

John Benjamins Publishing Company



This is a contribution from *English World-Wide 35:1*
© 2014. John Benjamins Publishing Company

This electronic file may not be altered in any way.

The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.

Permission is granted by the publishers to post this file on a closed server which is accessible to members (students and staff) only of the author's/s' institute, it is not permitted to post this PDF on the open internet.

For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com).

Please contact rights@benjamins.nl or consult our website: www.benjamins.com

Tables of Contents, abstracts and guidelines are available at www.benjamins.com

The dative alternation in South Asian English(es)

Modelling predictors and predicting prototypes

Tobias Bernaisch^{*}, Stefan Th. Gries[§] and Joybrato Mukherjee^{*}
Justus Liebig University, Giessen^{*} / University of California, Santa Barbara[§]

The present paper focuses on the modelling of cross-varietal differences and similarities in South Asian English(es) and British English at the level of verb complementation. Specifically, we analyse the dative alternation with GIVE, i.e. the alternation between the double-object construction (*John gave Mary a book*) and the prepositional dative (*John gave a book to Mary*) as well as their passivised constructions with regard to the factors that potentially exert an influence on this alternation in seven varieties of English. The South Asian varieties under scrutiny are Bangladeshi English, Indian English, Maldivian English, Nepali English, Pakistani English and Sri Lankan English, while British English serves as the reference variety. The patterns of GIVE are annotated according to the following parameters including potential predictors of the dative alternation: syntactic pattern and semantic class of GIVE; syntactic complexity, animacy, discourse accessibility and pronominality of constituents (cf. Gries 2003b; Bresnan and Hay 2008). The choices of complementation patterns are then statistically modelled using conditional inference trees and a random-forest analysis.

The results indicate that many of the predictors found to be relevant in British English are at play in the South Asian varieties, too. The syntactic pattern of GIVE is, in descending order, uniformly influenced by the predictors pronominality of recipient, length of recipient, semantic class of GIVE and length of patient. Interestingly, the predictor country is marginal in accounting for the dative alternation of GIVE across the varieties at hand. Based on this observation, we derive variety-independent protostructions, i.e. abstract combinations of (cross-varietally stable) features with high predictive power for a particular syntactic pattern, which we argue to be part of the lexicogrammatical “common core” (Quirk *et al.* 1985: 16) of English.

The implications of the present paper are twofold. While the order of the predictors regarding their influence on the dative alternation is clearly compatible with earlier studies (cf. e.g. Green 1974; Ransom 1979; Hawkins 1994; Gries

2003b), the stability of the order across varieties of English calls for a) a more fine-grained gradation of linguistic forms and structures at the lexis-grammar interface as indicators of structural nativisation and b) a revision of earlier verb-complementational findings specific to individual or groups of varieties of South Asian English.

Keywords: South Asian Englishes, dative alternation, GIVE, conditional inference trees, random forests, protostructions

1. Introduction

In corpus-based research into regional varieties of English, the focus has traditionally been on the description and analysis of features which may help to distinguish one particular variety from another. In contrast, a similar focus on linguistic features shared by a range of varieties of English has not yet been established in corpus-based research into World Englishes, notwithstanding some laudable exceptions (cf. e.g. Szmrecsanyi and Kortmann 2009). One of the reasons may be that until recently there were not enough truly comparable corpora of varieties of English available. With the release of an increasing number of regional components of the International Corpus of English (ICE), this has changed, leading to a growing interest in the description of aspects of unity and diversity at the same time across varieties of English (cf. e.g. Hundt and Gut 2012), be it on a global scale (cf. e.g. Sand 2004) or in a geographically more limited context (cf. e.g. Mukherjee and Gries [2009] with regard to Asian Englishes).

Against this background, the anglophone South Asian *Sprachraum* — including Bangladesh, India, the Maldives, Nepal, Pakistan and Sri Lanka — certainly constitutes an interesting case in point. For a long time, English in South Asia tended to be viewed as a fairly homogeneous variety based on Indian English, as is observable in Kachru's (passim) work. From the late 1980s onwards, this somewhat monolithic view of South Asian English has been complemented by a range of studies of specific features of individual South Asian Englishes (cf. e.g. Hartford [1989] on Nepali English syntax, Baumgardner [1993] on Pakistani English lexis and Rajapakse [2008] on the Burgher sociolect in Sri Lankan English). The advent of the South Asian Varieties of English (SAVE) Corpus (cf. Bernaisch *et al.* 2011) now provides an unprecedented database of 18 million words of acrolectal English language use with comparable components representing all six South Asian Englishes for the description and analysis of variety-specific features (including quantitative differences in the frequency and distribution of linguistic forms across varieties) on the one hand and pan-South Asian features (including

linguistic forms evenly distributed across Englishes in South Asia) on the other; cf. Section 2 for more details on the SAVE Corpus.

It goes without saying that differences between varieties of English manifest themselves in different degrees and at different levels of analysis. It is certainly true that the most obvious and transparent differences emerge in the areas of pronunciation and vocabulary: a speaker's regional accent and his/her choice of local lexical items are usually very good indicators of the variety at hand (in contrast to, say, his/her grammar, which tends to be relatively stable across varieties of English). However, if we are interested in more opaque and subtle differences between varieties of English at the level of the underlying language structure, the lexis-grammar interface is a particularly relevant area, as Schneider (2007) points out:

Innovations and distinctive structural properties of PCEs [= postcolonial Englishes] are frequently positioned at the interface between lexis and grammar, i.e. certain words but not others of the same word class prefer certain grammatical rules or patterns. The patterns as such are not new, nor are the words, but what is novel is the habitual association between them in specific varieties (Schneider 2007: 83).

It has been shown in a number of studies that, in general, many variety-specific forms and tendencies can be found in the lexicogrammar of South Asian Englishes, as for example shown by Sedlatschek (2009), Mendis (2010), Schilk (2011), or Zipp and Bernaisch (2012) for particle verbs and/or verb-complementational preferences in Indian and/or Sri Lankan English. Nevertheless, even within the lexis-grammar interface, different levels of granularity can be posited, ranging from the more concrete levels (e.g. *n*-grams as concrete and fixed multi-word combinations) to more abstract levels (e.g. transitivity of verbs). Depending on the level of granularity, the degree to which individual varieties of English differ from each other varies considerably (cf. Gries and Mukherjee 2010: 537–8).

A relevant area within the lexis-grammar interface where the correlation between the level of descriptive granularity on the one hand and the degree of intervarietal differences in South Asian Englishes on the other can be illustrated is verb complementation. Studies on the most concrete level of granularity take into account collocational routines apparent in verb-complementational patterns, one of which is e.g. "OFFER_vb + *prayer*" in montransitive constructions in Indian English (cf. Bernaisch 2013). *Prayer* does not collocate with OFFER_vb in montransitive patterns in either British English or Sri Lankan English, but their co-occurrence is relatively frequent in Indian English. Schilk (2011) studies collocations in the context of verb complementation in Indian English and identifies a set of collocational routines that differ between Indian English and British English. Thus, collocational routines embedded in verb-complementational patterns

constitute the most refined level in the context of verb complementation, at which differences between varieties of English abound.

Verb-complementational patterns as such relate to a more abstract level of granularity and various analyses have shown that there are clear differences between varieties of English at the level of verb-complementational patterns for specific verbs. For example, Schilk (2011) explores the verb-complementational profiles of GIVE, SEND and OFFER and identifies quantitative differences between Indian and British English, while Mukherjee (2008) shows differences in the complementation of GIVE between British English, Indian English and Sri Lankan English. The same holds true for collostructional routines, i.e. the association of a given verb (e.g. OFFER_vb) with a particular complementation pattern (e.g. the monotransitive construction): Mukherjee and Gries (2009) identify clear correlations between collostructional differences between various Asian varieties of English on the one hand and their evolutionary stage of development (according to Schneider's [2003, 2007] dynamic model of variety-formation) on the other.

Within the context of verb complementation, the degree of transitivity of verbs and verb groups constitutes a very abstract level of descriptive granularity (cf. Hopper and Thompson 1980). For example, Mukherjee and Schilk (2008) attest a higher level of transitivity for the semantically related verbs CONVEY, SUBMIT and SUPPLY in Indian English than in British English. However, the results of the corresponding follow-up study are not as clear-cut and also show a number of shared verb-complementational tendencies across South Asian varieties, in particular with regard to the complementation of the verb SUPPLY (cf. Schilk, Bernaisch and Mukherjee 2012).

When it comes to the most abstract level of descriptive granularity, i.e. the (strength of) predictors of certain verb-complementation patterns (such as length of constituents, animacy of constituents, discourse accessibility of constituents, *etc.*), Schilk *et al.* (2013) show that there are hardly any identifiable differences between Indian English, Pakistani English, and British English. In the light of this scale of granularity and the findings of previous studies, a general trend can thus be observed: the more concrete and fine-grained the verb-complementational analysis (e.g. specific collocations of a verb), the greater the tendency for differences between varieties of English to emerge; the more abstract and coarse the verb-complementational analysis (e.g. predicting factors influencing the choice of a verb-complementational pattern), the greater the tendency for cross-varietal homogeneity to prevail.

In the present paper we will focus and expand on the most abstract level of descriptive granularity in the area of verb complementation: our interest lies in the modelling of predictors for specific verb-complementational patterns across South Asian Englishes. Specifically, we will focus on predictors of the dative alternation with GIVE, the most prototypical ditransitive verb. In this context, the

present paper provides new insights from two perspectives. Firstly, from a descriptive perspective, we analyse data from the SAVE Corpus representing all six South Asian varieties of English. Secondly, from a methodological perspective, we triangulate a number of complex statistical models to verify the validity of our findings and step away from hitherto prevailing one-way approaches.

In the light of the aforementioned scale of granularity, our general hypothesis is (unsurprisingly) that the predictors of the complementation patterns of GIVE can be modelled relatively homogeneously across the six South Asian Englishes and British English. The remainder of the paper is structured as follows. In Section 2, the corpus data will be presented and the variables for which all instances of GIVE have been annotated are discussed. We will also elaborate on the statistical approach to the annotated data. In Section 3, the results of the quantitative-statistical analyses — using the methods of conditional inference trees and random forests — will be presented. On the basis of these results, we will propose the concept of cross-varietally stable, prototypical and construction-specific (protostructural) parameters for the dative alternation. Section 4 will provide a discussion and some concluding remarks.

2. Corpus data and data annotation

2.1 The South Asian Varieties of English (SAVE) Corpus

In the context of the research project “Verb complementation in South Asian Englishes: a study of ditransitive verbs in web-derived corpora,” funded by the German Research Foundation (*Deutsche Forschungsgemeinschaft*, Project No. DFG MU 1683/3., 2008–2011), a corpus of newspaper English covering all the six aforementioned South Asian Englishes has been compiled: the South Asian Varieties of English (SAVE) Corpus. Each of the six national components of the SAVE Corpus includes 3 million words, obtained from the online archives of two leading English-medium national newspapers, thus representing acrolectal written English usage of the variety concerned. The texts included in the 18-million-word SAVE Corpus have been produced by highly proficient users and have undergone several rounds of editing so that deviances from native Englishes in the Kachruvian Inner Circle cannot be viewed as learner mistakes, but as results of a process of structural nativisation. Given the absence of lexicographical accounts and full-fledged grammatical descriptions of most South Asian varieties of English (with Indian English standing out as the only fairly well-described variety), newspaper language fulfils an important standardising function and thus provides important insights into emerging local norms. Against this background, one particularly valuable asset of the SAVE Corpus is the fact that news agency reports have

Table 1. The SAVE Corpus (Bernaisch *et al.* 2011: 2).

Country	Newspaper	URL	Time span
Bangladesh	Daily Star	http://www.thedailystar.net	2003–2006
	New Age	http://www.newagebd.com	2005–2006
India	The Statesman	http://www.thestatesman.net	2002–2005
	The Times of India	http://www.timesofindia.indiatimes.com	2002–2005
Maldives	Dhivehi Observer	http://www.dhivehiobserver.com	2004–2007 [2008]
	Minivan News	http://www.minivannews.com	2004–2008
Nepal	Nepali Times	http://www.nepalitimes.com	2000–2007
	The Himalayan Times	http://www.thehimalayantimes.com	[2000] 2002–2008
Pakistan	Daily Times	http://www.dailytimes.com.pk	2002–2006
	Dawn	http://www.dawn.com	2002–2007
Sri Lanka	Daily Mirror	http://www.dailymirror.lk	2002–2007
	Daily News	http://www.dailynews.lk	2001–2005

been systematically removed, which ensures a high degree of variety-specificity of the data (cf. Bernaisch *et al.* 2011: 3). The SAVE Corpus is available in a plain-text version and in a POS-tagged version. Table 1 provides an overview of the design of the SAVE Corpus. The compilation and annotation of the SAVE Corpus is described in detail in the manual to the SAVE Corpus (cf. Bernaisch *et al.* 2011).

For intervarietal comparative analyses based on the SAVE Corpus, the periodicals section of the British National Corpus (BNC news) usually serves as a reference corpus, representing the present-day stage of the comparable newspaper genres in British English as the relevant historical input variety.

2.2 Coded variables

We retrieved 500 randomly selected instances of the verb GIVE from each of the six national components of the SAVE Corpus: Bangladesh (SAVE-BAN), India (SAVE-IND), the Maldives (SAVE-MAL), Nepal (SAVE-NEP), Pakistan (SAVE-PAK) and Sri Lanka (SAVE-SL). For reference purposes, we also retrieved 500 randomly selected instances of GIVE from BNC news.¹ Of the 3 500 instances, 2 971 cases turned out to be relevant in the sense that they could be coded according to all or most of the following variables:

1. The respective sets of corpus examples were extracted on the basis of all word forms of GIVE, i.e. *gave*, *give*, *given*, *gives*, *giving* (excluding non-verbal usages), and the samples were reduced to 500 instances using the random sampling option of *WordSmith Tools* (Version 4.0; Scott [1998]).

- TRANSITIVITY: the transitivity pattern of the verb: *ditransitive* (cf. (1)), *prepositional dative* (cf. (2)), *monotransitive* (cf. (3)), *intransitive* (cf. (4)), *monotransitive with indirect objects* (cf. (5)) plus separate codes for derivative patterns, e.g. passivised forms of various kinds (cf. (6)).²
 - (1) The mob gave **the criminals a good beating** and handed them over to police. <SAVE-BAN-DS_2004-05__pt20>
 - (2) His father gave **the money** to **his uncle**. <SAVE-IND-SM_2005-02-08>
 - (3) Funeral homes have plenty of brochures that give **detailed steps on how to plan your own funeral**. <SAVE-NEP-NT_2003-06-13>
 - (4) ... you have to give in order to receive. <SAVE-NEP-NT_2000-12-01>
 - (5) Undertaking the responsibility of revamping the gallery will give **the bank** to create options that would add to the avenues of learning in Sri Lanka. <SAVE-SL-DM_2005-01-29>
 - (6)
 - a. The PU VC announced that **the toppers** would be given **free education** if they chose to continue studying at PU. <SAVE-PAK-DT_2004-08-31>
 - b. **The assurance to set up the committee** was given by Patil to **a group of top social activists from Pune** on Tuesday. <SAVE-IND-TI_38406>
 - c. **The clearance for the trading company** was given by the cabinet on Saturday. <SAVE-IND-TI_38019>;
- COUNTRY: the component of SAVE (or BNC news) from which the concordance line was retrieved;
- RECLENGTH and PATLENGTH: the lengths of the recipient and the patient in words;
- RECANIMACY and PATANIMACY: the animacy of the recipient and/or the patient (*animate vs. inanimate*);
- REACCESSIBILITY and PATACCESSIBILITY: the accessibility of the recipient and/or the patient operationalised as to whether the recipient and/or the patient were mentioned in the preceding ten lines (*given vs. new*);
- RECPRONOMINALITY and PATPRONOMINALITY: whether the recipient and/or the patient was a pronoun or a lexical NP (*pronoun vs. np*);
- PATSEMANTICS: the semantics of the patient: abstract (as in *give him a hard time*), concrete (as in *give him a book*), or informational (as in *give him a warning*).

The annotation scheme follows, by and large, previous multifactorial studies of the dative alternation with GIVE, e.g. Gries (2003b), Bresnan and Hay (2008). The

2. The categorisation of verb-complementational patterns follows Mukherjee (2005).

annotated data were analysed in two different ways, which will be discussed in the next section.

2.3 Statistical analysis

For our statistical analysis, the above data set had to be trimmed given the relative rarity of some patterns. Specifically, we restricted the analysis to the four most frequent patterns of TRANSITIVITY: the active and passive versions of the ditransitive and the prepositional dative, which amount to 1 871 fully annotated concordance lines distributed as shown in Table 2.

Table 2. Observed frequencies (percentages) of the four transitivity patterns studied.

Pattern	Ditransitive	Prep. dative	Ditransitive passive	Prep. dative passive
Frequency (%)	905 (48.4%)	476 (25.4%)	349 (18.7%)	141 (7.5%)

Two different but related ways of exploring the data were pursued. First, we used the method of conditional inference trees (*ctree*) from the package *party* for R (cf. Hothorn et al. 2006 and R Development Core Team 2012). Conditional inference trees are a recursive partitioning approach towards classification and regression that attempt to classify / compute predicted outcomes / values on the basis of multiple binary splits of the data. Less technically, a data set is recursively inspected to determine according to which (categorical or numeric) independent variable the data should be split up into two groups to classify / predict best the known outcomes of the dependent variable: in our case, the dependent variable in question is TRANSITIVITY. This process of splitting the data up is repeated until no further split that would still sufficiently increase the predictive accuracy can be made, and the final result is a flowchart-like decision tree.

However, since classification trees sometimes fail to notice highly predictive interactions, we also fit a classification tree to our data that was based on the method of random forests (cf. Breiman 2001), specifically the implementation of Liaw and Wiener (cf. 2002, 2012) in the package *randomForest* for R. While this approach is also based on classification trees, it adds additional layers of randomness to the analysis.³ First, many different trees are constructed on different bootstrapped samples of the data. Second, each split in a tree chooses from only a subset of the available predictors, and the overall result amalgamates the multitude of trees that has been generated. The user has to specify only two parameters: the

3. This property of classification trees — traditional ones such as those generated by *rpart:rpart* and more advanced ones such as those generated by *party:ctree* — seems to be neglected at times. Consider Table (i).

number of predictors sampled at each split (we used the default setting, which amounts to 3) and the number of trees to grow (we chose 2000).

Both classification methods were fed the same model formula represented in (7), in which TRANSITIVITY is predicted on the basis of all the aforementioned predictors, i.e. all the variables for which the data have been coded. In the following section, we will present and discuss the results obtained from both approaches to the data.

$$(7) \text{ Transitivity} \sim \text{Country} + \text{RecLength} + \text{RecAnimacy} + \text{RecAccessibility} \\ + \text{RecPronominality} + \text{PatLength} + \text{PatAnimacy} + \text{PatAccessibility} + \\ \text{PatPronominality} + \text{PatSemantics}$$

3. Results

3.1 Conditional inference tree

The conditional inference tree returned a relatively high classification accuracy: 63.7% of all sentences included were classified correctly, which is highly

Table (i). Fictitious frequency distribution of three independent variables (1–3) and one dependent binary variable (4)

Variable 1	Variable 2	Variable 3	Variable 4: x	Variable 4: y
a	e	m	6	–
a	e	n	–	3
a	f	m	–	–
a	f	n	1	–
b	e	m	–	–
b	e	n	–	1
b	f	m	–	6
b	f	n	3	–

A moment's reflection will reveal two things: first, of the three independent variables, Variable 1 has the highest discriminatory power in isolation because it predicts Variable 4 correctly 70% of the time, whereas Variable 2 and Variable 3 only get 60% and 50% right respectively. Second, Variable 2 and Variable 3 together predict Variable 4 correctly 100% of the time: $e+m$ and $f+n \rightarrow x$ and $e+n$ and $f+m \rightarrow y$. However, the above classification trees do not 'see' that: they split on Variable 1 (because of the three variables, it has the highest classification accuracy) and can then not improve the tree anymore, ending up with 70% instead of 100% accuracy. A random forests type of analysis does also not necessarily see that Variable 2 \times Variable 3 perfectly predicts Variable 4, but its random sampling increases the chances that the importance of Variables 2 and 3 are recognised at least in part.

significantly different from the baseline percentage of 48.4% one would obtain by always just predicting the most frequent pattern ($p_{\text{binomial test}} < 10^{-40}$) and even much more different from the baseline percentage of 33.9% one would obtain from guessing randomly ($p_{\text{binomial test}} < 10^{-150}$). However, not all predictors are involved in the tree, which is shown in Figure 1 and should be read and interpreted as follows.

- one starts at the top node (node 1);
- at each node one goes to the left or the right to the next node (either node 2 on the left or node 5 on the right), which corresponds to assuming that the predictor mentioned in the node assumes the values superimposed onto the line leading to the next node;
- the above step is repeated until one reaches a terminal node, which provides the following information: (i) the number of the node and the number of the sentences which exhibit the combination of features by means of which one got to that node, and (ii) a bar plot of observed percentages of the four patterns (in the order *ditransitive*, *prepositional dative*, *ditransitive passive*, *prepositional dative passive*).

For example, node 3 shows that there are 443 cases with a pronominal recipient (the path from node 1 to node 2) and with abstract or informational patients (the path from node 2 to node 3); of these, about 85% involve the ditransitive pattern.

As Figure 1 indicates, the accessibility predictors did not make it into the tree, which is most likely due to the fact that these predictors are strongly correlated with others that did make it into the tree, such as RECPRONOMINALITY, RECLENGTH and PATLENGTH. Figure 1 reveals the significant patterns discussed in 1 through 9).

1. if the recipient is pronominal, then the ditransitive is preferred (nodes 1, 2, 3, 4), as in *He has really given me a chance to get some runs.* <SAVE-BAN-DS_2004-05__pt20>;
2. if the recipient is an NP and the patient is one word long, then the prepositional dative is preferred (nodes 1, 5, 6, 7, 8, 9), as in *Our expecting mothers (heroines) will give birth to babies and there will be newspapers and magazines flooded with the news;* <SAVE-IND-SM_2005-01-16>;
3. if the patient has ≤ 3 words and the recipient is an NP with > 5 words, then the prepositional dative is preferred (nodes 1, 5, 6, 10, 18), as in *All relevant stakeholders need to give more attention to the prevailing lack of awareness and sensitivity amongst the people in both urban and rural areas, which becomes a major handicap in rational use of water and in safeguarding its pollution.* <SAVE-PAK-DT_2004-08-24>;

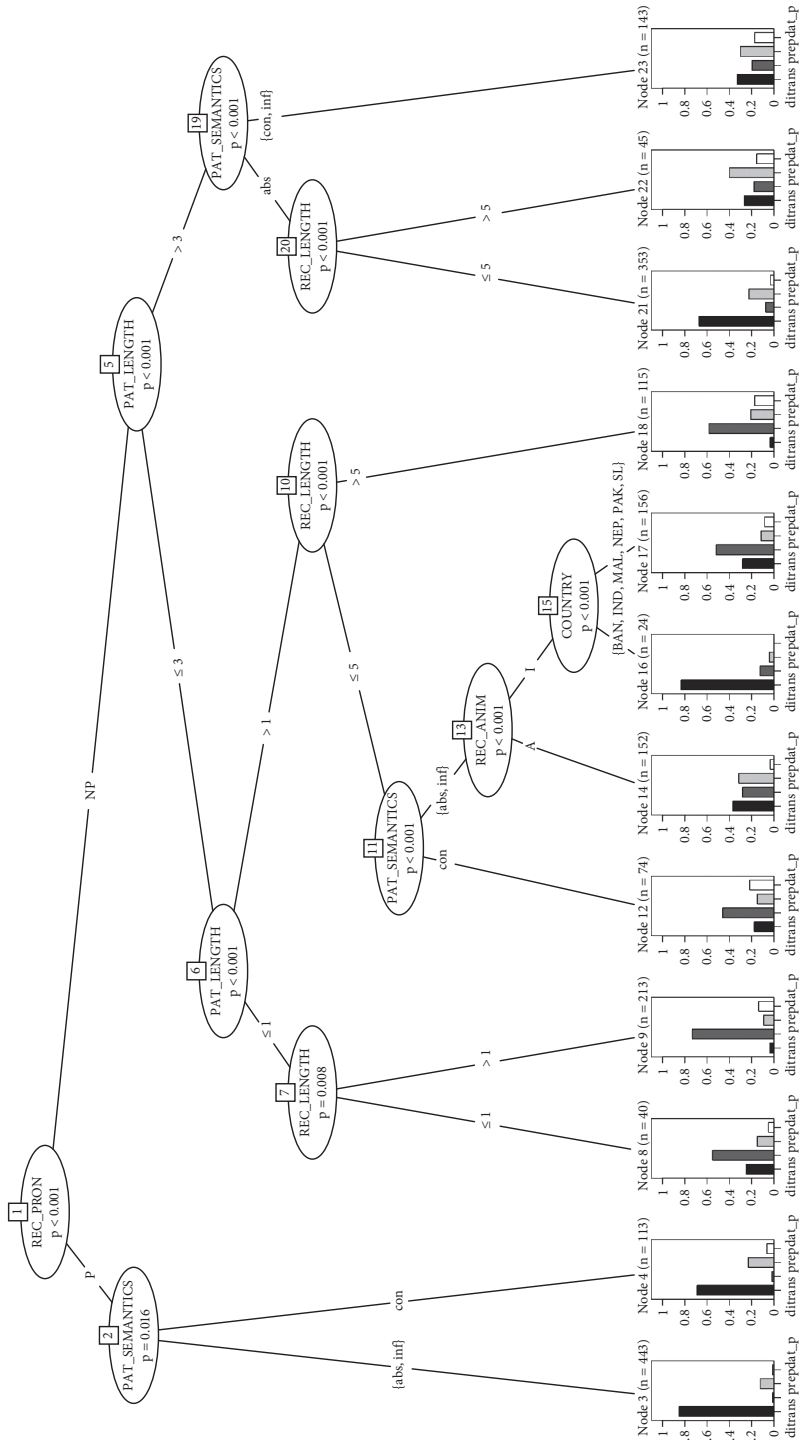


Figure 1. Conditional inference tree.

4. if the patient has ≤ 3 words and the recipient is an NP ≤ 5 words, and the clause denotes concrete transfer, then prepositional datives are preferred (nodes 1, 5, 6, 10, 11, 12), as in *His father gave the money to his uncle*. <SAVE-IND-SM_2005-02-08>;
5. if the patient has ≤ 3 words and the recipient is an NP with ≤ 5 words, the clause does not denote concrete transfer, and the recipient is animate, then ditransitives are preferred (nodes 1, 5, 6, 10, 11, 13, 14), as in *If, even after the Gujarat carnage, the people of the state gave "Mahatma" Modi a second term, they are welcome to their wisdom*. <SAVE-IND-SM_2005-03-29>;
6. if the patient has ≤ 3 words and the recipient is an NP with ≤ 5 words, the clause does not denote concrete transfer, and the recipient is inanimate, then British English strongly prefers ditransitives whereas South Asian Englishes prefer prepositional datives (nodes 1, 5, 6, 10, 11, 13, 15, 16, 17), as in *Fealy's try gave Newbridge the lead after Hayward had missed with two earlier attempts*. <BNC AKV> and *The citizens rallied round such kings who gave meaning to their existence through economic, religious and literary revivals*. <SAVE-SL-DN_2003-08-26>;
7. if the patient has > 3 words and is abstract and the recipient is an NP with ≤ 5 words, then ditransitives are preferred (nodes 1, 5, 19, 20, 21), as in *The as yet unpublished text, which I read this week, gives the rebels most of what they went to war for*. <SAVE-BAN-NA_2006-05-06>;
8. if the patient has > 3 words and is abstract and the recipient is an NP with > 5 words, then ditransitive passives are preferred (nodes 1, 5, 19, 20, 22), as in *However, he said the Minister had also directed that teachers who worked in uncongenial areas should be given a chance to work in a school closer to their residences*. <SAVE-SL-DM_2003-03-14>;
9. if the patient has > 3 words and is concrete or informational and the recipient is an NP, then ditransitives (and, a little less so, ditransitive passives) are preferred (nodes 1, 5, 19, 23), as in ... *and gives his audience a lesson on the evolution of American foreign policy ...* <SAVE-NEP-NT_2004-07-16> (and *Minister Choksy has been given a copy of the Chamber's response to the expectations of the Government as stated at the Development Forum*. <SAVE-SL-DM_2002-06-14>).

Given the length and complexity of the above, a different way of summarising these data is provided in Table 3.

Given these results, two related things are particularly noteworthy:

- The results are clearly compatible with previous studies of the dative alternation (Green 1974; Ransom 1979; Hawkins 1994; Gries 2003b, *etc.*) with regard to the more general and interrelated tendencies involved, namely short

Table 3. Summary of the preferred patterns of the conditional inference tree.

Ditransitives	Prepositional datives	Ditransitive passives
recipient is pronominal	recipient is an NP and patient is 1 word	
patient is short and recipient is a shortish NP and recipient is animate and the clause does not denote transfer or recipient is inanimate and the variety is British English	patient is short and recipient is a long NP or recipient is a shortish NP and the clause denotes transfer or recipient is inanimate and the variety is not British English	
patient is longish and abstract and recipient is shortish		patient is longish and abstract and recipient is long
patient is longish and not abstract and recipient is not abstract		

> long and given > new (but apparently only or mostly when givenness has morphosyntactic consequences in terms of length or pronominality because the accessibility variables are not part of the tree). It thus seems that these general tendencies, which are also captured by the interrelated principles of end-weight and end-focus (cf. Quirk *et al.* 1985: 1361–2), denote very robust principles across varieties of English, including non-native second-language varieties of English across South Asia.

- It is remarkable how little the varieties appear to differ with regard to the predictors' influence on the constructional choices. The predictor COUNTRY shows up only once in the tree (node 15). Even when it does occur, this is only at a very low — i.e. specialised — level, and indicates a very coarse differentiation: British English is distinguished from all other varieties in that it prefers ditransitives, as opposed to prepositional datives, with short abstract and informational patients and inanimate recipients. This combination of features is exemplified in (8).

- (8) a. the university names have been used to give **the letter 'spurious credibility'**. <BNC K54>
- b. Food Minister David MacLean is to visit the Newboulds factory in Startforth Lane, Riverside Estate, Middlesbrough, to give **the mizza** [= a kind of pizza] **the thumbs up** tomorrow. <BNC K4S>

The following section will discuss the results of the random-forest approach to the same data in order to (i) determine whether the conditional inference tree results can be taken at face value, (ii) test whether the predictor COUNTRY really plays such a subordinate rule in the data, and (iii) explore what the prototypical realisations of the four patterns under investigation are in our data.

3.2 Random forest

Our results show that the classification accuracy of the random-forest approach is even a bit higher than that of the conditional inference tree: 64.7% of the constructional choices are predicted correctly, which again is highly significantly better than the baseline percentage of 48.4% from always choosing the most frequent construction ($p_{\text{binomial test}} < 10^{-45}$) or choosing constructions randomly ($p_{\text{binomial test}} < 10^{-160}$). The construction predicted best is the ditransitive; the cross-validated accuracy was 64%.

3.2.1 Predictors, their importance, and their effects

The importance of predictors can be identified by determining the decrease in overall explanatory power that is caused by their deletion from the analysis. In this sense, the order of importance of predictors is as follows: PATLENGTH > RECPRONOMINALITY > RECLENGTH > PATSEMANTICS > COUNTRY. In other words, accessibility predictors again did not make an important (enough) contribution to the choice of a construction. Figure 2 summarizes how each of these predictors influences the predictions of the analysis. In each panel, the predicted probabilities of patterns are shown on the y -axis. In the upper three panels, where the graphically represented predictors are categorical (RECPRONOMINALITY, PATSEMANTICS and COUNTRY), the four predicted patterns are shown on the x -axis and the different levels of the predictors in question are shown with different line plots and means (with 95%-confidence intervals). In the lower two panels, where the graphically represented predictors are numeric (RECLENGTH and PATLENGTH), the predictors in question are shown with different smoothers:

- RECPRONOMINALITY (in the upper left panel): If the recipient is a pronoun, ditransitives are much more likely — otherwise, esp. prepositional datives are chosen.
- PATSEMANTICS (in the upper middle panel): if the patient is abstract, ditransitives are more likely; if it is informational, prepositional datives are slightly more likely; and if it is concrete, passives are most likely.
- COUNTRY (in the upper right panel): Ditransitives are predicted most strongly for the British English data, prepositional datives are predicted most strongly

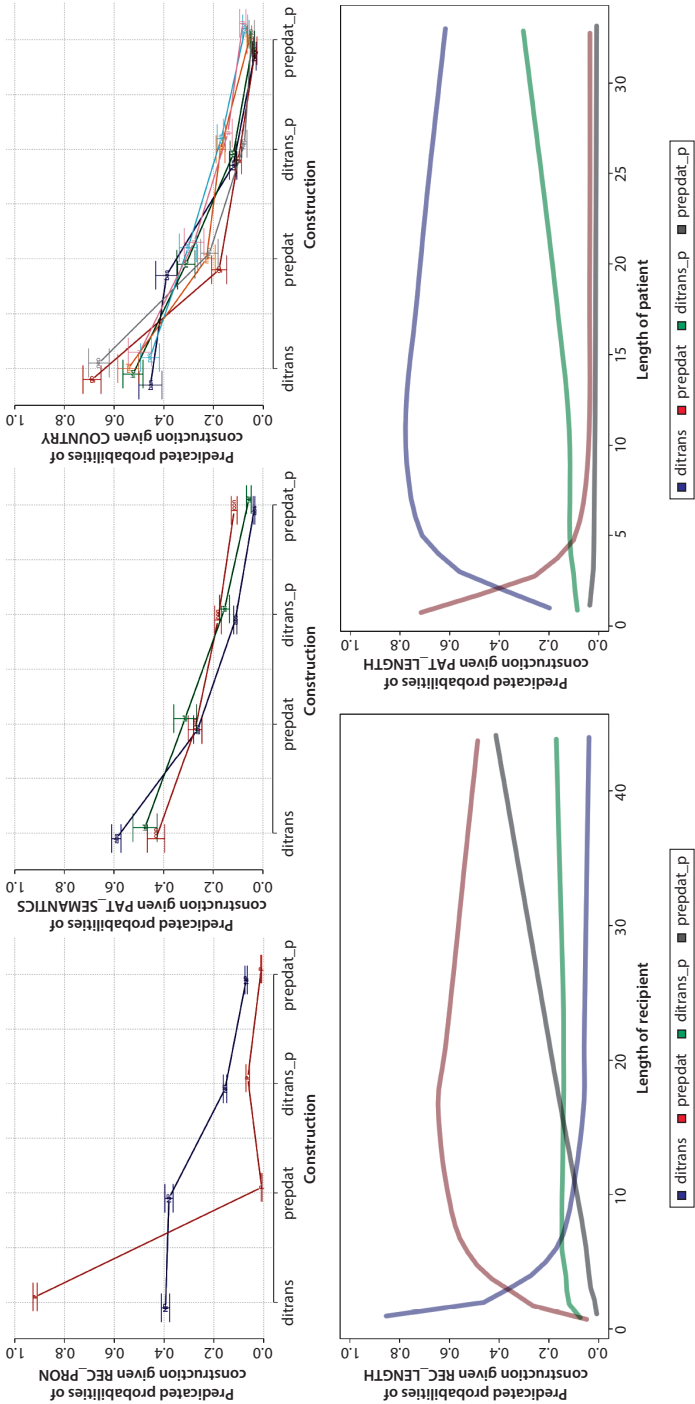


Figure 2. (Mean) predicted probabilities of the four patterns for the five most important predictors.

for the Bangladeshi English data, and there are no stronger patterns for the passives.

- RECLENGTH (the lower left panel): if the recipient is short, ditransitives are predicted; as soon as the recipient is longer than average, the more likely prepositional datives become, and the longer it becomes after that, the more likely passive prepositional datives also become.
- PATLENGTH (the lower right panel): if the patient is short, prepositional datives are predicted; as soon as the patient is of average length, ditransitives become more likely, and the longer patients become after that, the more also the probability of ditransitive passives increases.

In general, then, the results of the random-forest analysis are compatible with — and thus support — the conditional inference tree.

3.3 Prototypes and protostructions

An interesting follow-up that the present type of data and results allow is to explore what the analysis reveals to be the prototypical instantiations of the four constructions across the varieties under scrutiny. A similar approach has been pursued in Gries (2003b), which was the first corpus-based multifactorial study of the dative alternation. Gries (2003b) applied a linear discriminant analysis to instances of the dative alternation to (i) test and compare the predictive power of a whole set of independent variables at the same time (an approach also used in Gries [2003a] and that has since become a standard even if different statistical methods are used),⁴ and (ii) explore an objective way to identify the prototypical realisations of both constructions. With regard to (ii), Gries (2003b) ordered instances of constructions on a continuum based on their discriminant scores and found that this method revealed very plausible candidates for prototypical ditransitive constructions and prototypical prepositional dative constructions, as exemplified in (9) and (10) respectively.

- (9) going round beer festivals gave **me the idea of doing it for a living**
- (10) Jean Floud, A. H. Halsey, and F. M. Martin gave **a new impetus both to the study of these themes and to action upon them**

4. Strictly speaking, a linear discriminant analysis (LDA) requires multivariate normality, which the kind of corpus data typically used do not exhibit. However, results of LDAs are often very similar to those of, say, logistic regressions, and Jarvis (2011) compared a variety of classifiers and found that discriminant analyses were among the most successful.

For example, in (9), the recipient is very short, pronominal, and given, the patient is fairly long and complex, and the construction denotes something that can be viewed as metaphorical transfer. On the other hand, in (10), the patient is relatively short, especially when compared to the long phrase with two recipients, and the whole construction does not convey literal transfer (see Gries [2003a] for a similar approach to particle placement).

In our data, we can now extend Gries's (2003a, b) approach by including the passive versions of the two constructions. To this end, for each of the four constructions, we identified the 5% of the most certain predictions — i.e. the 5% of the predictions with the highest predicted probability of the construction in question — and compared the characteristics of these predicted constructions to the baseline of all the remaining data. As a result, we obtained the combinations of features for each construction that led to its being expected most strongly, which can reasonably be considered the prototypes of the constructions in question. Consider Table 4 for the results, where the four constructions whose features

Table 4. Prototypical instantiations of the four constructions as defined by the presence of highly predictive annotated characteristics.

Predictor	Ditransitive	Prep. dative	Voice
RecAnimacy	<i>animate</i>	<i>inanimate</i>	active
	<i>animate</i>	<i>animate</i>	passive
PatAnimacy	–	–	active
	–	–	passive
RecPronominality	<i>pronoun</i>	<i>NP</i>	active
	<i>NP</i>	<i>NP</i>	passive
PatPronominality	–	–	active
	–	<i>pronoun</i>	passive
RecAccessibility	<i>given</i>	<i>new</i>	active
	–	<i>new</i>	passive
PatAccessibility	<i>new</i>	<i>new</i>	active
	<i>given</i>	<i>given</i>	passive
RecLength	<i>shorter</i>	<i>longer</i>	active
	<i>longer</i>	<i>longer</i>	passive
PatLength	<i>longer</i>	<i>shorter</i>	active
	<i>longer</i>	<i>longer</i>	passive
PatSemantics	<i>abstract</i>	<i>abstract</i>	active
	<i>concrete (and informational)</i>	<i>concrete</i>	passive

one might want to focus on are highlighted in an alternating fashion. As Table 4 shows, each construction has two minimal-pair partners, so to speak: one with the other pattern but the same voice, one with the same pattern but the other voice. Interestingly, the prototypes show that (i) the constructional partners with the same voice are more different from each other and (ii) this difference is more pronounced in the active voice.

Examples (11) to (14) represent concrete realisations of the patterns of the dative alternation in active and passive voice to which the related predictive parameters in Table 4 apply. In (11) for instance, the recipient is animate ('tourists') and pronominal and short (*them*) and with high accessibility, with an abstract / metaphorical patient (*peek*), etc. The respective ditransitives in active and passive voice are exemplified in (11) and (12), while the prepositional datives are shown in (13) and (14).

- (11) The last stop will be Nasik, where tourists will be taken around vineyards and wineries to give **them a peek into the grape industry** ... <SAVE-IND-TI_37529>
- (12) **All 48 people who took an active part in last night's exercise** were given a **New Brighton stick of rock**. <BNC K3C>
- (13) Nepali nationalism was a Panchayati attempt to give **meaning to the imperial victories of the Gorkha kings** by exogenously introducing a common Nepali identity. <SAVE-NEP-HT_2001-06-01>
- (14) "It is Wakf land ... no deal can be made with it, nor can **it** be given to **anyone**," Board convener Mr SQR Ilyaas reportedly said after a meeting in Lucknow today. <SAVE-IND-SM_2003-06-2>

In essence, Table 4 thus shows typical (combinations of) parameters that predict to a very large extent specific instantiations of one of the four constructions at hand, as exemplified in (11) to (14) above. However, there is a possibility for extending this approach even further. In cognitive-linguistic studies, the notion of *prototype* is often defined as an abstract entity that unites the features with the highest cue validity for the construction at hand, a definition Gries (2003a, b) also applied in his related work. More concretely, the prototype of *bird* is not a robin / sparrow, but the abstract combination of characteristics that have a high cue validity (i.e. predictive power) for the category *bird* (a particular shape, a beak, feathers, ...), and then actually existing individual birds come closer or less close to that theoretical ideal.

Returning to the present constructional categories, above and as in Gries (2003a, b), we identified as prototypical sentences that (i) exhibited a large number of features that are probabilistically associated with one of the constructions and

(ii) that were attested in the data. However, there is a different and ultimately maybe more interesting alternative strategy, which involves taking the above definition of prototypes very literally. That is, much like the prototype of *bird* is not any one concrete bird but an amalgamation of characteristics with a high cue validity for the category *bird*, we can use the results of the statistical analyses to assemble the abstract combination of features that are associated with a particular construction, regardless of whether this combination of features is in fact instantiated in the data.

Why is this interesting? It is interesting because it makes it possible to identify a prototype even if not a single actual example exhibiting **all** the relevant characteristics is ever attested in the data. That is, apart from the concrete constructional instances in (11) to (14) above, which instantiate approximations to the prototype, there is what we might call a *protostruction*, an abstract combination of (cross-varietally stable) features with a high cue validity, or preference or predictive power, for a particular syntactic construction, or pattern.⁵ This approach would allow, for instance, to make comparisons of constructions in cross-varietally or otherwise different corpora even if a particular ditransitive passive in one corpus does not have exact corresponding counterparts in another corpus; the comparisons would be based on the abstract combinations of features and their constructional preferences / predictions, which can be reflected in classification trees, random forests, or, most frequently in current research, regression coefficients. From this quantitative view then, in a regression analysis, a protostructional ditransitive would be the combination of all the variable levels whose occurrence increases the odds / probability of a ditransitive, and the more these are stable across varieties and across corpora, the more we can speak of a protostruction and explore what it is that gives rise to that kind of stability. We therefore believe that this notion is useful because (i) it is more in line with the definition of prototypes in terms of cue validity that has been widely used in cognitive-linguistic circles, (ii) it constitutes a statistically rigorous operationalisation of constructional prototypes that extends Gries's earlier work, and (iii) it allows the analyst to explore interesting aspects of constructions such as the scope and stability of prototypical constructions as defined above (see also Sec. 4).

5. Note here that we are able to predict the protostruction for passive prepositional datives, although none of our examples exhibits all its protostructional features. (14) is the example instantiating most features with high cue validity for passive prepositional datives in our data, but it does not display all of them since PATSEMANTICS is not concrete and PATLENGTH is not longer.

4. Discussion and concluding remarks

The results presented in Section 3 allow for two conceptual and fundamental conclusions. Firstly, as initially hypothesised, variety-specificity plays only a negligible role in the modelling of the predictors of the dative alternation in active and passive voice across South Asian English(es) and British English. What the present study thus confirms is our understanding of predictors of the dative alternation as a fairly abstract level of verb-complementational preferences and tendencies — at this abstract level, varieties of English seem to be largely homogeneous and not subject to intervarietal differences in general or structural nativisation of second-language varieties in particular. Note that this is a more interesting finding than it may at first seem. This is because many of the predictors that both analytical approaches reveal as significantly affecting the choices of the argument structure constructions and voices can be, and have been in the past, related to the notion of processing effort (cf. Gries 2003a, b):

- length factors are related to processing in that, on the whole, longer material requires more processing cost than shorter material; thus, the results reflect the principle of short-before-long or end-weight;
- pronominality (as opposed to full lexical NPs) is related to processing in that, on the whole, pronouns are not only shorter (see above) but also a reliable indicator that their referents are considered given or at least accessible/inferable;
- semantic characteristics are related to processing in that (i) they are correlated with the degree of idiomaticity of the VP, (ii) they can be seen to reflect the degree with which the semantics of an utterance is compatible with that of the construction in general and the related notion that (iii) they are correlated with the choice of verb in the construction, and it is well known that verbs have strikingly different preferences for constructions (cf. Gries and Stefanowitsch [2004] in general and Mukherjee and Gries [2009] with regard to Asian Englishes).

Thus, while Mukherjee and Gries (2009) find that Asian Englishes exhibit a cline of collostructional nativisation of three Asian Englishes — that is, the preference of a verb to occur in the ditransitive in these Englishes is correlated with their evolutionary stage in Schneider's (2003, 2007) model — the present data show that such systematic lexico-constructional differences seem to develop independently of the above-mentioned processing-related factors that drive the constructional choices of speakers of the six Asian Englishes under scrutiny here. Note also that our findings are not in line with Hoffmann's (2011) general claim, based on his study of preposition placement across World Englishes, that second-language varieties of English tend to display stronger prototype and processing factors than

native varieties of English. Our observation is that at least as far as processing-related factors underlying speakers' choice of verb-complementational patterns are concerned, there is no such fundamental difference between the two types of Englishes. However, this does not mean that cognitive processing efforts are irrelevant to the process of structural nativisation in postcolonial Englishes. For example, Mukherjee and Hoffmann (2006) as well as Koch and Bernaisch (2013) identify sets of so-called "new ditransitives" in South Asian Englishes, i.e. verbs used in the double-object construction in South Asian English(es), but not in their historical input variety British English. These novel verb-construction associations are likely to have been derived by means of "semantico-structural analogy" (Mukherjee 2007: 175), which means that competent speakers of English as a second language "introduce new forms and structures into the English language on grounds of semantic and formal templates that already exist in the English language system" (Mukherjee 2007: 176). The utilisation of existing constructions as templates in the creation of new forms may thus be seen as another (possibly more variety-specific) processing-related factor in addition to the ones which we have shown to apply universally to the dative alternation in English.

Secondly, the similarity of the constructional choices and preferences across varieties extends to protostructions as abstract combinations of (cross-varietyally stable) features with a high predictive power for a particular syntactic construction: the protostructions of the ditransitive and the prepositional dative as derived from the random forests of the British and all South Asian data are, with the possible exception of PATSEMANTICS, very much compatible with what we know about the prototypical instantiations of the dative alternation in general, and this in spite of the fact that the examples studied were from written language, which arguably — but apparently not much — reflect online processing constraints less. Consequently, it seems that factors determining processing and, thus, ultimately constructional choices are widely applicable to all varieties of English and may therefore represent features shared by all Englishes. This is intriguing for two reasons: on the one hand, the speakers of South Asian Englishes usually learn English as a second or third language in a complex language-contact situation and after having acquired their indigenous first language. In spite of the typological differences between the enormous range of first languages across South Asia, all the Englishes that have emerged under and after British colonial rule obviously share with British English general processing principles and statistical tendencies underlying the dative alternation.

On the other hand, the fact that the processing-related factors all seem to be at work very similarly across the varieties may have important implications with regard to another dimension of language variation, namely genre. In a series of studies, Gries and Wulff (2013) and Wulff and Gries (2011) showed that German and Chinese learners' constructional choices in the genitive alternation are different

from each other in terms of processing-related variables. Since their data were from the ICLE corpus, one important methodological question that arises from such comparisons is how we can compile corpora of South Asian Englishes (and, of course, other varieties) that contain more / less-edited language, i.e. language use where local variety influences may roam more freely. This will be a necessary but also interesting aspect future research will have to address.

Finally, it needs to be stressed that the findings of the present study do not call into question earlier studies of verb-complementational differences between individual South Asian Englishes and British English which have referred to more concrete levels of descriptive granularity. For example, Hoffmann et al. (2011) have shown in a study based on a pilot version of the SAVE Corpus and the BNC that a number of light-verb constructions such as *give boost to* or *give chase to* are more characteristic of Indian English than British English. While this remains a valid observation, our present study has revealed that the choice between the ditransitive construction with GIVE and the prepositional dative with GIVE in active and passive voice seems to be governed by cross-varietally stable predictors. This overall finding is also in line with Bresnan and Hay's (2008) study of the dative alternation in American and New Zealand English; notwithstanding a few specific inter-varietal differences, they, too, observe largely homogeneous effects of predicting parameters and processing factors on the choice between constructions. In essence, then, the protostructions that we have identified in the present study seem to form part of the "common core" (Quirk *et al.* 1985: 16) of English lexicogrammar.

Against this background, future research needs to further scrutinise protostructions from temporal as well as (additional) spatial angles. In the light of studies such as Rohdenburg (2007), who depicts diverging trends in the verb-complementational profiles in British and American English across the 19th and 20th century, and Grimm and Bresnan (2009), who highlight a diachronic increase in the probability of double-object constructions in both American and British English texts, the degree to which protostructions are diachronically stable needs to be examined. Given that there seem to be verb-complementational differences across regional varieties which the paper at hand has not taken into consideration (e.g. in American and Australian English; cf. Bresnan and Ford 2010; Collins 1995), it is certainly desirable to extend the spatial scope of studies on protostructions by including additional native and second-language varieties.

References

- Baumgardner, Robert J. 1993. *The English Language in Pakistan*. Karachi: Oxford University Press.

- Bernaish, Tobias. 2013. "The verb-complementational profile of OFFER in Sri Lankan English". In Magnus Huber and Joybrato Mukherjee, eds. *Corpus Linguistics and Variation in English: Focus on Non-Native Englishes*. Helsinki: Research Unit for Variation, Contacts and Change in English, <http://www.helsinki.fi/varieng/journal/volume/13/bernaish>.
- , Christopher Koch, Joybrato Mukherjee and Marco Schilk. 2011. *Manual for the South Asian Varieties of English (SAVE) Corpus*. Giessen: Justus Liebig University, Department of English.
- Breiman, Leo. 2001. "Random forests". *Machine Learning* 45: 5–32. DOI: 10.1023/A:1010933404324
- Bresnan, Joan and Jennifer Hay. 2008. "Gradient grammar: an effect of animacy on the syntax of *give* in New Zealand and American English". *Lingua* 118: 245–59. DOI: 10.1016/j.lingua.2007.02.007
- and Marilyn Ford. 2010. "Predicting syntax: processing dative constructions in American and Australian varieties of English". *Language* 86: 186–213.
- Collins, Peter. 1995. "The indirect object construction in English: an informational approach". *Linguistics* 33: 35–49. DOI: 10.1515/ling.1995.33.1.35
- Green, Georgia M. 1974. *Semantics and Syntactic Irregularity*. Bloomington, IN: Indiana University Press.
- Gries, Stefan Th. 2003a. *Multifactorial Analysis in Corpus Linguistics: A Study of Particle Placement*. London, New York: Continuum.
- . 2003b. "Towards a corpus-based identification of prototypical instances of constructions". *Annual Review of Cognitive Linguistics* 1: 1–27. DOI: 10.1075/arcl.1.02gri
- and Joybrato Mukherjee. 2010. "Lexical gravity across varieties of English: an ICE-based study of *n*-grams in Asian Englishes". *International Journal of Corpus Linguistics* 15: 520–48. DOI: 10.1075/ijcl.15.4.04gri
- and Anatol Stefanowitsch. 2004. "Extending collocation analysis: a corpus-based perspective on alternations". *International Journal of Corpus Linguistics* 9: 97–129. DOI: 10.1075/ijcl.9.1.06gri
- and Stefanie Wulff. 2013. "The genitive alternation in Chinese and German ESL learners: Towards a multifactorial notion of context in learner corpus research". *International Journal of Corpus Linguistics* 18: 327–56. DOI: 10.1075/ijcl.18.3.04gri
- Grimm, Scott and Joan Bresnan. 2009. "Spatiotemporal variation in the dative alternation: a study of four corpora of British and American English". Paper presented at the Third International Conference Grammar & Corpora, University of Mannheim, 22 September 2009.
- Hartford, Beverley S. 1989. "Prototype effects in non-native English: object-coding in verbs of saying". *World Englishes* 8: 97–117. DOI: 10.1111/j.1467-971X.1989.tb00647.x
- Hawkins, John A. 1994. *A Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.
- Hoffmann, Sebastian, Marianne Hundt and Joybrato Mukherjee. 2011. "Indian English — an emerging epicentre? A pilot study on light verbs in web-derived corpora of South Asian Englishes". *Anglia* 129: 258–80. DOI: 10.1515/angl.2011.083
- Hoffmann, Thomas. 2011. *Preposition Placement in English: A Usage-Based Approach*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511933868
- Hopper, Paul and Sandra A. Thompson. 1980. "Transitivity in grammar and discourse". *Language* 56: 251–99.

- Hothorn, Torsten, Kurt Hornik and Achim Zeileis. 2006. "Unbiased recursive partitioning: a conditional inference framework". *Journal of Computational and Graphical Statistics* 15: 651–74. DOI: 10.1198/106186006X133933
- Hundt, Marianne and Ulrike Gut, eds. 2012. *Mapping Unity and Diversity World-Wide: Corpus-Based Studies of New Englishes*. Amsterdam, Philadelphia: Benjamins.
- Jarvis, Scott A. 2011. "Data mining with learner corpora: choosing classifiers for L1 detection". In Fanny Meunier, Sylvie de Cock, Gaëtanelle Gilquin and Magali Paquot, eds. *A Taste for Corpora: In Honour of Sylviane Granger*. Amsterdam, Philadelphia: Benjamins, 127–54.
- Koch, Christopher and Tobias Bernaisch. 2013. "Verb complementation in South Asian English(es): the range and frequency of 'new' ditransitives". In Gisle Andersen and Kristin Bech, eds. *English Corpus Linguistics: Variation in Time, Space and Genre*. Amsterdam: Rodopi, 69–89.
- Liaw, Andy and Matthew Wiener. 2002. "Classification and regression by randomForest". *R News* 2: 18–22.
- and ———. 2012. randomForest: Breiman and Cutler's random forests for classification and regression. Version 4.6.7. <<http://cran.r-project.org/web/packages/randomForest/index.html>> (accessed 22 Jan. 2013).
- Mendis, Dushyanthi. 2010. "Formality in academic writing: the use / non-use of phrasal verbs in two varieties of English". In Miguel F. Ruiz-Garrido, Juan C. Palmer-Silveira and Inmaculada Fortanet-Gomez, eds. *English for Professional and Academic Purposes*. Amsterdam: Rodopi, 11–24.
- Mukherjee, Joybrato. 2005. *English Ditransitive Verbs: Aspects of Theory, Description and a Usage-Based Model*. Amsterdam: Rodopi.
- . 2007. "Steady states in the evolution of New Englishes: present-day Indian English as an equilibrium". *Journal of English Linguistics* 35: 157–87. DOI: 10.1177/0075424207301888
- . 2008. "Sri Lankan English: evolutionary status and epicentral influence from Indian English". In Klaus Stierstorfer, ed. *Anglistentag 2007 Münster: Proceedings*. Trier: Wissenschaftlicher Verlag Trier, 359–68.
- and Marco Schilk. 2008. "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, 163–81.
- and Sebastian Hoffmann. 2006. "Describing verb-complementational profiles of New Englishes: a pilot study of Indian English". *English World-Wide* 27: 147–73. DOI: 10.1075/eww.27.2.03muk
- and Stefan Th. Gries. 2009. "Collostructional nativisation in New Englishes: verb-construction associations in the International Corpus of English". *English World-Wide* 30: 27–51. DOI: 10.1075/eww.30.1.03muk
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- R Development Core Team. 2012. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <<http://www.R-project.org>> (accessed 22 Jan. 2013).
- Rajapakse, Agra. 2008. "A descriptive analysis of the language of the Burghers of Sri Lanka". In Dinali Fernando and Dushyanthi Mendis, eds. *English for Equality, Employment and Empowerment: Selected Papers from the 4th International Conference of the Sri Lanka English Language Teachers' Association*. Colombo: SLELTA, 48–58.

- Ransom, Elizabeth. 1979. "Definiteness and animacy constraints on passives and double-object constructions in English". *Glossa* 13: 215–40.
- Rohdenburg, Günter. 2007. "Grammatical divergence between British and American English in the 19th and early 20th centuries". Paper presented at the Third Late Modern English Conference, University of Leiden, 1 September 2007.
- Sand, Andrea. 2004. "Shared morpho-syntactic features of contact varieties: article use". *World Englishes* 23: 281–98. DOI: 10.1111/j.0883-2919.2004.00352.x
- Schilk, Marco. 2011. *Structural Nativization in Indian English Lexicogrammar*. Amsterdam, Philadelphia: Benjamins.
- , Joybrato Mukherjee, Christopher F.H. Nam and Sach Mukherjee. 2013. "Complementation of ditransitive verbs in South Asian Englishes: a multifactorial analysis". *Corpus Linguistics and Linguistic Theory* 9: 187–225.
- , Tobias Bernaisch and Joybrato Mukherjee. 2012. "Mapping unity and diversity in South Asian English lexicogrammar: verb-complementational preferences across varieties". In Hundt and Gut, eds. 2012: 137–65.
- Schneider, Edgar W. 2003. "The dynamics of New Englishes: from identity construction to dialect birth". *Language* 79: 233–81.
- . 2007. *Postcolonial English: Varieties around the World*. Cambridge: Cambridge University Press.
- Scott, Mike. 1998. *WordSmith Tools*. Version 4.0. Oxford: Oxford University Press.
- Sedlatschek, Andreas. 2009. *Contemporary Indian English: Variation and Change*. Amsterdam, Philadelphia: Benjamins.
- Szmrecsanyi, Benedikt and Bernd Kortmann. 2009. "The morphosyntax of varieties of English worldwide: a quantitative perspective". *Lingua* 119: 1643–63.
- Wulff, Stefanie and Stefan Th. Gries. 2011. "The genitive alternation in German and Chinese L2 English". Paper presented at the American Association for Corpus Linguistics (AACL) 2011 Conference, Georgia State University, 8 October 2011.
- Zipp, Lena and Tobias Bernaisch. 2012. "Particle verbs across first and second language varieties of English". In Hundt and Gut, eds. 2012: 167–96.

Authors' addresses

Tobias Bernaisch
Department of English
Justus Liebig University Giessen
Otto-Behaghel-Str. 10C
35394 Giessen
Germany

Joybrato Mukherjee
Department of English
Justus Liebig University Giessen
Otto-Behaghel-Str. 10B
35394 Giessen
Germany

Tobias.J.Bernaisch@anglistik.uni-giessen.de

Mukherjee@uni-giessen.de

Stefan Th. Gries
Department of Linguistics
University of California, Santa Barbara
Santa Barbara, CA 93106-3100
United States of America

stgries@linguistics.ucsb.edu