Measuring analyticity and syntheticity in creoles*

Jeff Siegel, Benedikt Szmrecsanyi and Bernd Kortmann
University of New England (Australia) and Freiburg Institute for Advanced
Studies / University of Leuven / University of Freiburg

Creoles (here including expanded pidgins) are commonly viewed as being more analytic than their lexifiers and other languages in terms of grammatical marking. The purpose of the study reported in this article was to examine the validity of this view by measuring the frequency of analytic (and synthetic) markers in corpora of two different English-lexified creoles — Tok Pisin and Hawai'i Creole — and comparing the quantitative results with those for other language varieties.

To measure token frequency, 1000 randomly selected words in each creole corpus were tagged with regard to word class, and categorized as being analytic, synthetic, both analytic and synthetic, or purely lexical. On this basis, an Analyticity Index and a Syntheticity Index were calculated. These were first compared to indices for other languages and then to L1 varieties of English (e.g. standard British and American English and British dialects) and L2 varieties (e.g. Singapore English and Hong Kong English). Type frequency was determined by the size of the inventories of analytic and synthetic markers used in the corpora, and similar comparisons were made.

The results show that in terms of both token and type frequency of grammatical markers, the creoles are not more analytic than the other varieties. However, they are significantly less synthetic, resulting in much higher ratios of analytic to synthetic marking. An explanation for this finding relates to the particular strategy for grammatical expansion used by individuals when the creoles were developing.

Keywords: creole, analytic, synthetic, corpus, Hawai'i Creole, Tok Pisin, frequency

1.   Introduction

Expanded pidgins and creoles (referred to together in this article as creoles) are often characterized as being predominantly analytic (or isolating) languages as opposed to synthetic languages (Romaine 1988: 28–9; Parkvall 2008: 282; Bakker et al. 2011: 7–8). The implication is that they are more analytic than their lexifiers and other languages. The focus is generally on analyticity with regard to marking of grammatical categories. That is, when creoles are characterized as analytic, it means that when they have grammatical markers, these are generally free rather than bound morphemes. For example, in English-lexified Hawai'i Creole, past tense is usually indicated by the preverbal marker *wen* rather than the suffix *-ed*, as in *dey wen paint his skin* 'they painted his skin' (Morales 1988: 72).

While this view about creole analyticity is widespread, there has been up to now no rigorous research to verify it. This article reports on a preliminary study undertaken to begin rectifying this situation. The study examines quantitatively the use of analytic and synthetic marking in two different English-lexified creoles compared to other languages, and especially to varieties of the lexifier.

The article begins with a short account of previous research in this area, and then presents working definitions of analyticity and syntheticity. This is followed by information on the data and methodology used in the study. The next two sections report on the findings. A discussion section then offers some explanations for the findings, followed by the conclusions.

2.   Previous research

The impression that grammatical marking in creoles is predominantly analytic, and that creoles are more analytic than their lexifiers, often comes from a focus on particular linguistic features such as tense, modality (or mood) and aspect (TMA) marking in the verb phrase and plural and nominal possessive marking in the noun phrase. In English, for example, four out of five of these features are normally marked synthetically (with an affix or clitic): past tense (e.g. walk*ed*), progressive aspect (e.g. walk*ing*), plural (dog*s*), and nominal possession (the dog*'s* bone); only future is marked analytically (e.g. *will* walk). Siegel (2012) examined whether these five linguistic features are analytic or synthetic in the 18 creole grammars in Holm & Patrick (2007), based on spelling conventions (see Table 1.) He found that out of 90 possibilities in the creoles (5 times 18), these features are marked with an overt morpheme, bound or free, in 84 cases (or 93.3%). Out of these 84 cases, only a bound morpheme (i.e. synthetic marking) is used 11 times (or 13%) and a free morpheme (i.e. analytic marking) is used 73 times (or 87%). For four

Table 1. Synthetic versus analytic constructions for five features in 18 creoles. Shaded features are unmarked; boxed features are only synthetic; the others are analytic. N_1 = possessor; N_2 = possessum; * -*s* also exists.

	PAST	PROG	FUT/IRR	PL	POSS
Angolar	*ta V*	*θa ka V*	*kia V*	*anɛ N*	N_2 *(ri)* N_1
Berbice Dutch	*wa V*	*V-a(rɛ)*	*ma/si V*	*N-apu*	N_1 *si* N_2
Cape Verdean	*V-ba*	*(s)ta V*	*ad/al/a V*	*uns N**	N_1 *di* N_2
Guinea-Bissau Kriyol	*V-ba*	*(s)ta V*	*ta/na V*	*uns N**	N_1 *di* N_2
Dominican	*te V*	*ka V*	*ka alé V*	*se N*	$N_2 N_1$
Haitian Creole	*te V*	*ap V*	*a(va)/va V*	*N yo*	N_2 *(a/pa)* N_1
Jamaican Creole	*did/ben V*	*(d)a V*	*(a)go/wi V*	*N-dem*	$N_1 N_2$
Korlai	*ti V*	*(te) V-n*	*lə/tɛd V*	–	N_1 *su* N_2
Krio	*bin V*	*dè/dé pan V*	*go V*	*N dɛn*	N_1 *in/den* N_2
Nubi	*kan V*	*gi V*	*bi-V*	–	N_2 *ta* N_1
Ndjuka	*be V*	*e V*	*sa/o V*	*den N*	N_2 *fa* N_1
Nagamese	*V-ise/yse*	*V-yose*	*V-to*	*N-khan*	N_2 *laga* N_1
Negerhollands	*a V*	*le/lo V*	*lo/sa V*	*N sini*	N_2 *fan* N_1
Palenquero	*V-ba*	*ta V*	*tan V*	–	N_2 *ri* N_1
Papiamentu	*tabata V*	*ta V*	*lo/ta V*	*N nan*	N_2 *di* N_1
					N_1 *su* N_2
Seychellois	*ti V*	*(a)pe V*	*a(va) V*	*ban N*	$N_2 N_1$
Tok Pisin	*bin V*	*V i stap*	*bai V*	*ol N*	N_2 *bilong* N_1
Zamboangueño	*ya V*	*ta V*	*ay V*	*mana N*	N_2 *di* N_1

Table 2. Percentages of synthetic vs analytic marking for 5 features.

	synthetic	analytic
all 18 creoles	13	87
4 English-lexified creoles	5	95
English	80	20

English-lexified creoles (Jamaican Creole, Krio, Ndjuka, and Tok Pisin), only 1 out of 19 possibilities (5%) is synthetic compared to 4 out of 5 (8%) in the lexifier. The figures are summarized in Table 2.

Some information about the analyticity of creoles also came from the preliminary database for the *Atlas of Pidgin and Creole Language Structures* (Michaelis et al. 2013). Four of the five grammatical features mentioned above (all except future/irrealis) are specifically dealt with for 70 languages. With regard to the

expanded pidgins and creoles that grammatically mark past tense and progressive aspect, less than a quarter use an affix or clitic while more than two thirds use a separate word. Some use both, and other mechanisms such as stem change and reduplication. However, for tense and aspect, analytic marking is by far the most frequent strategy. With regard to plurals, analytic and synthetic marking occur almost equally in the creoles, but the presence of analytic marking appears to be much more common than in the lexifier languages. Synthetic marking in nominal possession is relatively common in lexifiers other than English; however, where such possessive markers occur in the creoles, about nine tenths are analytic, including those in most of the English-lexified creoles.

So it appears that on the basis of type frequencies of grammatical markers in these two sources of data, the generalization about the grammatical analyticity of creoles has some basis. On the other hand, these preliminary examinations have looked at only a small subset of grammatical markers, and cannot compare to the corpus-based studies measuring token frequencies of analyticity and syntheticity that have been done for a range of languages (Greenberg 1960) and varieties of English (Szmrecsanyi 2009; Szmrecsanyi & Kortmann 2009, 2011). It is the methodology of these studies that is adopted in the research on analyticity and syntheticity in two creoles reported on in this article.

3.   Defining analyticity and syntheticity

Here we offer precise definitions of analyticity and syntheticity as used specifically in this study (which are similar to those underpinning Szmrecsanyi 2009 and Szmrecsanyi & Kortmann 2009, 2011). We are interested, first of all, in the overt coding of grammatical information, so we are not concerned with derivational morphology. Thus, we focus on the use of either analytic constructions (analyticity) or synthetic constructions (syntheticity) to realize grammatical distinctions. Analytic constructions are commonly defined as those 'in which grammatical distinctions are expressed by use of separate auxiliary words' (Trask 1993: 15) and synthetic constructions as those 'in which grammatical distinctions are expressed by variation in the form of words' (p. 273). We also follow linguists such as Danchev (1992: 26) in using the term 'free' for the markers that carry grammatical meaning in formal analyticity and 'bound' for those in formal syntheticity. With these provisos in mind, we give the following operational definitions:

–   FORMAL GRAMMATICAL ANALYTICITY comprises all those coding strategies where grammatical information is conveyed by free grammatical markers or function words. Such markers are members of synchronically closed word

classes. In English, for example, these include conjunctions, determiners, pro-nouns, prepositions, infinitive markers, and modal verbs. Note that this defi-nition of analyticity and of what should count as a function word, is fairly un-controversial and in accordance with standard reference works (for instance, Bussmann, Trauth & Kazzazi 1996: 22,471).

–   FORMAL GRAMMATICAL SYNTHETICITY comprises all those coding strategies where grammatical information is signaled by bound grammatical markers. Specifically, we take bound grammatical markers to comprise verbal, nomi-nal, and adjectival inflectional affixes and allomorphies such as ablaut or tonal modification. Our model of morphological analysis is hence an item-and-pro-cess model (Hockett 1954) where grammatically marked forms are thought of as deriving from simple forms via some sort of process.[1]

An important distinction between free and bound grammatical markers is that free markers are distinct phonological words whereas bound markers are phono-logically attached to or dependent on another word. The writing systems of the languages we are considering in this study generally represent a free marker as a separate word and a bound marker as a part of another word. Thus, we use this conventional orthography as a basis for our analysis. Of course, this is not ideal, as there is not necessarily an unequivocal relationship between spelling conventions and language structure (Haspelmath 2011). However, as a detailed phonological and morphosyntactic analysis of each language's texts would not be feasible, it is the best option.

One potential problem for our definitions and reliance on orthography con-cerns clitics. By most definitions, a clitic is considered to differ from an affix but nevertheless to be phonologically bound to another word. However, clitics are frequently written as separate words, reflecting their syntactic characteristics. Furthermore, clitics are often phonologically reduced forms of free grammatical markers. In spoken English, for example, function words such as *the*, *a*, *of*, and *to* are most commonly reduced and act as clitics — e.g. *the* /ði:/ becomes /ðə/ or /ð/ before a consonant and /ði/ before a vowel (Dixon 2007: 577). Again, without detailed phonological analysis of the texts, it is impossible to determine which alternant was used — the free function word or the bound clitic. Thus, to be con-sistent, all grammatical markers represented orthographically as separate words are classified as free, even though they may sometimes function as bound clitics.

---

1. What does not enter into our notion of syntheticity, however, is the 'zero morpheme' con-struct postulated in some morphological approaches to deal with paradigmatic contrasts in fi-nite verb forms. We are, in fact, going to be interested in null marking, but only to the extent that null marking serves as an alternative to non-null synthetic (and also analytic) marking, not as an instantiation of synthetic marking.

On the other hand, some clitics do not have free alternants — for example, the 's possessive marker in English. These clitics are generally not written as separate words, and are thus classified as bound.

Finally, we would like to emphasize that we do not use the terms analyticity or syntheticity for the typological categorization of whole languages. (See Schwegler 1990 for a discussion of the use of these terms in linguistic typology.) While we may say, for example, that Language A is more analytic than Language B, this means that, based on representative samples, Language A exhibits a higher frequency of analytic coding strategies than Language B, measured in a consistent manner. Our notion of analyticity and syntheticity is thus a gradient one, rather than a categorical one.

4. Data

This section discusses the data sources used for the present study.

4.1 European languages other than English

To put our findings in a wider cross-linguistic context, we include data representing four spoken varieties of European languages — German, Italian, Russian, and Spanish:[2]

– *spoken German*: four interviews retrieved from various online sources totaling 10,200 words of running text
– *spoken Italian*: four dialogues retrieved from the *Corpora e Lessici dell'Italiano Parlato e Scritto* (http://www.clips.unina.it/) totaling 7,000 words of running text
– *spoken Russian*: two interviews retrieved from the website of a Russian radio station totaling 11,000 words of running text
– *spoken Spanish*: various texts retrieved from the UAM Spanish Treebank (*Laboratorio de Lingüística Informática Universidad Autónoma de Madrid*, http://www.lllf.uam.es/ESP/Treebank.html) totaling 12,100 words of running text.

---

2. These were the native languages of postgraduate students participating in the tutorial 'Research Design' at the University of Freiburg (summer 2007), and were therefore chosen as a convenience sample. Even so, the languages covered represent three different genera — Germanic, Romance, and Slavic — in the Indo-European language family. (Thanks go to these students for help in coding the data from their languages.)

4.2 Varieties of English

To profile spoken varieties of English, we tap a number of publicly available text corpora. The dataset used here partially overlaps with the datasets analysed in (Szmrecsanyi 2009; Szmrecsanyi & Kortmann 2009).

–   *Standard conversational British English*: the spoken-demographic section of the *British National Corpus* (BNC-S), a part-of-speech (POS) annotated corpus (Aston & Burnard 1998). The BNC-S contains about million words of running text.
–   *Standard conversational American English*: the *Switchboard* corpus, which is POS annotated (Godfrey, Holliman & McDaniel 1992). The *Switchboard* corpus contains about 3 million words of running text.
–   *Traditional (spoken) British English dialects*: (East Midlands dialects, North of England dialects, Scottish Lowlands dialects, Southeast of England dialects, Southwest of England dialects): a POS annotated version of the *Freiburg Corpus of English Dialects Sampler* (FRED-S) (Szmrecsanyi & Hernández 2007), which samples 1 million words of running text.
–   *Colonial* (New Zealand English), *language-shift* (Irish English) *and indigenized varieties of English* (East African English, Hong Kong English, Indian English, Jamaican English, Philippine English, Singapore English): the conversational section (register code s1a) of the *International Corpus of English* (ICE) (Greenbaum 1996). The s1a section of each ICE component contains about 360,000 words of running text.

In this study, standard (or general) spoken British and American English, New Zealand English, Irish English, and the traditional British dialects are grouped together as L1 varieties. The indigenized varieties, which have been heavily influenced by second language acquisition, are classified as L2 varieties.

4.3 Creoles

The two creoles examined here were chosen because they have been studied in detail by one of the authors. Nevertheless they have the advantage of representing opposite ends of the spectrum with regard to the degree of influence from the lexifier language, English, in terms of morphosyntax. At one end is Tok Pisin, which has very little influence and has adopted only one grammatical marker from English (plural *-s*, optionally used by some speakers for some words). In this way, it is similar to what are often called 'radical creoles', such as Saramaccan (Byrne 1987). At the other end is Hawai'i Creole, which has been heavily influenced by English, and has adopted many of its grammatical markers (e.g. optional *-s* plural, *-ing* progressive, *was* past tense copula).

3.3.1 *Tok Pisin*

Our spoken Tok Pisin corpus totals about 57,000 words of running text provided by Geoff Smith from his approximately 400,000 word corpus of spoken Tok Pisin, collected with one-to-one interviews from 1985 to 1993 (see Smith 2002). Example (1) is a passage from the corpus. Some typical examples of Tok Pisin grammatical marking that differ from English are the analytic markers *bin* (past tense), *bai* (future) and *blo* (< *bilong*) (possessive), and the synthetic transitive verb marker *-im* (all shown in boldface type).

(1) *Mi blong liklik ples Inlen Riva, na mi **bin** statim skul **blo** mi long naintin eiti tri, na mi wok**im** greid wan i go lo greid siks long Grin Riva iet, na mi put**im** fest chus **blo** mi ose lo Vanimo Aiskul, na mi kam olsem. So ol **bin** aksept**im** mi lo Vanimo Haiskul ose fest chus **blo** mi. Na nau mi wok**im** greid nain nau, mi gat, bikpla tingting osem **bai** mi mas kisim ofa **blo** mi long taim mi lus**im** greid ten long disla haiskul lo Vanimo. So tingting **blo** mi nau **bai** mi laik kisim wok osem polisman sapos mi kis**im** ofa **blo** mi lo neks yia naintin nainti tu tri, sori, so ating liklik tok **blo** mi em tasol.*
‘I'm from the small village of Inland River, and I started my schooling in 1983, and I did grade 1 to grade 6 at Green River, and I put my first choice [for secondary education] as Vanimo High School, and came like that. So they accepted me at Vanimo High School, my first choice. And now I'm doing grade 9; I've got the idea that I must get an offer [for a job] when I finish grade 10 at this high school in Vanimo. So my thought now is that I'd like to work as a policeman if I get an offer next year 1993. Sorry, this is all of my small story.’

3.3.2   *Hawaiʻi Creole*

The spoken Hawaiʻi Creole corpus we analyse totals 13,000 words of running text. The corpus was compiled by Aya Inoue from approximately 30 hours of recorded one-on-one interviews with ten speakers, conducted as part of the ‘Pidgin Today Project’ directed by Siegel[3] (see also Inoue 2007). Example (2) illustrates the kind of language sampled in the corpus. This passage shows several analytic grammatical markers that differ from English — i.e. *wen* (past tense), *gon* (future), and *fo* (complementizer) — as well as synthetic markers as in English (e.g. *-ed* [past tense] and *-s* [plural]).

---

3. The ‘Pidgin Today Project’ was the simplified name given to the USA National Science Foundation-funded project ‘External influences and internal variation in current Hawaiʻi Creole English’ (award number BCS03–45959). Transcriptions were done by the fieldworkers who made the recordings, and who generally used their own individual ways of spelling the language.

(2)   *We **wen** go da school counselor jus want**ed fo** get me into college because she*
        *thot ai was jus some moke da jus **gon** live up da street. Turn out she **wen** go*
        *luk at my SAT scoa**s** an dat ai could scoa fourteen hundred an den she was*
        laik oh 'wo::w' an den dad was laik 'yea::h yea::h'. well yu dass wat ai mean yu
        kno. Dass why even tho dey can speak yu kno good English an stuff laik dat an
        *talk gud grammar sometaim dey don't know how **fo** act. ya dey donno how **fo***
        act sometaims. an dey make plenny yu kno dey judge people laik dat ya? an
        ai tink its unfair. unfair? Ya, actually ai use dat to my advantage. well yeah.
        no ai mean laik dey tink so low an den ai go beat dem in da math in da math
        *problem an dey laik 'eh! Wat happen**ed**?'*

        'We went to the school counsellor in order to get me into college because
        she thought I was just some local layabout who was just going to live up the
        street. Turned out she looked at my SAT scores and saw that I could score
        1400 and then she said "Wow", and then dad said "Yeah, yeah". Well, that's
        what I mean you know. That's why even though they can speak good English
        and stuff like that that, and talk good grammar, sometimes they don't
        know how to act. Yeah, they don't know how to act sometimes. And they
        make plenty…you know, they judge people like that. And I think it's unfair.
        Unfair? Yeah, actually I use that to my advantage. Well, yeah, no, I mean that
        they think so low and then I beat them in math, in the math problem, and
        they go "Eh, what happened?"'

5.   General methodology

The main research question of this study is: How do the representative creoles
compare quantitatively to other languages — especially to L1 and L2 varieties of
the lexifier — in the frequency of tokens and types of analytic and synthetic gram-
matical marking? This section explains the methodology used to answer this ques-
tion.

In the most general terms, to obtain quantitative results, the present study
exploits part-of-speech (POS) annotation, where tokens in a corpus are tagged in
respect to their word class (this includes information on whether nouns, verbs,
adjectives, and certain pronouns carry inflections).[4]

The different data sources we are tapping in this study require slightly different
handling:

---

4. Our data on European languages other than English is based on very simple annotation
schemes with only three categories: (i) word carries a bound grammatical marker, (ii) word is a
function word, (iii) none of the above.

- *POS annotated English corpora*. In the case of those corpora which are POS-annotated in the first place — the BNC-S (Standard conversational British English), *Switchboard* (Standard conversational American English), and FRED-s (Traditional British English dialects) — the token frequencies we report derive from an exhaustive analysis of all the material sampled in the corpora based on the POS tag set of these corpora, which detail dozens of categories. (See Szmrecsanyi 2009 for details.) Our empirical discussion of marker inventory sizes derives from random 1,000 token samples.
- *Non-POS annotated English corpora*. In the case of the ICE subcorpora, which are not POS-annotated -annotated a priori, an algorithm selected 1,000 random decontextualized tokens (i.e. words) per variety studied. Subsequently, these tokens were annotated manually for their part of speech using the BNC tag set with a minor extension. (See Szmrecsanyi 2009 for details and for statistics on how robust findings deriving from 1,000 token random samples are.)
- *Hawai'i Creole*. A custom-made POS annotation scheme for Hawai'i Creole was developed by Aya Inoue, distinguishing between 94 POS categories. Subsequently, an algorithm selected 1,000 random decontextualized word tokens in the Hawai'i Creole texts, which were then annotated using the POS annotation scheme.
- *Tok Pisin*. A custom-made POS annotation scheme for Tok Pisin was developed by Geoff Smith, distinguishing between 55 POS categories. Subsequently, an algorithm selected 1,000 random decontextualized word tokens in the Tok Pisin texts, which were then annotated using the POS annotation scheme.

Each of the corpora from which 1,000 word tokens were randomly selected consists of several texts. The texts for the two creoles were all narratives, recorded in interviews in which people talked about their experiences. The texts for the varieties of English also included similar narratives, as well as interactional face-to-face conversations. Szmrecsanyi (2009: 332–340) examines text type variability in analyticity and syntheticity scores in standard British English, and demonstrates that the spoken text types are all remarkably similar, in stark contrast to written text types. Since text type differences are not significant as long as we look at spoken language, we can therefore safely assume that the samples we are investigating are comparable.

In line with the definitions presented above in Section 2, we next define four overarching categories, identifiable via part-of-speech annotation, into which individual word tokens in the corpus texts may be grouped (examples given are from Standard English).

1. *Analytic word tokens*: words that have a grammatical function — e.g. complementizers (as in *he thinks **that** he will go*), coordinating conjunctions (*he sleeps **and** she reads*), determiners (***the** house*), infinitive markers (*he needs **to** go*),

modals (*he **can** go*), negators (*she will **not** leave*), pronouns (e.g. *I, you, me*), prepositions (e.g. *of, in, at*), comparative and superlative quantifiers (*more* and *most*), and auxiliary verbs (as in *I **have** eaten lunch*).

2. *Synthetic word tokens*: words that contain verbal inflections (e.g. *I walk > he walk-**s***), nominal inflections (*one dog > two dog-**s***), and adjectival inflections (*small > small-**er***). We also include here stem modifications such as *sing > sang* and *goose > geese*, and other non-regular yet clearly bound grammatical markers. Each of these is counted as one instance of synthetic marking.

3. *Simultaneously analytic and synthetic word tokens* such as inflected auxiliary verbs, as in *he **has** eaten lunch*. Each of these is counted twice — once as analytic marking and once as synthetic.

4. *Purely lexical word tokens*, such as singular nouns. These are not associated with any grammatical marking and are therefore not counted as either analytic or synthetic.

On the basis of this categorization, items in the various POS annotation schemes are mapped to one of the above four categories.[5]

6. Token frequencies

In this section, we establish the text frequencies of the analytic and synthetic word tokens in the corpus material, and subsequently calculate indices in a fashion that is inspired by Joseph Greenberg's seminal (1960) paper on quantitative morphological typology.

6.1 Method

Two indices are used to represent frequencies:

– an ADJUSTED ANALYTICITY INDEX (henceforth: 'Analyticity Index'): the ratio of the number of free grammatical markers (i.e. function words) in a text (F) to the total number of content words (i.e. the total number of orthographically transcribed words in the text minus function words) in the text (W), normalized to the random sample size of 1,000 tokens. Hence: Analyticity Index = F/W × 1,000.

– an ADJUSTED SYNTHETICITY INDEX (henceforth: 'Syntheticity Index'): the ratio of the number of words in a text that bear a bound grammatical marker (B)

---

to the total number of content words (i.e. the total number of orthographically transcribed words in the text minus function words) in the text (W), normalized to the random sample size of 1000 tokens. Syntheticity Index = B/W × 1,000.

This exercise in index calculation is essentially the method proposed by Greenberg (1960), who in turn was inspired by Edward Sapir (1921). The reliance on 'words' (instead of grammatical markers) as the basic unit of analysis is mandated by Greenberg's method.

In previous work using such indices, Szmrecsanyi & Kortmann (2009, Table 5.5) compared the three types of varieties of English: traditional (low-contact) L1 varieties (e.g. English Midlands English), high-contact L1 varieties (e.g. New Zealand English) and indigenized L2 varieties (e.g. Singapore English). The indices were calculated on the basis of digitized corpora of spoken data from a subset of five varieties of each type. making a total of 15 varieties. From each variety, 1,000 orthographically transcribed words were randomly selected for analysis. Szmrecsanyi & Kortmann found that frequencies of both analytic and synthetic marking were statistically significantly higher for traditional, low-contact L1 varieties of English (Analyticity Index: 923; Syntheticity Index: 250) than for high-contact L1 varieties of English (Analyticity Index: 852; Syntheticity Index: 204). The frequencies for high-contact L1 varieties were in turn significantly higher than those for indigenized L2 varieties of English (Analyticity Index: 818; Syntheticity Index: 164).[6] As mentioned above, in the present study study the two L1 types are combined, resulting in a comparison again between three types of varieties: L1 varieties, indigenized L2 varieties, and creoles.

6.2 Tok Pisin and Hawai'i Creole in a cross-linguistic perspective

We begin by exploring the two creoles in comparison to other languages. Figure 1 plots Syntheticity Index scores (on the horizontal axis) against Analyticity Index scores (on the vertical axis). Thus Russian — quite expectedly — shows the most synthetic marking in our sample, scoring 1,172 Syntheticity Index points. This is another way of saying that in a sample comprising 1,000 content words in Russian, we find 1,172 inflections (that is, the average content word in Russian carries more than one bound grammatical marker). At the other end of the continuum, we find Tok Pisin with an Syntheticity Index score of just 128 points: so, in a sample of 1,000 Tok Pisin content words we find only 128 bound grammatical markers. In

---

6. Note that Szmrecsanyi & Kortmann (2009, Table 5.5) report unadjusted index scores. To make their findings comparable to the adjusted scores reported throughout the present study, we re-calculated the scores accordingly.
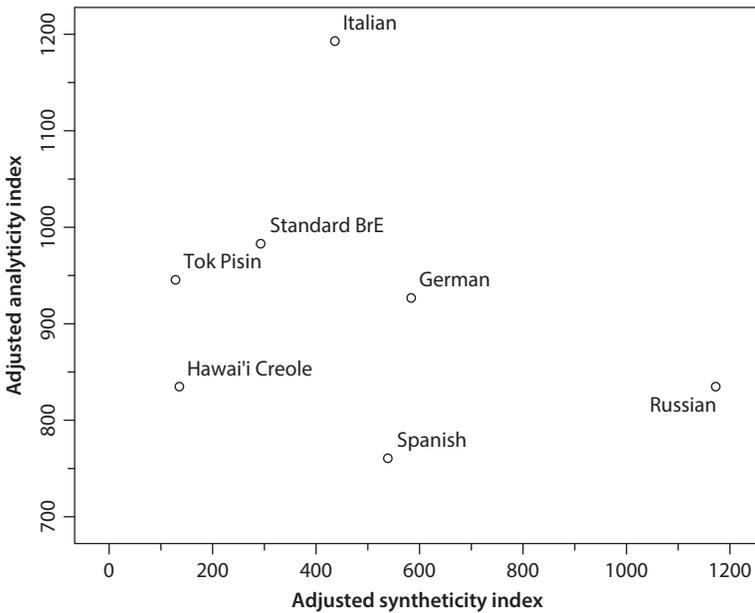
Figure 1. The crosslinguistic perspective: Analyticity Index scores against Syntheticity Index scores (focus on marker frequencies).

terms of analyticity, we note that Italian shows the most analytic marking in our sample: with an Analyticity Index score of 1,193, a sample of 1,000 Italian content words will come with 1,193 accompanying free grammatical markers (or function words).

In all, the crosslinguistic picture portrayed in Figure 1 is that Russian is the most synthetic in grammatical marking but not very analytic, whereas Italian is the most analytic but not very synthetic. With regard to the creoles, according to Figure 1, both Tok Pisin and Hawai'i Creole are not necessarily more analytic in grammatical marking than the other languages. In point of fact, with Analyticity Index scores of 946 and 835, respectively, Tok Pisin and Hawai'i Creole are less analytic than Italian and even slightly less analytic than Standard British English. However, with Syntheticity Index scores below 140, Tok Pisin and Hawai'i Creole are very clearly the least synthetic data points in the sample. Thus, in Figure 1 it is syntheticity that distinguishes the creoles from the other languages.

6.3 Tok Pisin and Hawai'i Creole vis-à-vis varieties of English

6.3.1 *All grammatical markers*
Figure 2 presents a more fine-grained view of the analyticity and syntheticity of Tok Pisin and Hawai'i Creole and plots these in comparison to both L1 and L2
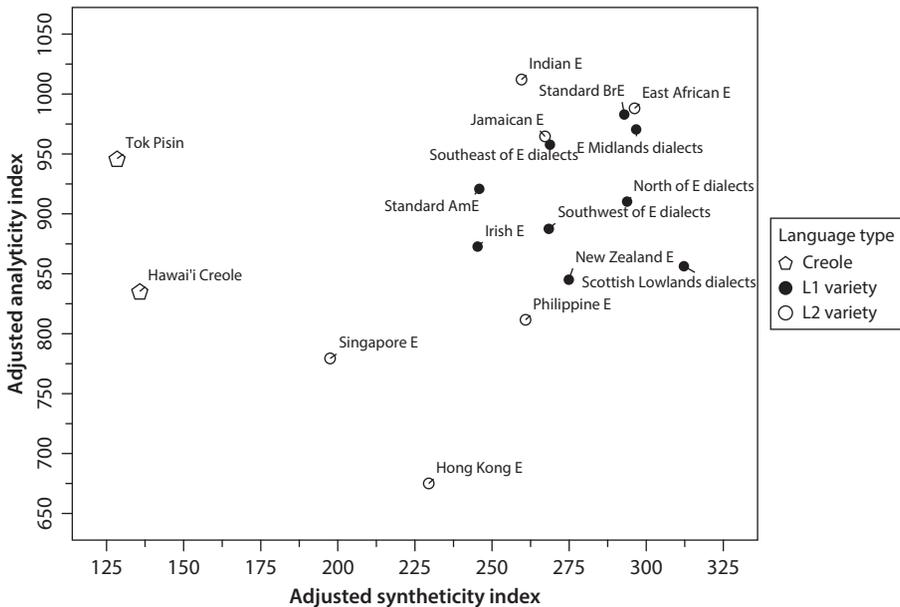
Figure 2. Tok Pisin and Hawaiʻi Creole vis-à-vis varieties of English: Analyticity Index scores against Syntheticity Index scores (focus on marker frequencies). All grammatical markers.

varieties of the lexifier, English.[7] Much as in Figure 1, we again find that the two creoles do not use analytic marking more frequently than L1 and L2 varieties, but that they use synthetic marking much less frequently. Tok Pisin has more analytic marking than Hawaiʻi Creole, and Hawaiʻi Creole has slightly more synthetic marking than Tok Pisin.

More specifically, the average Syntheticity Index for each of the three language types is creoles 132, L1 varieties 278, and L2 varieties 252. Again, these scores are to be interpreted as follows: in a sample comprising 1,000 content words, the creoles attest on average 132 bound grammatical markers, L1 varieties of English 278 bound grammatical markers, and indigenized L2 varieties of English 252 bound grammatical markers. A one-way ANOVA shows that language type is indeed a significant predictor of syntheticity scores ($p < .001$, $F = 24.02$).

The results also show that for all three types of language varieties, the Analyticity Index is much higher than the Syntheticity Index. However, the significantly lower frequency of synthetic markers but not analytic markers in the

---

7. Note that the scale of both indices in this figure is different from that in Figure 1. In this article, we use different scales in the figures for the sake of adequately plotting the observable variation. This is because the purpose of these figures is not to present the absolute scores in either dimension, but rather to show the coordinates of varieties and languages relative to each other.

Table 3. Analyticity Index scores, Syntheticity Index scores, and ratio of analyticity to syntheticity in varieties of English and creoles.

	Analyticity Index	Syntheticity Index	ratio of analytic to synthetic
Standard (spoken) BrE	983	293	3.4
Standard (spoken) AmE	921	246	3.7
New Zealand E	845	275	3.1
Scottish Lowlands dialects	856	312	2.7
North of England dialects	910	294	3.1
English Midlands dialects	970	297	3.3
Southeast of England dialects	958	269	3.6
Southwest of England dialects	888	268	3.3
Mean: L1 varieties	**916**	**281**	**3.3**
Indian E	1012	260	3.9
East African E	988	296	3.3
Hong Kong E	675	229	3.0
Irish E	873	245	3.6
Jamaican E	965	267	3.6
Philippine E	812	261	3.1
Singapore E	779	198	3.9
Mean: L2 varieties Total	**872**	**250**	**3.5**
Tok Pisin	946	128	7.4
Hawai'i Creole	835	136	6.1
Mean: Creoles	**890**	**132**	**6.8**

creoles results in a relationship between analytic and synthetic marking that differs from that of both L1 and L2 varieties. Table 3 gives the figures for Analyticity and Syntheticity Index scores and shows the ratios between them. The analytic to synthetic ratio for creoles of 6.8 — meaning that the frequency of analytic markers in creoles is almost seven times the frequency of synthetic markers — is approximately twice that of each of the other variety types (3.3 for L1 varieties and 3.5 for L2 varieties).

6.3.2 *Focus on markers related to the verb phrase*

In this section, we restrict attention to tense, modality, and aspect markers, agreement markers, infinitive markers, and negators — markers that have received a
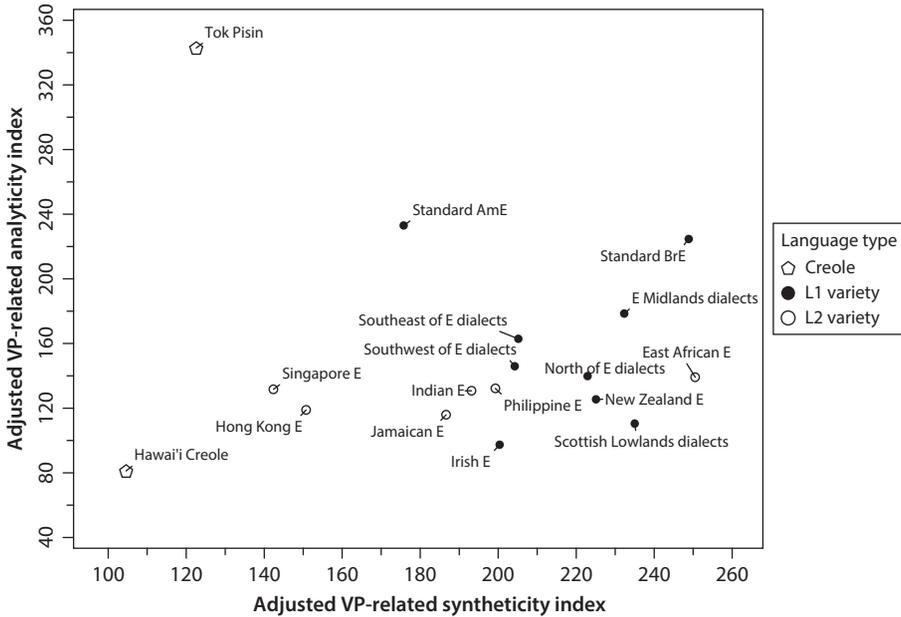
Figure 3. Tok Pisin and Hawai'i Creole vis-à-vis varieties of English: Analyticity Index scores against Syntheticity Index scores (focus on marker frequencies). Verb phrase-related grammatical markers only.

great deal of attention in creole studies.[8] Figure 3 shows that with this restriction in place, the frequency of analytic marking is higher for Tok Pisin than for all the other varieties, including Hawai'i Creole, which has a comparatively low frequency of analytic marking. Again, frequency of synthetic marking is lower for the creoles than for the other varieties.

The differences in analyticity among the three types of languages are once more not significant, but the differences in syntheticity are. The average Syntheticity Index scores for verb phrase marking are: creoles 114, L2 varieties 187, and L1 varieties 217. Language type is a significant predictor of Syntheticity Index scores for all three types ($p = .002$, $F = 10.72$).

6.3.3 *Focus on tense, mood or modality, and aspect (TMA)*

Since TMA marking and plural marking are often focused on in discussions of analyticity in creoles, we look specifically at these two areas. First we examine coding of TMA categories. Figure 4 shows that both creoles do not have higher levels of analyticity in TMA marking than varieties of their lexifier. However, it

---

8. Existential marking is a verb phrase feature in the two creoles, but it is often considered to be a noun phrase feature (using *there*) in varieties of English. Therefore it is not included here.
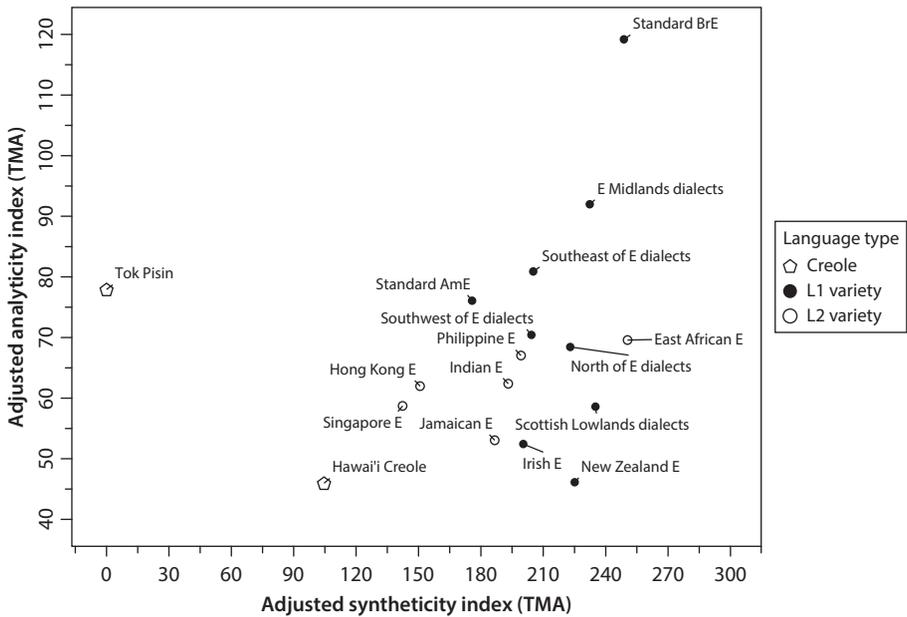
Figure 4. Tok Pisin and Hawaiʻi Creole vis-à-vis varieties of English: Analyticity Index scores against Syntheticity Index scores (focus on marker frequencies). Markers coding TMA only.

also shows that Tok Pisin clearly has an extremely low level of synthetic marking while Hawaiʻi Creole's level is only slightly less than that of L1 and L2 varieties.

6.3.4   *Focus on nominal plurality*
With regard to the analytic coding of nominal plurality (Figure 5), however, Tok Pisin is clearly way above other varieties, which display only synthetic grammatical marking in this area in the data. Again, however, the creoles show less synthetic marking than the other language types.

6.3.5   *Interim summary*
In terms of token frequencies, the two creoles examined here generally do not show significantly higher levels of analytic marking than the other varieties. However, the creoles have significantly lower levels of synthetic marking than other languages or L1 and L2 varieties of English. Thus, in this case low syntheticity does not necessarily mean high analyticity. On the other hand, the low level of synthetic marking in creoles results in a notably higher ratio of analyticity to syntheticity compared with varieties of English.
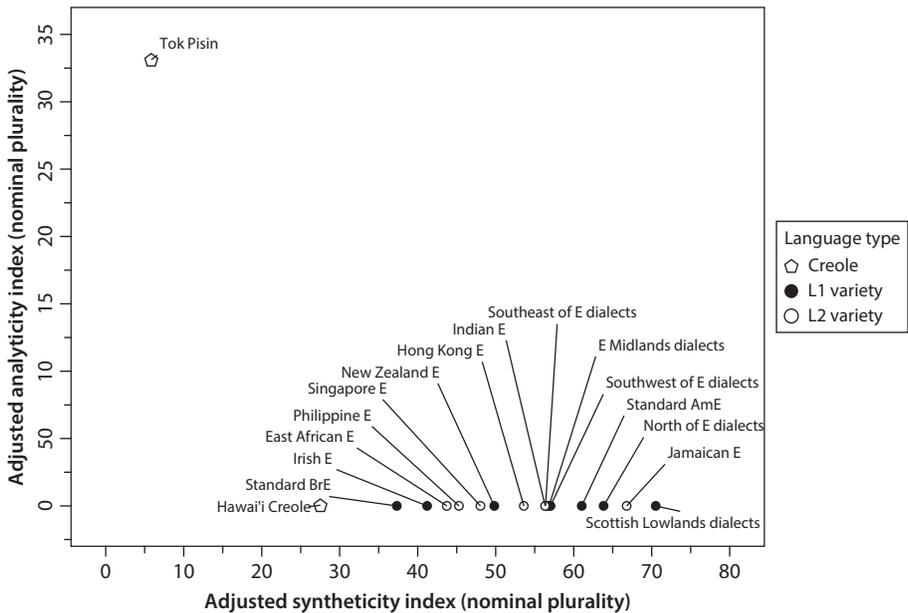
Figure 5. Tok Pisin and Hawai'i Creole vis-à-vis varieties of English: Analyticity Index scores against Syntheticity Index scores (focus on marker frequencies). Markers coding nominal plurality only.

7. Type frequencies

In this section, we take a complementary perspective: rather than determining the token frequencies of grammatical markers, we seek now to approximate the type frequencies — i.e. the sizes of the inventories of analytic and synthetic grammatical markers that are used in each corpus.

7.1 Method

The empirical method is a simple one — we essentially ask: 'In a random sample of 1,000 orthographically transcribed words, how many different analytic or synthetic markers do we find in the texts?' Here we look only at Tok Pisin and Hawai'i Creole in comparison to varieties of English, not in comparison to other languages. (See the Appendix for some representative marker inventories, including Tok Pisin's and Hawai'i Creole's.)

7.2 All grammatical markers

In Figure 6, we see that for all varieties, there are far more analytic markers than synthetic markers. The Tok Pisin text has relatively small inventories of both analytic and synthetic grammatical markers, attesting 46 different analytic marker types and 2 different synthetic marker types. Hawai'i Creole has a slightly smaller inventory of synthetic markers (8) than varieties of English (average inventory of synthetic markers: 11), but the difference is not significant.

Figure 6. Tok Pisin and Hawai'i Creole vis-à-vis varieties of English: total number of analytic types against total number of synthetic types (focus on inventory sizes). All grammatical markers.[9]

Table 4 gives these results in tabular form and presents ratios of analytic to synthetic types. The average ratio for the creoles of 16.3 (meaning that the inventory of analytic markers is 16 times as large as the inventory of synthetic makers) is more than twice as high as that of both the L1 and L2 varieties (7.3). Although the ratio for creoles is skewed by the large value for Tok Pisin (22.5), the value for Hawai'i Creole (10.1) is still higher than the other averages, and higher than that of any other individual L1 or L2 variety.

---

9. Note that data points for Standard BrE and Indian E have exactly the same coordinates, and thus only one dot is shown for both.

Table 4. Number of types of analytic and synthetic markers, and the ratio of analytic to synthetic markers in varieties of English and creoles.

	analytic marker types	synthetic marker types	total number of types	ratio of analytic to synthetic
Standard (spoken) BrE	78	10	88	7.8
Standard (spoken) AmE	71	10	81	7.1
New Zealand E	86	9	95	9.5
Scottish Lowlands dialects	66	9	75	7.3
North of England dialects	70	10	80	7.0
English Midlands dialects	72	11	83	6.5
Southeast of England dialects	73	12	85	6.1
Southwest of England dialects	68	10	78	6.8
Mean: L1 varieties	**73**	**10**	**83**	**7.3**
Indian E	78	9	87	8.7
East African E	83	11	94	7.5
Hong Kong E	60	13	73	4.6
Irish E	72	10	83	7.2
Jamaican E	91	10	101	9.1
Philippine E	70	11	81	6.4
Singapore E	67	9	76	7.4
Mean: L2 varieties	**74**	**10**	**85**	**7.3**
Tok Pisin	45	2	47	22.5
Hawaiʻi Creole	81	8	89	10.1
Mean: Creoles	**63**	**5**	**68**	**16.3**

7.3 Focus on markers related to the verb phrase

In this section, we again restrict attention to grammatical markers modifying verbs (lexical, modal, or auxiliary), infinitive markers, and negators. Figure 7 shows that L1 markers tend to be towards the synthetic pole. Again, Hawaiʻi Creole is fairly average, though its synthetic inventory is slightly smaller (6) than that of most varieties of English (average synthetic inventory size: 7). Tok Pisin has substantially fewer synthetic markers — just one, namely the transitive verb marker *-im* — than varieties of English, but it is among those with the most analytic markers (13).
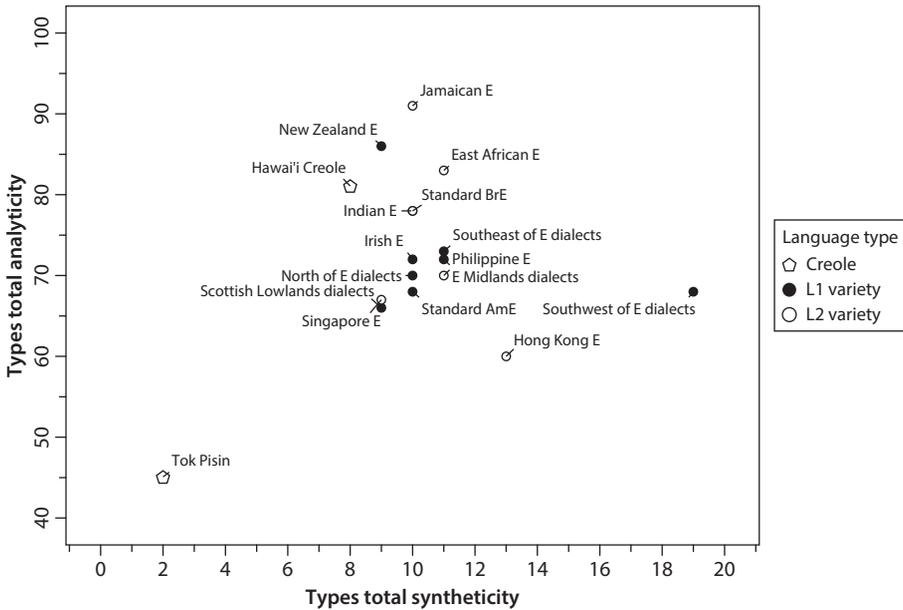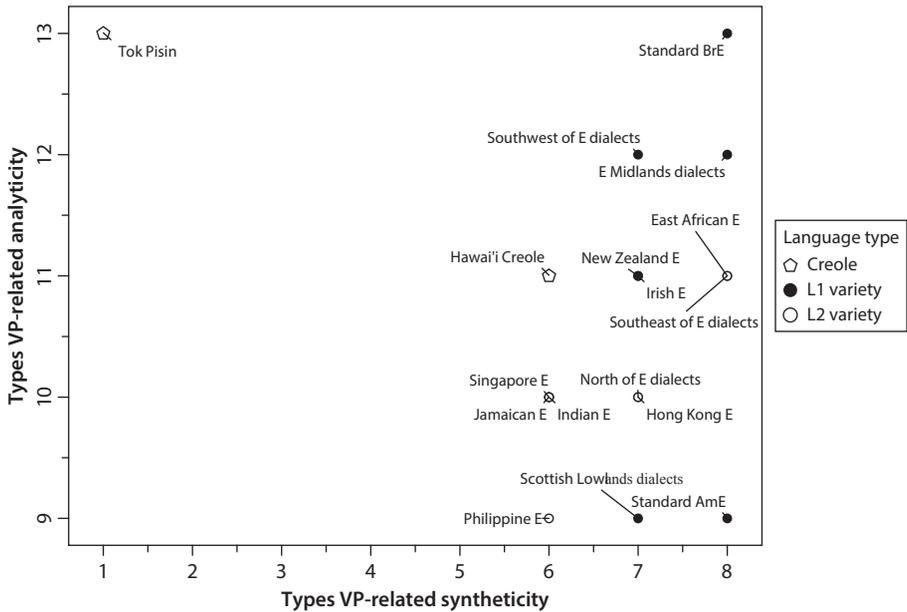
Figure 7. Tok Pisin and Hawai'i Creole vis-à-vis varieties of English: total number of analytic types against total number of synthetic types (focus on inventory sizes). Verb phrase-related grammatical markers only.

7.4 TMA and nominal plurality marking

With regard to the focus areas of TMA and nominal plural marking, the total number of analytic markers versus the total number of synthetic markers that occur in the data are shown in Tables 5 and 6.

Again, we see that Hawai'i Creole does not differ much from varieties of English. On the other hand, Tok Pisin has no synthetic TMA markers as opposed to a range of 6 to 8 in both L1 and L2 varieties of English. And Tok Pisin has an analytic plural marker (*ol* N) in contrast to varieties of English which use only synthetic marking for this function. (Hawai'i Creole also has an analytic plural marker *gaiz* [*guys*], but its use is less frequent and it was not among the 1,000 randomly selected words.)

7.5 Interim summary

With regard to the inventory of types of grammatical markers, Hawai'i Creole patterns close to varieties of the lexifier in all areas examined here while Tok Pisin has a much smaller inventory of both analytic and synthetic markers. But in areas that creolists often focus on, such as TMA and nominal plural marking, Tok Pisin has

Table 5. Tok Pisin and Hawaiʻi Creole vis-à-vis varieties of English: total number of analytic types against total number of synthetic types, TMA markers only (focus on inventory sizes).

	analytic types (TMA)	synthetic types (TMA)	total number of types
Tok Pisin	**5**	**0**	**5**
Philippine E	6	6	12
Indian E	7	6	13
Jamaican E	7	6	13
Singapore E	7	6	13
Scottish Lowlands dialects	6	7	13
Hawaiʻi Creole	**7**	**6**	**13**
Hong Kong E	7	7	14
North of E dialects	7	7	14
Irish E	8	7	15
New Zealand E	8	7	15
East African E	8	8	16
Southeast of E dialects	8	8	16
Southwest of E dialects	9	7	16
Standard BrE	9	8	17
Standard AmE	9	8	17
E Midlands dialects	10	8	18

a much higher proportion of analytic markers to synthetic markers than the other varieties. More significantly, the overall mean ratio of types of analytic markers to types of synthetic markers is higher for both creoles than for L1 as well as L2 varieties of English. This, of course, echoes the text frequency-based findings in Section 5 concerning the ratio of tokens of analytic markers to synthetic markers.

8.   Discussion

The findings described in the preceding two sections raise questions about the similarities and differences between the two creoles and the reasons for their higher ratios of analyticity to syntheticity. These are discussed in this section.

Table 6. Tok Pisin and Hawai'i Creole vis-à-vis varieties of English: total number of analytic types against total number of synthetic types, nominal plurality markers only (focus on inventory sizes).

	analytic types (nominal plurality)	synthetic types (nominal plurality)	total number of types
Singapore E	0	1	1
Scottish Lowlands dialects	0	1	1
E Midlands dialects	0	1	1
Southwest of E dialects	0	1	1
Indian E	0	2	2
East African E	0	2	2
Standard BrE	0	2	2
Jamaican E	0	2	2
New Zealand E	0	2	2
Philippine E	0	2	2
Tok Pisin	1	1	2
Standard AmE	0	2	2
Southeast of E dialects	0	2	2
North of E dialects	0	2	2
Hawai'i Creole	**0**	**2**	**2**
Hong Kong E	0	3	3
Irish E	0	3	3

8.1 Hawai'i Creole versus Tok Pisin

We have seen that Hawai'i Creole uses a greater inventory of both analytic and synthetic grammatical markers than Tok Pisin, but that with regard to TMA marking, for example, Tok Pisin's markers are exclusively analytic whereas Hawai'i Creole's are evenly analytic and synthetic. With such differences in type frequency, the question arises as to why the two creoles pattern together with regard to overall syntheticity in token frequency. First of all, their use of synthetic marking is significantly less than that of both L1 and L2 varieties of English. In the case of Tok Pisin, this can be attributed to the lower inventory of synthetic markers. In the case of Hawai'i Creole, more markers exist, but they are often not used — for example, -*s* present third person singular and -*ed* past tense, as illustrated in this passage from the corpus:

(3)   *And then when he **look** down, **get** this little new oriental teacher Miss Jane.*
Uhm. 'Try fo move da table up on the other end', so 'Miss Jane, I get om' and
*he **hold** da whole table ya, and she **know**. 'Ho Lenson you so strong so strong.'*
*You know, the child cause he **say** 'Mom was a good five minutes.' Good for him,*
*because he **get** plenny power that boy, and plenny love too. One moment he,*
*he, he **get** mad and next moment, he **forget** it; he was the only one wen **ride***
moped you know.
'And then when he looked down there was this little new oriental teacher,
Miss Jane. Um. "Please move the table up on the other end" [she said]. So,
[he said] "Miss Jane, I got it" and he held [i.e. lifted up] the whole table and
she knew. "Ho, Lenson, you're so strong so strong" [she said]. You know he
was still a child because he said "Mom was a good five minutes". Good for
him, he has plenty of power that boy, and plenty of love too. One moment he
gets mad and the next moment he forgets it. He was the only one who rode a
moped, you know.'

At the same time, Hawai'i Creole has several analytic markers that are not found
in varieties of English, such as the preverbal past tense marker *wen*. This can be
seen in the last sentence of the above example, and the following passage from the
corpus:

(4)   *De, a, oda gay. bat da gay, **wen** pul da nayf ae?... ay dono wat ay **wen** duu bat*
*huu ay **wen** bas da gay aap.*
'The other guy, but the guy pulled the knife, eh? … I don't know what I did,
but man I busted the guy up.'

Hawai'i Creole also has markers that are only analytic whereas their counterparts
in English are both synthetic and analytic — for example, the copula/auxiliary *stei*
(*stay*) versus *am/are/is/was/were*, as in the following examples from the corpus:

(5)   a.   *how come aunty **stay** laughing ova dea?*
'How come aunty is laughing over there?'
b.   *Traffic. Ho, when I **stay** Maui, traffic; you just drive.*
'Traffic. Man, when I was on Maui, traffic; you just drive and drive.'

While Tok Pisin has relatively few analytic or synthetic grammatical markers, these
are frequently used and often obligatory — for example, the predicate marker (or
subject referencing pronoun) *i*, the transitive suffix *-im*, and the prenominal plural
marker *ol* in this passage from the corpus:

(6)   *Bifo tru **i** gat wanpla sneik **i** stap lo ples bilong mipla **i** longpla tru na em **i** gat*
*— **ol** meri blongen em **ol** meri tru. Tasol em, em **i** no man tru, em sneik. Tasol*
*taim, sapos em **i** laik go mit**im** ol disla ol meri blongen em tan**im** em olsem*
*man tru, em bai go mit**im** — mit**im** ol. Sapos em **i** stap n ol lak kam bring**im***

*kaikai bloen ol mas, kam na pait**im** wanpla kudu pastem lo rot blong em na*
*ba em, em tan**im** em osem man, orait ol ken kam na luk**im** em.*
'Long ago there was a snake living in our village, it was really long, and it had
wives who were real women. But he wasn't a man, he was a snake. But when,
if he wanted to go meet his wives, he turned into a man and go meet them.
If he was there [by himself as a snake] and they wanted to come bring him
food, they had to first beat a drum on his road so that he'd turn into a man
and then they could come and see him.'

8.2 Accounting for analyticity in creoles

We have demonstrated that while the two creoles examined here are not more ana-
lytic than either L1 or L2 varieties of English in terms of token or type frequency,
they are clearly less synthetic and therefore show a higher ratio of analytic to syn-
thetic marking. These phenomena require some explanation.

To do so, we first look at the overall frequency of grammatical morphology
in the three types of varieties with the use of another index, the Grammaticity
Index, which basically sums up the indices of analytic and synthetic marking and
hence measures overall frequencies of grammatical making (see Szmrecsanyi &
Kortmann 2009; Szmrecsanyi 2009 for discussion):

– GRAMMATICITY INDEX: the ratio of the total number of grammatical markers
  (free grammatical markers (F) plus bound grammatical markers (B)) in a text
  to the total number of words in the text (W), normalized to a sample size of
  1,000 tokens. Hence: $GI = (F + B)/W \times 1{,}000$.

In a 'more is more complex' (Arends 2001: 180) spirit, the Grammaticity Index
may be interpreted to index overall levels of absolute-quantitative (Miestamo
2009: 81) linguistic complexity. The Grammaticity Indices for the three types of
varieties in this study are given in Table 7.

Table 7. Grammaticity Index (overall frequency of grammatical marking) for the three
variety/language types.

	Grammaticity Index
L1 varieties	622
L2 varieties	598
Creoles	541

Between-group differences are significant (one-way ANOVA: $p = .012$, $F = 6.221$).
The difference between Tok Pisin (GI: 552) and Hawai'i Creole (GI: 529) is not sig-
nificant, according to a 2×2 chi-square test of independence. Thus, on the basis of

the Grammaticity Index, the creoles are less complex than the L2 varieties, which are in turn less complex than the L1 varieties.[10]

The lower frequency of grammatical marking in creole languages is not surprising (e.g. see Aboh 2009). The most common explanation for this phenomenon is that it is a consequence of adult second language learning (e.g. Wekker 1996; DeGraff 1999, 2005; Chaudenson 2001; Mufwene 2001; McWhorter 2007). That is, at an early stage in the development of creoles, the lexifier language was a target that was being learned as a second language (L2) for use in wider communication. The learners were adults who had various other first languages (L1s), usually referred to in this context as the substrate languages. This is the 'imperfect second language learning theory' (Muysken & Smith 1995:10).

Studies of interlanguage in early L2 learning show that learners go through a stage when they do not use any grammatical morphology. This is called the 'content-word state' by van de Craats, Corver, & van Hout (2000) and the 'Basic Variety' (BV) by Klein & Perdue (1997), who summarize its structural features as follows: 'Strikingly absent from the Basic Variety are … free or bound morphemes with purely grammatical function' (1997:332). For example, instead of TMA markers, lexical items, such as adverbs, are used. In a survey of studies on the acquisition of tense and aspect, Bardovi-Harlig (2000) describes similar lexical means, rather than grammatical means, used to express temporality in the early stages of L2 learning. Creolists now seem to agree that varieties of the lexifier with features of the Basic Variety were instrumental in the development of creoles — whether these varieties had stabilized into a restricted pidgin or were 'pidginized interlanguages' spoken by individuals (DeGraff 1999:524). Although these simplified varieties eventually expanded to become more grammatically complex, the resultant creoles still demonstrate some 'residual simplicity' in the absence of some grammatical marking (Siegel 2008:49).

This may explain the significantly low Grammaticity Index for the two creoles compared to varieties of their lexifier. It does not explain, though, why the syntheticity component is significantly lower for creoles but the analyticity component is not, leading to a higher ratio of analytic to synthetic marking in the creoles. It could be that when individuals wanted to expand the restricted pidgin or their individual pidginized interlanguages, they continued to target the lexifier and acquired analytic grammatical markers, but not bound morphemes. According to Seuren & Wekker (1986) and Kusters (2003, 2008), the semantic transparency of analytic grammatical markers makes them easier to acquire than synthetic markers. And van de Craats et al. (2000:38) note that in second language acquisition,

---

10. With regard to Tok Pisin, this finding disagrees with that of Nichols (2009:121), who used different criteria for measuring complexity.

items with 'perceptual saliency' are learned first, so that free function morphemes are acquired before bound function morphemes.

This view may account for the acquisition of analytic over synthetic structures when there are alternatives in the lexifier. For example, French has two ways of marking future tense:

(7) French
  a. *Pierre manger-**a**.*
  b. *Pierre **va** manger.*
      'Pierre will/is going to eat.'

French-lexified creoles, such as Haitian Creole, have adopted only the analytic alternative *va* (see Table 1).

However, in the cases of English-lexified creoles, represented here by Hawaiʻi Creole and Tok Pisin, the analytic grammatical markers are often not found in the lexifier language — for example, the preverbal tense and aspect markers *wen* and *stei* (*stay*) in Hawaiʻi Creole and the prenominal plural and possessive marker *ol* and *bilong* in Tok Pisin.

Furthermore, if the ease of acquiring analytic grammatical morphemes as opposed to synthetic ones in L2 learning is the cause of the higher ratio of analytic marking in creoles, then we would expect similar ratios in L2 varieties. However, while we find that the Grammaticity Index for L2 varieties is lower than that of L1 varieties, as we would expect as the result of L2 learning, we do not find a very different ratio of analytic to synthetic marking. In fact, the mean ratios are very close — 3.3 for the L1 varieties and 3.5 for the L2 varieties, compared to 6.8 for the creoles. In their earlier study comparing L1 and L2 varieties of English, Szmrecsanyi & Kortmann (2009) similarly show that the Grammaticity Index is lower for L2 varieties, but this is the result of both the Analyticity Index and the Syntheticity Index being lower. This is very different from what we find in the creoles, where the level of analytic marking is comparable to that in L1 varieties while the level of synthetic marking is significantly lower. This requires some explanation.

Our explanation canters on the existence of two different strategies that can be used in the grammatical expansion of individual interlanguages or a restricted pidgin. Here we use the term 'morphological expansion' to refer to the development of either free or bound morphemes to mark grammatical distinctions. As described above, in early L2 learning L1 speakers target the L2, acquiring lexical items but not grammatical markers, and resulting in interlanguages resembling the Basic Variety (Klein & Perdue 1997). In some situations, common features of individuals' interlanguages may be conventionalized to form a restricted pidgin. This is Stage 1 of L2 learning (see Figure 8a).
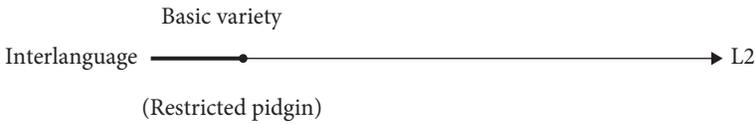
Basic variety

Interlanguage ———————————————————————————► L2

(Restricted pidgin)

Figure 8a. Stage 1 in L2 learning.

If the needs are only for basic communication, speakers may continue to use their own interlanguage versions of the Basic Variety, or the restricted pidgin. However, if the needs for communication expand — for example, with greater integration into the L2 community or the adoption of the L2 as a lingua franca among speakers of different L1s — individuals' interlanguages or the restricted pidgin also need to expand and this involves the development of grammatical morphology. This is Stage 2, and the expansion can occur in two different ways: continuing L2 acquisition with the L2 as the target or expansion without a target.

In the first strategy, continuing L2 acquisition, expansion occurs by L1 speakers learning grammatical morphemes from the L2 and incorporating them into their versions of the L2 (see Figure 8b).
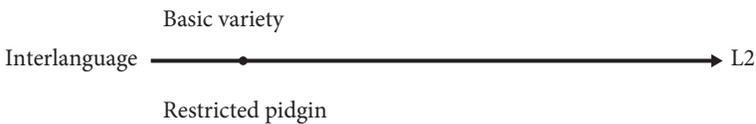
Basic variety

Interlanguage ———————————————————————————► L2

Restricted pidgin

Figure 8b. Stage 2, targeted morphological expansion (by continued L2 acquisition).

In the second strategy, expansion occurs not by acquiring grammatical morphemes from the L2, but rather by creating new grammatical morphemes — usually by giving grammatical functions to existing lexical items in individuals' interlanguages or the restricted pidgin (see Figure 8c).
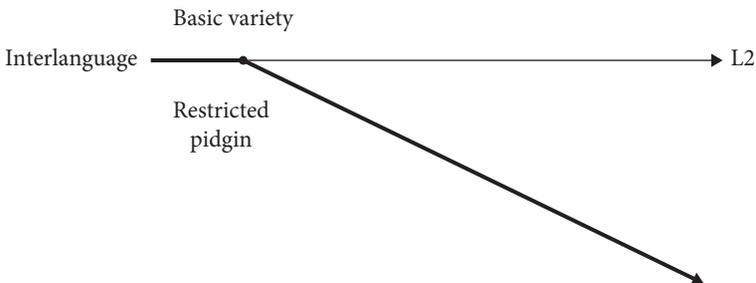
Basic variety

Interlanguage ————————————————————————————► L2

Restricted
pidgin

Figure 8c. Stage 2, untargeted morphological expansion.

These functions are most often based on those of grammatical morphemes in the L1, and thus this grammatical expansion is thought to occur unconsciously through the psycholinguistic process of transfer — more specifically, one kind of functional transfer (Siegel 2008): i.e. using a lexical item from the L2 as if it had the

properties of a syntactically congruent and semantically related grammatical mor-
pheme in the L1. This kind of transfer is recognized as a strategy used by speakers
of a second language to compensate for a lack of knowledge of grammatical fea-
tures felt to be necessary for communication (Kellerman 1995; Siegel 2003, 2008).
The end results look the same as the outcome of the process of grammaticalization,
but the determinants and pathways are different.

Many clear cases of this kind of functional transfer are described by Siegel
(2008). An example from Tok Pisin is the prenominal plural marker *ol*, as seen in
example 5 above and in these sentences from the corpus:

(8) a. *Mipla go lukim **ol** trasel.*
         'We go see turtles.'
     b. ***Ol** bigpla manggi kisim **ol** spia.*
         'The big guys got the spears.'

The marker *ol*, derived from the English word *all*, has taken on the properties of
prenominal markers in substrate languages such as Kuanua and Pala, as in these
examples from Mosel (1980:116):

(9) a. Kuanua: *a **umana** bul*
                  D PL   boy
                  'the boys'
     b. Pala: *a **bar** bulu*
                  D PL boy
                  'the boys'

In Hawai'i Creole, there is a negative possessive and existential marker *nomo(a)*
derived from English *no more*, as in these examples from the corpus:

(10) a. *Wii **nomo** kaa,*
          'We don't have a car.'
      b. *She **no mo'** da kine expensive kine taste buds.*
          'She doesn't have the expensive kind of taste buds.'
      c. ***No moa** aaktapus*
          'There aren't any octopus.'[11]

Here the phrase *no more* appears to have taken on the properties of the negative
possessive and existential marker *móuh* in Cantonese, one of the most influential
substrate languages, as shown in these examples from Matthews & Yip (1994:138,
283):

---

11. Note that transcribers who use an etymological orthography write this marker as two words.

(11) a. *Ngóh **móuh**   saai chín   la wo.*
          1SG   NEG.have all   money PRT PRT
          'I'm out of money.' ['I don't have any money.']
      b. ***Móuh**   yàhn   gaau ngóh Jūgmán.*
          NEG.have person teach 1SG   Chinese
          'There's no one to teach me Chinese.'

An important point here is that untargeted morphological expansion results in mainly analytic grammatical markers since functional transfer normally involves free lexical items acquiring grammatical properties. On the other hand, when expansion involves targeting L2 grammatical morphemes, it appears to result in a balance of analytic and synthetic markers similar to that in the L2.

Both strategies of morphological expansion occur in the formation of creoles and L2 varieties. However, the two types of contact languages differ in which strategy was predominant in their development. It appears that creoles have more frequently adopted the second strategy — i.e. creating new analytic grammatical morphemes through functional transfer — while L2 varieties have adopted the first — acquiring existing grammatical morphemes, both synthetic and analytic, from the L2. This would explain the higher proportion of analytic markers in the creoles in this study compared to both the L1 and L2 varieties.

In the case of Tok Pisin, there has been very little targeted expansion, most probably because of a lack of access to the L2 due to the small number of English speakers available when the language was expanding. In the case of Hawai'i Creole, there was significantly more access to the L2 through the large number of English speakers in Hawai'i plus widespread education in English. Thus, in addition to markers derived from functional transfer, Hawai'i Creole, unlike Tok Pisin, has many synthetic markers from English in its repertoire. However, for reasons of access or identity, these have not been fully incorporated in the language, and are often optional. This accounts for the low frequencies of synthetic marking in Hawai'i Creole, comparable to those in Tok Pisin, despite its comparatively large inventory of synthetic markers.

The idea that the creators of creoles did not always target the lexifier is not new (e.g. Baker 1995, 1997, 2002). But here we have evidence of how untargeted expansion could have occurred.

9.   Conclusion

This preliminary study has found that in terms of the frequency of tokens of grammatical markers occurring in spoken language, two English-lexified creoles

— Hawaiʻi Creole and Tok Pisin — are overall not more analytic than other languages, including both L1 and L2 varieties of their lexifier. On the other hand, both of the creoles use significantly less synthetic marking than the L2 varieties of their lexifier, which in turn use significantly less synthetic marking than L1 varieties.

With regard to type frequencies, the creoles do not have larger overall inventories of analytic grammatical markers than the varieties of their lexifier, and in fact, Tok Pisin's is smaller. However, in relation to the verb phrase, Tok Pisin is among the varieties with the most analytic markers. Furthermore, it is the only variety that has an analytic nominal plurality marker and does not have any synthetic TMA markers.

It is perhaps the concentration on creoles with little influence from the lexifier, such as Tok Pisin, and on particular areas of grammar, such as TMA marking, that have led to the impression that creoles are more analytic than other languages.

The most important findings of this study are twofold. First, for both the creoles compared to the L2 varieties and the L2 varieties compared to the L1 varieties, the lower frequency of synthetic grammatical marking is not matched by higher frequency of analytic marking. Thus, as Szmrecsanyi & Kortmann (2009: 74) point out in similar findings for L2 varieties, 'there is *no* trade-off between analyticity and syntheticity' [italics in original]. Second, however, the creoles' significantly lower levels of synthetic marking are not matched by similar lower levels of analytic marking, as they are for L2 varieties compared to L1 varieties. Thus, although the Grammaticity Indices for the creoles are significantly lower than for both the L2 and L1 varieties, this is due only to the lower levels of syntheticity, and not to lower levels of analyticity. As a consequence, the ratios of analytic marking to synthetic marking for the creoles, based on both token and type frequencies, are significantly higher than those for the varieties of English.

The explanation for this finding relies on the premise that creoles developed from individual interlanguages which had little if any grammatical morphology or from a restricted pidgin based on such interlanguages. During a later stage of morphological expansion, individuals made predominant use of a strategy that did not involve targeting the lexifier and acquiring its existing grammatical morphemes — whether synthetic or analytic. Rather, they unconsciously created new grammatical morphemes by applying the functions of grammatical morphemes in their first language to lexical items of the lexifier that they perceived to be syntactically and semantically similar. Since these new grammatical markers were formerly free content morphemes, they were nearly all analytic.

Of course, this is only a preliminary study, with only two creoles and one lexifier. Clearly further research is needed to examine analyticity and syntheticity in additional creoles and with different lexifiers, and if there are similar findings, to see whether the explanations proposed here have more general applicability.

Appendix: Sample Marker inventories

NOTE: The following lists all refer to random samples of 1,000 orthographically transcribed words.

Standard Spoken British English (BNS-S)
The following list derives from a 1,000 random word token sample taken from BNC text KDY, a text with matches fairly well the overall BNS-S's Analyticity Index and Syntheticity Index scores.

– 64 noun phrase-related and other analytical markers: *it, I, the, you, a, she, that, he, and, in, they, when, all, her, of, cos, to, your, at, but, his, if, nothing, on, we, what, as, for, or, about, every, my, no, this, through, up, according, after, any, because, by, down, half, into, its, like, me, more, most, much, off, one, own, so, somebody, something, them, under, whatever, where, while, who, with, worth*
– existential *there* (analytic)
– 13 verb phrase-related analytical markers: *not, to, will, can, might, should, let's, must, used (to), would* plus auxiliary *do, be, have*
– 2 noun phrase-related synthetic markers: plural *-s*, i-mutation plural
– 8 verb phrase-related synthetic markers:
    – 2 past tense markers: regular *-ed*, lexically conditioned allomorphies (this includes *be, do*, and *have* forms)
    – the *-ing* marker
    – 4 past participle markers: regular *-ed*, lexically conditioned allomorphies (this includes *be, do*, and *have* forms), *-en*, and zero
    – the 3rd person singular *-s* marker

Dialects in the English Midlands (FRED-S-MID)
The following list derives from a 1,000 random word token sample taken from FRED-S text NTT005, a text with matches fairly well the overall Analyticity Index and Syntheticity Index scores in the FRED-S English Midlands material.

– 59 noun phrase-related and other analytical markers: *I, you, and, the, a, it, that, but, they, of, to, in, we, them, for, he, if, some, this, where, at, me, more, what, when, an, as, before, no, or, these, about, anything, by, everything, from, front, him, mi/my, on, till, with, anybody, her, how, near, nobody, now, off, one, ours, out, plenty, since, somebody, something, than, us, which*
– existential *there* (analytic)
– 12 verb phrase-related analytical markers: *not, used, to, would, could, can, might, must, should* plus auxiliary *do, be, have*
– 3 noun phrase-related synthetic markers: possessive *-s*, comparative suppletion, plural *-s*
– 8 verb phrase-related synthetic markers:
    – 3 past tense markers: regular *-ed*, lexically conditioned allomorphies (this includes *be, do*, and *have* forms), zero

- the *-ing* marker
- 3 past participle markers: regular *-ed*, lexically conditioned allomorphies (this includes *be*, *do*, and *have* forms), *-en*
- the 3rd person singular *-s* marker

Singapore English (ICE-SIN)
- 56 noun phrase-related and other analytical markers: *I, you, the, it, we, so, and, if, that, a, they, what, in, he, of, but, her, like, me, or, to, for, his, she, this, us, your, all, as, at, because, how, something, through, who, about, from, my, no, one, them, where, which, with, across, before, between, during, him, on, someone, than, their, till, unless, when*
- existential *there* (analytic)
- 10 verb phrase-related analytical markers: *to, not, can, should, would, could, will* plus auxiliary *do, be, have*
- 3 noun phrase-related synthetic markers: adjectival comparative *–er*, suppletive adjective comparison, plural *-s*
- 6 verb phrase-related synthetic markers:
    - 2 past tense markers: regular *-ed*, lexically conditioned allomorphies (this includes *be, do*, and *have* forms)
    - the *-ing* marker
    - 2 past participle markers: regular *-ed*, *-en*
    - the 3rd person singular *-s* marker

Tok Pisin
- 32 noun phrase-related and other analytical markers: *afta, aninit, antap, bihain, bikos, bilon-gen, dispela, em, frant, husat, insaid/insait, klostu, lo, long, mi, mipela, mitupela, na, ol, olsem, sapos, taim, tambolo, tasol, tupela, wanem/wonem, wantaim, we, yu, yumi, yupela, yutupela*
- existential marker *gat/nogat*
- 13 verb phrase-related analytical markers: *i, bilong, ia, ol, bin, pinis, wok (long), bai, save (+V), mas, no, iet, olsem*
- 1 noun phrase-related synthetic marker: nominal plural *-s*
- 1 verb phrase-related synthetic marker: transitive verb marker *-im*

Hawai'i Creole
- 67 noun phrase-related and other analytical markers: *a, aal, aan/aen/an/and, aan, abawt, aefta, aes/aez, ai/ay, akraas, all, at, awey, bat/but, bawt, biifo, cause/cuz, da/the, daet/dat, daet/dat, dass, dat, dem, dey, deyr, diiz, dis/this, evriting, fo/for, fram, haw/how, he/hi/hii, her, him, hiz, I, in, into, it, ladaet, layk/like, may/my, me/mii, of/ov/av, on, or, over/ova, she/shii, some, somebody, then, they, this, to/tuu, us, wan, wan, wat/what, we/wii, wea, wen, wen, when, who, why, you/yu/yuu, your/yoa/yo, yuunowadaymiin*
- existential markers: *there, get, had*
- 11 verb phrase-related analytical markers: *fo/foa/for, tuu/to, stay, could, can/kaen, use/used/yuustuu, gon, haed, wen, naat/not, no*
- 2 noun phrase-related synthetic markers: nominal plural *-s*, plural *-en* (*chiljrin*)
    - 6 verb phrase-related synthetic markers:
    - 2 past tense markers: regular *-ed*, lexically conditioned allomorphies (this includes *be, do*, and *have* forms)
    - the *-ing* marker

– 2 past participle markers: regular *-ed*, lexically conditioned allomorphies (this includes *be*, *do*, and *have* forms)
– the 3rd person singular *-s* marker

References

Aboh, Enoch O. 2009. Simplicity, simplification, complexity and complexification: Where have all the interfaces gone? In Enoch O. Aboh & Norval Smith (eds.), *Complex processes in new languages*, 1–25. Amsterdam: John Benjamins.

Arends, Jacques. 2001. Simple grammars, complex languages. *Linguistic Typology* 5(2/3). 180–182.

Aston, Guy & Lou Burnard. 1998. *The BNC handbook: Exploring the British National Corpus with SARA*. Edinburgh: Edinburgh University Press.

Baker, Philip. 1995. Motivation in creole genesis. In Philip Baker (ed.), *From contact to creole and beyond*, 3–15. London: University of Westminster Press.

Baker, Philip. 1997. Directionality in pidginization and creolization. In Arthur K. Spears & Donald Winford (eds.), *The structure and status of pidgins and creoles*, 91–109. Amsterdam: John Benjamins.

Baker, Philip. 2002. No creolisation without prior pidginisation. In Terry Crowley & Jeff Siegel (eds.), *Studies in creole linguistics in memory of Chris Corne, 1942-1999 (Te Reo 44)*, 31–50. Auckland: Linguistic Society of New Zealand.

Bakker, Peter, Aymeric Daval-Markussen, Mikael Parkvall & Ingo Plag. 2011. Creoles are typologically distinct from non-creoles. *Journal of Pidgin and Creole Languages* 26(1). 5–42. 10.1075/jpcl.26.1.02bak.

Bardovi-Harlig, Kathleen. 2000. *Tense and aspect in second language acquisition: Form, meaning, and use*. Oxford: Blackwell.

Bussmann, Hadumod, Gregory Trauth & Kerstin Kazzazi. 1996. *Routledge Dictionary of language and linguistics*. New York: Routledge.

Byrne, Francis. 1987. *Grammatical relations in a radical creole: Verb complementation in Saramaccan*. Amsterdam: John Benjamins.

Chaudenson, Robert. 2001. *Creolization of language and culture (revised in collaboration with Salikoko S. Mufwene)*. London: Routledge.

Craats, Ineke van de, Norbert Corver & Roeland van Hout. 2000. Conservation of grammatical knowledge: On the acquisition of possessive noun phrases by Turkish and Moroccan learners of Dutch. *Linguistics* 38. 221–314. 10.1515/ling.38.2.221.

Danchev, Andrei. 1992. The evidence for analytic and synthetic developments in English. In Matti Rissannen, Ossi Ihalainen, Terttu Nevalainen & Irma Taavitsainen (eds.), *History of Englishes: New methods and interpretations in historical linguistics*, 25–41. Berlin, New York: Mouton de Gruyter.

DeGraff, Michel. 1999. Creolization, language change, and language acquisition: An epilogue. In Michel DeGraff (ed.), *Language creation and language change: Creolization, diachrony, and development*, 473–543. Cambridge, MA/London: MIT Press.

DeGraff, Michel. 2005. Morphology and word order in 'creolization' and beyond. In Guglielmo Cinque & Richard S. Kayne (eds.), *The Oxford handbook of comparative syntax*, 293–372. Oxford: Oxford University Press.

Dixon, Robert M. W. 2007. Clitics in English. *English Studies* 88. 574–600. 10.1080/00138380701566102.

Garrote, Marta & Antonio Moreno (project directors). 1999. *UAM Spanish Treebank*. Madrid: Laboratorio de Lingüística Informática, Universidad Autónoma de Madrid. http://www.lllf.uam.es/ESP/Treebank.html. (20 September, 2013.)

Godfrey, John J., Edward C. Holliman & Jane McDaniel. 1992. SWITCHBOARD: Telephone speech corpus for research and development. *IEEE International Conference on Speech, and Signal Processing, ICASSP-92*. 517–520.

Greenbaum, Sidney (ed.). 1996. *Comparing English worldwide: The International Corpus of English*. Oxford: Clarendon.

Greenberg, Joseph H. 1960. A quantitative approach to the morphological typology of language. *International Journal of American Linguistics* 26(3). 178–194. 10.1086/464575.

Haspelmath, Martin. 2011. The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica* 45(1). 31–80. 10.1515/flin.2011.002.

Hockett, Charles F. 1954. Two models of grammatical description. *Word* 10. 210–231.

Holm, John & Peter L. Patrick (eds.). 2007. *Comparative creole syntax: Parallel outlines of 18 creole grammars*. London: Battlebridge.

Inoue, Aya. 2007. Copula patterns in Hawaiʻi Creole: Creole origin and decreolization. In Magnus Huber & Viveka Velupillai (eds.), *Synchronic and diachronic perspectives on contact languages*, 199–212. Amsterdam: John Benjamins.

Kellerman, Eric. 1995. Crosslinguistic influence: Transfer to nowhere? *Annual Review of Applied Linguistics* 15. 125–150 10.1017/S0267190500002658.

Klein, Wolfgang & Clive Perdue. 1997. The basic variety (or: Couldn't natural languages be much simpler?). *Second Language Research* 13. 301–347. 10.1017/S0267190500002658.

Kusters, Wouter. 2003. *Linguistic complexity: The influence of social change on verbal inflection*. Uthrcht: LOT.

Kusters, Wouter. 2008. Complexity in linguistic theory, language learning and language change. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, 3–22. Amsterdam: John Benjamins.

Leoni, Federico Albano (project director). 2000. *Corpora e Lessici dell'Italiano Parlato e Scritto (CLIPS)*. Napoli: Università di Napoli. http://www.clips.unina.it/ (20 September, 2013.)

Matthews, Stephen & Virginia Yip. 1994. *Cantonese: A comprehensive grammar*. London/New York: Routledge.

McWhorter, John. 2007. *Language interrupted: Signs of non-native acquisition in standard language grammars*. Oxford/New York: Oxford University Press.

Michaelis, Susanne, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.). 2013. *Atlas of pidgin and creole language structures online*. Munich: Max Planck Digital Library. http://apicsonline.info/languages/64.

Miestamo, Matti. 2009. Implicational hierarchies and grammatical complexity. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language complexity as an evolving variable*, 80–97. Oxford: Oxford University Press.

Mosel, Ulrike. 1980. *Tolai and Tok Pisin: The influence of the substratum on the development of New Guinea Pidgin*. Canberra: Australian National University (Pacific Linguistics B-73).

Morales, Rodney. 1988. *The speed of darkness*. Honolulu: Bamboo Ridge Press.

Mufwene, Salikoko S. 2001. *The ecology of language evolution*. Cambridge: Cambridge University Press .10.1017/CBO9780511612862.

Muysken, Pieter & Norval Smith. 1995. The study of pidgin and creole languages. In Jacques Arends, Pieter Muysken & Norval Smith (eds.), *Pidgins and creoles: An introduction*, 1–14. Amsterdam: John Benjamins.

Nichols, Johanna. 2009. Linguistic complexity: A comprehensive definition and survey. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language complexity as a variable concept*, 110–125. Oxford: Oxford University Press.

Parkvall, Mikael. 2008. The simplicity of creoles in a cross-linguistic perspective. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, 265–285. Amsterdam: John Benjamins.

Romaine, Suzanne. 1988. *Pidgin and creole languages*. London: Longman.

Sapir, Edward. 1921. *Language: An introduction to the study of speech*. New York: Harcourt, Brace and Co.

Schwegler, Armin. 1990. *Analyticity and syntheticity: A diachronic perspective with special reference to Romance languages*. Berlin: Mouton de Gruyter. 10.1515/9783110872927.

Seuren, Pieter A. M. & Herman Wekker. 1986. Semantic transparency as a factor in creole genesis. In Pieter Muysken & Norval Smith (eds.), *Substrata versus universals in creole genesis*, 57–70. Amsterdam: John Benjamins.

Siegel, Jeff. 2003. Substrate influence in creoles and the role of transfer in second language acquisition. *Studies in Second Language Acquisition* 25. 185–209. 10.1017/S0272263103000093.

Siegel, Jeff. 2008. *The emergence of pidgin and creole languages*. Oxford/New York: Oxford University Press.

Siegel, Jeff. 2012. Accounting for analyticity in creoles. In Bernd Kortmann & Benedikt Szmrecsanyi (eds.), *Linguistic complexity: Second language acquisition, indigenization, contact*. Berlin: Mouton de Gruyter. 10.1515/9783110229226.35.

Smith, Geoff P. 2002. *Growing up with Tok Pisin: Contact, creolization, and change in Papua New Guinea's national language*. London: Battlebridge.

Szmrecsanyi, Benedikt. 2009. Typological parameters of intralingual variability: Grammatical analyticity versus syntheticity in varieties of English. *Language Variation and Change* 21(3). 319–353. 10.1017/S0954394509990123.

Szmrecsanyi, Benedikt & Nuria Hernández. 2007. *Manual of information to accompany the Freiburg Corpus of English Dialects Sampler ('FRED-S')*. URN: *urn:nbn:de:bsz:25-opus-28598.*Freiburg: University of Freiburg. http://www.freidok.uni-freiburg.de/volltexte/2859/ (20 September, 2013.)

Szmrecsanyi, Benedikt & Bernd Kortmann. 2009. Between simplification and complexification: Non-standard varieties of English around the world. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language complexity as an evolving variable*, 64–79. Oxford: Oxford University Press.

Szmrecsanyi, Benedikt & Bernd Kortmann. 2011. Typological profiling: Learner Englishes versus indigenized L2 varieties of English. In Joybrato Mukherjee & Marianne Hundt (eds.), *Exploring second-language varieties of English and learner Englishes: Bridging a paradigm gap*, 167–187. Amsterdam: John Benjamins.

Trask, Robert L. 1993. *A dictionary of grammatical terms in linguistics*. London/New York: Routledge.

Wekker, Herman. 1996. Creolization and the acquisition of English as a second language. In Herman Wekker (ed.), *Creole languages and language acquisition*, 139–149. Berlin: Mouton de Gruyter.

Authors' addresses

Jeff Siegel
Linguistics, BCSS
University of New England
Armidale, NSW 2351
Australia

jsiegel@une.edu.au

Benedikt Szmrecsanyi
University of Leuven
Department of Linguistics
Blijde-Inkomststraat 21
PO Box 03308
B-3000 Leuven, Belgium

benszm@kuleuven.be

Bernd Kortmann
University of Freiburg
Department of English
Rempartstr. 15
79098 Freiburg
Germany

Bernd.Kortmann@anglistik.uni-freiburg.de