

Table A1. Summary of Mixed-Effects Model, Community Sample (Predictions for Marking)

	Estimate	SE	z-Value	<i>p</i>
(Intercept)	0.14	1.56	0.09	.929
lex_aspectachievement	1.18	0.40	2.98	.003
lex_aspectactivity	0.17	0.45	0.38	.703
lex_aspectstative	1.52	0.53	2.88	.004
verb_classv.fin.reg	1.60	0.75	2.13	.033
verb_classsyll.reg	0.72	0.55	1.30	.194
verb_classirregular	1.09	0.41	2.68	.007
verb_classd-g-h-m-s	2.45	0.65	3.75	<.001
gramm_aspectperfective	1.09	0.35	3.12	.002
persistenceunmarked	0.12	0.56	0.22	.825
persistenceunmarked	-1.04	0.60	-1.74	.081
adverbialduration	-2.02	1.61	-1.26	.209
adverbialfrequency	-1.36	1.64	-0.83	.407
adverbialpoint of time	-1.13	1.44	-0.78	.436
adverbialthen	-2.54	1.44	-1.77	.077
adverbialnone	-1.47	1.41	-1.04	.297
genderwomen	0.76	0.28	2.70	.007

Declaration of Conflicting Interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research is supported by the Australian Research Council, Discovery Grant DP130103935, CI Robert Mailhammer.

ORCID iDs

Stephanie Hackert  <https://orcid.org/0000-0002-0880-229X>

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

Diana Wengler  <https://orcid.org/0000-0002-1848-0474>

Notes

1. By “high-contact varieties,” we refer to varieties of English that have been shaped by contact between speakers of English and speakers of languages other than English. High-contact varieties, in other words, either have or had a significant number of users for whom English is or was a second language. While technically creoles could thus also be designated “high-contact” varieties, we restrict this term to languages that emerged in highly multilingual contact settings that also involved forced population migration or displacement, as in trans-Atlantic slavery or indenture in the Indian and Pacific Oceans. “Traditional” L1 varieties are “[t]raditional, regional non-standard mother-tongue varieties, e.g. East Anglian English [. . .] and Newfoundland English, Appalachian English and Ozark English” (Kortmann & Lunkenheimer 2012:3). Standard English comes in many forms today, but each of these forms is codified to at least some extent and serves as the language of choice in most public, formal discourse.
2. The teachers were both Indigenous and non-Indigenous, and some interviews were also conducted by “the Perth-based research team,” but Malcolm (1996:149) does not explore this dimension of variation further.
3. A reviewer wondered whether we took into account the interlocutor as a factor possibly prompting accommodation by participants. We did, in fact, but with the exception of the pear story, where participants talked to an in-group member, the interviews were always conducted by researchers. This means that what we are describing here is more of a “light” variety than a “heavy” one. We do not think this matters, as long as we do not assume that this is (a) the only kind of English spoken by any participant and (b) the most basilectal version of English used on Croker Island.

References

- Andersen, Roger W. & Yasuhiro Shirai. 1994. Discourse motivations for some cognitive acquisition principles. *Studies in Second Language Acquisition* 16. 133-156.
- Bakir, Murtadha. 1986. Sex differences in the approximation to standard Arabic: A case study. *Anthropological Linguistics* 28(1). 3-9.
- Baranowski, Maciej & Danielle Turton. 2020. TD-deletion in British English: New evidence for the long-lost effect. *Language Variation and Change* 32. 1-23.

- Bardovi-Harlig, Kathleen & Llorenç Comajoan-Colomé. 2020. The aspect hypothesis and the acquisition of L2 past morphology in the last 20 years: A state-of-the-scholarship review. *Studies in Second Language Acquisition* 42. 1137-1167.
- Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67. 1-48.
- Bayley, Robert. 1996. Competing constraints on variation in the speech of adult Chinese learners of English. In Robert Bayley & Dennis R. Preston (eds.), *Second language acquisition and linguistic variation*, 97-120. Amsterdam: Benjamins.
- Bickerton, Derek. 1975. *Dynamics of a creole system*. Cambridge: Cambridge University Press.
- Bickerton, Derek. 1981. *Roots of language*. Ann Arbor: Karoma.
- Biewer, Carolin. 2015. *South Pacific Englishes: A sociolinguistic and morphosyntactic profile of Fiji English, Samoan English and Cook Islands English*. Amsterdam: Benjamins.
- Bohmann, Axel & Adesoji Babalola. 2023. Verbal past inflection in Nigerian English: A case for sociolinguistic compound vision. In Guyanne Wilson & Michael Westphal (eds.), *New Englishes, new methods*, 16-41. Amsterdam: Benjamins.
- Britain, David. 2003. Exploring the importance of the outlier in sociolinguistic dialectology. In David Britain & Jenny Cheshire (eds.), *Social dialectology: In honour of Peter Trudgill*, 191-208. Amsterdam: Benjamins.
- Bybee, Joan. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10. 425-455.
- Capell, Arthur & Heather E. Hinch. 1970. *Maung grammar, texts and vocabulary*. The Hague: Mouton.
- Caudal, Patrick & Robert Mailhammer. 2022. Linear lengthening in Iwaidja: An event-quantifying intonation at the phonology to semantics/pragmatics interface. *Languages* 7. 1-23.
- Chafe, Wallace L. (ed.). 1980. *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.
- Comrie, Bernard. 1976. *Aspect: An introduction to the study of verbal aspect and related problems*. Cambridge: Cambridge University Press.
- Di Biase, Bruno, Satomi Kawaguchi & Yumiko Yamaguchi. 2015. The development of English as a second language. In Camilla Bettoni & Bruno Di Biase (eds.), *Grammatical development in second languages: Exploring the boundaries of processability theory*, 85-115. Italy: EUROSLA.
- Eckert, Penelope. 2000. *Linguistic variation as social practice*. Oxford: Blackwell.
- Evans, Nicholas. 2003. *Bininj Gun-wok: A pan-dialectal grammar of Mayali, Kunwinjku and Kune*. Canberra: Pacific Linguistics.
- Filip, Hana. 2012. Lexical aspect. In Robert I. Binnick (ed.), *The Oxford handbook of tense and aspect*, 721-751. Oxford: Oxford University Press.
- Fox, John & Sanford Weisberg. 2019. *An R companion to applied regression*. 3rd edn. Thousand Oaks, CA: Sage.
- Gries, Stefan Th. 2015. The most under-used method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora* 10. 95-125.
- Gut, Ulrike. 2009. Past tense marking in Singapore English verbs. *English World-Wide* 30. 262-277.
- Hackert, Stephanie. 2004. *Urban Bahamian Creole: System and variation*. Amsterdam: Benjamins.
- Hackert, Stephanie. 2008. Counting and coding the past: Circumscribing the variable context in quantitative analyses of past inflection. *Language Variation and Change* 20. 127-153.

- Hackert, Stephanie, Catherine Laliberté & Diana Wengler. 2024. Past inflection around the world: A cross-variety analysis of New Englishes. *Lingua* 307. 103776. <https://www.sciencedirect.com/science/article/pii/S0024384124001050> (22 August, 2024).
- Harkins, Jean. 1994. *Bridging two worlds: Aboriginal English and cross-cultural understanding*. St. Lucia: University of Queensland Press.
- Hawkins, Roger & Sarah Liszka. 2003. Locating the source of defective past tense marking in advanced L2 English speakers. *Language Acquisition and Language Disorders* 30. 21-44.
- Heller, Benedikt, Tobias Bernaisch & Stefan Th. Gries. 2017. Empirical perspectives on two potential epicenters: The genitive alternation in Asian Englishes. *ICAME Journal* 41. 111-144.
- Hothorn, Torsten, Frank Bretz & Peter Westfall. 2008. Simultaneous inference in general parametric models. *Biometrical Journal* 50. 346-363.
- James, Gareth, Daniela Witten, Trevor Hastie & Robert Tibshirani (eds.). 2013. *An introduction to statistical learning, with applications in R*. New York, NY: Springer.
- Jankowski, Bridget L. & Sali A. Tagliamonte. 2022. He come out and give me a beer but he never seen the bear: Vernacular preterites in Ontario dialects. *English World-Wide* 43. 267-296.
- Kirkpatrick, Andy & Sophiaan Subhan. 2014. Non-standard or new standards or errors? The use of inflectional marking for present and past tenses in English as an Asian lingua franca. In Sarah Buschfeld, Thomas Hoffmann, Magnus Huber & Alexander Kautzsch (eds.), *The evolution of Englishes: The dynamic model and beyond*, 386-400. Amsterdam: Benjamins.
- Kortmann, Bernd & Kerstin Lunkenheimer. 2012. Introduction. In Bernd Kortmann & Kerstin Lunkenheimer (eds.), *The Mouton world atlas of variation in English*, 1-11. Berlin: de Gruyter Mouton.
- Labov, William, Paul Cohen, Clarence Robins & John Lewis. 1968. *A study of the non-standard English of Negro and Puerto Rican speakers in New York City*. New York, NY: Columbia University.
- Leimgruber, Jakob R. E., Jun Jie Lim, Wilkinson Daniel Wong Gonzales & Mie Hiramoto. 2021. Ethnic and gender variation in the use of Colloquial Singapore English discourse particles. *English Language and Linguistics* 25. 601-620.
- Lüdecke, Daniel. 2018. Ggeffects: Tidy data frames of marginal effects from regression models. *Journal of Open Source Software* 3(26). 772.
- Lüdecke, Daniel, Mattan S. Ben-Shachar, Indrajeet Patil, Philip Waggoner & Dominique Makowski. 2021. Performance: An R package for assessment. *Journal of Open Source Software* 6(60). 3139.
- Mailhammer, Robert. 2021. *English on Croker Island: The synchronic and diachronic dynamics of contact and variation*. Berlin: De Gruyter Mouton.
- Malcolm, Ian. 1996. Observations on variability in the verb phrase in aboriginal English. *Australian Journal of Linguistics* 16. 145-165.
- Mufwene, Salikoko S. 2001. *The ecology of language evolution*. Cambridge: Cambridge University Press.
- Munro, Jennifer & Ilana Mushin. 2016. Rethinking Australian Aboriginal English-based speech varieties: Evidence from Woorabinda. *Journal of Pidgin and Creole Languages* 31. 82-112.
- Patrick, Peter. 1991. Creoles at the intersection of variable processes: -t, d deletion and past-marking in the Jamaican mesolect. *Language Variation and Change* 3. 171-189.
- Patrick, Peter. 1999. *Urban Jamaican creole: Variation in the mesolect*. Amsterdam: Benjamins.

- Poplack, Shana. 2000. Introduction. In Shana Poplack (ed.), *The English history of African American English*, 1-34. Oxford: Blackwell.
- Poplack, Shana & Sali Tagliamonte. 2001. *African American English in the diaspora*. Oxford: Blackwell.
- Pym, Noreen & Bonnie Larrimore. 1979. *Papers on Iwaidja phonology and grammar*. Darwin: SIL.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.
- Rodriguez Louro, Celeste & Glenys Collard. 2021. Australian Aboriginal English: Linguistic and sociolinguistic perspectives. *Language and Linguistics Compass* 2021. 1-12.
- Schiffirin, Deborah. 1981. Tense variation in narrative. *Language* 57. 45-62.
- Schneider, Edgar W. 2007. *Postcolonial English*. Cambridge: Cambridge University Press.
- Schreier, Daniel. 2005. *Consonant change in English worldwide: Synchrony meets diachrony*. Houndmills: Palgrave Macmillan.
- Schultze-Berndt, Eva, Felicity Meakins & Denise Angelo. 2013. Kriol. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *The survey of pidgin and creole languages*, vol. 1, *English-based and Dutch-based languages*, 241-251. Oxford: Oxford University Press.
- Smith, Carlota S. 1997. *The parameter of aspect*. 2nd edn. Dordrecht: Kluwer.
- Strobl, Carolin, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin & Achim Zeileis. 2008. Conditional variable importance for random forests. *BMC Bioinformatics* 9. 307 <http://www.biomedcentral.com/1471-2105/9/307> (31 October, 2020).
- Szmrecsanyi, Benedikt. 2006. *Morphosyntactic persistence in spoken English: A corpus study at the intersection of variationist sociolinguistics, psycholinguistics, and discourse analysis*. Berlin: Mouton de Gruyter.
- Tagliamonte, Sali A. 2013. Comparative sociolinguistics. In J. K. Chambers & Natalie Schilling (eds.), *The handbook of language variation and change*. 2nd edn., 128-156. Chichester: Wiley-Blackwell.
- Tagliamonte, Sali A. & R. Harald Baayen. 2012. Models, forests and trees of York English: *Was/were* variation as a case study for statistical practice. *Language Variation and Change* 24. 135-178.
- Tagliamonte, Sali A. & Shana Poplack. 1993. The zero-marked verb: Testing the creole hypothesis. *Journal of Pidgin and Creole Languages* 8. 171-206.
- Tomaschek, Fabian, Peter Hendrix & R. Harald Baayen. 2018. Strategies for addressing collinearity in multivariate linguistic data. *Journal of Phonetics* 71. 249-267.
- Troy, Jakelin. 1994. *Melaleuka: A history and description of New South Wales Pidgin*. Canberra: Australian National University PhD dissertation.
- Wickham, Hadley. 2016. *Ggplot2: Elegant graphics for data analysis*. New York, NY: Springer.
- Winford, Donald. 1992. Back to the past: The BEV/creole connection revisited. *Language Variation and Change* 4. 311-357.
- Wolfram, Walt. 1969. *A sociolinguistic description of Detroit Negro speech*. Washington, DC: Center for Applied Linguistics.
- Wolfram, Walt & Deborah Hatfield. 1984. *Tense marking in second language learning: Patterns of spoken and written English in a Vietnamese community*. Washington, DC: Center for Applied Linguistics.
- Zuur, Alain F., Elena N. Ieno, Neil J. Walker, Anatoly A. Saveliev & Graham M. Smith. 2009. *Mixed effects models and extensions in ecology with R*. New York, NY: Springer.

Author Biographies

Stephanie Hackert is Professor of English Linguistics at the University of Munich (LMU). Her research interests center around pidgins and creoles, variationist sociolinguistics, corpus linguistics, recent grammatical change, and historical discourse analysis. Her publications include monographs on urban Bahamian Creole English (Benjamins, 2004) and the history of the English native speaker (de Gruyter, 2012).

Catherine Laliberté is a postdoctoral researcher in English Linguistics at the University of Munich. Her book on the use of English by Panamanians of West Indian descent was published by John Benjamins in 2023. Her research interests include World Englishes and language in film.

Robert Mailhammer is Professor of Linguistics at Western Sydney University, having held positions in Europe, the U.S., and in Australia. Robert has published widely with a focus on the history of the Germanic and the Australian languages as well as language contact.

Diana Wengler is a postdoctoral researcher at the University of Regensburg, Germany. Her Ph.D. in English Linguistics at LMU Munich on historical audio recordings from the Bahamas dealt with the recent history of Bahamian Creole as well as with issues of dialect representation and the phonology-grammar interface in language change.

Ronia Zeidan has completed a Bachelor of Arts (Honours), majoring in Linguistics, at Western Sydney University. She works as editor and writer for the company BigPicture in Sydney.