John Benjamins Publishing Company

Exploring epicentres empirically

Focus on South Asian Englishes

Stefan Th. Gries and Tobias Bernaisch

University of California, Santa Barbara / Justus Liebig University Giessen

The present paper studies the dative alternation with GIVE, i.e. the alternation between the double-object construction (e.g. *John gave Mary a book*) and the prepositional dative (e.g. *John gave a book to Mary*), in relation to the norms underlying this constructional choice in six South Asian Englishes. Via *Multifactorial Prediction and Deviation Analysis with Regression* (MuPDAR) including random effects, we identify (i) factors triggering different constructional choices in South Asian Englishes in comparison to British English and (ii) the linguistic epicentre of English in South Asia with regard to the dative alternation. We are able to show that discourse accessibility of patient and recipient as well as pronominality of recipient are actuators of structural nativisation in South Asian Englishes and — in agreement with a more general sociolinguistic approach — find via a bottom-up approach that Indian English may be regarded as the linguistic epicentre of English for South Asia.

Keywords: South Asian Englishes, British English, Indian English, epicentre, dative alternation, MuPDAR (regression modelling)

1. Introduction

Research into the structures of South Asian Englishes (SAEs) has so far largely focussed on Indian English (IndE) for historical, demographic and data-related reasons. With the arrival of British settlers, the English language took root in India towards the end of the 16th century. In the centuries to follow, English was functionally and structurally expanded and became an integral part of the Indian linguistic scenery (see Mukherjee 2007: 164–171). After the demise of the British Raj in 1947 and the subsequent Indian independence, Pakistan split from India in the very same year and, in turn, Bangladesh from Pakistan in 1971, thus founding two new and younger SAEs with a common IndE ancestry.

In Sri Lanka, English arrived approximately two centuries later than in India when "the British East India Company annexed the Maritime Provinces of Ceylon in 1796" (Gunesekera 2005: 11). Despite historical similarities in the status and functions of English between India and Sri Lanka as evident from e.g. culture-laden, western-oriented colonial language policies in the form of Macaulay's ([1835] 1965) *Minute on Indian Education* and the Colebrooke Report of 1831/32 in Sri Lanka (see Yogasundram 2008: 238), the history of Sri Lankan English is unique among the SAEs. Sri Lanka is characterised by a long period of British occupation, but does not have an Indian legacy like the other countries which formed part of the British Raj.[1] Nepali English and Maldivian English should — despite strong ties with the British Empire for largely economic reasons — be considered peripheral exemplars of postcolonial Englishes in Schneider's (2007) sense given that they were at no point part of the British Raj. Consequently, the English language in both countries cannot be considered to be as institutionalised as in Bangladesh, India, Pakistan and Sri Lanka today. Given the historical facts that IndE is (i) the oldest South Asian English and (ii) the ancestor of two other South Asian varieties, IndE — the largest second-language variety world-wide (see e.g. Tully 1997) — has attracted more linguistic attention than other SAEs.

This interest in IndE is also reflected in the corpus data currently available for its description, e.g. the Kolhapur Corpus (see Shastri, Patilkulkarni and Shastri 1986) or the Indian component of the International Corpus of English. The South Asian Varieties of English (SAVE) Corpus (see Bernaisch et al. 2011), the empirical database of this study, features 18 million words of English newspaper texts from Bangladesh, India, the Maldives, Nepal, Pakistan and Sri Lanka and — with the Sri Lankan component of the International Corpus of English still in the making — is so far the only large-scale dataset for the study of SAEs other than IndE. Accordingly, the structures of IndE have already been described much more systematically than those of the remaining SAEs (see e.g. Sedlatschek (2009) on IndE lexis, morphosyntax, lexicogrammar and syntax, Schilk (2011) on its verb-complementational profiles and Lange (2012) on its non-canonical syntax and focus structures). Each of these corpus-based studies provides empirical evidence of nativised IndE forms located on different structural levels. However, once these structurally nativised elements have been described, the next question is which structural and speaker-related factors can be held accountable for the structural intricacies of present-day IndE. Thus, one main focus of this study will be the identification of factors significantly influencing (differences in) constructional choices with the dative alternation in SAEs as opposed to British English (BrE).

---

1. For a more detailed description of the development of English in India and Sri Lanka, see Mukherjee (2007) and Bernaisch (2015), respectively.

In the light of its historical, demographic and (partly resulting socio-) linguistic prominence, it is hardly surprising that IndE has been modelled as a linguistic epicentre of English for South Asia (see Leitner 1992: 225). India is also South Asia's geographical centre making it a direct neighbour to many other South Asian countries and thus rendering IndE an ideal candidate for a model for the surrounding varieties. In the tradition of English as a pluricentric language (see Kloss 1978: 66–67), the notion of a linguistic epicentre can be described as follows:

> [t]he consensus definition of what an epicentre is so far involves two dimensions: a variety can be regarded as a potential epicentre if it shows endonormative stabilization (i.e. widespread use, general acceptance and codification of the local norms of English) […] on the one hand, and the potential to serve as a model of English for (neighbouring?) countries on the other hand.[2] (Hundt 2013: 185)

However, to what extent do these two dimensions of epicentres apply to IndE? With regard to the former dimension, endonormative stabilisation, Mukherjee (2010: 220) posits for IndE that "[w]ith regard to all categories, ranging from the historical and political situation to sociolinguistic and linguistic aspects, […] the situation today represents a typical case of endonormative stabilisation". The by now relatively large body of academic writing on the structures of IndE including the highly influential work by Kachru (1983, 2005), the predominantly introspective usage guides devised by Nihalani et al. (1979, 2004), and more recent empirical studies (see e.g. Sedlatschek 2009; Schilk 2011; Lange 2012) clearly depict IndE as a distinct linguistic organism with its own variety-specific norms. Investigations of the usage of IndE highlighting its cross-domain pervasiveness (see Hohenthal 2003) and attitudinal studies indicating that Indian speakers have a positive attitude towards their local variety of English (see Bernaisch and Koch fc.) support — in conjunction with said structural observations — Mukherjee's (2010) assessment of IndE as an endonormatively stabilised variety.

Still, with regard to the latter dimension, the potential of IndE serving as a structural model for neighbouring SAEs, research has so far remained rather inconclusive. It is true that some studies on SAEs (see e.g. Hoffmann, Hundt and Mukherjee 2011; Bernaisch and Lange 2012; Hundt, Hoffmann and Mukherjee 2012; Koch and Bernaisch 2013) have incorporated the topic of IndE as a possible epicentre of English for South Asia, but they have mainly inferred interpretations "from degrees of similarity between a specific dominant variety on the one hand (i.e. BrE or Indian English) and peripheral varieties on the other (e.g. Sri Lankan

---

2. In their discussions of epicentre identification, Leitner (1992) and Pakir (2001) put more emphasis on the fact that epicentres must have developed their own norms as opposed to their norm-providing potential for other varieties.

English and Pakistani English)" (Hoffmann, Hundt and Mukherjee 2011:261) on the basis of (statistically significant) higher or lower frequencies of a structural feature. These studies depict feature-related end-products of a potential epicentral configuration in SAEs, i.e. structural similarities and/or differences that may have emerged via the varying degrees with which individual SAEs adopt an assumed epicentre (established on sociolinguistic rather than quantitative structural grounds) as a structural model for a given feature. However, the relatively limited number of studies has not yet produced an empirically reliable picture as to which variety could be considered a model for SAEs. In the course of this paper, we will therefore explore the potential of six SAEs to serve as linguistic epicentres of English in South Asia.

From research on focus marking with *itself* in SAEs (see Bernaisch and Lange 2012), there are indications that IndE assumes the role of an exonormative model for other SAEs, but the authors of the study refrain from formulating such a conclusion on the basis of the synchronic nature of their data and the focus on one structural feature only. Hoffmann, Hundt and Mukherjee (2011:277) are also careful for similar reasons when interpreting their results for light-verb constructions (e.g. GIVE sb. *a kiss*) in SAEs and conclude that "we would have to look at many more forms and structures in the lexicogrammar of [SAEs] in order to be able to assess the epicentre status of IndE on a solid basis".

Still, there are also structural investigations into SAEs which "did not find clear-cut evidence of IndE being an epicentre for SAEs" (Hundt, Hoffmann and Mukherjee 2012:162) or where other SAEs show a higher degree of structural innovation than IndE. This is the case with new ditransitives, i.e. verbs used in double-object constructions in New Englishes, but not in their historical input variety BrE, where, in comparison to IndE, data from Pakistan, Nepal and Sri Lanka display more new ditransitives (see Koch and Bernaisch 2013:76). Thus, Hundt (2013:186) is certainly correct in summarising that the present state of research on the epicentre hypothesis in English is characterised by the lack of "empirical evidence that would allow us to make more than educated guesses as to the role of AusE [Australian English] and NZE [New Zealand English] or, in fact, any other variety as new epicentres".

The underlying issue that may have so far impeded the empirical identification of linguistic epicentres in various regions around the globe is the seemingly exclusive focus on surface structures and their degree of similarity between an assumed epicentre and the varieties in its close physical proximity. Admittedly, cross-varietal structural similarities or differences can, as elaborated above, certainly be the results of epicentral configurations — but they do not need to be. Particularly with SAEs, there are a number of possible sources of structural convergence across national boundaries such as shared first languages (e.g. varieties of Tamil in the

South of India and the North of Sri Lanka), shared language families (e.g. Indo-Aryan languages throughout South Asia) or shared contact with varieties outside Asia (e.g. American or BrE). Thus, structural convergence could appear like the outcome of an epicentral configuration when its origins may in fact be quite different. Consequently, for empirical investigations of the epicentre theory, results of surface-structure analyses in isolation are relatively unreliable indicators, since they are highly likely to also mirror (distorting) effects of factors other than epicentral constellations.

So what are alternatives to investigating linguistic epicentres if not via relying solely on the surface structures of the varieties concerned? There seems to be consensus that epicentres have two characteristics: (i) they are endonormatively stabilised and (ii) they can function as a model for other varieties (possibly) in their vicinity (see Hundt 2013: 185). To put it differently, if (i) the structural realisations of a variety are governed by indigenous norms and (ii) the degree of similarity between the norms of this endonormatively stabilised variety and the norms of the other varieties in the region is high, it is likely that this variety functions as a regional linguistic epicentre. Consequently, it may be fruitful to directly analyse and compare their underlying norms, the results of which are similarities and/or differences in surface-structure choices.

There are different approaches towards norms. One way would be to explore the similarity of relative frequencies of surface structures, but structural choices can be equally frequent for very different reasons. For instance, the double-object construction is likely to occur when (i) the recipient is pronominal (as in *He gave her a book*) or (ii) the recipient is non-pronominal, the patient longer than three words, the semantics of the patient abstract and the recipient shorter than or equal to 5 words (as in *He gave his daughter the freedom to come home late*; see Bernaisch, Gries and Mukherjee 2014: 17). An alternative way is, therefore, required that can take linguistic and contextual characteristics of situations in which speakers take linguistic choices into consideration. This speaker-related probabilistic knowledge of structures is best reflected in probabilistic constellations of linguistic and contextual characteristics. In our study, we therefore understand the notion of norm as the complex probabilistic knowledge that makes speakers use certain linguistic patterns. Consequently, the study of these norms seems the most appropriate means to (i) tap into the question of how different SAEs differ from each other and (ii) identify epicentral constellations in SAEs.

In the present paper, we explore a novel approach to the identification of norm orientations and eventually epicentres in SAEs, which should also be applicable to other second- and first-language varieties in allegedly epicentral constellations as is the case e.g. in Southeast Asia or the Pacific region where Singapore and Australian English have been modelled as epicentres of English respectively (see

Leitner 1992). In keeping with Meyerhoff and Niedzielski's (2003: 544) "norms for good practice" in cross-varietal studies, we have chosen a structural object of investigation for which we have reliable evidence that "the variable under investigation is not only superficially the same but also that the same linguistic constraints regulate their use" (Hundt 2013: 184), namely the dative alternation with GIVE, i.e. the alternation between the double-object construction as in *John gave Mary a book* and the prepositional dative *John gave a book to Mary*. For SAEs (also in comparison to BrE), it has been shown that (i) pronominality, constituent length and semantic class are cross-varietally stable factors in predicting constructional choices with the dative alternation with GIVE and (ii) variety-specificity has a subtle, but nevertheless significant effect on the choice of one construction over the other (see Bernaisch, Gries and Mukherjee 2014).

In sum, the present paper first addresses the question of how exactly the dative alternation in SAEs differs from the historical input variety BrE, and then extends previous work in two ways: on the one hand, instead of adopting the null hypothesis that a corpus is sufficiently homogeneous internally to disregard its hierarchical compilation structure (e.g. files nested into registers nested into modes etc.), we take the hierarchical structure of the corpus being studied into consideration (see Gries 2006; Gries and Deshors 2015) by accounting for variation in the data that is due to the different newspapers that make up the corpus. On the other hand, we are applying a new quantitative method of *Multifactorial Prediction and Deviation Analysis with Regression* (MuPDAR), which was originally developed for learner corpus research, to the study of New Englishes in comparison to their historical input variety. Further, we demonstrate how the MuPDAR approach can be extended to corpus-based variety research to iteratively determine the best candidate for epicentre status of a set of varieties in a rigorous and bottom-up fashion, namely by identifying the South Asian English model displaying the highest degree of similarity with those of the remaining SAEs.

The remainder of the paper is structured as follows: In Section 2, we present (i) the corpus data, (ii) the coding that has been applied to capture the structural realisations of the dative alternation and the factors potentially influencing these realisations and (iii) how the MuPDAR approach can be applied in the study of postcolonial Englishes. In Section 3, we provide the results of the MuPDAR analysis by isolating factors where SAEs systematically deviate from BrE choices when all structural and contextual factors are considered at the same time. The empirical identification of a probable epicentre for SAEs is documented in Section 4, and Section 5 offers a summary of the results, a discussion and some concluding remarks.

2.   Methodology

2.1   The corpus data

The SAVE Corpus comprising 3 million words of newspaper English from Bangladesh, India, the Maldives, Nepal, Pakistan and Sri Lanka, respectively, is the empirical basis of the present study.[3] For each national component of SAVE, texts from two local English-medium newspapers were collected and news agency reports subsequently removed systematically to ensure a high degree of variety-specificity of the data (see Bernaisch et al. 2011: 3). Given that the newspaper articles selected were produced and at later stages edited by highly proficient, acrolectal users of the varieties concerned, the structural characteristics evident from analyses in SAVE cannot be dismissed as learner mistakes, but should be viewed as features of the varieties under scrutiny. Given the lack of reference works (e.g. dictionaries, grammars, etc.) and empirical databases for the individual SAEs (with Indian and Sri Lankan English as the only major exceptions), the texts in SAVE are particularly relevant for the study of epicentres of English in South Asia since local newspaper language may provide means of linguistic orientation for speakers and — on a more abstract level — standardisation for so far largely uncodified varieties. Consequently, SAVE, the structure of which is presented in Table 1, is ideally geared towards the study of norm orientation in SAEs, although it is understood that newspaper language — independent of how central this genre may be for principally undocumented varieties — is a first, register-specific approach that will need to be complemented with studies of texts from a wider range of genres for the individual SAEs.

Table 1. The SAVE Corpus (see 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

Table 1. (*continued*)

Country	Newspaper	URL	Time span
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

From each national component of SAVE, 500 examples of the verb GIVE were randomly extracted.[4] To retrieve comparable BrE reference data, we used the daily news section of the British National Corpus and created two randomly sampled and non-overlapping sets of 250 instances of GIVE (BNC news I/II). Out of the 3500 instances, 1381 could be coded on the basis of all or most of the following variables for our subsequent analysis, which were chosen due to their relevance in earlier multifactorial studies on the dative alternation with GIVE in general (see e.g. Gries 2003; Bresnan and Hay 2008) and in SAEs in particular (see Bernaisch, Gries and Mukherjee 2014):[5]

- Variety: the variety from which the example was taken;
- Paper: the newspaper from which the example was taken;
- RecLength and PatLength; the length of recipient and patient in graphemic words, which were then converted into the LogLenDiff, the logged difference RecLength-PatLength;
- RecAnimacy and PatAnimacy: the animacy of recipient and patient (*animate* vs. *inanimate*);
- RecAccessibility and PatAccessibility: the accessibility of recipient and patient in terms of whether the recipient or patient concerned was mentioned in the preceding ten lines (*given* vs. *new*);
- RecPronominality and PatPronominality: the realisation of recipient and patient in pronominal or non-pronominal form (*pronoun* vs. *np*);

---

4. The respective word forms of GIVE, i.e. *gave*, *give*, *given*, *gives*, *giving* (excluding non-verbal usages) were extracted and the samples reduced to 500 instances via random sampling in WordSmith Tools (Version 4.0; Scott 1998).

5. To ensure a certain degree of unanimity of the coding of the accessibility-related variables, a second coder coded a small sample of the corpus examples with regard to RecAccessibility and PatAccessibility; all instances but one were coded identically, indicating a high degree of interrater reliability.

- – PatSemantics: the semantic class of GIVE in accordance with the patient (*abstract* as in "give him happiness", *concrete* as in "give him my keys" or *informational* as in "give him a message");
- – Transitivity: the verb-complementational pattern of the verb: *ditransitive* (1) vs. *prepositional dative* (2).[6]

  (1) A clean sweep may give us some consolation for their last defeat in Zimbabwe. <SAVE-BAN-NA_2006-12-08>

  (2) The police force gave gun salute to the departed leader before his son Bijoyshree lit the funeral pyre. <SAVE-IND-SM_2004-10-05>

2.2 Applying the MuPDAR approach to SAEs

A lot of (earlier) studies on varieties of English may be characterised as involving relatively basic statistical analyses. In these studies, variety-specific structural features may be discovered and described for the first time, which means that factors influencing the usage of these features still need to be delineated. This limits the applicability of more complex statistical models at this early stage of description. While simpler statistical tools are still quite common in much of corpus linguistics, there is now a growing recognition in corpus research that once influential features guiding structural choices are identified and multifactorial approaches are thus feasible, more complex methods must be employed to adequately model these structural features. It is especially in learner corpus research and in variety research that more advanced statistical tools are now becoming more frequent, such as multifactorial regression modelling, classification trees and other multivariate methods. For instance, in learner corpus research, studies such as Gries and Wulff (2013) and Deshors and Gries (2014) have (i) demonstrated how likely it is that monofactorial studies overlook important patterns in the data and (ii) exemplified the power of regression analyses that feature the speakers' L1s as a(n interacting) predictor. The situation is similar in variety research, where more studies are now adopting multifactorial regressions as the main analytical tool (see Schilk et al. 2013; Bernaisch, Gries and Mukherjee 2014, and others).

This move towards more advanced methods has led to a welcome wealth of new results. However, a new methodological development in these areas attempts to provide yet another interesting perspective on the kinds of data that learner corpus research often deals with, namely data where one subset of the data constitutes, so to speak, a default reference level or a target; in many cases of learner

---

6. We chose Mukherjee (2005) as the reference for the coding of the verb-complementational patterns.

corpus research, for instance, BrE can be seen as a target that foreign language learners may ultimately aspire to, while in variety research, BrE is also often taken as a central point of comparison to describe to what extent a (regional) variety has developed recurrent structural patterns different from those of the historical input variety. This new method is called MuPDAR,[7] and its central focus is to attempt to answer two questions:

– In a particular (linguistic) context where a speaker of a variety other than the default/target/input variety is making a linguistic choice, what would a speaker of the default/target/input variety do? For regional variety corpus research that would mean: in the context that a speaker of a non-British variety is in when making a lexical and/or grammatical choice, what would a BrE speaker choose?
– What factors are responsible for the non-BrE speaker making choices other than what is predicted for the BrE speaker?

MuPDAR involves the following three steps:

– Fitting a regression R_1 that predicts the choices that speakers of the target/reference level (typically, native speakers of the reference variety) make with regard to the phenomenon in question;
– applying the results of R_1 to the other speakers in the data (typically, learners or speakers of institutionalised second-language varieties) to predict for each of their data points what the native speaker of the reference variety would have done in their place;
– fitting a regression R_2 that explores how the other speakers' choices differ from those of the speakers of the target/reference variety.

In this study, we propose to extend this method from learner corpus research to the comparison of native speakers of a language L and current speakers of different regional varieties of L at different levels of emancipation from a reference variety of L. Thus, the two questions relevant to the study of SAEs are

– In the (linguistic) context that a speaker of, say, IndE is in when making a linguistic choice, what would a native speaker of BrE do?
– What factors are responsible for the variety speaker of, say, IndE making choices other than what is predicted for the native speaker of the historical source variety?

---

7. This method was developed in Gries and Adelman (2014), Gries and Deshors (2014), and Wulff and Gries (2015) for subject realisation in Japanese, *may* vs. *can* by native speakers and French/Chinese learners of BrE, and prenominal adjective order by native speakers and German/Chinese learners of BrE respectively.

In this paper, we applied the MuPDAR approach precisely as discussed above. We first fit a logistic regression R_1 that predicted the constructional choices of the reference level (BrE) on the basis of all predictors and their pairwise interactions as described in the previous section. Then, we determined whether the regression results in a good fit, i.e. a good classification accuracy, which we explored by computing the classification accuracy as well as an R^2 value.

Secondly, we applied the regression equation from R_1 to the data points of all other varieties. Thus, we obtained, for every choice that a user of these other varieties made, a prediction of what the BrE user would have chosen, essentially answering the question "in this situation, what would a user of the reference variety have produced?" This allowed us to create a variable VarietySpecificity stating for every data point whether the user of the other varieties made the predicted choice.

Thirdly, we performed a model selection process with regression R_2, which had the binary variable VarietySpecificity as its dependent variable, i.e. we tried to determine which factors give rise to South Asian users not making the same choice as the reference variety of BrE, or, in other words, which factors eventually give rise to structural nativisation of the dative alternation in SAEs. Given the current data, this process yielded several problems with collinearity (high Variance Inflation Factors [*VIF*s]), which ruled out any automatic model selection or backwards model selection. We therefore undertook a forward model selection process: Our first model only predicted VarietySpecificity on the basis of all main effects, and then we identified in six subsequent steps the addition of which predictors would make the model most significantly better (using likelihood ratio tests and the Akaike Information Criterion) without inflating the model's *VIF*s. This process was terminated when no predictor could be added to make the model significantly better. For the final model, we then computed its classification accuracy, C value and R^2 to assess its quality of fit, and visualised the significant predictors' effects.

A final important point to mention is that we implement another methodological step that most corpus research does not undertake (but see Gries 2015, and Gries and Deshors 2015). This step involves using regression techniques that take the hierarchical structure of corpus data into consideration. Many corpora come with a hierarchical structure such that *files* (e.g. S1A-001 in the International Corpus of English) are nested into *sub-registers* (e.g. private spoken dialogue), which are nested into *registers* (e.g. spoken dialogue), which are nested into *modes* (e.g. spoken). This means that each of these hierarchical levels of corpus organisation accounts for some of the variability in the data and, therefore, also makes the data violate the assumption of the independence of data points of most statistical tests. In our data, we have a less complex hierarchical structure, but still some:

each South Asian variety is represented by data from two newspapers, which means that *newspaper* is nested into *variety*, and we model this accordingly. This is not just playing with ever more advanced methods — what this achieves is two-fold: First, all regression results will be more precise because variability that is due to the corpus structure will be partialled out. Second, it allows us to see whether the implicit assumption of much variety research — much corpus research in fact — namely that the corpus is homogeneous enough to allow for meaningful comparisons is justified or not (see Gries 2006 for an early discussion of this). Thus, in R_1, we include the two components of BNC news I and II as a random effect (for reasons of sample size, we are only including varying intercepts for these), while in R_2, we include newspaper nested into varieties as random effects (again, just as varying intercepts).[8]

3.   Case study 1: Identifying factors of structural nativisation in SAEs

3.1 Regression R_1 on the BrE data

The results of R_1 fit on the BrE data were very promising: The model achieved a classification accuracy of 93.7 per cent, which is significantly higher than one would attain by choosing constructions randomly ($p_{\text{binomial}} < 10^{-18}$). Reassuringly, each construction was more often predicted correctly than incorrectly, which justified proceeding further.

3.2 The application of R_1 to the SAE data

Applying the BrE regression equation to the situations the users of the other varieties were in also yielded a relatively good prediction accuracy. Specifically, 77 per cent of the choices of the other varieties' users were predicted correctly, which is again significantly better than the random baseline ($p_{\text{binomial}} < 10^{-60}$). Finally, for each SAE data point, we determined how much the SAE user's choice differed from the predicted BrE choice and collected those deviations in a variable called VARIETYSPECIFICITY:

–   If the SAE user made the choice predicted for a BrE user, VARIETYSPECIFICITY was set to 0;

---

8. We did not include varying intercepts for file names because there were too many files that contributed only very small numbers of data points. Because of this skewed distribution, an attempt at modelling these data resulted in convergence warnings which, if ignored, led to the same model as if file names were not included.

–   if the SAE user did not make the choice predicted for a BrE user,
    VARIETYSPECIFICITY was set to 0.5 minus the predicted probability of the
    prepositional dative.

That means, if VARIETYSPECIFICITY is > 0, the SAE user chose a prepositional dative
although a BrE user would have chosen a ditransitive, and if VARIETYSPECIFICITY
is < 0, the SAE user chose a ditransitive although a BrE user would have chosen a
prepositional dative. Thus, VARIETYSPECIFICITY reveals two things: (i) whether an
SAE user made the choice a BrE user would have made (by the sign) and (ii) if not,
how much the SAE user's choice deviated from that of the BrE user (by the size of
the deviation from 0).

   For example, a Bangladeshi English (BanE) user produced the sentence shown
in (3) with a ditransitive and, according to R_1, a BrE user would have, too; thus,
VARIETYSPECIFICITY for this example is 0. On the other hand, a BanE user pro-
duced the sentence shown in (4) with a prepositional dative, where, according to
R_1, a BrE user would have produced a ditransitive structure instead, "gave Jammu
and Kashmir a special status", presumably because both patient and recipient are
three words long, but the recipient is definite whereas the patient is indefinite.
Thus, VARIETYSPECIFICITY for this example is > 0 (0.4495, to be precise).

(3)   An early strike by Hanif gave Bangladesh Sporting full three points in the
      Pioneer Football Super League […]. <SAVE-BAN-DS_2006–10_pt2>

(4)   […] agenda of building a temple at Ayodhya and of deleting Article 370 that
      gave a special status to Jammu and Kashmir. <SAVE-BAN-DS_2005–05_
      pt13>

3.3 Regression R_2 on the SAE data

The results of R_2, i.e. a regression that tried to predict whether users of SAEs would
make the choices that BrE users would be predicted to make or not, were again
very good. The final model does not suffer from overdispersion ($p > 0.92$), exhibit-
ed very little collinearity problems (one $VIF \approx 5.4$, one $VIF \approx 3.2$, all others < 3) and
was highly significant (Likelihood Ratio chi-squared = 251.3, df = 14, $p < 0.001$).
It achieved a classification accuracy of 77.2 per cent, which is again significantly
higher than the random-choice baseline ($p < 10^{-8}$). $R^2_{marginal}$ and $R^2_{conditional}$ were
0.56 and 0.563, respectively. We will now discuss three significant highest-level
predictors of R_2 (space precludes an exhaustive discussion of all findings): first the
fixed effects, then the varying intercepts.

3.4 Discussion

3.4.1   *The fixed-effects structure of R_2: selected interactions*

Figure 1 represents two significant interactions, PatAccessibility:LogLenDiff and RecAccessibility:LogLenDiff. The *x*-axis represents LogLenDiff (slightly jittered horizontally) while the *y*-axis represents VarietySpecificity (slightly jittered vertically).[9] Each point — with colours representing the levels of Pat/RecAccessibility — represents an SAE choice, the two ×'s represent the bivariate means of the red and blue points. The left panel shows that

– SAE users typically make the BrE choice (see the many points at $y=0$) and are particularly likely to make BrE choices when LogLenDiff differs strongly from 0 (when abs(LogLenDiff) > 2, there are few non-BrE-like choices) and most non-BrE-like choices occur at $x=0$;
– when the patient is given, SAE users make about equally non-BrE-like choices of prepositional dative or ditransitives (the red × is close to $y=0$) whereas when the patient is new (i.e. BrE users prefer ditransitives), they are much more likely to choose non-BrE-like prepositional datives (the blue × has a high *y*-value). In other words, compared to BrE users, the strength of the cue 'new patient' for ditransitives is markedly lower for SAEs.

The right panel shows the corresponding findings for RecAccessibility: with given recipients, there is no strong pattern in how the SAE users make non-BrE choices, but with new recipients, the SAE users choose more non-BrE-like prepositional datives. This is interesting because, in BrE, new recipients prefer prepositional datives. However, it seems as if the strength of the cue 'new recipient' for prepositional datives is notably stronger for SAE users: they use them in contexts where BrE users might use a ditransitive even with a new recipient (because of other characteristics of that verb phrase). A case in point is shown in (5): The recipient ("fascists like Falwell") is new and the BanE user produced a prepositional dative, but a BrE speaker is predicted to use a ditransitive instead.

(5)   Of course, I am not proposing that we should give the valuable space of
      Daily Star to fascists like Falwell. <SAVE-BAN-DS_2003–06_pt29>

---

9. *Jittering* refers to adding small amounts of random noise to the data points to avoid overplotting in graphs.
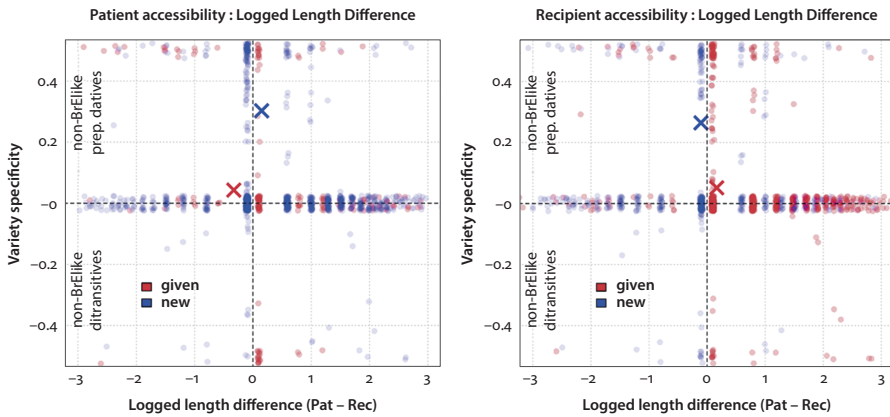
Figure 1. The interactions of PATACCESSIBILITY:LOGLENDIFF and RECACCESSIBILITY: LOGLENDIFF

The last interaction to discuss is shown in Figure 2. Here, the *x*-axis represents the four combinations of two levels of PATACCESSIBILITY (G = given, N = new) and two levels of RECPRONOMINALITY (P = pronominal, NP = non-pronominal form); the black ×'s represent the column-wise means. The plot shows that SAE norms are, in some areas, compatible with the information-structural conditions of the dative alternation in BrE: when the recipient is pronominal, SAE speakers' choices are with very few exceptions BrE-like and particularly so when the patient is new, because then both cues strongly favour ditransitives. However, variety-specificity
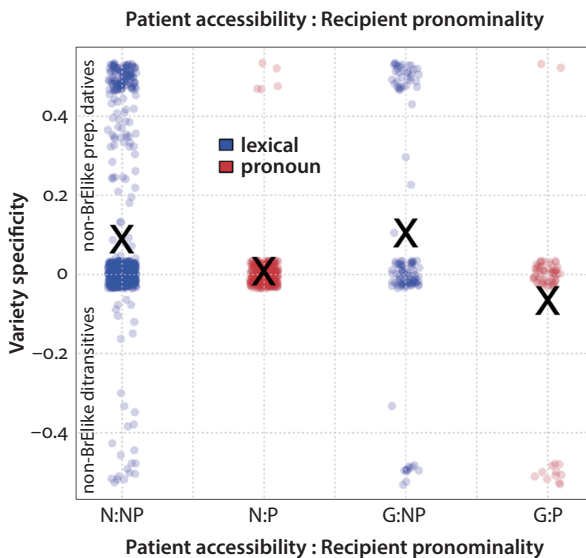

Figure 2. The interaction of PATACCESSIBILITY:RECPRONOMINALITY

of SAEs becomes more evident with non-pronominal recipients. These would typically, but not deterministically, favour prepositional datives, but the SAE users make proportionally more choices of prepositional datives that BrE users would not make (again, because of other characteristics of the verb phrase). Thus, we again have a case where SAE users assign more weight to the strength of a cue in their constructional choices than users of BrE.

Example (6) is a case in point. The recipient is animate and non-pronominal and the patient is new (in this light-verb use) but while this IndE user chose a prepositional dative, a BrE user is predicted to choose a ditransitive, "… who gave Sir William Jones a moral boost".

(6)   It was Warren Hastings who gave a moral boost to Sir William Jones […]
      <SAVE-IND-SM_2004-02-03>.

3.4.2   *The random-effects structure of R_2*

An important aspect of the data concerns the random-effects structure of the regression equation. Given the small sample size, we only included varying intercepts for Variety/Paper in the model, i.e. we allowed both varieties and papers within varieties to vary in terms of their baseline of users making BrE-like choices. While it would be interesting to also include varying slopes in the analysis (for instance, to allow the effects of regression predictors to vary for different files or speakers), this awaits larger data sets. However, given the fact that only very few corpus studies ever include multi-level effects, it is still a reasonable first step. In addition, the results show it is a very important one. Consider Figure 3, which
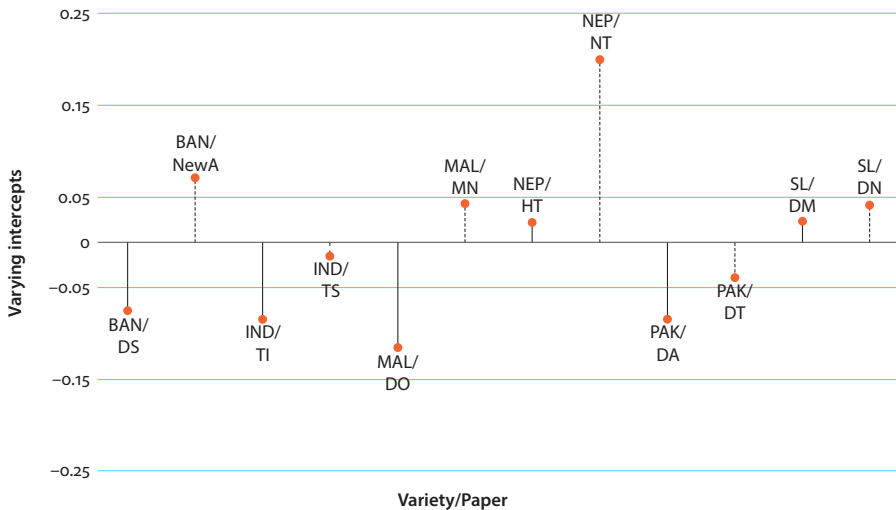

Figure 3. Varying intercepts for Variety/Paper (abbreviations as in Table 1)

shows for each newspaper and the variety it represents the intercept adjustments on the *y*-axis. Positive/negative adjustments signal that a particular Variety/ Paper was characterised by more/less BrE-like constructional choices.

This simple graph has a very important implication regarding the within- and between-homogeneity of corpora (see Gries 2006). The results show that, with regard to the dative alternation with GIVE at least, some of the SAVE components are much more homogeneous than others. For instance, for India, Pakistan and Sri Lanka, the adjustments for both newspapers go in the same direction, meaning that, given the overall trend in the data, both newspapers in each of these varieties behave alike and thus compare well. This is the null hypothesis of most such studies — that each corpus (part) is internally homogeneous (enough) to justify its comparison to another whole corpus. However, as demonstrated comprehensively by Gries (2006), this null hypothesis may also not hold, and if it does not and the analysis neither checks whether it holds or statistically controls for it by including multi-level structure, then results can be very misleading.

In our study, the three other varieties represent such cases. For Bangladesh and the Maldives, we find that the two newspapers constituting each corpus part in fact behave very differently from each other — one much more like the BrE predictions (the positive adjustments for BAN/NEWA and MAL/MN) and one much less like the BrE predictions (the negative adjustments for BAN/DS and MAL/ DO). For Nepal, both adjustments are positive, but one of them is the smallest positive adjustment (NEP/HT), the other is the largest (NEP/NT). Studies that pretend that each variety corpus is one homogeneous lump of data without testing whether that is the case, or controlling for it, run the risk of making false generalisations because of the variability in the data — data that seem to reflect what happens in one variety may in fact just be data about a particular newspaper (and its editorial policies?). It is true that, in the present case, the adjustments are relatively small so the damage from this would be negligible, but, again, this can only be known after the right type of analysis has been undertaken (see Gries and Deshors 2015 for conceptually similar implications). Thus, it seems reasonable for future (variety) corpus research to take the hierarchical level of its data into consideration, and multi-level modelling of the kind presented here is one way of doing this.

In sum, we have shown that the combination of the MuPDAR approach with multi-level modelling provides very precise and very fine-grained results, especially when compared to studies that do not (yet) account for multifactorial or multi-level effects. Hopefully, the present work will be an incentive for more applications of these methods.

4.   Case study 2: Exploring epicentres in SAEs

In this second case study, we are making a methodological proposal with regard to how epicentres can be identified empirically. Epicentres have "the potential to serve as a model of English for (neighbouring?) countries" (Hundt 2013:185). If we operationalise the notion of "model" as a set of norms governing surface structure choices, the degree of similarity of the norms of an assumed epicentre with the norms of the varieties potentially under epicentral influence will be indicative of the extent to which the assumed epicentre can be considered to serve as a model for the varieties potentially under epicentral influence. In other words, the higher the degree of similarity between the norms of the assumed epicentre and the varieties potentially under epicentral influence, the more evidence we have for an epicentral constellation. In a first step, each of the varieties studied here will assume the role of a potential epicentre of English in South Asia to find the variety where the degree of similarity between the norms of the assumed epicentre and the remaining six varieties is highest or, to put it differently, to find the variety which best serves as a model for the other varieties. Once this variety has been detected on an empirical basis, we will revisit earlier research on this variety to establish to what extent it can be considered endonormatively stabilised — the second central criterion for epicentre status (see Hundt 2013:185). In other words, we propose an approach that, in a bottom-up and data-driven way derived from the MuPDAR procedure above, tries to answer the question "with which variety is the degree of similarity between the norms of this variety and the norms of the other varieties (in the region) the highest?" If this turned out to be an endonormatively stabilised variety, it could be argued that this variety may be the linguistic epicentre for SAEs.

4.1   Method

In an iterative procedure, each of the seven varieties — BrE and the six South Asian varieties — was considered the reference level for a first regression R_1. That is to say, for each variety,

–   we fitted a generalised linear mixed-effects regression predicting the choice of one of the two constructions in one variety on the basis of the following predictors: RecLength and PatLength, RecAnimacy and RecPronominality (PatAnimacy and PatPronominality had to be excluded because, in each, one level accounted for more than 99 per cent of the data), PatSemantics, RecAccessibility and PatAccessibility, and varying intercepts for Paper;
–   we used the resulting regression equation (without the varying intercepts for Paper) to predict constructional choices in the other six varieties. We

computed two values: (i) how often in percent a user of one of the other six varieties made the predicted choice; (ii) the vector VARIETYSPECIFICITY as explained above, i.e. how much the user's actual choice was in line with what was predicted to be chosen, and converted it to its absolute value (because we are interested in the sizes of the deviations, not so much their directions); the latter set of values is more precise than the former because they do not just dichotomise the outcome — correct prediction or not — but also indicate how much off a prediction was;

– finally, we computed for each variety (i) the sum of all correctly predicted percentages and (ii) the sum of all deviations.

The results from this procedure allowed us to determine which variety serves best to predict, on average, the choices of the users of the other varieties, which would be the variety/varieties that give/s rise to (i) the largest summed percentage of correct predictions and (ii) the smallest summed deviations.

4.2 Results

We obtained for each variety two values, a sum of all correct predictions and a sum of all deviations when one uses that variety to predict the others. These can be represented in dotcharts as seen in Figure 4.

In spite of the exploratory nature of this proposal, the results are rather convincing. First, in both plots we can see that the BrE data behave quite differently from the South Asian varieties: using the BrE data to predict the dative alternation in the South Asian varieties leads to worse prediction accuracies than if one tried to predict South Asian varieties from any other South Asian variety.

Second, we can see that, jointly, the results appear to point to IndE as the linguistic epicentre for SAEs as far as the dative alternation is concerned. In the upper panel, it scores the second highest value only to be bested narrowly by Pakistani English, but in the more precise resolution of the lower panel, IndE is the variety from which the others are predicted with the smallest summed deviation, followed by Maldivian and then Pakistani English. The implications of these results will be discussed in more detail in Section 5.

In sum, while this particular application was not very broad — we only looked at one constructional alternation and one verb — it was deep: we looked at many predictors of that alternation and even included newspaper-specific variation. The results are encouraging and we hope to inspire further bottom-up exploration of epicentres for different phenomena.
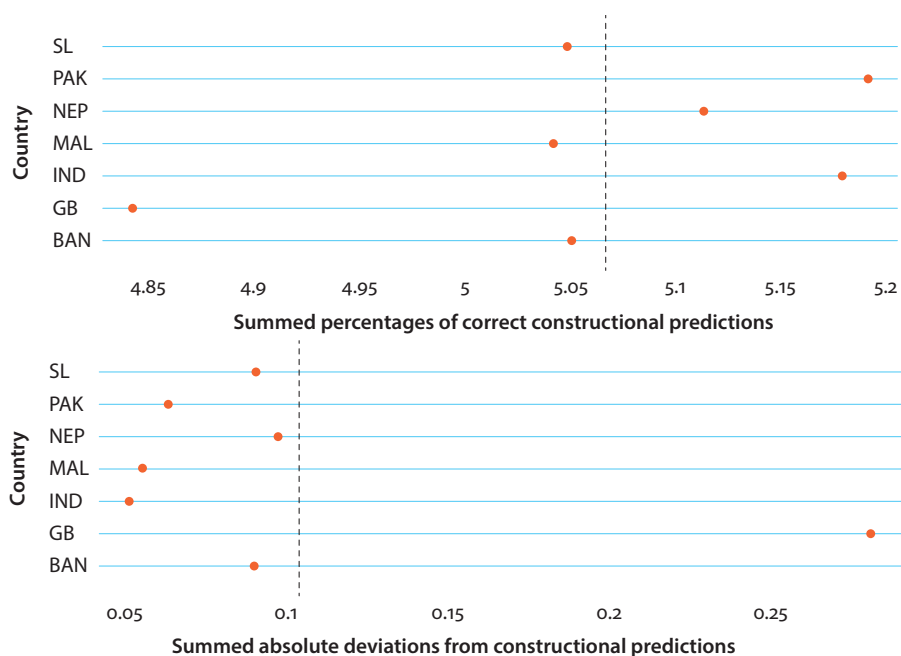
Figure 4. Results of epicentre exploration. Upper panel: sums of correctly-predicted percentages; lower panel: sums of absolute deviations from predictions. Dashed vertical lines represent means

5. Discussion and concluding remarks

In this paper, we empirically explored the epicentre hypothesis in SAEs. We identified the general focus on the surface structures instead of the norms guiding the choice of these surface structures as a potential reason as to why the empirical study of epicentres of English in South Asia and elsewhere has so far remained relatively vague from an empirical angle. In order to delineate these underlying norms, we investigated the dative alternation with GIVE in acrolectal newspaper language representing the varieties concerned. We geared the MuPDAR approach, which had formerly exclusively been used to describe structural deviations of learners from their target variety and factors causing these deviations, to the study of institutionalised second-language varieties in relation to their historical input variety. This was done by comparing the constructional choices users of SAEs make to those of BrE users while at the same time taking other factors (e.g. animacy of recipient/patient, semantic class, etc.) that are potentially influencing the choice into consideration. As a result, we established which and to what extent factors lead users of a South Asian variety to make a choice diverging from that of the BrE user. To put it differently, we identified factors behind structural

nativisation for the dative alternation in SAEs, which, among others, entail (i) discourse accessibility of recipients and patients and (ii) RecPronominality. We also accounted for random effects due to the hierarchical structure of the SAVE Corpus — a methodological step that will hopefully become a standard soon to avoid possibly misleading generalisations rooted in corpus variability.

Discourse accessibility and RecPronominality led to constructional choices in SAEs that are different from BrE. Both factors are relatively closely related in that a patient or recipient is much more likely to be realised as a pronoun in case it is given. Of course, impersonal pronouns such as *one* in (7) constitute a universal exception to this rule: a pronominal realisation is chosen although the abstract referent has not been previously mentioned. With these examples, however, the referents of pronouns are easily retrievable — this is not necessarily the case in SAEs (see Example 8).

(7)   Put simply, computer games are games on computers. But this does not give one a complete idea of what it actually is. <SAVE-IND-SM_2003-04-25>

(8)   I always tell them you know about Sri Lankan women especially the girls [how they are looked after like flowers] and I tell that one. (taken from Herat 2006:74)

In her study on pronominal *one*, Herat (2006:71) shows "that in SLE [= Sri Lankan English] *one* can have missing antecedents, which are not recoverable from the linguistic context", or in other words, that, in addition to e.g. impersonal pronouns, there are more contexts like (8) where a pronoun may be used although discourse accessibility is not given. These observations thus support the identification of discourse accessibility and RecPronominality as actuators of structural nativisation in SAEs. Against this background, it will certainly be fruitful to investigate whether the same actuators of structural nativisation also hold in studies of (i) the dative alternation outside South Asia and (ii) other alternations such as the genitive alternation in South Asia and elsewhere. With studies complementing the varieties examined here with non-South Asian varieties to establish how close or distant they are to the set of South Asian models and the BrE model, we expect varieties outside South Asia to be worse models for SAEs than IndE.

In the second case study, we used each of the seven varieties studied as an epicentral candidate and examined their compatibility with the constructional choices in the remaining varieties. With this approach, the underlying norms of the IndE model for the dative alternation with GIVE were most compatible with the constructional choices of the other varieties.

On the basis of Mukherjee's (2007:170) assessment that

> present-day Indian English can be viewed as a 'phase 4 variety' in Schneider's (2003) model: the process of nativization (in the sense of transplanting English to India and consolidating it in the new environment) is more or less over; the variety is now largely endonormatively stabilized […]

and the close fit between the norms of the dative alternation in varieties outside India and the respective IndE norms, it seems reasonable to profile IndE as the linguistic epicentre of SAEs at least for this alternation. The ensuing question then is: how can linguistic epicentres actually exert influence on the structures of other varieties? Nepali English is an intriguing case here. The history of Nepal does not show British colonisation or missionary activities, which is why Nepal mainly relied on English-language teaching and curricula from India. Nepali people with prestigious jobs in e.g. schools or governmental bodies received their English training in Indian institutions (see Kachru 1994: 548) and upon their return to Nepal functioned as linguistic multipliers of the norms of English they had adopted in India. In the light of this, IndE — and not BrE — should probably be regarded as the historical input variety of Nepali English. This characterises Nepali English as a "postcolonial English squared" (Bernaisch and Lange 2012: 13) in the sense that it was a postcolonial English which served as the historical input variety — a likely explanation for the structural similarities between both varieties today. Considering the central role India has played in the history of the South Asian *Sprachraum*, its demographic key figures (e.g. its number of English speakers) and its (possibly most advanced) sociolinguistic profile among SAEs, the empirical identification of IndE as the linguistic epicentre for SAEs is certainly plausible and in line with (less data-driven) assumptions in earlier publications on the subject.

When speakers under epicentral influence are aware of a certain structure originating in the epicentre, the attitudes these speakers have towards the epicentral variety may also become more relevant for the constructional choices made. For South Asia, it could be shown that Sri Lankan speakers of English show a slightly negative attitude towards IndE (see Bernaisch 2012), which could possibly preempt a potential epicentral influence of markedly IndE structures. Given that the norms underlying the dative alternation, however, probably operate below speakers' linguistic awareness, attitudinal profiles are likely to be of secondary relevance in explaining the results of the present study, while systematic explorations of cross-varietally shared substrate influences may have a higher explanatory power.

We hope to have provided an answer to Hundt's (2013: 184) question as to whether "the data have to be diachronic or whether it might also be possible to verify the epicentral status of a variety on the basis of synchronic evidence only". As we have shown, it is certainly possible to identify linguistic epicentres with synchronic corpus data, but a fuller account would also examine the possible diffusion

of certain structures and their underlying norms from the epicentre to other varieties in its vicinity. Structural diffusion is a diachronic process and thus necessitates diachronic corpus data for institutionalised second-language varieties, which are still largely lacking — the main reason as to why the study of a potential cross-varietal epicentral influence has so far not been sufficiently explored empirically. In a nutshell, synchronic data are sufficient to identify linguistic epicentres, but diachronic data are needed to study their seismic waves.

References

Bernaisch, Tobias. 2012. "Attitudes towards Englishes in Sri Lanka". *World Englishes* 31: 279–291. doi:10.1111/j.1467-971X.2012.01753.x

Bernaisch, Tobias. 2013. *The Lexis and Lexicogrammar of Sri Lankan English: A Corpus-Based Study of the Nativisation of a Link Language*. Amsterdam: John Benjamins.

Bernaisch, Tobias, and Christopher Koch. fc. "Attitudes towards Englishes in India". *World Englishes*.

Bernaisch, Tobias, and Claudia Lange. 2012. "The Typology of Focus Marking in South Asian Englishes". *Indian Linguistics* 73: 1–18.

Bernaisch, Tobias, Christopher Koch, Joybrato Mukherjee, and Marco Schilk. 2011. *Manual for the South Asian Varieties of English (SAVE) Corpus: Compilation, Cleanup Process, and Details on the Individual Components*. Giessen: Justus Liebig University.

Bernaisch, Tobias, Stefan Th. Gries, and Joybrato Mukherjee. 2014. "The Dative Alternation in South Asian English(es): Modelling Predictors and Predicting Prototypes". *English World-Wide* 35: 7–31. doi:10.1075/eww.35.1.02ber

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: 245–259. doi:10.1016/j.lingua.2007.02.007

Deshors, Sandra C., and Stefan Th. Gries. 2014. "A Case for the Multifactorial Assessment of Learner Language: The Uses of *May* and *Can* in French-English Interlanguage". In Dylan Glynn, and Justyna Robinson, eds. *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*. Amsterdam: Benjamins, 179–204. doi:10.1075/hcp.43.07des

Gries, Stefan Th. 2003. "Towards a Corpus-Based Identification of Prototypical Instances of Constructions". *Annual Review of Cognitive Linguistics* 1: 1–27. doi:10.1075/arcl.1.02gri

Gries, Stefan Th. 2006. "Exploring Variability within and between Corpora: Some Methodological Considerations". *Corpora* 1: 109–151. doi:10.3366/cor.2006.1.2.109

Gries, Stefan Th. 2015. "The Most Underused Statistical Method in Corpus Linguistics: Multi-Level (and Mixed-Effects) Models". *Corpora* 10: 95–125.

Gries, Stefan Th., and Allison S. Adelman. 2014. "Subject Realization in Japanese Conversation by Native and Non-Native Speakers: Exemplifying a New Paradigm for Learner Corpus Research". In Jesús Romero-Trillo, ed. *Yearbook of Corpus Linguistics and Pragmatics 2014*. Berlin: Springer, 35–54.

Gries, Stefan Th., and Sandra C. Deshors. 2014. "Using Regressions to Explore Deviations between Corpus Data and a Standard/Target: Two Suggestions". *Corpora* 9: 109–136. doi:10.3366/cor.2014.0053

Gries, Stefan Th., and Sandra C. Deshors 2015. "EFL and/vs. ESL? On How to Properly Bridge the Paradigm Gap". *International Journal of Learner Corpus Research* 1: 130–159 .

Gries, Stefan Th., 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: 327–356. doi:10.1075/ijcl.18.3.04gri

Gunesekera, Manique. 2005. *The Postcolonial Identity of Sri Lankan English*. Colombo: Katha Publishers.

Herat, Manel. 2006. "Substitute *One* in Sri Lankan English". *Leeds Working Papers in Linguistics* 11 <http://www.leeds.ac.uk/arts/info/125154/leeds_working_papers_in_linguistics_and_phonetics/1949/volume_11_2006> (accessed December 18, 2014).

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: 258–280. doi:10.1515/angl.2011.083

Hohenthal, Annika. 2003. "English in India: Loyalty and Attitudes". *Language in India 3*(5) <http://www.languageinindia.com/may2003/annika.html> (accessed December 18, 2014).

Hundt, Marianne. 2013. "The Diversification of English: Old, New and Emerging Epicentres". In Daniel Schreier, and Marianne Hundt, eds. *English as a Contact Language*. Cambridge: Cambridge University Press, 182–203.

Hundt, Marianne, Sebastian Hoffmann, and Joybrato Mukherjee. 2012. "The Hypothetical Subjunctive in South Asian Englishes: Local Developments in the Use of a Global Construction". *English World-Wide* 33: 147–164. doi:10.1075/eww.33.2.02hun

Kachru, Braj B. 1983. *The Indianization of English: The English Language in India*. Delhi: Oxford University Press.

Kachru, Braj B. 1994. "English in South Asia". In Robert Burchfield, ed. 1994. *The Cambridge History of the English Language. Vol 5: English in Britain and Overseas: Origins and Developments*. Cambridge: Cambridge University Press, 497–553.

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

Kloss, Heinz. 1978. *Die Entwicklung neuer germanischer Kultursprachen seit 1800* (2nd ed.). Düsseldorf: Schwann.

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 – Selected Papers from ICAME 32*. Amsterdam: Rodopi, 69–89.

Lange, Claudia. 2012. *The Syntax of Spoken Indian English*. Amsterdam: John Benjamins. doi:10.1075/veaw.g45

Leitner, Gerhard. 1992. "English as a Pluricentric Language". In Michael Clyne, ed. *Pluricentric Languages: Differing Norms in Different Nations*. Berlin: Mouton de Gruyter, 179–237.

Macaulay, Thomas B. [1835] 1965. "Minute on Indian Education". In Henry Sharp, ed. *Selections from Educational Records, Part I: 1781–1839*. Delhi: National Archives of India, 107–117.

Meyerhoff, Miriam, and Nancy Niedzielski. 2003. "The Globalisation of Vernacular Variation". *Journal of Sociolinguistics* 7: 534–555. doi:10.1111/j.1467-9841.2003.00241.x

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

Mukherjee, Joybrato. 2007. "Steady States in the Evolution of New Englishes: Present-Day Indian English as an Equilibrium". *Journal of English Linguistics* 35: 157–187. doi:10.1177/0075424207301888

Mukherjee, Joybrato. 2010. "Corpus-Based Insights into Verb-Complementational Innovations in Indian English: Cases of Nativised Semantico-Structural Analogy". In Alexandra N. Lenz, and Albrecht Plewnia, eds. *Grammar between Norm and Variation*. Frankfurt: Peter Lang, 219–241.

Nihalani, Paroo, Ray K. Tongue, and Priya Hosali. 1979. *Indian and British English: A Handbook of Usage and Pronunciation*. Delhi: Oxford University Press.

Nihalani, Paroo, Ray K. Tongue, Priya Hosali, and Jonathan Crowther. 2004. *Indian and British English: A Handbook of Usage and Pronunciation* (2nd ed.). Delhi: Oxford University Press.

Pakir, Anne. 2001. "The Voices of English-Knowing Bilinguals and the Emergence of New Epicentres". In Vincent Ooi, ed. *Evolving Identities. The English Language in Singapore and Malaysia*. Singapore: Times Academic, 1–11.

Schilk, Marco. 2011. *Structural Nativization in Indian English Lexicogrammar*. Amsterdam: John Benjamins. doi:10.1075/scl.46

Schilk, Marco, 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. doi:10.1515/cllt-2013-0001

Schneider, Edgar W. 2003. "The Dynamics of New Englishes: From Identuty Construction to Dialect Birth." *Language* 79: 233–281.

Schneider, Edgar W. 2007. *Postcolonial English: Varieties Around the World*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511618901

Scott, Mike. 1998. *WordSmith Tools* (Version 4.0). Oxford: Oxford University Press.

Sedlatschek, Andreas. 2009. *Contemporary Indian English: Variation and Change*. Amsterdam: John Benjamins. doi:10.1075/veaw.g38

Shastri, S.V., C.T. Patilkulkarni, and Geeta S. Shastri. 1986. *Manual of Information to Accompany the Kolhapur Corpus of Indian English, for Use with Digital Computers*. Kolhapur: Shivaji University.

Tully, Mark. 1997. "English: An Advantage to India?". *ELT Journal* 51: 157–164. doi:10.1093/elt/51.2.157

Wulff, Stefanie, and Stefan Th. Gries. 2015. "Prenominal Adjective Order Preferences in Chinese and German L2 English: A Multifactorial Corpus Study". *Linguistic Approaches to Bilingualism* 5: 122–150.

Yogasundram, Nath. 2008. *A Comprehensive History of Sri Lanka: From Prehistory to Tsunami* (2nd ed.). Colombo: Vijitha Yapa Publications.

Authors' addresses

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

stgries@linguistics.ucsb.edu

Tobias Bernaisch (corresponding author)
Department of English
Justus Liebig University Giessen
Otto-Behaghel-Str. 10B
35394 Giessen
Germany

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