Structural nativization, typology and complexity: noun phrase structures in British, Kenyan and Singaporean English

THOMAS BRUNNER

Department of English and American Studies, University of Regensburg
(Received 5 February 2013; revised 27 June 2013)

Two much-cited explanations for linguistic innovations in varieties of New Englishes are cross-linguistic influence (see Gut 2011) and simplification (see Schneider 2007: 82). Using these two notions as starting points, the present study seeks to detect effects of structural nativization in noun phrase (NP) modification in two varieties of English whose substrate languages differ strongly from a typological point of view: Singaporean and Kenyan English. The results yielded by the comparison of random samples extracted from the relevant components of the *International Corpus of English* in the first part of the study show striking correspondences between the preferred NP structures in the varieties at hand and NP structures in the local languages concerned, which, in the light of Mufwene's (2001, 2008) ecological theory of language change, can be interpreted as effects of language contact. The second part of the study shows that the NPs from the three varieties also differ in terms of variables which can be viewed as measures of NP complexity. What is more, the different degrees of complexity found in the samples correspond closely to predictions about the evolutionary status of the varieties at hand made by Schneider's (2007) Dynamic Model.

1 Introduction

It is well known that postcolonial varieties of English have developed linguistic innovations which set them apart from their input varieties spoken in Great Britain or the United States. The levels of language affected by structural nativization in New Englishes range from pronunciation and lexical items, where changes may be immediately evident, to syntax, where novel linguistic structures frequently take the shape of statistical tendencies, which may be indicative of shifting preferences among speakers (see Schneider 2007: 87).

As regards the linguistic causes of such instances of language change in New Englishes, one explanatory model invoked frequently is transfer from local languages (see Schneider 2007: 89; Mesthrie & Bhatt 2008: 44–5). This line of research has spawned a wide range of studies on various linguistic variables across New Englishes (see e.g. Gut 2007; L. Lim 2007; Bao & Khin Khin Aye 2010; Mesthrie 2012). However, there have been only a few studies in this domain which combine the analysis of linguistic influences from indigenous languages in particular contact varieties of English with the broad comparative focus of language typology, i.e. which compare varieties in several contact scenarios involving typologically distinct languages (see Schneider 2007: 89; Szmrecsanyi 2009: 31).

This corpus-based study will help to fill this gap by analysing the influences from indigenous languages on modification patterns of noun phrases (NPs) in Kenyan

and Singaporean English. A comparison between these two varieties is particularly promising as the NP structures of the indigenous languages in Kenya and Singapore differ greatly from a typological point of view: as will be set out in detail in section 2, NPs in the local languages of Singapore tend to be head-final, which means that all modifiers are positioned before the NP head, while the NPs in Kenyan languages are largely head-initial, which means that modifiers follow the NP heads. The corpus analysis below will seek to ascertain to what extent such typological features of indigenous languages are reflected in the NPs of contact varieties.

A second explanatory model for innovations in contact varieties of English which has been advanced many times is simplification: such varieties are known to prefer simpler linguistic structures to more complex ones. Taking this as a starting point, the second part of the article will explore whether NPs in the two contact varieties at hand differ from British English in terms of syntactic complexity.

Section 2 gives an overview of the language situation and the linguistic ecologies in Kenya and Singapore. In section 3, two hypotheses about the distributions of NP structures in the corpora at hand will be developed. Section 4 describes the methods used in the corpus analysis. In section 5, the corpus results pertaining to Hypothesis 1 will be presented and discussed. Section 6 deals with Hypothesis 2, and section 7 provides a summary of the main findings.

2   The languages in Singapore and Kenya

According to an ecological view of language change (see Mufwene 2001, 2008; Ansaldo 2009a, 2009b), language change in such contact varieties as Kenyan and Singaporean English is brought about by speakers selecting and replicating features from the feature pool available to them. The feature pool comprises all models of pronunciation and the morphological and syntactic rules contributed by the languages in contact (see Mufwene 2001: 197–8). It stands to reason that the more typological differentiation there is to be found in a feature pool, the more likely it is that the languages involved will be subject to contact-induced change (see Ansaldo 2009a: 136) or, as Gut (2011: 104–5) has termed it, 'cross-linguistic influence'. The language situations in both Singapore and Kenya are highly salient cases in point, as both speech communities are multilingual, with the vast majority of speakers speaking more than one language on a regular basis. Section 2.1 will prepare the ground for the corpus study by providing a concise description of the varieties of English spoken in Singapore and Kenya. Sections 2.2 and 2.3 will give detailed accounts of the feature pools present in Singapore and Kenya, with a particular focus on the NP structures contributed by the local languages present in the areas.

2.1   Singaporean and Kenyan English

Singaporean and Kenyan English belong to the contact varieties of English which have come to be known as New Englishes. According to Schneider's (2007) *Dynamic*

Model of the evolution of Postcolonial Englishes, such varieties have gone through a uniform process of language change comprising five stages, in which sociopolitical developments in the speaker communities of the new varieties and a growing sense of identity interact with a characteristic process of linguistic evolution (see Schneider 2007: 56).

Kenyan English is assumed to have reached stage 3 ('nativization') by now, which is the phase marked by the 'heaviest effects on the restructuring of the English language itself' (Schneider 2007: 44), i.e. a stage where linguistic innovations of various kinds occur at all levels of linguistic organization, including the development of a characteristic accent, heavy lexical borrowing and nativization at the level of lexicogrammar (see Schneider 2007: 45–7).

Singaporean English, in contrast, has arguably reached phase 4 of the model ('endonormative stabilization'), in which a sense of political disengagement from the mother country contributes to a sense of linguistic independence (see Schneider 2007: 48–9). Singaporean English is a stable, relatively homogeneous, highly entrenched variety, which is nowadays spoken as a home language by 23 per cent of households in Singapore (see Department of Statistics Singapore 2010: ix).

2.2   The feature pool of Singapore English

The language policies of successive governments in Singapore have been aiming at 'English-knowing bilingualism' (Pakir 1991; see Kachru 1983: 99). Singaporeans are expected to be proficient in English and one of the three other official languages: Mandarin Chinese, Malay, or Tamil. This is reflected in the distribution of the languages spoken in Singaporean homes. As mentioned, English is used as a home language in a sizeable 23 per cent of Singaporean homes. It is only outdone by Mandarin, which over a third of Singaporeans (35.0 per cent) speak on an everyday basis. Other home languages are varieties of Chinese such as Hokkien and Cantonese (23.8 per cent), Malay (14.1 per cent) and Tamil (0.9 per cent) (see Department of Statistics Singapore 2010: ix).

In order to gain an understanding of the language-contact situation in Singapore, however, it does not suffice to take into account the languages spoken nowadays. According to the *Founder Principle* (see Mufwene 2001: 28–9), the languages which can be expected to be particularly influential in language-contact scenarios are those spoken in the early days of a community. So Standard Malay, which is spoken by today's Malay population, can be disregarded, since in the formative years of Singapore English, it was mainly used in formal contexts (see L. Lim 2007: 453). From a historical point of view, the main players in the contact dynamics in Singapore would have been (see L. Lim 2007: 452):

- Varieties of Chinese
  - ○ Hokkien
  - ○ Mandarin
  - ○ Cantonese

Table 1. *Noun phrase structures in Hokkien, Cantonese and Mandarin according to WALS (see Haspelmath 2005)*

	Cantonese	Hokkien	Mandarin
Gen – N	✓	✓	✓
Adj – N	✓	✓	✓
Num – N	✓	✓	✓
Dem – N	✓	✓	✓
Rel – N	✓	✓	✓

⋏ Contact varieties of Malay
  ○ Bazaar Malay
  ○ Baba Malay

In what follows, the noun phrase structures which these languages have contributed to the feature pool in Singapore will be considered.

2.2.1 NP structures contributed by varieties of Chinese
The sentences in (1)–(2) show the positions of attributive adjectives and relative clauses in Cantonese NPs:

(1) Léih māmih   haih go hóu hou-haak   ge yàhn.
    your mummy is   CL very hospitable LP person
    'Your mum's a really hospitable person.'[1]

    (Matthews & Yip 2011: 450)

(2) Gó   dī   sīk   Gwóngdūng-wá ge hohksāang háau   dāk
    those CL   know Cantonese   that students examine ADV
    well   a-bit
    hó   dī
    'The students who know Cantonese did better (on the exam).'

    (Matthews & Yip 2011: 326)

    This NP template applies for all three varieties of Chinese in question, as table 1 confirms: in Cantonese, Hokkien and Mandarin, all modifiers and determiners precede the NP head. There are, as a rule, no elements which occur after the head.

2.2.2 NP structures contributed by contact varieties of Malay
While, as has been argued, Standard Malay is not considered to have had a major impact on the contact ecology of Singapore (see L. Lim 2007: 453), a significant role has been played by two contact varieties of Malay: Baba Malay and Bazaar Malay. Baba Malay

---

[1] CL is short for 'noun classifier'; LP designates a 'linking particle'.

used to be the mother tongue of the Peranakan community[2] located in the Straits of Singapore (see S. Lim 1988: 3; L. Lim 2010: 453). Bazaar Malay, in contrast, is not as firmly associated with one region but is best described as an 'interethnic lingua franca' (Ansaldo 2009b: 60), as it has never had native speakers and is arguably less focused as a variety than Baba Malay (see Ansaldo 2009b: 166; Bao & Khin Khin Aye 2010: 156). The lexicon of both Baba and Bazaar Malay is primarily Malay.

The morphosyntax of these languages, however, is heavily marked by language contact and typological convergence with Sinitic languages, particularly Hokkien (see S. Lim 1988: 25), which merits special attention from the point of view of this study: the NPs found in Singapore Bazaar Malay[3] (see Khin Khin Aye 2010) and Baba Malay (see S. Lim 1988: 47) display both head-initial patterns, which are derived from Malay (see (3)), and head-final patterns, which are due to language contact with varieties of Chinese (see (4)).

(3) Saya tengok ini malau manyak bagus.
    1SG look.at DEM monkey many good
    'I saw a very good monkey.' (Bazaar Malay)

(4) Cikgu pukul ini jahat budak.
    teacher beat DEM naughty child
    'The teacher beat this naughty child.' (Bazaar Malay)

    (Khin Khin Aye 2010: 9)

Interestingly, the relevant studies on Bazaar Malay as well as Baba Malay agree that speakers of those varieties strongly prefer head-final patterns.[4] Head-initial patterns are rare and tend to sound archaic (see Khin Khin Aye 2010: 7–13 for Bazaar Malay). S. Lim's (1988: 47) informants for Baba Malay, for instance, opt for the Hokkien-derived pre-head position of determiners instead of the post-head position in 82.7 per cent of cases and prefer pre-head adjectives over post-head adjectives in 65.8 per cent of cases. Pakir (1986: 142) documents a very similar tendency in her study of Baba Malay.

Taken together, it can be argued plausibly that the dominant NP structure in the feature pool in Singapore is head-final. As has been demonstrated, this structure of NPs is categorical in the three varieties of Chinese spoken in the region, and can be considered the numerically dominant variant in the NPs contributed by Baba Malay and Bazaar Malay.

---

[2] The forebears of the Peranakans (Babas) were settlers from southern China who settled in the Straits region of Singapore (see L. Lim 2010: 330) and produced their own contact variety of Malay, which has become known as Baba Malay (see Bao & Khin Khin Aye 2010: 1).

[3] The relevant studies on contact varieties focus on the variety of Bazaar/Baba Malay spoken today. It would certainly be highly promising to look at the historical variety of Bazaar Malay, but in the absence of historical data, present-day Bazaar Malay is probably 'as close as we can get to a specific BZM [Bazaar Malay, T. B.] variety' (Ansaldo 2009b: 161).

[4] The head-initial patterns in Bazaar Malay are derived from Standard Malay, whereas the head-final patterns are due to language contact with Hokkien (see Khin Khin Aye 2010: 13–14).

Table 2. *NP structures in languages spoken in Kenya according to WALS (see Haspelmath 2005)*

Family	Language	N – Adj	N – Num	N – Dem	N – Rel
Niger-Congo	Bukusu				
	Dabida				
	Gusii				
	Kamba	√		√	√
	Kikuyu	√		√	√
	Luyia	√		√	
	Swahili	√	√	√	√
Afroasiatic	Dahalo				
	Oromo (Boraana)	√	√	√	√
	Rendille				
	Oromo (Waata)				
Nilo-Saharan	Luo	√	√	√	√
	Massai	√	√	✗	√
	Nandi	√	√	suffix	√
	Pokot	√	√	suffix	√
	Turkana	√	√	√	√

2.3 The feature pool in Kenya

The multiethnicity of Kenya's society is matched by an enormous diversity of languages. Assumptions as to the number of languages vary a great deal from author to author, ranging from 34 (see Whiteley 1974: 27) to as many as 50 (see Githiora 2008: 236) or 70 (see Skandera 2003: 17).

Fortunately, in spite of the diversity of languages, the typological properties of noun phrases of these languages can be reduced to a common denominator quite safely. It is well known that on the African continent, languages 'exhibit a greater tendency to place modifiers after nouns than languages in other parts of the world' (Dryer 2011: 287), with the proportions of language genera with head-initial NP order at a striking 81 per cent in Africa (as opposed to 16 per cent in Europe) (see Dryer 2011: 290). The tendency towards postmodification is, as Dryer (2011: 307–8) summarizes, 'particularly well pronounced among Niger-Congo, Nilo-Saharan and Chadic languages' – the first two of which cover languages which are spoken by approximately 95 per cent of Kenyans (see Githiora 2008: 237). Table 2, which shows the NP-internal word order of languages spoken in Kenya which are recorded in *WALS* (see Haspelmath 2005), confirms the overall trend towards head-initial patterns.[5]

It is to be expected that a particularly significant contribution in terms of language contact comes from Swahili. Swahili has served as a lingua franca in the region since long before the advent of English, and even during colonial rule, the British followed a

[5] √ denotes 'feature present' – ✗ 'feature not present' – grey shading 'not recorded in *WALS*' (Haspelmath 2005).

trifocal language policy of promoting English for the upper echelons of society, Swahili as lingua franca and the corresponding African mother tongue for local communication (see Schmied 2008: 153). (5)–(7) show the positions of adjective phrases, numerals, possessives and demonstratives in Swahili NPs.

(5) nguo   nyekundu mbili
    dresses red   two
    'two red dresses'

(6) nguo   zake nyekundu mbili
    dresses her red   two
    'her two red dresses'

(7) matunda mabivu haya
    fruit   ripe   these
    'these ripe fruits'

    (Polomé 1967: 143–4)

In terms of noun phrase structures, the Kenyan feature pool is thus much the opposite of the feature pool in Singapore: while in the latter, noun phrases are determined and modified by constituents before the head, the former is marked by a high incidence of modifiers positioned after the NP head.

3 Hypotheses

As far as the NPs of the varieties of English in the two contact varieties are concerned, the following hypotheses will be addressed in the corpus analyses in sections 5 and 6.

Hypothesis 1 is based on the assumption that through cross-linguistic influence, the NPs of local languages in Kenya and Singapore may have influenced the structure of NPs in the two contact varieties of English spoken in these regions.

> ☊ Hypothesis 1: The frequency of head-initial NPs will be higher in Kenyan English than in Singapore English. Conversely, head-final NPs will be more frequent in Singaporean English than in Kenyan English.

The starting point of the second hypothesis is the shared status of Kenyan and Singaporean English as postcolonial varieties of English. In analyses of grammatical innovations in such varieties, it has been claimed time and again that these varieties have a tendency to use simpler structures than native varieties. Well-known phenomena explained along such lines are e.g. the abandonment of vocalic distinctions in phoneme systems (for instance, the coalescence of the KIT and the FLEECE or the FOOT/GOOSE vowels in Colloquial Singapore English; see Bao 1998: 156), a widespread tendency to omit inflectional endings or e.g. the overgeneralization of plural endings to mass nouns (see Schneider 2007: 89; Mufwene 2001). This tendency towards simplification has been generally attributed to 'universal laws of ontogenetic second-language acquisition' (Schneider 2007: 89): in many relevant speech communities such as Singapore and

Kenya, English functions as a second language alongside other local languages and is acquired relatively late in life. This ties in with research on second-language acquisition, where it is well known that there is a strong correlation between a variety of measures of syntactic complexity and the skills of language learners (see Lu 2010; Ortega 2003). With respect to the sample to be analysed below, the following hypothesis can thus be formulated:

> ⚞ Hypothesis 2: NP structures in Singapore and Kenyan English will be simpler than NPs in British English.[6]

4   Methods

Hypotheses 1 and 2 will be tested by analysing corpus data taken from the two postcolonial varieties. For purposes of comparison, corpus evidence from British English, which is the input variety common to the two contact varieties, will be considered as well. The following sections set out the methods used in the corpus analysis: section 4.1 deals with the choice of corpora, 4.2 with the retrieval of data, sections 4.3 and 4.4 operationalize the variables used to test Hypotheses 1 and 2, and 4.5 describes statistical tools used for the interpretation of corpus data.

4.1   *Choice of corpora*

In order to facilitate comparison between the three varieties in question, it was decided to tap into matching components of the *International Corpus of English* (ICE; see Nelson 1996).[7] More specifically, all data were drawn from the section containing spontaneous conversations (texts S1A-001–S1A-090), as this register is arguably the least stylized and can therefore be expected to be susceptible to contact-induced language change. The section comprises roughly 180,000 words in ICE-GB and ICE-SING. ICE-EA provides data from Tanzania and Kenya; however, only the Kenyan subsection of the corpus was used. It should be noted that the Kenyan subsection corresponding to S1A in ICE-GB and ICE-SING has 53,967 words only (see Hudson-Ettle & Schmied 1999: 54). Considering the high text frequency of noun phrases, though, the difference in corpus size should not pose a serious problem.

---

[6] As spelt out in more detail in section 4.4 below, simplicity/complexity in this respect is understood as the degree of hierarchical organization of NPs.

[7] The design of the ICE corpora has received ample treatment (see e.g. Nelson 1996). Suffice it to say that the chief goal of the ICE project has been to create a set of comparable 1-million-word corpora in order to provide a tool for the comparison of both first- and second-language varieties of English (see Greenbaum 1991: 84). Speakers and writers who have contributed texts to the corpus have at least secondary education (see Nelson 1996: 27).

Table 3. *Numbers of potential NPs (retrieved automatically)*

Corpus parts		Number of potential NPs	Frequency per 1000 words
ICE-GB	S1A-001–S1A-090	64141	356.3
ICE-SING	S1A-001–S1A-090	71201	395.6
ICE-EA	S1A-001-K–S1A-030-K	21050	390.1

Table 4. *Relevant and irrelevant NPs in the random sample*

	ICE-GB	ICE-SING	ICE-EA	Sum
Total random sample	2,200	2,200	2,200	6,600
Irrelevant	860	1,006	844	2,710
Relevant	1,340	1,194	1,356	3,890

4.2   Data retrieval

In order to extract sufficient numbers of NPs for the corpus analysis in a time-effective way, a semi-automatic retrieval mechanism was developed. As a first step in the extraction process, the corpus parts were part-of-speech tagged using the *CLAWS* 4 tagging system (see Garside & Smith 1997).[8] Subsequently, all words having the tag of word classes which can potentially be NP heads (i.e. all types of nouns, all pronoun classes, numerals etc.) were retrieved by means of a script implemented in the *R* environment (R Development Core Team 2008), which lays all the putative NPs out as a table. Table 3 shows the numbers of putative NPs yielded by this procedure for each of the three corpora along with normalized frequencies.[9]

As the next step, random samples of 2200 NPs were drawn from each corpus (see table 4), which were subsequently post-edited, weeded out by hand, and annotated. It is important to note that only lexical NPs were taken into account. All types of pronouns as well as proper names were disregarded from the outset since they are modified very

---

[8] According to the manual (see Garside & Smith 1997) which comes with *CLAWS*, its tagging engine reaches an accuracy of part-of-speech (POS) assignment of 96–97 per cent. It could be assumed that the tagger, which is originally geared towards an analysis of British English, might be less apt for second-language varieties used in the present study. Van Rooy & Schäfer's (2003: 836) study, however, shows that even in non-native data such as the *Tswana Learner English* corpus, an impressive 96.26 per cent of POS tags were assigned correctly. Manual spot checks have been conducted on part of the tagged material used for this study, and the consistency rate was indeed found to be up to this standard.

[9] It is important to note that the frequencies in table 3 should not be taken at face value since the output of the extraction routine unavoidably contains false hits, which can only be weeded out after the random sample has been drawn. It is certainly striking that Singaporean English and Kenyan English should have a higher NP incidence than British English. Whether this constitutes a stable tendency, though, would have to be established in a separate study focusing specifically on this question. The normalized frequencies are based on the respective word counts (180,000 words for ICE-GB and ICE-SING; 53,967 words for the Kenyan component of ICE-EA; see Hudson-Ettle & Schmied 1999: 54).

Table 5. *NP patterns (see de Haan 1993: 86)*

Type	NP pattern				Examples
prem+post	(DET)	PREM+	HD	POM+	*the able-bodied members of the Mike Heafy group* <ICE-GB:S1A-001 #52:1:A>
post	(DET)		HD	POM+	*the work that I was involved in* <ICE-GB:S1A-001 #31:1:B>
prem	(DET)	PREM+	HD		*a physical language* <ICE-GB:S1A-001 #36:1:B>
none	(DET)		HD		*the beginning* <ICE-GB:S1A-001 #43:1:A>

DET 'determiner' – PREM 'premodifier' – HD 'head' – POM 'postmodifier'

rarely. What is more, under the following three conditions, lexical nouns in the random sample were excluded from the analysis: (1) if they did not have a function within a regular syntactic structure, i.e. stood alone in elliptical utterances (e.g. *One box* as a reply to the question *How many walnut cakes you want to order* <ICE-SIN:S1A-006#49:1:A>); (2) if they belonged to fixed, conventionalized collocations such as *part* in *take part in*, since such NPs cannot be modified freely; (3) if their syntactic status and/or meaning was uninterpretable because of self-corrections, slips of the tongue or disfluencies, which is not infrequent in spontaneous spoken language. The number of irrelevant NPs in each corpus part can be seen in table 4.

4.3   How to test Hypothesis 1 – definition of variables

The two variables which will be analysed in order to unearth possible typological influences in NPs of Kenyan and Singaporean English are: (1) the frequencies of pre- and postmodified NPs in the sample and (2) the lengths (in words) of the pre- and postmodifiers of the NPs extracted from the corpora. These variables will be defined and operationalized in sections 4.3.1 and 4.3.2.

4.3.1   NP patterns
In order to establish the usage frequencies of head-final and head-initial NPs in the varieties at hand, all NPs in the random sample were classified into the following four structural patterns, as suggested by de Haan (1993: 86) (see table 5).

The most basic noun phrases do not have a modifier and consist of a head only (type 'none'). NPs which have one or several premodifiers are filed under 'prem', which, typologically speaking, represents a head-final pattern. NPs in category 'post', which have one or several postmodifiers, represent head-initial NPs. Finally, NPs which are both pre- and postmodified are categorized as 'prem+post'. As the classification

is geared specifically towards an analysis of NP modifiers, the use of determiners is optional in all of these patterns.

The patterns 'prem' and 'post' are particularly germane to Hypothesis 1, as the pattern 'prem', which is exemplified in (8)–(11), represents a head-final pattern:

(8) the most uh high-class place <ICE-SIN:S1A-057#119:1:A>
(9) a white-bashing session <ICE-SIN:S1A-034#97:1:A>
(10) a very civilised Kenyan <ICE-EA:S1A012K>
(11) property development exercise <ICE-EA:S1A015K>

The pattern 'post', by contrast, stands for a head-initial pattern, as can be seen from (12)–(15).

(12) this place which was a factory outlet <ICE-SIN:S1A-003#174:1:C>
(13) a sense of environmental awareness <ICE-SIN:S1A-070#68:1:A>
(14) an interview yeah for management trainees at EAI <ICE-EA:S1A001K>
(15) women talking about politics <ICE-EA:S1A027K>

The frequencies of these modification patterns in the random sample will help to establish to what extent preferences for either head-final or head-initial NPs surface in the English of speakers of the two contact varieties. The analysis of Patterns 1 and 4 will be postponed to section 6.1.1, as they are more relevant with respect to Hypothesis 2.

4.3.2 Lengths of pre- and postmodifiers

Another variable which will help to shed light on the relationship between premodification and postmodification (i.e. between head-initial and head-final NPs) in the varieties at hand is the length of modifiers. If speakers are particularly inclined to use one type of modifier, they may show a tendency to use longer modifier types (e.g. clausal postmodifiers) and may also exploit the potential of individual modifier types more fully by packing more information into them or using more elaborate structures. It can be hypothesized that, for example, exceptionally long postmodifiers consisting of multiple nested claused such as (16), which are quite unusual for spontaneous spoken language, are more likely to occur in a variety of English whose substrate favours head-initial NPs. Conversely, very long premodifiers (see (17)) might be preferred in a dominantly head-final feature pool.

(16) the problems uh with which women have to contend uh in order to produce any writings because of the many roles that they have to play <ICE-EA:S1A028K>
(17) some classical music sort of thing <ICE-SIN:S1A-046#133:1:B>

For each of the pre- and postmodified NPs in the sample, the length (in words) of the string of modifiers was determined. If an NP contained concatenated modifiers as in (18), the lengths of all postmodifiers were summed up; hence, a total length of 9 would have been recorded for the two postmodifiers in (18).

(18) everything for human existence that they need like a car <ICE-EA:S1A016K>

In order to test Hypothesis 2, the NPs in the sample will be ranked according to their complexity by means of variables which can be viewed as metrics of NP complexity under a very basic definition of complexity: according to Givón (2009: 2), complexity is to be understood as 'increased hierarchic organization; that is, an increase in the number of hierarchic levels within a system' (see Givón 2009: 2).

As will be argued below, the variables defined in sections 4.4.1–4.4.2 can be interpreted quite straightforwardly in terms of simplicity vs complexity.[10]

4.4.1   Frequencies of NP patterns

The NP patterns used in section 4.3.1 will be considered again in this section. Here, however, the focus will be on the simplest NP pattern 'none' as against all other NP patterns, which involve modifiers. With respect to Givón's definition of complexity cited in section 4.4, it stands to reason that heavily modified NPs such as (19) are to be considered syntactically more complex than unmodified NPs – adding modifiers to a simple NP undoubtedly increases the degree of 'hierarchic organization' (Givón 2010: 2) within the NP.

(19) the curious and delightful thing about working with a dance group <ICE-GB:S1A-004 #6:1:B>

4.4.2   Levels of embedding

Another variable which will be used to assess NP complexity is the degree of embedding of NPs. NP are recursive structures, which means that they do not only occur at the clause level but can also be part of modifiers within other noun phrases. As can be seen in (20), NPs occurring at the level of the clause were assigned the level 0 (e.g. *this growth of interest in CD ROM data* in (20)). NPs occurring in a modifier (e.g. a prepositional phrase) of a level-0-NP were assigned the level 1 (e.g. *interest in CD ROM data* in (20)). NPs in modifiers of level-1 NPs, in turn, were filed as level-2 NPs (e.g. *CD ROM data* in (20)), and so forth. Section 6.1.2 will detail the frequencies of NPs at different levels of embedding in the varieties at hand.

(20) It does raise a point also about you know this [this growth of [interest in [[CD ROM]$_3$ data]$_2$]$_1$]$_0$ <ICE-GB:S1A-029 #358:1:A>

Given Givón's definition cited in section 4.4, this variable can be viewed as another marker of NP complexity: quite obviously, in terms of their 'degree of hierarchical organization' NPs into which other NPs are embedded have to be ranked higher than NPs without nominal embedding. Further examples of NPs containing other NPs embedded at various levels are shown in (21)–(23).

---

[10] It is no easy task to relate these dimensions of complexity to the debates about language complexity currently going on in the literature (see Miestamo, Sinnemäki & Karlsson 2008a; Kortmann & Szmrecsanyi 2009; Sampson, Gil & Trudgill 2009), since those debates 1. usually do not focus on hierarchical complexity (see Miestamo, Sinnemäki & Karlsson 2008b: ix) and 2. largely try to assess the complexity of *langue* rather than *parole* and 3. mostly focus on issues of inflectional morphology.

Table 6. *A contingency table*

	Category a	Category b
Category c	20	40
Category d	60	80

(21) an emphasis on the use of the whole of the body and the use of breath <ICE-GB:S1A-004 #84:1:B>

(22) the strength of my English language competency <ICE-SIN:S1A- 071#55:1:A>

(23) What percentage of uh the population in let's say the students in this hall <ICE-EA:S1A009K>

4.5   Statistical methods

In the corpus analyses in sections 5 and 6, frequency data involving two nominal or categorical variables will be subjected to *Hierarchical Configural Frequency Analyses* (*HCFAs*).[11]

While a chi-square test would only yield one *p*-value for a contingency table such as table 6, an *HCFA* takes into consideration every possible configuration of variables in the contingency table (i.e. categories a–c, a–d, b–c, b–d in table 6) and determines whether the frequency in the cells concerned is higher or lower than expected under the null hypothesis (see Gries 2009: 240–52). In all bar charts used below, greater-than and less-than signs inside the bars indicate the results of the *HCFAs*. For instance, '>***' would mean that the respective frequency is higher than expected under the null hypothesis; asterisks indicate the significance levels established by the *HCFA* test: $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***).[12]

Interval data such as length differences were subjected to (monofactorial) *ANOVAS* (and post-hoc *TukeyHSD*s) in order to establish whether their means are significantly different (see Gries 2009: 275–83). If the requirements of *ANOVAS* were not met, *Kruskal-Wallis* tests (along with post-hoc *Wilcoxon rank sum tests)* were employed as non-parametric variants of *ANOVAS* (see Gries 2009: 283).

5   Testing Hypothesis I: typological influences on NP structures

The following section documents the results of the corpus analysis conducted in order to test Hypothesis 1, which predicts typological influence from local languages on modification patterns of NPs in ICE-EA and ICE-SING.

---

[11] The *HCFAs* have been computed by means of an *R* script developed by Stefan Th. Gries (2004).

[12] Configurations have only been included if the coefficient of pronouncedness *Q* (see Gries 2009: 249) was larger than 0.01.

Table 7. *NP patterns in ICE-GB, ICE-SING and ICE-EA – raw frequencies and percentages*

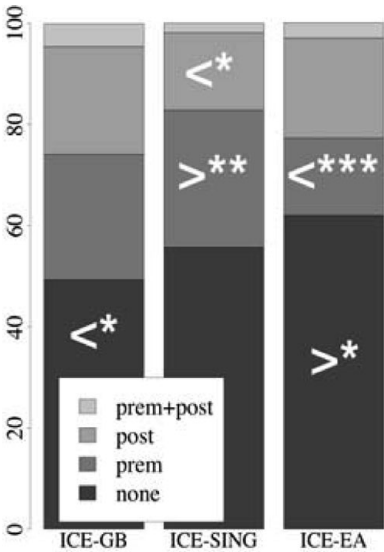|	ICE-GB		ICE-SING		ICE-EA		
	N	%	N	%	N	%	Sum
prem+post	44	4.5	17	1.9	32	3	93
post	208	21.3	135	15.3	214	19.8	557
prem	242	24.8	240	27.1	165	15.3	647
none	481	49.3	492	55.7	670	62	1,643
Sum	975		884		1,081		2,940

Figure 1. NP patterns in ICE-GB, ICE-SING and ICE-EA – percentages

5.1   Results

5.1.1   NP patterns

Table 7 and figure 1 show the distribution of NP patterns (see de Haan 1993: 86; also section 4.3.1 above) across the random samples taken from the three corpora.[13]

---

[13] Three exclusion criteria apply for this distribution: (1) Only NPs functioning as immediate constituents of clauses have been taken into account here. All NPs embedded into other NPs (see section 4.4.2) were excluded since they are arguably subject to different functional constraints. (2) Out of all premodifier types mentioned distinguished by Bache (2000: 239), only descriptive modifiers (such as *very good* in *very good ball sense* <ICE-SIN:S1A-077#145:1:A>) and classifiers (*ball* in *very good ball sense* <ICE-SIN:S1A-077#145:1:A>) are counted as modifiers. Specifiers (such as *same* in *the same garment* <ICE-SIN:S1A-037#98:1:A>) have been left out of the equation as their function is very close to that of determiners and hence subject to much stronger grammatical constraints than the use of descriptors and classifiers.

Table 8. *Numbers of premodified NPs in the modified random sample*

Variety	Premodified NPs	Mean length	Coeff. of variation
ICE-GB	286	1.29	0.45
ICE-SING	257	1.25	0.48
ICE-EA	197	1.16	0.47

As can be seen, NPs in spontaneous speech are generally simple, which is reflected in high proportions of the simplest pattern 'none' across varieties. Complex NPs which are both pre- and postmodified are infrequent.

From the point of view of Hypothesis 1, however, the most significant observation concerns the relationship between patterns 'prem' and 'post'. While being almost evenly distributed in ICE-GB, they occur in clearly different proportions in the contact varieties: ICE-SING displays a significantly higher frequency ($p < 0.01^{**}$) of premodification and a significantly lower proportion of postmodified NPs ($p < 0.05^*$). In ICE-EA, the reverse relationship is obtained: postmodification is more frequent than premodification, and premodified NPs occur at a significantly low rate ($p < 0.001^{***}$). It can be concluded that speakers of Singapore English prefer premodification (i.e. head-final structures) and are biased against postmodification (i.e. head-initial structures), while the reverse is true for speakers of Kenyan English: they strongly disprefer premodifiers and use postmodified nouns more frequently.

5.1.2 Modifier length
As was argued in section 4.3.2, such preferences may also be reflected in the lengths of pre- and postmodifiers.

5.1.2.1 Length of premodifiers   Premodifiers in spontaneous speech tend to be very short. As shown in table 8, the average length in the sample does not exceed 1.3 words.

If one looks closely, a very slight trend is visible in the data: the mean value for ICE-EA is lower than for ICE-GB and ICE-SING. Even though the differences between these means are not significant ($p = 0.058$, F = 2.85, df = 2) according to an *ANOVA*, a post-hoc *TukeyHSD*, which tests for the significance of the differences between individual factor pairs, shows that the difference between the means of ICE-GB and ICE-EA is significant (see table 9).

This tendency tallies with the dispreference of premodification in ICE-EA observed in 5.1.1 and corroborates the impression that speakers of Kenyan English are biased against premodification: not only do they use premodifiers rarely, they also use shorter premodifying strings than in British English.[14]

---

[14] This result should be taken with a grain of salt, since, as can be seen, only the length differences between the input variety and the contact variety are significant, not the differences between the two contact varieties.

Table 9. *TukeyHSD post-hoc comparison of the means of lengths of premodifiers in ICE-GB, ICE-SING, and ICE-EA*

	ICE-EA	ICE-GB
ICE-GB	$p = 0.048^*$	–
ICE-SING	$p = 0.23$	$p = 0.73$

Table 10. *Numbers of postmodified NPs in ICE-GB, ICE-SING, and ICE-EA*

Variety	Postmodified NPs	Mean length	Coeff. of variation
ICE-GB	252	4.27	0.81
ICE-SING	152	3.77	0.80
ICE-EA	246	4.53	0.99

5.1.2.2 Length of postmodifiers In the next step, the postmodifying portions of the NPs in the samples were analysed for length differences.

As can be seen from figure 2, in all three varieties, few postmodifiers are longer than 5 words. The dispersions differ subtly, however, with ICE-SING showing a slightly lower upper quartile and ICE-EA showing a high lower quartile. ICE-EA has a higher coefficient of variation[15] than the other varieties, which indicates that the lengths in ICE-EA are slightly more dispersed than in ICE-GB and ICE-SING (see figure 2 and table 10).

The average lengths of the postmodifying string, however, differ notably: they are 4.27 words and 4.53 words in ICE-GB and ICE-EA respectively, as opposed to only 3.77 words in ICE-SING (see table 10). According to a *Kruskal-Wallis* test, this difference is significant at $p = 0.02$ ($K = 8.21$, df = 2).[16] By means of a pairwise U-test (see Gries 2009: 218–26), the significance of length differences between individual varieties can be computed.

It emerges from table 11 that while not differing markedly from their common input variety British English, there is a significant difference between the postmodifier lengths of the NPs in ICE-SING and ICE-EA.

Obviously, while the Kenyan bias against premodification shortens premodifiers, the Singaporean preference for premodification does not lead to longer premodifiers than in the other two varieties.

[15] The variation coefficient is a measure of dispersion which is computed by dividing the standard deviation by the arithmetic mean. Comparing raw standard deviations across several samples would not be advisable, since the standard deviation is dependent on the mean value of a particular distribution (see Gries 2009: 117).

[16] Running an ANOVA is not an option here since according to a Bartlett test (see Gries 2009: 277), the variances of the lengths are not normally distributed. The *Kruskal-Wallis* rank sum test is a nonparametric alternative to ANOVAs (see Gries 2009: 283).

Table 11. *Pairwise comparison of varieties (Wilcoxon rank sum test)*

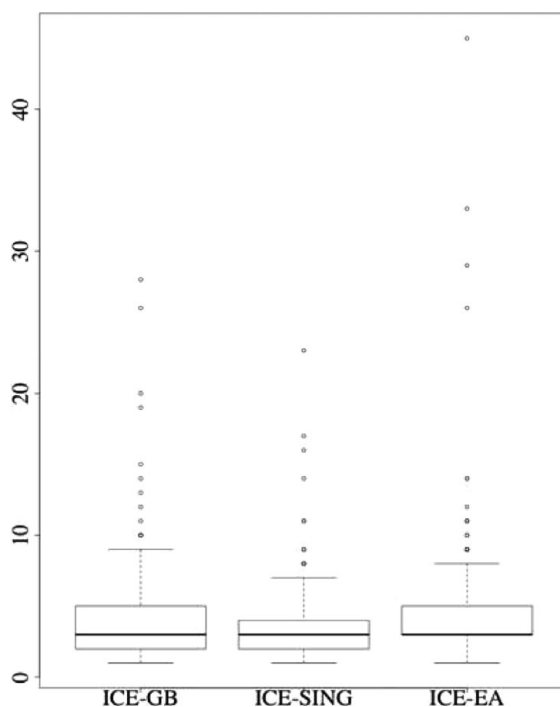|	ICE-GB	ICE-SING
ICE-SING	$p = 0.111$	–
ICE-EA	$p = 0.293$	$p = 0.012^*$

Figure 2. Box plot: lengths of postmodifiers in ICE-GB, ICE-SING and ICE-EA

These effects, too, can be seen as repercussions of the contact-induced tendencies established in section 5.1.1: the bias against postmodification found for Singapore English also results in shorter postmodifiers.

5.2   Discussion: focus on Hypothesis I

The main findings so far clearly confirm Hypothesis 1, which, for Singapore English, predicts a preference for head-final structures. In comparison to the three other NP patterns, in this variety, premodified, i.e. head-final, NPs are used with a significantly higher frequency than expected under the null hypothesis. Head-initial, postmodified NPs are significantly less frequent. Under a usage-based view of language, such corpus frequencies can be viewed as indicators of cognitive entrenchment (see Bybee 2006:

714–15). NPs as structural patterns can be thought of as a category exhibiting prototype effects, with the members most frequent in usage constituting the central members of the category (see Bybee 2006: 726). Thus, within the category of NPs in speakers of Singapore English, the pattern 'prem' is a highly prototypical NP pattern, while the pattern 'post' is marginal.

For Kenyan English, in contrast, Hypothesis 1 predicts a preference for head-initial structures, which is borne out by the sample: in ICE-EA, the incidence of premodified (= head-final) NPs is significantly lower than expected under the null hypothesis, which, cognitively speaking, is indicative of the marginal position of the pattern 'prem' in the category of NP in speakers of Kenyan English. The NP pattern which occupies a prototypical, central position in the category of NP in Kenyan English is the pattern 'post' (see Bybee 2006: 726–8).

In addition to that, the evidence for Hypothesis 1 has been corroborated by a comparison of the mean lengths of pre- and postmodifiers. The preference for postmodification in Kenyan English not only leads to a higher frequency of postmodified NPs but also to significantly longer postmodifiers than in ICE-SING, since their functional potential is exploited more fully. At the same time, the aversion to premodifiers in ICE-EA results in particularly short premodifiers in comparison to the input variety ICE-GB.

In order to gain a deeper understanding of these tendencies from the point of view of language change, it is useful to consider the factors which are known to influence the selection of features from the feature pool. It has frequently been claimed that one factor underlying pattern selection in contact varieties is markedness: as pointed out by Mufwene (2008: 150), unmarked features are frequently preferred over marked ones. In this sense, however, markedness is defined relative to a specific contact situation – the markedness of a feature does not depend on its crosslinguistic frequency, but on its relation to other features in the same feature pool:

> the term unmarked applies . . . to the fact that in a particular setting of language 'acquisition' and/or change . . . an option has been favored by a variety of structural and other, external ecological factors, for instance simplicity, generality, frequency, semantic transparency, salience and congruence between features of the 'lexifier' and of the substrate languages.

It is plausible to assume that due to their high frequency in the feature pool, head-final NPs are unmarked sensu Mufwene in comparison to postmodified NPs in the Singaporean feature pool, which makes it more likely for them to be replicated by speakers of other languages in the ecology. Head-initial structures are less likely to win in the 'competition' of features and are therefore selected less frequently in the variety of English spoken in the region.

In Kenya, the reverse situation is found. The feature pool is awash with head-initial NPs, which makes them the unmarked option and leads speakers to opt for head-initial structures in English more frequently. Accordingly, the ecology disfavours the use

Table 12. *Levels of NP embedding in ICE-GB, ICE-SING and ICE-EA*

	ICE-GB		ICE-SING		ICE-EA		
	N	%	N	%	N	%	Sum
Level 0	975	72.8	884	74.0	1081	79.7	2,940
Level 1	316	23.6	285	23.9	243	17.9	844
Level 2	39	2.9	25	2.1	29	2.1	93
Level 3	9	0.7	0	0.0	3	0.2	12
Level 4	1	0.1	0	0.0	0	0.0	1
Sum	1,340		1,194		1,356		3,890

of head-final structures, so that the chances of premodified NPs for replication are diminished.

6   Testing Hypothesis 2: differences in complexity

Beyond the typological effects identified so far Hypothesis 2 predicts different degrees of NP complexity in the varieties at hand.

6.1   Results

In sections 6.1.1–6.1.3, the variables defined in sections 4.4.1–4.4.2 will be dealt with in turn, which will help to shed light on whether Hypothesis 2 stands up to scrutiny.

6.1.1   Frequencies of NP patterns

Figure 1 above suggests a striking cline between varieties: the rate of the simplest (unmodified) NP pattern rises from ICE-GB, where they are significantly underrepresented, via ICE-SING to a peak in ICE-EA, where they are significantly more frequent than in the other corpora. Thus, in the least advanced variety, Kenyan English, the preference for the simplest NP structure is most pronounced. There are many more modified (and hence more complex) NPs in British English, and Singapore English covers the middle ground.

6.1.2   NP embedding

As can be seen from figure 3 and table 12, most NPs occur directly at the clause level. Levels of embedding beyond 0 are rather infrequent. However, there is one striking difference between the varieties: in terms of level 1, the data suggest a split between ICE-GB, ICE-SING on the one hand, which have both slightly over 23 per cent of NPs at level 1, and ICE-EA on the other, in which only 17.9 per cent of NPs are at level 1, which, according to the HCFA procedure, constitutes a significantly lower frequency.

Table 13. *Variety patterns suggested by the data*

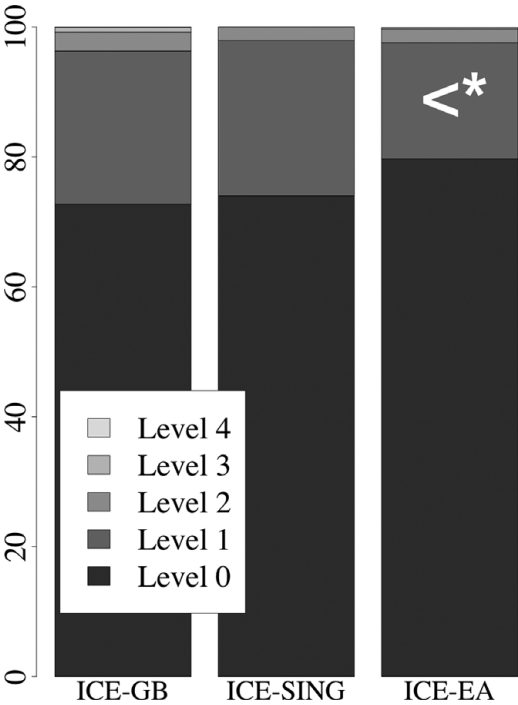|	Distribution			Patterns		
(1)	share of pattern 'none'	ICE-GB	<	ICE-SING	<	ICE-EA
(2)	NP embedding	ICE-GB	&	ICE-SING	vs	ICE-EA

Figure 3. Levels of NP embedding in ICE-GB, ICE-SING and ICE-E

The proportions of NPs at levels 2, 3 and 4, in contrast, are so infrequent that it is difficult to make out any meaningful differences between varieties.

6.2 *Discussion: focus on Hypothesis 2*

Hypothesis 2 predicts differences in the degree of NP complexity between British English and the contact varieties. Table 13 shows the patterns of varieties suggested by the two distributions analysed in section 6.

Distribution (1) describes a gradient from ICE-GB via ICE-SING to ICE-EA, while distribution (2) suggests a split between ICE-GB and ICE-SING on the one hand and ICE-EA on the other hand. Thus, in both cases, British English displays the most complex structures and Kenyan English turns out to be least complex. Singapore

English once covers the middle ground between British and Kenyan English and once corresponds with British English.

These distributions confirm Hypothesis 2 on several counts. British English, the only L1 variety in the sample, displays the highest degree of complexity in both distributions.

Kenyan English is least complex in both distributions, which is plausible from the point of view of the Dynamic Model, considering that it represents the least advanced variety in the present study. For speakers of such a variety, simplicity is an important principle of pattern selection as English is acquired relatively late and has the role of an L2 alongside other indigenous languages. As argued in section 3, psycholinguistic processes of second-language acquisition can be assumed to play more of a role here and lead to the preference for simpler structures. This is also plausible from the point of view of SLA research, where NP-related complexity measures are well-recognized: in Lu's (2010: 491) study on syntactic complexity in second-language acquisition, for instance, the rate of 'complex nominals'[17] turns out to be a particularly significant predictor of the progress made by Chinese learners of English as a foreign language.

As for the second contact variety in the sample, it is highly suggestive for Singapore English to once cover the middle ground between British and Kenyan English (see distribution (1)) and to correspond once with the native variety (see distribution (2)). It could be argued that in a variety which is as advanced as Singapore English, where more and more children are growing up speaking English as a first language (Schneider 2007: 157), the impact of simplification strategies typical of SLA is diminished, so that speakers opt more frequently for more complex structures.[18]

On the whole, thus, distributions (1) and (2) not only confirm Hypothesis 2, which refers to complexity differences between the native variety and the contact varieties, but also precisely bear out the evolutionary rankings assumed for the contact varieties, since Kenyan and Singaporean English pattern exactly as predicted by the Dynamic Model.

It should be stressed, however, that this study expressly focuses on the complexity/simplicity of NPs only. It does not seek to assess the complexity of the linguistic systems of the two varieties of New Englishes in their entirety. It is more than likely that in the bigger picture, there will be trade-offs in terms of complexity: for instance, simplicity in NPs may be counterbalanced by complexity in terms of a higher processing load on the part of the listener (see Hawkins 2009: 253–60).

---

[17] In Lu's study, 'complex nominals' are defined as NPs involving '(i) nouns plus adjective, possessive, prepositional phrase, relative clause, participle, or appositive, (ii) nominal clauses and (iii) gerunds and infinitives in subject position' (Lu 2010: 483).

[18] L. Lim (2004) has suggested that there is evidence that Singapore English is in fact about to move into phase 5 ('differentiation') of the Dynamic Model, which would be consonant with the findings on complexity in the present study.

7   Conclusion

The modification patterns of NPs are a little-researched area of variation in New Englishes. The aim of this study was twofold: on the one hand, it sought to document influences of indigenous languages on the modification patterns of NPs in Kenyan and Singaporean English. On the other hand, it looked for evidence of complexity reduction as an aspect of structural nativization in New Englishes.

On both counts, the hypotheses developed above are borne out by the results of the corpus analysis. In Singapore English, premodified NPs are significantly overrepresented; in Kenyan English, postmodifiers are more frequent than premodifiers, with premodifying patterns significantly underrepresented. This confirms that in both varieties, the use of NP modification patterns in English is palpably influenced by the patterns which have the highest type frequency in the feature pools concerned and therefore buttresses Mufwene's (2001, 2008) and others' ecological view of language contact. From a bird's eye perspective, the results might also be taken as evidence that due to contact with local languages, the varieties of English spoken in Kenya and Singapore are slowly and steadily 'developing a distinct new character of their own by consistently selecting forms and patterns that conform to an overarching language type' (Schneider 2007: 89) – similar evidence of contact-induced language change has been found e.g. for the system of aspect markers in Singapore English (see Bao 2005). The comparative approach adopted conforms to what Gut (2011: 117) has identified as part of a 'best practice methodology' (Gut 2001: 112) for the study of cross-linguistic influences in varieties of New Englishes: rather than relying on a comparison of contact varieties with its input varieties only, the present study compares two varieties with typologically distinct substrate languages and also takes into account the corresponding structures of indigenous languages.

The study has also looked for correlates of the evolutionary stages of varieties assumed in the Dynamic Model (see Schneider 2007), namely differences in terms of structural complexity of NPs. Evidence for this hypothesis was found in terms of the proportions of simple (unmodified) NPs and the levels of NP embedding in the corpus. The tendency for simplicity has been found to be, broadly speaking, least pronounced in British English and most pronounced in Kenyan English, with Singapore English covering the middle ground. This ranking might be due to simplification processes typical of second-language acquisition, which are more pronounced in Kenyan English than in Singaporean English. The varieties therefore pattern exactly as predicted by the Dynamic Model.

It is worth mentioning that such results have not only been documented for ICE section S1A. In a further study by the author, which will feed into a larger project on NP syntax in New Englishes (see Brunner forthcoming), both main results of the present study – typological influences and complexity differences – have been replicated in a written ICE genre (W1A: student essays). An area not covered by the present study which would merit further work is the use of different kinds of formal realizations of

pre- and postmodifiers. It is to be expected that in this domain, too, typological factors will play a role. Further work on such problems might help to establish the usage preferences of NPs in contact varieties of English in more detail.

Author's address:
Department of English and American Studies
University of Regensburg
93040 Regensburg
Germany
thomas.brunner@sprachlit.uni-regensburg.de

References

Ansaldo, Umberto. 2009a. The Asian typology of English: Theoretical and methodological considerations. *English World-Wide: A Journal of Varieties of English* 30, 133–48.

Ansaldo, Umberto. 2009b. *Contact languages: Ecology and evolution in Asia*. Cambridge and New York: Cambridge University Press.

Bache, Carl. 2000. *Essentials of mastering English: A concise grammar*. Berlin: Mouton de Gruyter.

Bao Zhiming. 1998. The sounds of Singapore English. In Joseph A. Foley, Thiru Kandiah, Bao Zhiming, Anthea F. Gupta, Lubna Alsagoff, Ho Chee Lick, Lionel Wee, Ismail S. Talib & Wendy Bokhorst-Heng (eds.), *English in new cultural contexts: Reflections from Singapore*, 152–74. Singapore: Oxford University Press.

Bao Zhiming. 2005. The aspectual system of Singapore English and the systemic substratist explanation. *Journal of Linguistics* 41, 237–67.

Bao Zhiming & Khin Khin Aye. 2010. Bazaar Malay topics. *Journal of Pidgin and Creole Languages* 25, 155–71.

Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.

Brunner, Thomas. Forthcoming. The noun phrase in Kenyan and Singaporean English. PhD dissertation, University of Regensburg.

Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language* 82, 711–33.

Department of Statistics Singapore. 2010. Census of population 2010. www.singstat.gov. sg/pubn/popn/c2010acr.pdf (6 November 2012).

Dryer, Matthew. 2011. Noun-modifier order in Africa. In Osamu Hieda, Christa König & Hirosi Nakagawa (eds.), *Geographical typology and linguistic areas: With special reference to Africa*, 287–311. Amsterdam and Philadelphia: John Benjamins.

Garside, Roger & Nicholas Smith. 1997. A hybrid grammatical tagger: CLAWS 4. In Roger Garside, Geoffrey Leech & Tony McEnery (eds.), *Corpus annotation: Linguistic information from computer text corpora*, 102–21. London: Longman.

Githiora, Chege. 2008. Kenya: Language and the search for a coherent national identity. In Andrew Simpson (ed.), *Language and national identity in Africa*, 235–52. Oxford and New York: Oxford University Press.

Givón, Talmy. 2009. Introduction. In Talmy Givón & Masayoshi Shibatani (eds.), *Syntactic complexity. Diachrony, acquisition, neuro-cognition, evolution*, 1–19. Amsterdam and Philadelphia: John Benjamins.

Greenbaum, Sidney. 1991. The development of the International Corpus of English. In Karin Aijmer & Bengt Altenberg (eds.), *English corpus linguistics. Studies in honour of Jan Svartvik*, 83–91. London: Longman.

Gries, Stefan Th. 2004. HCFA 3.2 – A Program for Hierarchical Configural Frequency Analysis. www.linguistics.ucsb.edu/faculty/stgries/research/overview-research. html#PublicationsEditing (6 November, 2012).

Gries, Stefan Th. 2009. *Statistics for linguistics with R: A practical introduction*. Berlin: Mouton de Gruyter.

Gut, Ulrike. 2007. First language influence and final consonant clusters in the new Englishes of Singapore and Nigeria. *World Englishes* 26, 346–59.

Gut, Ulrike. 2011. Studying structural innovations in new English varieties. In Joybrato Mukherjee & Marianne Hundt (eds.), *Exploring second-language varieties of English and learner Englishes: Bridging a paradigm gap*, 101–24. Amsterdam and Philadelphia: John Benjamins.

Haan, Pieter de. 1993. Noun phrase structure as an indication of text variety. In Andreas H. Jucker (ed.), *The noun phrase in English: Its structure and variability*, 85–106. Heidelberg: Winter.

Haspelmath, Martin (ed.). 2005. *The world atlas of language structures*. Oxford: Oxford University Press.

Hawkins, John Alan. 2001. The role of processing principles in explaining language universals. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals: An iInternational handbook*, vol. 20, 360–9. Berlin: Mouton de Gruyter.

Hawkins, John Alan. 2009. An efficiency theory of complexity and related phenomena. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language complexity as an evolving variable*, 252–68. Oxford and New York: Oxford University Press.

Hudson-Ettle, Diana & Tore Nilsson. 2002. Orality and noun phrase structure in registers of British and Kenyan English. *ICAME Journal* 26, 33–61.

Hudson-Ettle, Diana & Josef Schmied. 1999. *Manual to accompany the East African component of the International Corpus of English*. Chemnitz: Chemnitz University of Technology.

Jucker, Andreas H. 1992. *Social stylistics: Syntactic variation in British newspapers*. Berlin: Mouton de Gruyter.

Kachru, Braj B. 1983. *The Indianization of English: The English language in India*. Oxford, New York and Melbourne: Oxford University Press.

Khin Khin Aye. 2010. Identity and variations in the syntax of Singapore Bazaar Malay noun phrase. International Conference on Minority and Majority: Language, Culture and Identity (ICMM). *Malaysian Association of Modern Languages & Centre for Language Studies Universiti Malaysia Sarawak*. www.mymla.org/icmm2010papers (6 November 2010).

Kortmann, Bernd & Benedikt Szmrecsanyi. 2009. World Englishes between simplification and complexification. In Thomas Hoffmann, Lucia Siebers & Edgar W. Schneider (eds.), *World Englishes: Problems, properties and prospects*, 265–85. Amsterdam and Philadelphia: John Benjamins.

Li, Charles N. & Sandra A. Thompson. 1981. *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.

Lim, Lisa. 2004. Sounding Singaporean. In Lisa Lim & Joseph A. Foley (eds.), *Singapore English: A grammatical description*, 19–56. Amsterdam and Philadelphia: John Benjamins.

Lim, Lisa. 2007. Mergers and acquisitions: On the ages and origins of Singapore English particles. *World Englishes* 26, 446–73.

Lim, Lisa. 2010. Peranakan English in Singapore. In Daniel Schreier, Peter Trudgill, Edgar W. Schneider & Jeffrey P. Williams (eds.), *The lesser-known varieties of English: An introduction*, 327–47. Cambridge: Cambridge University Press.

Lim, Sonny. 1988. Baba Malay: The language of the Straits-born Chinese. *Papers in Western Austronesian Linguistics* 3, 1–61.

Lu, Xiaofei. 2010. Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics* 15, 474–96.

Matthews, Stephen & Virginia Yip. 2011. *Cantonese: A comprehensive grammar*, 2nd edn. London: Routledge.

Mesthrie, Rajend. 2012. Deletions, antideletions and complexity theory, with special reference to Black South African and Singaporean Englishes. In Bernd Kortmann & Benedikt Szmrecsanyi (eds.), *Linguistic complexity: Second language acquisition, indigenization, contact*. Berlin: Mouton de Gruyter.

Mesthrie, Rajend & Rakesh Mohan Bhatt 2008. *World Englishes: The study of new linguistic varieties*. Cambridge: Cambridge University Press.

Miestamo, Matti, Kaius Sinnemäki & Fred Karlsson. 2008a. *Language complexity: Typology, contact, change*. Amsterdam and Philadelphia: John Benjamins.

Miestamo, Matti, Kaius Sinnemäki & Fred Karlsson. 2008b. Introduction. The problem of language complexity. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, viii–xiv. Amsterdam and Philadelphia: John Benjamins.

Miller, Jim & Regina Weinert. 1998. *Spontaneous spoken language: Syntax and discourse*. Oxford: Clarendon Press.

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

Mufwene, Salokoko S. 2008. *Language evolution: Contact, competition and change*. London: Continuum.

Nelson, Gerald. 1996. The design of the corpus. In Sidney Greenbaum (ed.), *Comparing English worldwide: The International Corpus of English*, 27–35. Oxford: Clarendon.

Ortega, Lourdes. 2003. Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics* 24, 492–518.

Pakir, Anne Geok-In Sim. 1986. A linguistic investigation of Baba Malay. Unpublished PhD dissertation, University of Hawaii.

Pakir, Anne Geok-In Sim. 1991. The range and depth of English-knowing bilinguals in Singapore. *World Englishes* 10, 167–79.

Polomé, Edgar C. *Swahili language handbook*. Washington, DC: Center for Applied Linguistics.

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

R Development Core Team. 2008. R: A language and environment for statistical computing. www.r-project.org (6 November 2012).

Ravid, Dorit & Ruth. A. Berman. 2010. Developing noun phrase complexity at school age: A text-embedded cross-linguistic analysis. *First Language* 30, 3–26.

Rescher, Nicholas. 1998. *Complexity: A philosophical overview*. New Brunswick and London: Transaction Publishers.

Rooy, Bertus van & Lande Schäfer. 2003. An evaluation of three POS taggers for the tagging of the Tswana Learner English Corpus. *Lancaster University Centre for Computer Corpus Research on Language Technical Papers*, 835–44.

Sampson, Geoffrey, David Gil & Peter Trudgill. 2009. *Language complexity as an evolving variable*. Oxford and New York: Oxford University Press.

Schmied, Josef. 2008. East African English (Kenya, Uganda, Tanzania): Phonology. In Rajend
    Mesthrie, Bernd Kortmann & Edgar W. Schneider (eds.), *Varieties of English*, vol. 4: *Africa,
    South and Southeast Asia*, 150–63. Berlin: Mouton de Gruyter.

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

Skandera, Paul. 2003. *Drawing a map of Africa*. Tübingen: Narr.

Szmrecsanyi, Benedikt. 2009. Typological parameters of intralingual variability: Grammatical
    analyticity versus syntheticity in varieties of English. *Language Variation and Change* 21,
    319–53.

Technical report: CLAWS part-of-speech tagger. University Centre for Computer Corpus
    Research on Language, Lancaster University (UCREL).

Whiteley, Wilfred H. 1974. The classification and distribution of Kenya's African languages.
    In Wilfred H. Whiteley (ed.), *Language in Kenya*, 13–64. Nairobi: Oxford University Press.