The lexicogrammar of present-day Indian English

Corpus-based perspectives on structural nativisation

Joybrato Mukherjee
Justus Liebig University of Giessen

The present paper puts into perspective four areas in which new local norms have emerged in the lexicogrammar of Indian English, the largest institutionalised second-language variety of English world-wide: (1) collocations, (2) new prepositional verbs, (3) new ditransitive verbs, and (4) verb-complementational patterns. At the descriptive level, it will be shown that corpus-based research provides new insights into quantitative and qualitative aspects of on-going structural nativisation at the lexis-grammar interface of Indian English. At the methodological level, it will be argued that in research into New Englishes well-balanced standard-size corpora can be fruitfully combined with very large collections of text obtained from the world-wide web, i.e. web-derived corpora.[1]

1. Introduction: Indian English as an endonormatively stabilised variety

Indian English is a paradigmatic example of institutionalised second-language varieties of English, which are described by Kachru (1985) as follows:

> The institutionalized second-language varieties have a long history of acculturation in new cultural and geographical contexts; they have a large range of functions in the local educational, administrative, and legal systems. The result of such uses is that such varieties have developed nativized discourse and style types

> and functionally determined sublanguages (registers), and are used as a linguistic
> vehicle for creative writing in various genres.   (Kachru 1985: 211)

Although English in India is a largely non-native language, it is – as in many other former colonial territories – institutionalised because it has been retained as an official language that fulfils important intranational functions, e.g. in administration and politics, on TV and in the press, in school-education and at universities. This process of institutionalisation has been accompanied by the Indianisation of the English language (cf. Kachru 1983): English has been adapted to the new local context by its indigenous users so that new forms and structures have developed in phonetics and phonology, in morphology, in the lexicon, in syntax, in style and discourse. There is a rich body of literature on the local norms of English that have been shaped in India. Comprehensive overviews of such aspects of structural nativisation in Indian English at the aforementioned levels are provided, for example, by Kachru (1983), Shastri (1988), Kachru (1994), Mehrotra (1998), Nihalani et al. (2004) and Kachru (2005).

As local Indian norms have emerged and are now widely accepted as part of an Indianised variety of English, present-day Indian English can be viewed as being endonormatively stabilised. In Schneider's (2003) evolutionary model of the development of New Englishes, endonormative stabilisation represents the fourth phase in a sequence of five stages which is posited to be characteristic of the emergence of new – native and non-native – varieties of English world-wide:

> As the English language has been uprooted and relocated throughout colonial
> and postcolonial history, New Englishes have emerged by undergoing a funda-
> mentally uniform process which can be described as a progression of five charac-
> teristic stages: FOUNDATION, EXONORMATIVE STABILIZATION, NATIV-
> IZATION, ENDONORMATIVE STABILIZATION, DIFFERENTIATION.
>   (Schneider 2003: 243)

Without going into detail, it should be pointed out that the succession from one stage to the next is essentially based on two interrelated factors: (1) group-interaction, and (2) identity-construction. The two groups that Schneider (2003) distinguishes are the indigenous people (i.e. the "IDG strand") and the settlers (i.e. the "STL strand"). Typically, the social and communicative interaction between the two strands becomes more and more intense in colonial contexts, so that a new hybrid identity is created, which, at the linguistic level, manifests itself in a new variety of English:

> The stages and strands 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)

In a recent macrosociolinguistic study, Schneider's (2003) innovative model has been applied to the historical development and the present-day situation of English in India (cf. Mukherjee 2007). Table 1 lists the four parameters that are relevant to the description of the five stages and summarises the various criteria according to which Indian English can be considered an almost prototypical case of endonormative stabilisation. Note that for nearly all criteria of endonormative stabilisation, the present-day situation of English in India can be marked with a "+". It is only with regard to the question whether the STL strand and the IDG strand are intricately interwoven in today's India that English in India opts out of the prototypical phase-IV scenario ("−") simply because there are hardly any descendants of British settlers left in present-day India. It seems that this particular criterion is not applicable to IDG-strand-dominated varieties of English such as Indian English (cf. Schilk 2006: 280).

In the present paper, I will focus on those structural effects that are characteristic of the endonormative stabilisation of present-day Indian English that cluster around the lexis-grammar interface. Such lexicogrammatical deviances from British English, the historical input variety, have been widely neglected in corpus-based research into Indian English (and, one may add, many other New Englishes as well) until very recently, laudable exceptions such as Shastri (1996) and Olavarría de Ersson and Shaw (2003) notwithstanding. More specifically, I will provide an overview of corpus-based findings with regard to four areas in which

Table 1. Endonormative stabilisation of present-day Indian English: Parameters and criteria

Parameter	Criterion	+/−
History and politics	Post-independence?	+
	Self-dependence?	+
Identity construction	STL/IDG strands interwoven?	−
	New nation with panethnic identity?	+
Sociolinguistics of contact, use, attitudes	Acceptance of local norms?	+
	Positive attitude to local variety?	+
	Literary creativity?	+
Linguistic developments, structural effects	Stabilisation of a new variety?	+
	Codification (e.g. dictionaries)?	+
	Relative homogeneity of local norms?	+

local norms have emerged in Indian English lexicogrammar: (1) collocations, (2) new prepositional verbs, (3) new ditransitive verbs, and (4) verb-complementational patterns. At the methodological level, I will argue the case for the use of both traditional standard-size corpora and much larger web-derived databases for research into phenomena at the lexis-grammar interface.

2. Structural nativisation at the lexis-grammar interface of Indian English

2.1 Collocations

There is no unanimous agreement on how to precisely define collocations. Specifically, the definitions range from statistical and frequency-based understandings of collocations (most notably in the corpus-linguistic tradition, cf. e.g. Sinclair's (1991) and Stubbs's (2001) approaches) to phraseological and more functionally-oriented definitions (as, for example, in Cowie's (1998) and Nesselhauf's (2005) work). In the following, I will keep to a more frequency-based understanding of collocation which typically underpins corpus-based quantitative research and which is borne out by the following descriptions of the nature of collocation:

> Collocation is the occurrence of two or more words within a short space of each other in a text.   (Sinclair 1991: 170)

> By *collocation* I mean a relationship of habitual co-occurrence between words [...].   (Stubbs 1995: 23)

These definitions highlight the fact that – especially from a quantitative point of view – collocation can be viewed as an inherently statistical concept that can only be described in a viable manner after the analysis of large amounts of spoken and written language in use, i.e. corpora, so that "habitual co-occurrences" of "words within a short space of each other" become identifiable. The word at the centre of a concordance-based search is called the collocational node word.

Table 2 lists all nouns, adjectives and adverbs for which Nihalani et al. (2004) posit a collocational deviation from British English. The middle column gives the number of occurrences and the normalised frequencies (per million words) in ICE-India; in the right-hand column the corresponding data from the British component (ICE-GB) can be found.[2]

---

2. The total numbers of words for the two corpora refer to the numbers calculated by *Word-Smith Tools*. Version 4.00 (Scott 1998).

Table 2. Some collocational node words in ICE-India and ICE-GB (cf. Schilk 2006)

Collocational node word	ICE-India [1,136,849 words]		ICE-GB [1,061,264 words]	
	Instances	Frequency [pmw]	Instances	Frequency [pmw]
affable	0	0.00	0	0.00
apt	2	1.76	5	4.71
chaste	3	2.64	1	0.94
class	260	228.70	240	226.15
cost	158	138.98	166	156.42
educational	**67**	**58.93**	**38**	**35.81**
extermination	2	1.76	1	0.94
far	504	443.33	597	562.54
illicit	15	13.19	0	0.00
implementation	42	36.94	27	25.44
postal	**19**	**16.71**	**6**	**5.65**
practitioner	3	2.64	14	13.19
rank	41	36.06	11	10.36
resourceful	1	0.88	0	0.00
scope	49	43.10	30	28.27
snag	7	6.16	3	2.83
superb	14	12.31	14	13.19
trophy	57	50.14	0	0.00
unique	56	49.26	32	30.15
wavy	2	1.76	0	0.00

The collocational differences between ICE-India and ICE-GB suggested by Nihalani et al. (2004) on intuitive grounds have been put to the test by Mukherjee (2005b) and Schilk (2006). In the following, I will sketch out some of our findings for two adjectival collocational node words, namely *educational* and *postal*.[3] These are the node words marked by boldface in Table 2. On the one hand, in both cases structural nativisation at the level of collocations can be traced in Indian English. On the other hand, the examples show that while intuition-based descriptions may often be in line with corpus findings, corpus data may also contradict intuitive accounts at times.

With regard to the use of the adjective *educational*, Nihalani et al. (2004) claim the following:

---

3. For more detailed information (including observations based on the older standard-size corpora, i.e. the Lancaster-Oslo/Bergen (LOB) of written British English and the Kolhapur Corpus of written Indian English), see Schilk (2006).

> When reference is being made to people or institutions, [...] BS [= British Standard English] usually prefers the noun form as attributive: 'education officer' and 'education ministry'. The adjectival -al form seems to be more frequent in IVE [= Indian Varieties of English] generally. (Nihalani et al. 2004:70)

The second claim about the overall frequency of *educational* is confirmed by the data given in Table 2. The first claim implies that the co-occurrence of *educational* and words that refer to persons or institutions is more frequent in Indian English than in British English. Table 3 vindicates that there is a greater tendency in Indian English for *educational* to co-occur with persons and institutions than in British English. With the *caveat* that the figures are relatively low, the distribution is significant at the five-percent level and thus captures a contrastive collocational difference between Indian and British English.

In the case of *educational*, it thus seems that the collocational differences between British and Indian English as depicted by Nihalani et al. (2004) are corroborated by the ICE data.

For the use of *postal*, too, Nihalani et al. (2004) claim that there is a collocational difference between Indian and British English:

> [IVE:] 'I enclose postal stamps to the value of 75 paise.'
> The phrase 'postal stamps' would not be found in BS where 'postage stamp' is the only possibility, though usually shortened to 'stamps'. (Nihalani et al. 2004:142)

However, in none of the 19 instances of *postal* in ICE-India does the collocate *stamp* occur, as Table 4 shows. As *postal stamp* does not occur in ICE-GB and in

Table 3. *Educational* – ICE-India vs. ICE-GB

educational	ICE-India		ICE-GB	
	Instances	Perc. of cases	Instances	Perc. of cases
+ [people/institutions]	20	29.8%	4	10.5%
– [people/institutions]	47	70.2%	34	89.5%

Table 4. *Postal* – right-hand collocates in ICE-India

postal	ICE-India		ICE-GB	BNC
	Instances	Perc. of cases		
+ course	5	26.3%	collocations not	
+ bill	4	21.1%	attested	
+ articles	3	15.8%		
+ stamp	0	0.0%		
+ [other collocates]	7	36.8%	6	726

the 100-million-word BNC either, there seems to exist no collocational difference between Indian and British English.

While the collocation *postal stamp* thus seems to be equally uncommon in Indian and British English, Table 4 nevertheless allows an interesting observation about the combinability of *postal* in Indian and British English: all right-hand collocates of *postal* that occur recurrently in ICE-India are not attested in ICE-GB or in the BNC. It thus transpires that *postal course*, *postal bill* and *postal articles* mark collocations that are highly typical of Indian English.

In the light of our findings so far, it seems to be necessary and fruitful to delve into collocational differences between Indian English and other New Englishes on the one hand and their parent varieties on the other in much more detail in future research. In this context, it is essential that existing intuition-based descriptions be checked against large amounts of natural text as available through corpora.

2.2   New prepositional verbs

Another area at the lexis-grammar interface in which new forms have emerged in Indian English refers to prepositional verbs. Nihalani et al. (2004) list prepositional verbs that are acceptable in Indian English but that are not used in British English because the preposition is considered to be superfluous in native varieties. For example, while *discuss about sth.* and *visit to so.* are admissible in Indian English, the corresponding British English forms are the one-word verbs *discuss sth.* and *visit so.* (cf. Nihalani et al. 2004: 66; 191). Table 5 shows that such "new" prepositional verbs can indeed be found in corpus data, albeit sporadically. Also, it is quite clear – as in many other cases in Indian English (cf. Mukherjee 2002) – that the new forms do not simply replace the established one-word verbs, but represent new variants that are used alongside the native forms.

It should be noted that the new forms in Indian English have emerged on grounds of analogies that Indian users have drawn between semantic and formal templates that already exist in the English language on the one hand and the new prepositional verbs on the other. For example, the emergence of *discuss about sth.*

Table 5. Some new prepositional verbs in Indian English

New prepositional verb	Instances in ICE-India
approach to	2
comprise of	8
discuss about	14
order for	2
visit to	2

in Indian English is based on two already existing templates in the British English input variety: (1) the prepositional verb *talk about* to which *discuss* is closely related semantically; (2) from a more formal point of view, the frequent collocation of the derivative noun *discussion* which typically occurs in the combination with *about* (i.e. as *discussion about*). Examples (1) to (5) are taken from ICE-India and illustrate the use of the five new prepositional verbs listed in Table 5. Underneath each of the examples the semantic and formal templates (as forms that are also perfectly acceptable in British English) that might have triggered the emergence of the new prepositional verbs are given.

(1) Yeah Yeah uh what manner you people are **approaching to** the higher authorities? <ICE-India S1A-083>
licensed by: BrE *appeal to* (v.), *approach to* (n.)

(2) ... the train will **comprise of** forty vans each with a capacity of nine <ICE-India S2B-001>
licensed by: *consist of* (v.), *be comprised of* (v. pass.)

(3) I had decided to **discuss about** these oddities today in our regular meeting but unfortunately ... <ICE-India W1B-017>
licensed by: *talk about* (v.), *discussion about* (n.)

(4) Now when you **ordered for** the beer did you get the beer? <ICE-India S1B-061>
licensed by: *ask for* (v.), *order for* (n.)

(5) He has **visited to** the following study areas at respective dates during his fellowship period <ICE-India W1B-020>
licensed by: BrE *go to* (v.), *visit to* (n.)

Being licensed by semantic and formal templates, the emergence of new prepositional verbs follows a rational impetus. Specifically, it can be viewed as a case of nativised semantico-structural analogy, which is defined by Mukherjee and Hoffmann (2006) as follows:

> [Nativised semantico-structural analogy is] a process by means of which non-native speakers of English as a second language are licensed to introduce new forms and structures into the English language because corresponding semantic and formal templates already exist in the English language system.
>
> (Mukherjee and Hoffmann 2006: 166f.)

In the following section, I will present another case of nativised semantico-structural analogy in Indian English, namely new ditransitive verbs.

Table 6. Some new ditransitives in ICE-India

New ditransitive verb	Instances in ICE-India
convey	2
furnish	1
inform	4
present	1
provide	24

2.3   New ditransitive verbs

As Olavarría de Ersson and Shaw (2003), Mukherjee and Hoffmann (2006) and Hoffmann and Mukherjee (2007) show, corpus data reveal that there are quite a few verbs – i.e. new ditransitives – that cannot be complemented with two object noun phrases in British English but that can be used in this basic ditransitive pattern in Indian English. However, as most of the new ditransitives are relatively rare, only few of them can be detected in the 1-million-word ICE-India corpus. Table 6 provides a list of new ditransitives in ICE-India.[4]

Some illustrative examples of the five verbs listed in Table 6 in the basic ditransitive pattern are given in (6) to (10).

   (6)   The envoy also **conveyed** the French Prime Minister Mr Rokha's invitation ... <ICE-India S2B-004>

   (7)   Can you **furnish** me Dr. Shastri's address? <ICE-India W1B-006>

   (8)   I will be very glad if you kindly **inform** me the final and exact schedule of selection committe meetings, ... <ICE-India W1B-030>

   (9)   ... and Congress-I ... cannot **present** us a viable government with the help of which it can rule the whole of the country <ICE-India S1A-005>

   (10)   Since they could not be taken back into the army he had asked the State Government to **provide** them other jobs <ICE-India W2C-001>

As new ditransitives represent a low-frequency phenomenon, it would be useful to compile corpora that are much larger than ICE-India. As such corpora of Indian English are not available, an obvious alternative to classic standard corpora is the utilisation of the Internet as a resource for corpus-linguistic research. Hoffmann (2007) introduces a methodology of how to obtain very large databases

---

4. Note, however, that provide is frequently complemented by two object noun phrases in some native varieties, especially American English (cf. Quirk et al. 1985:1210; Mukherjee 2001:299).

Table 7. New ditransitives in *The Statesman Archive*

Verb	Instances	Verb	Instances
advise	10	*notify*	1
brief	1	*present*	18
confer	3	*print*	1
despatch/dispatch	1	*provide*	217
explain	2	*put*	2
father	1	*remind*	4
gift	26	*rob*	4
impart	8	*submit*	1
inform	4	*supply*	15
intimate	1	*threaten*	1

from on-line archives that are available on the Internet. Mukherjee and Hoffmann (2006) have applied this methodology to the on-line archive of the Calcutta-based national daily newspaper *The Statesman* and derived from the world-wide web a 31-million-word corpus of articles from *The Statesman*: *The Statesman Archive*. In this web-derived corpus, many more new ditransitives in Indian English can be identified. Table 7 includes all verbs that are attested in the basic ditransitive pattern in *The Statesman Archive* but that are not acceptable as ditransitive verbs in British English.

The most frequent new ditransitive verb in *The Statesman Archive* is *gift*. While according to some native speakers, *gift* may be accceptable in British English in the context of sports commentaries, as in example (11), the usages exemplified in (12) and (13) are clearly innovations introduced by Indian users of English.

(11)   He was forced to bring down Nabi in the danger zone after **gifting** him the ball ... <The Statesman 2003-12-12>

(12)   Delay means serious risk of **gifting** Islamabad a talking point. <The Statesman 2002-10-26>

(13)   She said she wanted to **gift** him a dream. <The Statesman 2003-02-17>

Again, it needs to be pointed out that new ditransitives do not emerge out of the blue, but that they are based on logical and plausible analogies that Indian users draw between the meaning and complementation of existing templates and new forms. More specifically, as *gift* is semantically closely related to the most proto-typical ditransitive verb *give* in the sense that both *gift* and *give* refer to TRANS-FER events with the typical ditransitive meaning "X causes Y to receive Z" (cf. Goldberg 1995, 2006), the extension of the basic ditransitive pattern to the verb *gift* makes perfect sense. In essence, then, we are dealing with another case of na-

tivised semantico-structural analogy here: it is an analogy between the semantics and syntax of *give* and other established ditransitive verbs on the one hand and semantically closely related verbs such as *gift* on the other.

It should not go unmentioned that the new ditransitives in present-day Indian English are not forms that have been retained from ealier stages of British English. In order to test the superstrate retention hypothesis, Hoffmann and Mukherjee (2007) compiled a diachronic web-derived corpus with 23.5 million words from the on-line Gutenberg Archive (cf. <http://www.gutenberg.org/>) including texts from 1650 to 1900, i.e phases I to III in the evolution of Indian English (cf. Section 1). The overwhelming majority of new ditransitives in Indian English (e.g. *advise*, *convey*, *gift*) cannot be found in the basic ditransitive pattern in this database. In fact, of all new ditransitives identified in *The Statesman Archive* it is only the verbs *inform* and *put* that are also attested in the basic ditransitive pattern in the Gutenberg texts. It thus seems that most of the new ditransitives in Indian English are not archaic forms that were once fully institutionalised in British English when English was transported to India, but that they represent the output of a genuinely creative process on the part of Indian users of English after English had already been uprooted and relocated on the Indian subcontinent.

2.4   Verb-complementational patterns

While in the previous section the focus has been on the range of verbs that is associated with a particular pattern (i.e. the basic ditransitive pattern with two object noun phrases) in Indian English and in British English, I will now zoom in on the complementary perspective: the frequency and distribution of the various complementation patterns of an individual verb in the two varieties, for example ditransitive verbs such as *give*. From this perspective, too, there are differences between Indian and British English. A useful concept in this context is the notion of verb-complementational profile. As visualised in Figure 1, the concept of verb-complementational profile includes two related aspects: (a) the range (and frequencies) of the patterns of an individual verb in a variety; (b) the range of verbs with which an individual pattern is associated in a variety.

Figure 1 visualises the concept of verb-complementational profile by focusing on the ditransitive verb *give* and the basic ditransitive pattern. On the one hand, *give* is typically used in the basic ditransitive pattern with both objects realised as object noun phrases (e.g. *John gave Mary a book*), but it can also be used in other patterns, e.g. the *to*-phrase variant (e.g. *John gave a book to Mary*) and passive constructions (e.g. *Mary was given a book [by John]*). On the other hand, each individual pattern can be used with specific verbs in a given variety. For example,
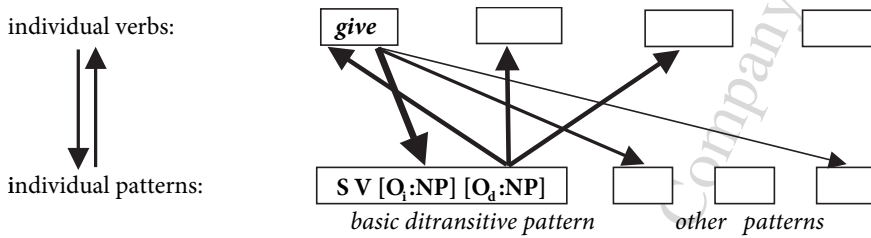
Figure 1. Two complementary aspects of a verb-complementational profile:
Verbs and associated patterns vs. patterns and associated verbs

as discussed in Section 2.3, while *inform* is not acceptable in the basic ditransitive
pattern in present-day British English, it is admissible in this pattern in Indian
English. Generally speaking, all varieties of English are potentially characterised
by differences between their verb-complementational profiles with regard to in-
dividual verbs and/or entire verb classes.

Verb complementation is increasingly considered as a significant area in
which varieties of English develop their own local norms:

> Verb complementation is an all-pervading structural feature of language and
> thus likely to be more significant in giving a variety its character than, for ex-
> ample, lexis. (Olavarría de Ersson and Shaw 2003: 138)

A very interesting example of an incipient verb-complementational divergence
between Indian and British English is provided by the prototypical high-fre-
quency ditransitive verb *give*, as Mukherjee and Hoffmann's (2006) corpus-based
study shows. Their analysis of the complementation patterns of *give* and other
ditransitive verbs is based on a categorisation of all patterns with which a given
ditransitive verb is attested in five basic types.[5] These basic types are described
and exemplified in (14) to (18).[6]

(14) a. I (S) GIVE [O_i:NP] [O_d:NP]
       b. On Tuesday members of Parliament **gave** the government their
           overwhelming support <ICE-GB:S2B-030 #54>

(15) a. II (S) GIVE [O_d:NP] [O_i:PP$_{to}$]
       b. I meant to **give** it to you earlier <ICE-GB:S1A-022 #176>

---

5.   The framework for the categorisation of all the patterns of give is taken over from Mukher-
jee (2005a).

6.   Examples (14) to (17) are taken from ICE-GB, while (18) is a rare pattern that could only
be found in the much larger BNC.

(16) a. $\boxed{\text{III (S) GIVE } [O_d\text{:NP}] \; \Theta_i}$

　　 b. he wanted physical love and I couldn't **give** that
　　　　 <ICE-GB:S1A-050 #184>

(17) a. $\boxed{\text{IV (S) GIVE } \Theta_i \; \Theta_d}$

　　 b. The other major point he raises is in addressing the question of 'why
　　　　 give in the first place?' <ICE-GB:W1A-011 #94>

(18) a. $\boxed{\text{V (S) GIVE } [O_i\text{:NP}] \; \Theta_d}$

　　 b. ... well she did [...] she did **give** Mrs <anonymised_last_name> didn't
　　　　 she? <BNC:KE6 2368-70>

Type I is the most basic type of ditransitive complementation with both objects realised as noun phrases. Type II is the pattern in which the indirect object is realised as a *to*-phrase and placed after the direct object. In type III, the indirect object is not made explicit, in type IV both objects are deleted, and in type V it is the direct object that is not made explicit. Note that from each of the five basic patterns, it is possible to derive various structurally related patterns, e.g. fronted constructions, passive constructions and relative clause structures. The categorisation is based on the basic cognitive assumption that *give* retains its ditransitive meaning "X causes Y to receive Z" in all instances and patterns.[7]

　　Figure 2 gives an overview of the distribution of most frequent patterns of *give* in ICE-India and ICE-GB. The data in Figure 2 show that while in British English the type-I pattern[8] (i.e. the basic ditransitive pattern with two object noun phrases) is the most frequent pattern of *give*, in Indian English it is the type-III pattern that occurs most frequently, i.e. the seemingly "monotransitive" pattern, cf. (16). Additionally, the type-II pattern (i.e. the *to*-prepositional pattern) occurs much more frequently in Indian English than in British English. Given the importance of *give* for the encoding of prototypical transfer events, the differences between Indian English and British English with regard to the three most frequent patterns of *give* are quite remarkable. Note also in this context that Hoffmann and Mukherjee (2007), by making use of data from the *Representative Corpus of Historical English Registers* (ARCHER, cf. Biber et al. 1994), have been able to show that the frequency and distribution of the complementation patterns of *give* has remained very stable from the mid 17th century to the 20th century in British English: again, the divergence between present-day Indian English and

---

7.　For a full list of the patterns with which *give* is associated in ICE-GB and ICE-India, see Mukherjee (2005a) and Mukherjee and Hoffmann (2006).

8.　Types IP, IIP, IIIP and IIIPb refer to passive patterns that are derived from the basic active types I, II and II, respectively.
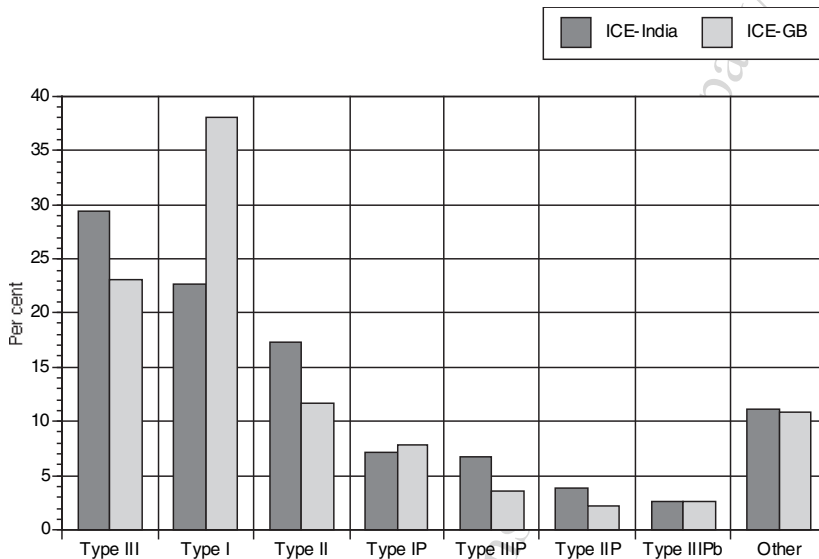
Figure 2. Complementation of give in ICE-India and ICE-GB: Focus on the most frequent patterns

British English can, therefore, not be explained in terms of superstrate retention but seems to be based on processes of change bound to Indian English.

While in the case of new prepositional verbs (cf. Section 2.2) and new ditransitives (cf. Section 2.3) nativised semantico-syntactic analogy can be seen as a rational force that has motivated the emergence of new forms in Indian English, the reasons for incipient changes in the frequency and distribution of the complementation patterns of a specific verb are much more multifaceted. A cultural explanation for the differences between Indian and British English has been offered by Olavarría de Ersson and Shaw (2003):

> Northern European cultures could have been more influenced by subjectivism, and see the individual as being at the center of the world, while South Asian cultures might tend to view the individual as a part or a small object in a larger whole. If this were so, when offered the choice between two syntactic structures that focus either on what is provided or on the recipient to express more or less the same thing, BrE speakers would be likely to profile the recipient more frequently in their use of language than their Indian counterparts do, whereas IndE speakers will be more likely to use the structure which profiles what is provided rather than the structure where the recipient is profiled.
>
> (Olavarría de Ersson and Shaw 2003: 159)

This general cultural difference might well explain the larger proportion of the type-III pattern in ICE-India, in which the recipient of the transfer event is not made explicit, i.e. not "profiled". As Mukherjee and Hoffmann (2006) have shown, however, it runs counter to the larger proportion of the type-II pattern in Indian English, as in the type-II pattern the recipient is in end-focus position and thus clearly represents a "profiled" entity.

It is more likely that there are various language-internal reasons for the complementational differences visualised in Figure 2. For example, new light-verb constructions in Indian English including *give* (e.g. *give development to their career*, *give explanation to the problem* instead of the simplex root verbs in *develop their career*, *explain the problem*) can be viewed as one reason for the higher proportion of type-II constructions.

Also, new verb-noun collocations in Indian English including *give* (e.g. *give a problem*, *give a complaint* instead of *cause a problem*, *make a complaint*) may well be one cause for the higher proportion of the type-III pattern in Indian English.[9] It will be a major challenge for future research to identify more language-internal reasons for the complementational differences that have emerged between Indian and British English.

3.   Concluding remarks: Large and small corpora as complementary databases

I hope to have shown in the present paper that the lexis-grammar interface is an area in which many processes of structural nativisation in Indian English can be detected, for example with regard to collocations, new prepositional verbs, new ditransitive verbs and verb complementation. All the four areas are lexicogrammatical phenomena that have so far been neglected in research into New Englishes in general and Indian English in particular. One of the reasons is the fact that large corpora of the size of, say, the 100-million-word BNC are not available for New Englishes – and for the analysis of many lexicogrammatical forms emerging in structural nativisation (e.g. the identification of new ditransitives), 1 million

---

9.   From a language-acquisitional perspective, one could also argue that high-frequency verbs such as *give* are overused by Indian users of English, who typically acquire English as an additional language in educational contexts. This is also corroborated by the fact that the total number of occurrences of *give* is much higher in ICE-India (1797 instances) than in ICE-GB (1064 instances). However, and notwithstanding the potential relevance of language-acquisitional theories to the description of New Englishes (cf. Williams 1987), the overuse perspective, which is strongly linked to learner corpus linguistics, might easily result in the overall and unwelcome view of Indian English as a deficient learner variety.

words are very often simply not enough. The present paper has shown that in spite of the fact that Internet data are fraught with problems (e.g. with regard to the representativeness of a web-derived corpus and the authorship of texts published on the web) a way forward could be the compilation of well-defined web-derived corpora with texts from on-line text archives.

This said, I should not be understood to mean, however, that small corpora like the 1-million-word ICE-components are not useful for research into the lexis-grammar interface – quite on the contrary. It would be highly desirable, for example, to have not only ICE-GB available in a fully tagged and parsed version, but also other ICE corpora such as ICE-India. The syntactic parsing of ICE-GB makes it possible to carry out statistically elaborated analyses of the associations between individual words and syntactic patterns. Specifically, I am thinking here of collostructional analyses, as suggested by Stefanowitsch and Gries (2003) and Gries and Stefanowitsch (2004):

> Collostructional analysis always starts with a particular construction and investigates which lexemes are strongly attracted or repelled by a particular slot in the construction (i.e. occur more frequently or less frequently than expected).
>
> (Stefanowitsch and Gries 2003: 214)

With regard to the complementation of ditransitive verbs, a collostruction can be described as the association of a ditransitive verb (the "collexeme", e.g. *give*) and its various complementation patterns (the "collostructs", e.g. the basic ditransitive construction = type I). The calculation of the collostruction strength, i.e. the strength of the association between the collexeme and the collostruct, can be based on the Fisher exact test (cf. Gries 2004). Table 8 shows the results of the collostructional analysis of the association between the five high-frequency ditransitive verbs *ask*, *give*, *offer*, *send* and *tell* and the various complementation patterns with which these verbs are associated according to the ICE-GB parsing scheme (i.e. "ditransitive" ~ type I, "complex-transitive" ~ type II, "monotransitive" ~ type III, "dimonotransitive" ~ type V, "intransitive" ~ type IV, cf. Nelson et al. 2002).[10] As the ditransitive construction is particularly frequently attested with the five verbs at hand, a distinction has been made between active and passive voice. The plus sign indicates collostructional attraction (the higher the figure, the stronger the association between the verb and the complementation pattern); the minus

---

10. Note that the system of complementation patterns in the parsing scheme of ICE-GB is surface-structure-oriented and, thus, fundamentally different from the view that all patterns of, say, *give* are ditransitive patterns. However, there are clear correspondences between the complementation types in ICE-GB and the five pattern types exemplified in (14) to (18).

Table 8. Collostructional attraction (+) and repulsion (−) in ICE-GB (−log Fisher exact, 10, based on *Coll. Analysis 3*, cf.Gries 2004)

Complementation type in ICE-GB	*ask*	*give*	*offer*	*send*	*tell*
ditransitive (active)	58.0 +	Inf. +	20.6 +	68.9 +	Inf. +
ditransitive (passive)	13.0 +	117.5 +	21.2 +	3.8 +	61.4 +
complex-transitive	5.8 −	11.8 −	0.9 −	5.8 +	9.5 −
monotransitive	15.0 −	0.6 −	10.3 +	10.6 +	97.5 −
dimonotransitive	73.2 +	0.2 −	0.5 +	0.3 +	260.4 +
intransitive	3.1 −	104.2 −	20.1 −	32.7 −	65.0 −

sign stands for collostructional repulsion (the lower the figure, the stronger the repulsion).

Table 8 illustrates that while some patterns are particularly strongly associated with specific verbs (e.g. the dimonotransitive pattern with *tell* as, for example, in *I told you*), other patterns may be strongly repelled by particular verbs (e.g. the intransitive pattern by *give*, but not so much by *ask*). Note also that the basic ditransitive pattern shows different preferences for the various verbs in Table 8: while the ditransitive (active) is very strongly associated with *give* and *tell* (which is also in line with intuition), this is not so much the case for *offer*.

Even without entering into a detailed discussion of the findings, Table 8 makes it clear that the results of a collostructional analysis provide a goldmine of observations for the description of the strength of word-pattern associations at the lexis-grammar interface. Unfortunately, however, collostructional analyses of this kind are only possible for fully parsed corpora such as ICE-GB. In order to describe lexicogrammatical differences between, say, British and Indian English in terms of collostructions, it would be necessary to annotate ICE-India according to the same parsing scheme as ICE-GB. It is in this context, for example, that small 1-million-word corpora such as the ICE components have a very important role to play. It is thus to be hoped that the parsing of ICE-India – and possibly other ICE corpora – will not remain wishful thinking.

References

Biber, D., Finegan, E. & Atkinson, D. 1994. ARCHER and its challenges: Compiling and exploring a representative corpus of historical English registers. In *Creating and Using English Language Corpora*, U. Fries, G. Tottie & P. Schneider (eds), 1–14. Amsterdam: Rodopi.

Cowie, A. P. 1998. *Phraseology: Theory, Analysis, Applications*. Oxford: The Clarendon Press.

Goldberg, A. E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago IL: University of Chicago Press.

Goldberg, A. E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: OUP.

Gries, S. T. 2004. *Coll. Analysis 3: A Program for R for Windows 2.x*.

Gries, S. T. & Stefanowitsch, A. 2004. Extending collostructional analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics* 9(1): 97–129.

Hoffmann, S. 2007. From web-page to mega-corpus: The CNN transcripts. In *Corpus Linguistics and the Web*, M. Hundt, N. Nesselhauf & C. Biewer (eds), 69–85. Amsterdam: Rodopi.

Hoffmann, S. & Mukherjee, J. 2007. Ditransitive verbs in Indian English and British English: A corpus-linguistic study. *Arbeiten aus Anglistik und Amerikanistik* 32 (1): 5–24.

Kachru, B. B. 1983. *The Indianization of English. The English Language in India*. Delhi: OUP.

Kachru, B. B. 1985. Institutionalized second-language varieties. In *The English Language Today*, S. Greenbaum (ed.), 211–226. Oxford: Pergamon.

Kachru, B. B. 1994. English in South Asia. In *The Cambridge History of the English Language,* Vol. V, R. Burchfield (ed.), 497–553. Cambridge: CUP.

Kachru, B. B. 2005. *Asian Englishes: Beyond the Canon*. Hong Kong: Hong Kong University Press.

Mehrotra, R. R. 1998. *Indian English: Texts and Interpretation*. Amsterdam: John Benjamins.

Mukherjee, J. 2001. Principles of pattern selection: A corpus-based case study. *Journal of English Linguistics* 29 (4): 295–314.

Mukherjee, J. 2002. Norms for the Indian English classroom: A corpus-linguistic perspective. *Indian Journal of Applied Linguistics* 28(2): 63–82.

Mukherjee, J. 2005a. *English Ditransitive Verbs: Aspects of Theory, Description and a Usage-Based Model*. Amsterdam: Rodopi.

Mukherjee, J. 2005b. 'All mine, mine alone...': Emerging local norms in Indian English lexico-grammar. Paper presented at the University of Zurich, 26 May 2005.

Mukherjee, J. 2007. Steady states in the evolution of New Englishes: Present-day Indian English as an equilibrium. *Journal of English Linguistics* 35(2): 157–187.

Mukherjee, J. & Hoffmann, S. 2006. Describing verb-complementational profiles of New Englishes: A pilot study of Indian English. *English World-Wide* 27(2): 147–173.

Nelson, G., Wallis, S. & Aarts, B. 2002. *Exploring Natural Language: Working with the British Component of the International Corpus of English*. Amsterdam: John Benjamins.

Nesselhauf, N. 2005. *Collocations in a Learner Corpus*. Amsterdam: John Benjamins.

Nihalani, P., Tongue, R. K., Hosali, P. & Crowther, J. 2004. *Indian and British English: A Handbook of Usage and Pronunciation*. 2nd edn. Delhi: OUP.

Olavarría de Ersson, E. O. & Shaw, P. 2003. Verb complementation patterns in Indian Standard English. *English World-Wide* 24(2): 137–161.

Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

Schilk, M. 2006. Collocations in Indian English: A corpus-based sample analysis. *Anglia* 124 (2): 276–316.

Schneider, E. W. 2003. The dynamics of new Englishes: From identity construction to dialect birth. *Language* 79(2): 233–281.

Scott, M. 1998. *WordSmith Tools*. Version 4.00. Oxford: OUP.

Shastri, S. V. 1988. The Kolhapur Corpus of Indian English and work done on its basis so far. *ICAME Journal* 12: 15–26.

Shastri, S. V. 1996. Using computer corpora in the description of language with special reference to complementation in Indian English. In *South Asian English: Structure, Use, and Users*, R. J. Baumgardner (ed.), 70–81. Urbana IL: University of Illinois Press.

Sinclair, J. M. 1991. *Corpus, Concordance, Collocation*. Oxford: OUP.

Stefanowitsch, A. & Gries, S. T. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2): 209–243.

Stubbs, M. 1995. Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of Language* 2(1): 23–55.

Stubbs, M. 2001. *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.

Williams, J. 1987. Non-native varieties of English: A special case of language acquisition. *English World-Wide* 8(2): 161–199.