

Collostructional nativisation in New Englishes

Verb-construction associations in the International Corpus of English*

Joybrato Mukherjee and Stefan Th. Gries

Justus Liebig University / University of California, Santa Barbara

The present paper investigates the strength of verb-construction associations across various New Englishes on the basis of comparable corpora. In contrast to previous studies into verb complementation in New Englishes, we start off from three basic constructions in English — the intransitive, the monotransitive and the ditransitive construction — and analyse the co-occurrences of the three constructions and a wide range of verbs. The present study is based on the Hong Kong, the Indian, and the Singapore components of the International Corpus of English (ICE) because the three varieties represent markedly different stages in the process of the evolution of New Englishes with British English as the historical input variety. Our quantitative analysis includes multiple distinctive collexeme analyses for the different varieties. The results show, *inter alia*, that, firstly, processes of structural nativisation of New Englishes can also be observed at the level of verb-construction associations, which can be subsumed under the notion of “collostructional nativisation”, and that, secondly, there are identifiable intervarietal differences between British English and New Englishes as well as between individual New Englishes. In general, there is a correlation between the evolutionary stage of a New English variety and its collostructional nativisation: The more advanced a New English variety is in the developmental cycle, the more dissimilar its collostructional preferences are to British English.

* The order of authors is arbitrary. The present paper goes back to a poster presented at the 29th Conference of the International Computer Archive of Modern and Medieval English (ICAME 29) in Ascona/Switzerland. We would like to thank the participants in the poster session for constructive criticism and stimulating discussions. We are also very grateful to Christopher Koch, who helped us immensely in analysing the corpus data and coding many thousands of data points, and to Rosemary Bock, who proof-read the final version of the manuscript. The usual disclaimers apply.

Keywords: British English, Hong Kong English, Indian English, Singapore English, varieties, verbs, constructions, corpora, distinctive colloceme analysis, evolution of New Englishes

1. Introduction: Towards a description of collostructional nativisation in the evolution of New Englishes

There is unanimous agreement that the English language has been undergoing complex processes of acculturation in many colonial and post-colonial contexts, including, for example, former British colonies in Asia which have retained the English language after independence (cf. Kachru 2005). In the course of these processes, “English has been appropriated by its non-European users and changed to reflect their own experiences” (Mair 2008: 235). An integral part of the processes of acculturation are linguistic changes, with new forms and structures emerging at the level of vocabulary and syntax (e.g. due to loanwords and transfer from local languages) and new norms evolving in phonology and intonation (e.g. with regard to the range of consonant clusters and intonation contours). The linguistic changes can be subsumed under the notion of structural nativisation, i.e. “the emergence of locally characteristic linguistic patterns” (Schneider 2007: 5–6). The various aspects of structural nativisation at the levels of phonology and morphosyntax are well-documented for many New Englishes (cf. e.g. Kortmann *et al.* 2004). However, only recently has it been noted that structural nativisation not only refers to entirely new and innovative forms and structures in individual varieties, but also covers quantitative differences between varieties of English in the use of forms and structures that belong to the common core (cf. Quirk *et al.* 1985: 16) that is shared by all Englishes. Such quantitative differences in usage are not immediately accessible to intuition and can only be identified by analysing very large amounts of natural data, i.e. large and representative computerised corpora such as the International Corpus of English (ICE). For example, Mukherjee and Hoffmann’s (2006) corpus-based study of verb-complementation patterns in Indian English has shown that, while in British English (BrE) the most frequent complementation pattern of the verb *give* is the ditransitive construction (*give someone something*), in Indian English — the largest second-language variety of English world-wide — the same verb is used most frequently with the monotransitive construction (*give something*, i.e. with no recipient being made explicit). Also, the complex-transitive construction (*give something to someone*) is more frequently attested in Indian English than in British English. On grounds of such quantitative findings, Schneider (2007) thus notes: “These are stable and noteworthy results, and it is worth pointing out that they operate way below the level of linguistic awareness:

without quantitative methodology no observer would have expected such differences to exist” (Schneider 2007: 87).

For Indian English, a large machine-readable corpus of English had already been compiled in the 1980s, namely the Kolhapur Corpus with one million words from 500 texts originally published in 1978, covering 15 genres of written Indian English. It is therefore not surprising that some quantitative corpus-based studies of verb-complementation patterns were conducted for Indian English in the 1980s and 1990s already (cf. e.g. Shastri 1988, 1996). With the completion of the Indian component of ICE in 2002, a new one-million word corpus covering written and spoken Indian English became available. Also, the Internet has been utilised to create large web-derived corpora of New Englishes, e.g. newspaper corpora of many million words of acrolectal standard written usage (cf. e.g. Hoffmann and Mukherjee 2007). It is worth pointing out, however, that all corpus-based studies of verb-complementation patterns in New Englishes have so far focused on individual verbs or relatively small classes of verbs, e.g. pelt verbs such as *pelt*, *shower* and *pepper* (cf. Olavarria de Ersson and Shaw 2003) and transfer-caused motion verbs such as *convey*, *submit* and *supply* (cf. Mukherjee and Schilk 2008) in Indian English. For Hong Kong English and Singapore English, for which ICE corpora have also become available recently, large-scale quantitative analyses of complementational preferences of verbs and verb classes have, to our knowledge, not yet been carried out.

The development of a new variety of English is very much dependent on the specific socio-cultural and historical context as well as the local language contact situation. Although the emergence of a new variety is, thus, an essentially unique process, it has recently been suggested that in spite of many variety-specific processes and features the evolution of New Englishes follows a fundamentally uniform pattern world-wide: This basic assumption is at the basis of Schneider’s (2003, 2007) dynamic model of the evolution of New Englishes. It would go beyond the scope of the present paper to discuss the evolutionary model by Schneider (2003, 2007) and its implications in detail here. Suffice it to say that the evolutionary pattern which he posits for the development of New Englishes consists of five phases, which — at the risk of some oversimplification — can be summarised as follows:

Phase 1 — Foundation: In this initial phase, the English language is transported to a new (colonial) territory.

Phase 2 — Exonormative stabilisation: There is a growing number of English settlers / speakers in the new territory, but the language standards and norms are still determined by the input variety and are, thus, usually oriented towards British English.

Phase 3 — Nativisation: The English language becomes an integral part of the local linguistic repertoire as there is a steady increase in the number of competent bilingual L2 speakers of English from the indigenous population, and undergoes a characteristic restructuring process labelled “structural nativization”.

Phase 4 — Endonormative stabilisation: After Independence, English may be retained as a/an (co-)official language and a medium of communication for a more or less wide range of intra-national contexts (e.g. administration and the press, academia and education); in this phase a new variety of English emerges with generally accepted local standards and norms.

Phase 5 — Differentiation: Once a New English variety has become endonormatively stabilised, it may develop a wide range of regional and social dialects.

In essence, the progression through the evolutionary cycle is primarily motivated by — and based on — two interrelated factors: group-interaction and identity-construction. Both factors go back to the existence of — and interaction between — the indigenous population (the IDG strand) and the new settler community (the STL strand):

The stages and strands [i.e. the STL strand and the IDG strand] of this process are ultimately caused by and signify reconstructions of group identities of all participating communities, with respect to the erstwhile source society of the colonizing group, to one another, and to the land which they jointly inhabit. (Schneider 2003: 244)

The underlying idea is that with growing (communicative) interaction between the STL strand and the IDG strand a new hybrid identity of the two groups comes into existence which manifests itself in the formation of a new variety of English along the evolutionary line. It should be noted that the evolutionary stages represent idealised states, and there may be considerable overlap between subsequent phases. Schneider (2003, 2007) discusses the history and the present-day situation of many varieties and shows that the evolutionary model can be mapped, *mutatis mutandis*, onto a wide range of postcolonial Englishes world-wide.

The present paper focuses on the structural nativisation at the level of verb complementation in the English language in Hong Kong, India and Singapore. Specifically, we are interested in quantitative differences between the three Asian Englishes with regard to the frequencies of — and preferences for — certain verb-construction associations. In our study, the quantitative description of such verb-construction associations is based on collostructional analysis as introduced by Stefanowitsch and Gries (2003: 214): “Collostructional analysis always starts with a particular construction and investigates which lexemes are strongly attracted to or repelled by a particular slot in the construction (i.e. occur more frequently or less frequently than expected)”.

The constructions that we take into account are the ditransitive construction (e.g. with the verb *give* as in *she gave him the book*), the intransitive construction (e.g. with the verb *happen* as in *it happened*) and the monotransitive construction (e.g. with the verb *find* as in *he found the keys*). It is our aim to shed new light on a hitherto neglected aspect of structural nativisation in the evolution of New Englishes, which we refer to as collostructional nativisation.

For a comparative corpus-based analysis of collostructional nativisation across New Englishes, Hong Kong English, Indian English and Singapore English provide interesting and suitable contexts because, firstly, they represent different stages in the evolutionary cycle according to Schneider's (2003, 2007) model; secondly, the historical input for all of them was British English; and, thirdly, comparable corpora with the same size and the same design are available for all of them (i.e. the ICE corpora). The research questions that we would like to address in the present paper are the following: (1) Are there significant quantitative differences between New Englishes in Asia with regard to verb-construction associations? (2) If so, do the differences correlate with the evolutionary stages of the individual varieties? (3) Does structural nativisation at the lexicogrammatical level include a process of collostructional nativisation?

The plan of the paper is as follows: In the next section, we will briefly describe the evolutionary stages of the three varieties and sketch out the comparable corpus data that we will make use of in the present study. In Section 3, we will introduce the methodology of collostructional analysis¹ that we have applied as well as some of the principles of data coding. In Section 4, we will present the results of the collostructional analysis, which we will discuss in Section 5. Finally, in Section 6 we will offer some concluding remarks and prospects for future research.

2. English in Hong Kong, India and Singapore: Evolutionary stages and corpus data

2.1 Hong Kong English

In Schneider's (2003, 2007) evolutionary model, the present-day situation of English in Hong Kong is marked by many features that are characteristic of Phase 3, while some features of Phase 2 can still be observed. The English language came to Hong Kong in the middle of the 19th century after the island of Hong Kong had become a British colony as a result of the first Opium War in 1841/42. After

1. The notion of "collostruction" and "collexeme analysis" will be explained in greater detail in Section 3.1.

World War II, a new bilingual middle class came into being, largely because of the economic rise of Hong Kong. As in many other colonial and postcolonial countries, the economic elite tended to acquire English in an English-medium education system. In the 1970s the introduction of Anglo-Chinese secondary schools changed this orientation towards elitist bilingualism into mass/folk bilingualism: In 2001, 43% of the population of Hong Kong claimed a knowledge of English (cf. Bolton 2003: 85–8). This process has led to the development of a local variety of English, and the local characteristics of Hong Kong English at the various levels of analysis, including vocabulary and grammar, are well documented (cf. e.g. Bolton 2002, 2003). However, the overall attitude towards the local variant of English is not unanimously positive, especially after the handover of Hong Kong in 1997, and linguists disagree on whether Hong Kong English is a fully-fledged nativised variety as in other postcolonial contexts:

[U]nlike many post-colonial societies such as India, Nigeria and the Philippines, English is rarely used by Chinese Hongkongers for intra-ethnic communication. There is thus no societal basis for any nativised variety of “Hong Kong English”. Instead the norms of correctness as referenced in the key domains of education, government, business and law follow those of standard English varieties, especially British English... (Li 1999: 95)

It thus seems that to some extent an exonormative orientation, which is characteristic of Phase 2, is still present. Together with the uncertain future of Hong Kong’s language repertoire in the framework of a Special Administrative Region of the People’s Republic of China, it is best to view present-day Hong Kong English as being marked both by Phase 2 features and by Phase 3 features.

2.2 Indian English

English was transported to India in the 17th century when the East India Company began to explore the Indian subcontinent and set up trading posts in the East, South and West. The British victory in the Battle of Plassey in 1757 marks the beginning of the British Empire in India. Afterwards, an increasing number of British settlers, soldiers and missionaries came to India so that the interaction between the colonisers and the local population became more and more intense. Of particular importance for the growing entrenchment of the English language in the indigenous population and the rise of a bilingual English-speaking elite amongst the Indians is Macauley’s famous Minute of 1835, in which he advocates an English-medium education for Indians wishing to work as civil servants for the colonial administration and which soon became the basis for the language-education policy of the colonial government. After Independence in 1947, the English

language was retained as a co-official language alongside Hindi and has developed into the largest second-language variety of English world-wide with approximately 35 to 50 million competent L2 speakers. English is used for a wide range of intranational communicative purposes, serves as a pan-Indian link language and is used by an ever-increasing number of authors for their creative fiction writing (cf. Rushdie and West 1997). The local features of Indian English at the levels of pronunciation, vocabulary, grammar and style have been described in detail by various scholars, e.g. Kachru (1983), Mehrotra (1998) and Nihalani *et al.* (2004). There is also a growing body of corpus-based research into Indian English, e.g. in the area of lexicogrammar (e.g. Schilk 2006; Mukherjee and Hoffmann 2006). While there is general consensus on the status of Indian English as a fully-fledged nativised variety in its own right, there is some disagreement on where exactly to plot Indian English in the evolutionary process of variety-formation. Schneider (2007: 171–3) argues that Indian English is essentially a Phase 3 variety with some early symptoms which “are foreshadowing endonormative stabilization, but they are disputable or weak”. On the other hand, Mukherjee (2007) discusses a range of socio-cultural and linguistic markers indicating that present-day Indian English has already reached Phase 4 in spite of some undeniable remnants of Phase 3. In the present study, we therefore view Indian English as a variety of English that represents the transition period from Phase 3 to Phase 4.

2.3 Singapore English

Without any doubt, Singapore English is the most advanced variety among the New Englishes under scrutiny from an evolutionary perspective. It is amazing that the rapid development of English in Singapore from the foundation phase to the present situation as an endonormatively stabilised variety in Phase 4 took place in less than 200 years. In 1819, the British East India Company set up a trading post on a sparsely inhabited island which later developed into the mega-city of Singapore. Throughout the 19th century and the first half of the 20th century people from various countries and with different ethnic, religious and linguistic backgrounds settled down in Singapore, the most important groups being Chinese, Malay, Tamil and English people. After Independence in 1965, the new government pursued a steadfast policy of promoting the use of English as a link-language of the multilingual and multicultural city-state. This clear language policy in favour of English, accompanied by the establishment of an English-medium education system, and the tremendous economic success of Singapore have resulted in a dramatic increase in the number of competent (bilingual) speakers of English, a rapid widening of the domains and discourse types in which English has become the default language of communication, including informal contexts at home in

an increasing number of households (cf. Gupta 1998; Ooi 2001). It is obvious that this process has been marked by “a vibrant process of structural nativization”, as Schneider (2007: 158) notes. The local features of Singapore English, which are present at all linguistic levels, including pronunciation, vocabulary and grammar, are not only well documented (cf. e.g. Lim 2004; Low and Brown 2005), but are also fully accepted as linguistic markers of a new pan-ethnic identity:

By now Singapore has clearly reached phase 4 of the cycle. The country’s unique, territory-based, and multicultural identity construction has paved the way for a general acceptance of the local way of speaking English as a symbolic expression of the pride of the Singaporeans in their nation. (Schneider 2007: 160)

In the present study, we thus view Singapore English as representing a fully endo-normatively stabilised Phase 4 variety.

2.4 ICE data

For the comparative analysis of collostructional nativisation in the three Asian Englishes we will use the Hong Kong, Indian and Singapore components of the *International Corpus of English* (ICE-HK, ICE-IND and ICE-SIN, respectively). We will also use the British component (ICE-GB), as it represents the present-day situation of the historical input variety for all Asian Englishes, namely British English. The ICE project was launched in the mid-1990s (cf. Greenbaum 1996). Its aim is to provide a family of comparable corpora with the same size and the same design, representing all major varieties of English world-wide and including both native varieties of English in the inner circle (e.g. American English, British English, New Zealand English) and institutionalised second-language varieties in the outer circle (e.g. the Asian Englishes under discussion in the present paper). The design of each one-million word ICE corpus is sketched out in Figure 1. Each ICE corpus includes 60% spoken texts and 40% written texts, which are usually taken from the early 1990s. While the overall size of any single ICE corpus is relatively small, the major advantage of an ICE-based analysis is the comparability of the data and the results of the analysis across varieties on grounds of the shared corpus design.

Figure 2 offers a visualisation of the intervarectal comparison of the three New Englishes that represent three markedly different stages in the evolutionary cycle. The arrows from left to right in Figure 2 show that the input variety of British English is a diachronically changing reference point. When comparing ICE-HK, ICE-IND and ICE-SIN with ICE-GB, we nevertheless assume that present-day British English is still a useful and valid baseline for the assessment of the evolutionary stage of the three Asian Englishes. While one could argue that for any comparison of a New English variety with its historical input variety a

SPOKEN (300 texts)	
Dialogue (180)	
<i>Private</i> (100)	direct conversations (90), telephone calls (10)
<i>Public</i> (80)	classroom lessons (20), broadcast discussions (20), broadcast interviews (10), parliamentary debates (10), legal cross-examinations (10), business transactions (10)
Monologue (120)	
<i>Unscripted</i> (70)	spontaneous commentaries (20), unscripted speeches (30), demonstrations (10), legal presentations (10)
<i>Scripted</i> (50)	broadcast news (20), broadcast talks (20), non-broadcast talks (10)
WRITTEN (200 texts)	
Non-printed (50)	
<i>Students' texts</i> (20)	students' untimed essays (10), students' examination scripts (10)
<i>Letters</i> (30)	social letters (15), business letters (15)
Printed (150)	
<i>Informational writing</i> (100)	academic (40), popular (40), press reports (20)
<i>Instructional writing</i> (20)	administrative/regulatory (10), skills and hobbies (10)
<i>Persuasive writing</i> (10)	press editorials (10)
<i>Creative writing</i> (20)	novels and stories (20)

Figure 1. Design of ICE corpora (cf. Nelson, Wallis and Aarts 2002: 307–8)

diachronic corpus of the native variety is needed (cf. e.g. Hoffmann and Mukherjee's 2007 study of ditransitive verbs in Indian and British English), we suggest for the present study that a synchronic comparison of the Asian components of ICE and ICE-GB still makes sense: As indicated in Figure 2, we assume that the

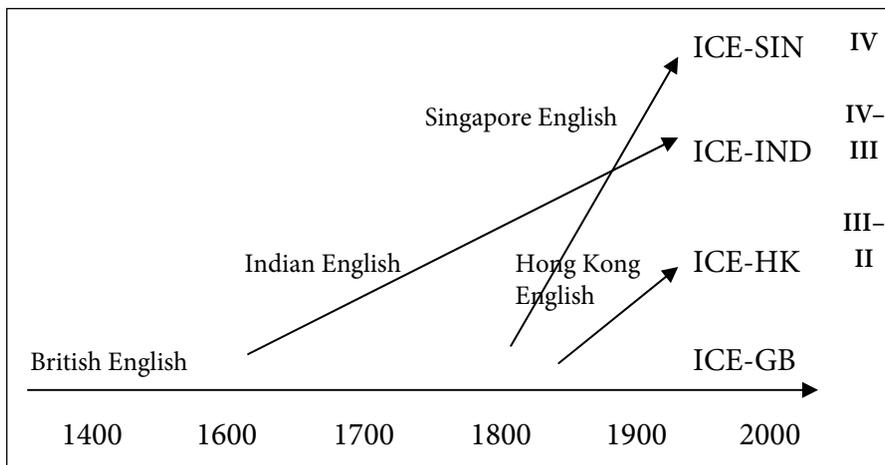


Figure 2. Comparing evolutionary stages of different Asian Englishes with ICE corpora

more advanced a New English variety is in its evolution, the more dissimilar it is to present-day British English.

It should be noted that for the analysis of collocations, ICE-GB also provides a useful starting-point from a methodological perspective as it is fully syntactically parsed. It is thus possible to automatically retrieve all instances of the ditransitive, the intransitive and the monotransitive construction and all frequent co-occurrences with individual verbs in ICE-GB, which can then be tested against the other ICE corpora. Thus, the ICE-GB data can be used as reference data, which makes it possible to extend collocational analysis to other untagged and unparsed ICE corpora (cf. Mukherjee *fc.*).

In the following section, we will turn to the methodology of our intervarietal comparison of collocations and discuss the various steps involved in the data coding and the collocational analysis.

3. Methodology: Collocational analysis and data coding

We proceeded in three steps. First, using separate collexeme analyses, we computed the association of all the verbs that are used at least once in the ICE-GB in an intransitive, monotransitive, or ditransitive pattern (cf. Section 3.1). Second, using multiple distinctive collexeme analyses, we determined for a subset of these verbs which of the three patterns each verb is attracted to, or repelled by, in each New English corpus and how strong this attraction/repulsion is (cf. Section 3.2). Third, we made pairwise comparisons of the results between the British English data on the one hand and the New Englishes on the other hand (cf. Section 3.3 and 4). The following sections will outline these steps in more detail.

3.1 The reference data from ICE-GB: Corpus analyses for the three patterns

Multiple distinctive collexeme analysis is one method from the family of methods known as collocational analysis. These methods are an extension of the familiar collocational approaches that study the association of two or more words to each other (or their repulsion from each other). Specifically, collocational analysis investigates the attraction of each one of many words *W* to a particular syntactically defined slot (or the repulsion of a word from the syntactic slot at hand) in a particular syntactic pattern *P* (cf. Stefanowitsch and Gries 2003), hence the name: a blend of collocation and construction. These methods are based on an assumption shared by many recent theories that can be subsumed under the headings of construction grammar (cf. Goldberg 1995, 2006) in particular or usage-based cognitive grammar in general (cf. Langacker 1987), *viz.* the assumption that there

is no strict dividing line between the lexicon on the one hand and grammar/syntax on the other. Thus, just as strong/significant collocates of a particular node word allow us to pinpoint functional characteristics of the node word — functional is understood broadly here to encompass matters of meaning and discourse/pragmatic function — strong/significant collexemes of a particular syntactic pattern allow us to pinpoint functional characteristics of the pattern. This method as well as some of its extensions have been applied to a large variety of patterns of different degrees of schematicity, in different languages, and to different purposes; the statistics they provide have been useful in studies on the syntax-semantics interface in general, but also more specifically in the domains of syntactic priming, second language learning, *etc.* Since this kind of approach plays a crucial role in the discussion below, we would like to briefly illustrate its logic.

Just like most collocational approaches, the simplest collostructional method — collexeme analysis — is based on a two-by-two table of co-occurrence frequencies of the kind represented in Table 1.

Table 1. Schematic representation of co-occurrence frequencies of word *W* and pattern *P*

	Pattern <i>P</i>	Not pattern <i>P</i>	Totals
Word <i>W</i>	<i>a</i>	<i>b</i>	<i>a+b</i>
not word <i>W</i>	<i>c</i>	<i>d</i>	<i>c+d</i>
Totals	<i>a+c</i>	<i>b+d</i>	<i>n=a+b+c+d</i>

The study of syntactic pattern *P* requires a table showing each word *W* that is attested at least once in *P*. The number of times *W* occurs in *P* is denoted by *a*; the number of times *W* occurs elsewhere is denoted by *b*; the number of times *P* occurs without *W* is denoted by *c*; the number of times patterns other than *P* occur with words other than *W* is denoted by *d*. In practice, this means that for each *W* and *P*, the analyst retrieves the frequencies *a*, *a+b*, *a+c*, and *n* from the corpus, and fills the remaining cells on the basis of these values (i.e. *b* is then the difference between *a+b* and *a*, *etc.*). Crucially, these observed frequencies are compared to the frequencies expected by chance, i.e. the frequencies expected if there was no relation between *W* and *P*, and, just as crucially, the discrepancy between the observed frequencies and the expected frequencies is evaluated and quantified by means of a statistical test such as, for instance, the *p*-value of the Fisher-Yates exact test (cf. Stefanowitsch and Gries 2003: 238–9, note 6 for discussion). However, the *p*-values are intuitively somewhat difficult to interpret: (1) while *p*-values range from 0 to 1, the theoretically interesting part of that range is only the smallest five percent of it: $0 \leq p \leq 0.05$; (2) seemingly counter-intuitively, an increase of significance is only reflected in a larger number for the negative exponent: $9.7\text{E}-10$ is in fact a smaller number than $5.4\text{E}-8$ although 9.7 is larger than 5.4. For this reason

and others discussed in Gries, Hampe and Schönefeld (2005: 671–2, note 13), we follow the by now established practice of reporting the mathematically equivalent $-\log_{10}p$.

The method we used here is an extension of simple collexeme analysis. This extension, multiple distinctive collexeme analysis, differs from collexeme analysis in that it does not just look at how different words relate to one and the same pattern, but at which of three or more patterns different words are attracted to, or repelled by, and how strongly. For an example, consider Table 2, which lists how often the verb lemma *talk* was attested in the three relevant patterns in the ICE-GB.

Table 2. Frequencies (absolute and relative) of the verb lemma *talk* in three patterns in the ICE-GB

Ditransitive	Intransitive	Monotransitive	Totals
0 (0%)	472 (95.74%)	21 (4.26%)	493 (100%)

It is immediately obvious that *talk* is preferably used intransitively, but to quantify exactly how strong this preference is, one needs to compare the observed frequency of *talk* with the frequency that is expected given the overall frequencies of the three patterns in the same corpus; these are listed in Table 3.

Table 3. Frequencies (abs. and rel.) of three patterns in the ICE-GB

Ditransitive	Intransitive	Monotransitive	Totals
1743 (1.769%)	32391 (32.776%)	64486 (65.455%)	98520 (100%)

In other words, approximately 9 occurrences of *talk* as a lemma should be in the ditransitive (1.769% of 493), approximately 162 occurrences of *talk* as a lemma should be in the intransitive (32.776% of 493), and c. 323 occurrences of *talk* as a lemma should be in the monotransitive (65.455% of 493). To compare how much each observed frequency differs from each expected one, three separate binomial tests are computed, each of which answers the following question: Given that c. 9, 162, and 323 occurrences of *talk* were expected in the ditransitive, intransitive and monotransitive, respectively, how likely is it that 0, 472 and 21 instances, respectively, will actually be found?

Given the large number of words that may occur in any given pattern and the highly complex and unwieldy computations involved in the binomial tests, it is not feasible to perform these tests by hand; we used an R script *Coll.analysis* (cf. Gries 2004 and R Development Core Team 2008) for the analysis. As input for each word *W*, this program requires the observed frequencies in the three patterns as well as the overall frequencies of the patterns, and returns a list of the words *W* plus how strongly they are attracted to, or repelled by, each pattern.

In order to provide these frequencies for the three patterns under consideration here, we used the ICECUP retrieval software that comes with the ICE-GB corpus (version 3.0) to retrieve:

- all the verbs in verb phrases tagged as ditransitive;
- all the verbs in verb phrases tagged as intransitive;
- all the verbs in verb phrases tagged as monotransitive.

This provided us with the overall frequencies of the three patterns (as listed in Table 3). In the next step, for each pattern we extracted the exact verb forms from the concordance lines and replaced each of them by their corresponding lemma. Finally, for all forms of each of 3 387 verb lemmas, we determined how often they were attested in which construction and how each observed frequency differed from the expected one. For the above example of the verb *talk*, for instance, we obtained the results summarised in Table 4; the listed values are $\log_{10} P_{\text{binomial test}}$ which were set to positive or negative depending on whether the observed frequency was larger or smaller than the expected frequency.

Table 4. Multiple distinctive collexeme strengths of the verb lemma *talk* in three patterns in the ICE-GB

Ditransitive	Intransitive	Monotransitive
-3.8219	195.6157	-185.0817

As can be gleaned from the data in Table 4, it turns out that the verb *talk* strongly prefers only the intransitive pattern and disprefers ditransitives and monotransitives. Similar kinds of results were obtained for all 3 387 verb lemmas included in this part of the analysis so that we were able to determine which verbs prefer which pattern(s) most strongly and which verbs did not exhibit any substantial preference (in spite of, say, at least a moderate overall frequency of occurrence).

3.2 The data from ICE-HK, ICE-IND and ICE-SIN: Corpus analyses for the three patterns

In order to determine to what degree, if any, the New Englishes differ from the British English variety in the way verbs are (dis)preferred in the three syntactic patterns, we first picked 59 verbs for the comparative analysis. Specifically, we picked 15 verbs that were very strongly attracted to one of the three patterns in the British English variety, and the verbs for each of these are listed in (1) to (3). However, the verb *be* was later discarded from the group of intransitive verbs because in the vast majority of cases it functions as an auxiliary verb. Thus only 14 intransitive verbs entered into the analysis.

- (1) **ditransitive verbs:** *give, tell, ask, send, offer, show, convince, remind, cost, assure, inform, teach, allow, lend, persuade*
- (2) **intransitive verbs:** *(be,) go, come, look, work, talk, live, occur, sit, move, speak, stay, wait, fall, happen*
- (3) **monotransitive verbs:** *have, want, make, use, take, do, see, find, need, mean, produce, bring, put, provide, involve*

Since it was most likely that the verbs listed in (1) to (3) — i.e. verbs that are strongly associated with a pattern in British English — would exhibit rather similar preferences in the New Englishes, we also included an additional set of 15 verbs which were reasonably frequent in ICE-GB but not strongly attracted to any of the three patterns under consideration; these verbs are listed in (4), and we hypothesised that these would be verbs more likely to exhibit variation across the varieties.

- (4) **neutral verbs:** *fine, reassure, call, cook, command, surprise, serve, prescribe, pass, bet, file, quote, purchase, permit, feed*

In the second step, we had to determine how strongly these verbs are attracted to the three syntactic patterns. However, unlike ICE-GB, the corpora for New Englishes are not parsed, which is why we had to devise a sampling strategy that provided us with a comparable yet unbiased and stratified sample of verb matches. To that end, we sampled several thousand tokens of the above-mentioned verb lemmas from ICE-HK, ICE-IND and ICE-SIN. Using R scripts, we then identified the overall frequencies of the 59 verb lemmas in each New English corpus. We decided on a sample size of 3000 verbs per New English corpus and computed how many tokens of each verb would have to be included in our 3000-verb sample so that the token frequencies of the verbs in the sample would be proportional to the overall frequencies of the verbs in the corpus. For example, we found the overall number of tokens of all 59 verb lemmas in ICE-HK to be 36901, of which 106 were instances of the lemma *allow*. These 106 instances correspond to 0.29% of all the relevant verbs, which means that we decided to manually check as many instances of *allow* in ICE-HK until we had 0.29% of 3000 verbs $\approx 8.6 \approx 9$ instances of *allow* that could unambiguously be defined as either ditransitive, intransitive, or monotransitive. The same logic was applied to all other verbs in all three New Englishes corpora. As a result, some verbs had to be discarded because their overall frequencies in the corpora were so low that a proportional sample of 3000 verbs did not yield a desired sample of 1 or more; we will address these cases below. In total, we manually checked 11 487 matches until all desired sample sizes per verb per corpus were obtained. Every instance of all verb lemmas was categorised either into one of the constructional categories (ditransitive, intransitive,

monotransitive) or left uncoded because (a) it represented another construction (e.g. the copular construction), (b) it represented a non-canonical variant of one of the three constructions under scrutiny (e.g. an agentless passive form) and/or (c) because it was not possible to decide on the constructional category (e.g. in cases of elliptical structures in spontaneous spoken speech). The examples in (5) to (7) exemplify cases of ditransitive, intransitive and monotransitive constructions, respectively, that were taken into consideration.

- (5) a. If you want to know this way, send me another email
(ICE-HK W1B-009) DITR
b. Didn't I tell you yesterday my brother came at ten thirty
(ICE-IND S1A-052) DITR
- (6) a. Or he'll come earlier
(ICE-IND S1A-095) INTR
b. ...during weekend we never cook
(ICE-HK S1A-023) INTR
- (7) a. So it means we want to find what is the total surface area of the prism
(ICE-IND S1B-013) MONOTR
b. There was also another 4 hour job helping some researchers at the med school move furniture.
(ICE-SIN W1B-015) MONOTR

As our reference data were taken from the ICE-GB data, we coded the ICE-HK, ICE-IND and ICE-SIN data according to the parsing scheme of ICECUP, the retrieval software for ICE-GB. The ICECUP parsing scheme, which is largely based on the descriptive apparatus of the *Comprehensive Grammar of the English Language* (Quirk *et al.* 1985), is described in detail in the ICE-GB handbook (cf. Nelson, Wallis and Aarts 2002).

In a third step, we did three separate multiple distinctive collexeme analyses, one for each New English corpus to, firstly, determine each verb's constructional preference in each corpus and, secondly, compare the results of all varieties with an eye to variety-specific differences in complementation patterns. The results of these steps are discussed in the following section.

4. Results

The methodology sketched out in the previous sections yields a wealth of quantitative results, which we cannot comprehensively discuss here. The overall structure of the results is represented schematically in Table 5, which lists all verbs in the

rows and all constructions in all corpora in the columns. The italicised two-letter sequences in the main body of Table 5 represent distinctive collexeme strengths of the kind exemplified in Table 4 above.

Table 5. Schematic overview of the results for all verbs and constructions in all corpora

Verb	Ditransitive				Intransitive				Monotransitive			
	GB	HK	IND	SIN	GB	HK	IND	SIN	GB	HK	IND	SIN
verb ₁	<i>ab</i>	<i>cd</i>	<i>ef</i>	<i>gh</i>	<i>ij</i>	<i>kl</i>	<i>mn</i>	<i>op</i>	<i>qr</i>	<i>st</i>	<i>uv</i>	<i>wx</i>
verb ₂	<i>yz</i>	<i>ac</i>	<i>df</i>	<i>gi</i>	<i>jl</i>	<i>mo</i>	<i>pr</i>	<i>su</i>	<i>vx</i>	<i>ya</i>	<i>bc</i>	<i>de</i>
verb ₃	<i>fg</i>	<i>hi</i>	<i>jk</i>	<i>lm</i>	<i>no</i>	<i>pq</i>	<i>rs</i>	<i>tu</i>	<i>vw</i>	<i>xy</i>	<i>za</i>	<i>bd</i>
...

In Section 4.1, we will provide a brief sketch — largely as a graphic overview — of the numerical results and how the degrees of attraction and repulsion compare across the ICE corpora. In Section 4.2, we will then adopt a more simplistic approach and focus only on the results that emerge when each verb’s (dis)preferences are noted in a binary fashion.

4.1 Within-construction results

The first step was to compare the four corpora within each construction separately. To that end, we computed three cluster analyses — one for each construction — in which the four corpora were compared in terms of how strongly the 59 verbs were attracted to, or repelled by, the construction in question. More technically, this means that each of the corpora is defined by the vector of distinctive collex-

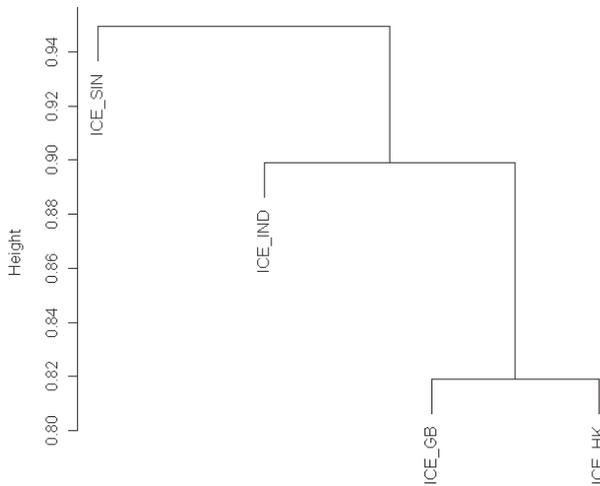


Figure 3. Dendrogram for the ditransitive data

eme strengths in its column in Table 5 and that then these vectors of numbers get compared. In order to prevent huge distinctive collexeme strengths from distorting the results, we chose the correlational measure of the cosine as the measure of similarity on the basis of which the columns are compared, and the clustering was done according to Ward's method, a reliable standard method that is conceptually similar to analyses of variance (cf. Ward 1963). Consider Figure 3 for the results with regard to the ditransitive.

There is a relatively straightforward cluster structure in Figure 3: The verb preferences in ICE-GB are most similar to those of ICE-HK. The joint verb preferences in this cluster in turn are more similar to the verb preferences in ICE-IND than to those in ICE-SIN, but there is already quite some distance between the

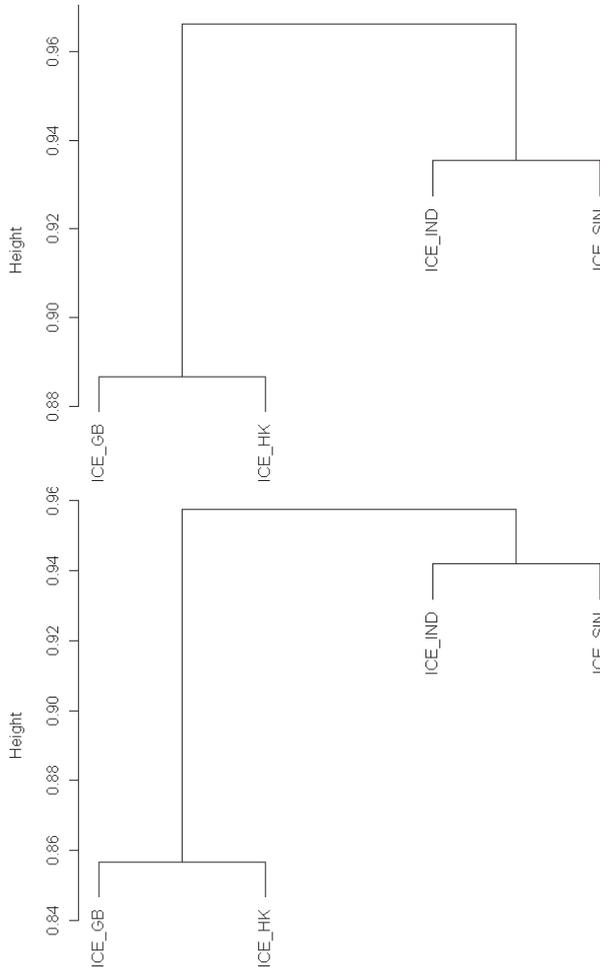


Figure 4. Dendrograms for the intransitive (above) and monotransitive (below) data

two. Finally, ICE-SIN is added — again at quite some distance — to the cluster of verb preferences made up of ICE-GB, ICE-HK, and ICE-IND.

The results for the intransitives and the monotransitives are virtually identical and also similar to those for the ditransitives. Consider Figure 4.

In both cases, the verb preferences in the historical input variety of British English are clearly more similar to those found in ICE-HK. The difference from the ditransitive data then is that the verb preferences in ICE-IND are more similar to those found in ICE-SIN — not to the cluster of ICE-GB and ICE-HK.

The overall picture is fairly clear. ICE-GB and ICE-HK are most similar to each other in terms of verbs' constructional preferences, and, within the present sampling space, ICE-IND and ICE-SIN are rather dissimilar to ICE-GB and ICE-HK. In the following section, we will abstract away from these correlational results the differences in the complementational (dis)preferences of the 59 verb lemmas in a binary fashion in order to identify shared trends and different trends across varieties.

4.2 Binary results

On a higher level of abstraction, it is useful to disregard the actual sizes of the differences and concentrate only on whether a verb prefers or disprefers a particular pattern in the four corpora. For expository reasons, we will begin with the verbs that, in ICE-GB, strongly prefer intransitives and monotransitives. The verbs from these two groups exhibit a striking conformity of (dis)preferences. All the verbs that prefer intransitives and disprefer ditransitives and monotransitives in ICE-GB behave exactly the same way across the New Englishes — with the exception of *speak*, which in ICE-HK and ICE-SIN also prefers monotransitives. An even more homogeneous result is obtained for the group of monotransitives: All the verbs that prefer monotransitives in ICE-GB also prefer monotransitives in the other corpora.

Let us now turn to the ditransitives and the neutral verbs, for which we found markedly different preferences. Consider Table 6, which lists the verbs that are strongly ditransitive in ICE-GB and whether they each prefer (+) or disprefer (–) to occur with ditransitives (d), intransitives (i), or monotransitives (m) in the other corpora (even if only slightly). For instance, in ICE-GB, *give* is only preferred in the ditransitive and dispreferred in the other two patterns, while, in ICE-IND, *give* is preferred both in the ditransitive and the monotransitive, and dispreferred in the intransitive; the empty cells indicate those verbs whose frequencies in the corpus were too low for them to be included in the sample (cf. Section 3.3).

The data in Table 6 show that, on the whole, these verbs prefer the ditransitive not only in ICE-GB, but also in the New Englishes. In particular, ICE-HK and

Table 6. Collostructional preferences of verbs that strongly prefer ditransitives in ICE-GB

Verb	ICE-GB	ICE-HK	ICE-IND	ICE-SIN
<i>give</i>	d+ i- m-	d+ i- m-	d+ i- m+	d+ i- m-
<i>tell</i>	d+ i- m-	d+ i- m-	d+ i- m-	d+ i- m-
<i>ask</i>	d+ i- m-	d+ i- m-	d+ i- m+	d+ i- m+
<i>send</i>	d+ i- m+	d+ i- m+	d+ i- m+	d+ i- m-
<i>offer</i>	d+ i- m+	d+ i- m+	d+ i- m+	d+ i- m-
<i>show</i>	d+ i- m+	d+ i- m+	d+ i- m+	d+ i- m+
<i>convince</i>	d+ i- m-	d- i- m+	d+ i- m-	-
<i>remind</i>	d+ i- m-	d+ i- m-	d- i- m+	d+ i- m-
<i>cost</i>	d+ i- m-	d+ i- m+	d- i- m+	d+ i- m+
<i>assure</i>	d+ i- m-	-	-	-
<i>inform</i>	d+ i- m-	d- i- m+	d+ i- m-	-
<i>teach</i>	d+ i- m-	d+ i- m-	d+ i- m-	d+ i+ m-
<i>allow</i>	d+ i- m-	d+ i- m+	d- i- m+	d+ i- m+
<i>lend</i>	d+ i- m-	-	d+ i- m-	d- i- m+
<i>persuade</i>	d+ i- m-	-	-	-

ICE-SIN are very similar, marked by diverging preferences only in the cases of two verbs (*convince* and *inform*) and one verb (*lend*), respectively. ICE-IND differs only by one more verb from the preferences observed in ICE-GB (*remind*, *cost* and *allow*). Interestingly, the three New Englishes all differ from the collostructional preferences in ICE-GB with regard to different verbs.

With regard to (dis)preferences of the intransitive, the results for the verbs in Table 6 are even more similar: With the single exception of *teach* in ICE-SIN, all the verbs disprefer intransitives. The picture is much more varied, however, once we look at the monotransitives: Most verbs — *give*, *ask*, *send*, *offer*, *convince*, *remind*, *cost*, *inform*, *allow*, *lend* — pattern differently in at least one of the New Englishes, if not all.

Let us now turn to the neutral verbs which do not exhibit a strong preference for any of the constructions under scrutiny in ICE-GB. They are listed in Table 7. A close inspection of the data reveals that the only thing that is systematic in Table 7 is that not a single verb exhibits the same preferences in all four corpora. For example, in ICE-GB, *fine* prefers ditransitives and disprefers monotransitive — but in ICE-SIN such preferences are reversed. Note also, for example, that *call* prefers ditransitives and intransitives, but in ICE-HK and ICE-IND, it disprefers the very same patterns. The group of neutral verbs is thus (and unsurprisingly so) marked by a very high degree of intervarectal variation in the (dis-)preferences of individual verbs for specific constructions.

Table 7. Collostructional preferences of neutral verbs with no strong preferences in ICE-GB

Verb	ICE-GB	ICE-HK	ICE-IND	ICE-SIN
<i>fine</i>	d+ i- m-	-	-	d- i- m+
<i>reassure</i>	d+ i- m+	-	-	-
<i>call</i>	d- i+ m-	d- i- m+	d- i- m+	d+ i+ m-
<i>cook</i>	d+ i+ m-	d- i+ m-	d- i+ m-	d- i+ m-
<i>command</i>	d+ i- m+	-	d- i- m+	d- i- m+
<i>surprise</i>	d+ i- m+	d- i- m+	d- i- m+	d- i- m+
<i>serve</i>	d- i- m+	d- i- m+	d- i+ m-	d- i- m+
<i>prescribe</i>	d+ i- m+	-	-	-
<i>pass</i>	d- i+ m-	d- i+ m-	d- i- m+	d- i+ m-
<i>bet</i>	d+ i- m+	d- i- m+	-	d- i- m+
<i>file</i>	d+ i- m+	d- i- m+	d- i- m+	d- i- m+
<i>quote</i>	d+ i- m+	d- i+ m-	d- i- m+	d- i- m+
<i>purchase</i>	d+ i- m+	d- i- m+	d- i- m+	d- i- m+
<i>permit</i>	d+ i- m+	d- i- m+	d- i- m+	d- i+ m-
<i>feed</i>	d- i- m+	d- i+ m-	d- i+ m-	d- i+ m-

From all the individual results for each of the New Englishes, the question arises how the three New English varieties compare to British English in general and with regard to the evolutionary cycle in particular. In order to answer this question, we cross-tabulated all the verb preferences in ICE-GB, pairing them off with each of the New Englishes. The results are shown in Table 8 and can be summarised as follows: Across all 59 verb lemmas, 38 verbs display the same (dis)preferences for the three constructions in ICE-GB and ICE-HK, 37 verbs display the same (dis)preferences for the three constructions in ICE-GB and ICE-IND, and 34 display the same (dis)preferences for the three constructions in ICE-GB and ICE-SIN. This means that the more advanced the New English variety is in the evolutionary cycle, the more dissimilar it is to present-day British English with regard to collostructional preferences; the NA frequencies belong to verbs we could not compare because of their non-occurrence in our sample.

In the following section, we will discuss all the above-mentioned results with a view to how they relate to Schneider's (2003, 2007) theory of the emergence of New Englishes.

Table 8. Collostructional (dis-)preferences of all verbs in pairwise corpus comparisons

ICE-GB vs. ICE-HK			ICE-GB vs. ICE-IND			ICE-GB vs. ICE-SIN		
same	different	NA	same	different	NA	same	different	NA
38	14	7	37	16	6	34	19	6

5. Discussion

Our findings show that the New Englishes under scrutiny — Hong Kong English, Indian English and Singapore English — fall into two groups, with Hong Kong English being more similar and Indian English and Singapore English being less similar to the input variety of British English. In the context of Schneider's (2003, 2007) evolutionary model, on-going structural nativisation in lexicogrammar can thus also be observed at the level of verb-construction association in terms of what we wish to label collostructional nativisation. In this context, it is worth pointing out that a slight cline of collostructional stability, as it were, can be described: While verbs that tend to be used monotonically and intransitively in ICE-GB show, by and large, the same preferences in ICE-HK, ICE-IND and ICE-SIN, there is a substantial degree of variation within the group of verbs that are strongly associated with the ditransitive in British English. In particular, there are various verbs in this group that do not prefer the ditransitive construction in one of the Asian Englishes (e.g. *convince* in Hong Kong English, *cost* in Indian English and *lend* in Singapore English); in all of these cases the verbs at hand prefer the monotransitive construction in the Asian English variety. That is to say, if a verb is repelled by the ditransitive construction in the course of the evolution of a New English variety, it tends to be attracted to the monotransitive construction instead.

A particularly interesting finding relates to the verb *give*, which is not only the most frequent verb in the group of ditransitives, but also the verb that is most prototypical — from a cognitive-linguistic point of view — of the transfer event (X causes Y to receive Z, cf. Goldberg 1995) to which the ditransitive construction refers. Our collostructional analysis shows that in ICE-IND, the verb *give* displays not only a preference for the ditransitive construction, but also for the monotransitive construction. This is in line with previous findings (cf. Mukherjee and Hoffmann 2006: 152) and indicates that in the group of ditransitive verbs, collostructional nativisation can also be observed for high-frequency verbs. It should be interesting to see in future follow-up studies what the reasons for the changing preferences are. The preference of *give* for the monotransitive construction in Indian English (IndE), for example, may be based on new light-verb constructions in Indian English (e.g. IndE *give a problem* for BrE *cause a problem*, IndE *give a complaint* for BrE *make a complaint*, cf. Mukherjee and Hoffmann 2006: 155).

The most significant overall finding of our collostructional analysis is the fact that a steady cline of dissimilarity to the historical input variety can be described at the level of shared (dis-)preferences of the 59 verbs for the three constructions: In Hong Kong English 38 verbs display the same (dis-)preferences as British English, in Indian English 37 verbs and in Singapore English 34 verbs. This cline correlates with Schneider's (2003, 2007) model as Hong Kong English represents Phase 2/3, Indian

English Phase 3/4 and Singapore English Phase 4 of the evolutionary cycle. Thus, the picture at large is that the more advanced a New English variety is in its evolution, the more dissimilar it is to British English at the level of collocations.

6. Concluding remarks and prospects for future research

The present paper allows for some general conclusions at the descriptive and methodological levels. At the descriptive level, our findings show that the process of structural nativisation indeed entails an appropriation of the language system in the new socio-cultural context at the level of verb-construction associations, i.e. collocational nativisation: The preferences of a given verb for a particular construction may change in the course of the evolution of a new variety. In our findings for 59 high-frequency verbs in present-day English, this can be observed in particular for verbs that have a strong preference for the ditransitive in British English. At the methodological level, we hope to have shown that collocational analysis can be fruitfully applied to research into lexicogrammatical differences between varieties of English: The collocational approach allows for a statistically sound, corpus-based description of the strength of verb-construction associations across varieties. The family of ICE corpora is ideally suited for such quantitative comparisons as they display the same corpus design.

Our study also opens up various avenues for follow-up studies of the lexicogrammar of New Englishes. Future research into lexicogrammatical differences between New Englishes based on collocational analyses of ICE corpora should include more constructions and more varieties of English in order to shed further light on the question whether there is a correlation between the evolution of New Englishes on the one hand and an increasing dissimilarity to the present-day version of the historical input variety on the other. Also, it is desirable to have a closer look at potential reasons for on-going collocational nativisation in New Englishes, e.g. typological factors and/or L1 interference that may trigger changes in a verb's preferences for certain constructions. Finally, we suggest that ICE-based research be complemented with analyses of large databases of New Englishes (e.g. by compiling web-derived corpora, cf. Hoffmann 2007) and, if possible, of diachronic databases of both the New Englishes and the historical input variety covering various stages of the colonial past and, thus, of the evolutionary process of variety formation.

References

- Bolton, Kingsley. 2003. *Chinese Englishes: A Sociolinguistic History*. Cambridge: Cambridge University Press.
- , ed. 2002. *Hong Kong English: Autonomy and Creativity*. Aberdeen: Hong Kong University Press.
- Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago/IL: University of Chicago Press.
- . 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Greenbaum, Sidney, ed. 1996. *Comparing English Worldwide: The International Corpus of English*. Oxford: Clarendon.
- Gries, Stefan Th. 2004. *Coll.analysis 3. A Program for R for Windows*.
- and Anatol Stefanowitsch. 2004. "Extending collostructional analysis: A corpus-based perspective on 'alternations'". *International Journal of Corpus Linguistics* 9: 97–129.
- , Beate Hampe and Doris Schönefeld. 2005. "Converging evidence: Bringing together experimental and corpus data on the association of verbs and constructions". *Cognitive Linguistics* 16: 635–76.
- Gupta, Anthea Fraser. 1998. "The situation of English in Singapore". In Joseph A. Foley, Thiru Kandiah, Bao Zhiming, Anthea Fraser Gupta, Lubna Alsagoff, Ho Chee Lick, Lionel Wee, Ismail S. Talib and Wendy Bokhorst-Heng, eds. *English in New Cultural Contexts: Reflections from Singapore*. Singapore: Oxford University Press, 106–26.
- Hoffmann, Sebastian. 2007. In Marianne Hundt, Nadja Nesselhauf and Carolin Biewer, eds. "From webpage to mega-corpus: The CNN transcripts". *Corpus Linguistics and the Web*. Amsterdam: Rodopi, 69–85.
- and Joybrato Mukherjee. 2007. "Ditransitive verbs in Indian English and British English: A corpus-linguistic study". *Arbeiten aus Anglistik und Amerikanistik* 32: 5–24.
- Kachru, Braj B. 1983. *The Indianization of English: The English Language in India*. Delhi: Oxford University Press.
- . 2005. *Asian Englishes: Beyond the Canon*. Hong Kong: Hong Kong University Press.
- Kortmann, Bernd, Kate Burridge, Rajend Meshtrie, Edgar W. Schneider and Clive Upton, eds. 2004. *A Handbook of Varieties of English*. Vol. 2: *Morphology and Syntax*. Berlin: Mouton de Gruyter.
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar*. Vol. 1: *Theoretical Prerequisites*. Stanford/CA: Stanford University Press.
- Li, David C.S. 1999. "The functions and status of English in Hong Kong: A post-1997 update". *English World-Wide* 20: 67–110.
- Lim, Lisa, ed. 2004. *Singapore English: A Grammatical Description*. Amsterdam, Philadelphia: Benjamins.
- Low, Ee Ling and Adam Brown. 2005. *English in Singapore: An Introduction*. Singapore: McGraw-Hill.
- Mair, Christian. 2008. *English Linguistics: An Introduction*. Tübingen: Gunter Narr.
- Mehrotra, Raja R. 1998. *Indian English: Texts and Interpretation*. Amsterdam, Philadelphia: Benjamins.
- Mukherjee, Joybrato. 2007. "Steady states in the evolution of New Englishes: Present-day Indian English as an equilibrium". *Journal of English Linguistics* 35: 157–87.

- . fc. “The lexicogrammar of present-day Indian English: Corpus-based perspectives on structural nativisation”. In Ute Römer and Rainer Schulze, eds. *Exploring the Lexis-Grammar Interface*. Amsterdam, Philadelphia: Benjamins.
- and Sebastian Hoffmann. 2006. “Describing verb-complementational profiles of New Englishes: A pilot study of Indian English”. *English World-Wide* 27: 147–73.
- and Marco Schilk. 2008 (in print). “Verb-complementational profiles across varieties of English: Comparing verb classes in Indian English and British English”. In Terttu Nevalainen, Irma Taavitsainen, Päivi Pahta and Minna Korhonen, eds. *The Dynamics of Linguistic Variation: Corpus Evidence on English Past and Present*. Amsterdam, Philadelphia: Benjamins.
- Nelson, Gerald, Sean Wallis and Bas Aarts. 2002. *Exploring Natural Language: Working with the British Component of the International Corpus of English*. Amsterdam, Philadelphia: Benjamins.
- Nihalani, Paroo, Ray K. Tongue, Priya Hosali and Jonathan Crowther. 2004. *Indian and British English: A Handbook of Usage and Pronunciation*. New Delhi: India Oxford University Press.
- Olavarría de Ersson, Eugenia O. and Philip Shaw. 2003. “Verb complementation patterns in Indian Standard English”. *English World-Wide* 24(2): 137–61.
- Ooi, Vincent B.Y., ed. 2001. *Evolving Identities: The English Language in Singapore and Malaysia*. Singapore: Times Academic Press.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech und Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- R Development Core Team. 2008. *R: A Language and Environment for Statistical Computing*. Vienna/Austria: R Foundation for Statistical Computing. <http://www.r-project.org> (accessed 19 Sept. 2008).
- Rushdie, Salman and Elizabeth West, eds. 1997. *The Vintage Book of Indian Writing 1947–1997*. London: Vintage.
- Schilk, Marco. 2006. “Collocations in Indian English: A corpus-based sample analysis”. *Anglia* 124: 276–316.
- Schneider, Edgar W. 2003. “The dynamics of new Englishes: From identity construction to dialect birth”. *Language* 79: 233–81.
- , Kate Burridge, Bernd Kortmann, Rajend Mesthrie and Clive Upton, eds. 2004. *A Handbook of Varieties of English*. Vol. 1: *Phonology*. Berlin: Mouton de Gruyter.
- . 2007. *Postcolonial English: Varieties Around the World*. Cambridge: Cambridge University Press.
- Shastri, S.V. 1988. “The Kolhapur Corpus of Indian English and work done on its basis so far”. *ICAME Journal* 12: 15–26.
- . 1996. “Using computer corpora in the description of language with special reference to complementation in Indian English”. In Robert J. Baumgardner, ed. *South Asian English: Structure, Use, and Users*. Urbana/IL: University of Illinois Press, 70–81.
- Stefanowitsch, Anatol and Stefan Th. Gries. 2003. “Collostructions: Investigating the interactions of words and constructions”. *International Journal of Corpus Linguistics* 8: 209–43.
- Ward, Joe H. 1963. “Hierarchical grouping to optimize an objective function”. *Journal of the American Statistical Association* 58: 236–44.

Authors' addresses

Joybrato Mukherjee
Justus Liebig University, Gießen
Department of English
Otto-Behaghel-Str. 10B
35394 Gießen
Germany
Joybrato.Mukherjee@anglistik.uni-giessen.de

Stefan Th. Gries
University of California, Santa Barbara
Department of Linguistics
Santa Barbara, CA 93106-3100
United States of America
stgries@linguistics.ucsb.edu