

John Benjamins Publishing Company



This is a contribution from *Grammatical Change in English World-Wide*.

Edited by Peter Collins.

© 2015. 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

Cultural keywords in context

A pilot study of linguistic acculturation in South Asian Englishes

Joybrato Mukherjee & Tobias Bernaisch

Justus Liebig University Giessen

The present study focuses on lexicogrammatical routines of South Asian Englishes that are associated with so-called ‘cultural keywords.’ These routines are particularly significant manifestations of the overarching process of the linguistic acculturation of the English language in new postcolonial settings. Specifically, we make use of the South Asian Varieties of English Corpus in order to compare acrolectal Indian, Pakistani and Sri Lankan English with regard to typical noun-verb collocations linked to three cultural keywords shared by all three South Asian Englishes: *government*, *terror* and *religion*. This pilot study offers a way of describing the effects of diachronic divergence in the formation of South Asian Englishes although comparable historical corpora of English in South Asia are not (yet) available.

Keywords: cultural keywords; South Asian Englishes; collocations; divergence; lexicogrammar

1. Introduction: Structural nativisation and linguistic acculturation

The institutionalisation of New Englishes has often been captured by the term and concept of ‘nativisation’ (cf. Kachru 1985). Broadly speaking, nativisation includes three levels (cf. Mukherjee 2007: 98ff.):

1. The functional level, which captures the increasing range of “nativised discourse and style types and functionally determined sublanguages (registers), and [the use of English] as a linguistic vehicle for creative writing in various genres” (Kachru 1985: 211);
2. The attitudinal level, which embraces the increasing acceptance of – and growing positive attitude towards – the English language that undergoes a transformation from a once foreign language to a localised language;

3. The structural level, which addresses “the emergence of locally characteristic linguistic patterns” (Schneider 2007: 5f.), including newly evolving norms and standards at all descriptive levels – ranging from phonetics and phonology to grammar and discourse pragmatics. (Lexico-)grammatical change is, thus, an integral part of structural nativisation continuously deriving the respective present-day variety-specific structural profiles of postcolonial Englishes – the interim products of in most cases centuries-long structure-related diachronic developments.

In a wider context, the process of nativisation at all three aforementioned levels leads to the formation of a new variety of English.

Variety-formation is always linked to identity-construction. By positing an evolutionary model of the emergence of postcolonial Englishes, Schneider (2003, 2007) establishes a framework which rests on the central assumptions that, firstly and generally, the construction of social identity is essential for language use and change and that, secondly and specifically, the construction of new and hybrid social identities in colonial and postcolonial contexts is at the very heart of the formation of New Englishes with their unique creative potentials and distinct linguistic patternings. In other words, identity-construction and variety-formation are intricately intertwined. Note that it is this intricate relationship that has motivated many authors of fiction in postcolonial literatures like Chinua Achebe to use and shape the English language in unprecedented ways: “I feel that the English language will be able to carry the weight of my African experience. But it will have to be a new English, still in full communion with its ancestral home but altered to suit its new African surroundings.” (Achebe 1975: 434).

Schneider (2007) relates the construction and reconstruction of new and hybrid social identities in colonial contexts, based on the growing interaction between the settlers and the indigenous population, to the concepts of ‘accommodation’ (cf. Giles 1984) and ‘negotiation’ (cf. Thomason 2001). In fact, we are dealing with a process of slowly and steadily integrating the English language into the local linguistic repertoire and, at the same time, of reshaping English usage (in the sense of structural nativisation). ‘Accommodating’ the system to a new English-speaking context and ‘negotiating’ standards and norms in a newly emerging Anglophone speech community are undeniably relevant. However, much as they are related and overlap, we view identity-construction, accommodation and negotiation as individual aspects of a much more basic and general cultural and psychological process, namely ‘linguistic acculturation’, a concept which has a long tradition in anthropology (cf. Johnson 1943; Basso 1967). Specifically, we use the term ‘linguistic acculturation’ to refer to the process of cultural entrenchment of the English language in newly emerging Anglophone speech communities in postcolonial contexts. The manifestation of the linguistic acculturation of English in a postcolonial context is the nativisation of the English language, resulting

in a new variety of English that serves as a localised means of communication and a vehicle for local identity-construction. It appears that the characteristic features of a specific postcolonial context will have a bearing on the process of linguistic acculturation and, thus, on the structural nativisation of English in case of highly culture-related items.

South Asia is a subcontinent that harbours a range of postcolonial contexts in which new national varieties of English have been shaped over the past centuries, especially in the post-Independence period when English was retained – for a variety of reasons – as a *de jure* or *de facto* (co-)official language and a medium for a range of communicative functions with a persistently high level of prestige. Research into New Englishes world-wide has profited immensely from the availability of large computerised corpora. For example, the Kolhapur Corpus of written Indian English (1978) and the Indian component of the International Corpus of English (2002) have provided a wealth of data to describe structural nativisation in the largest South Asian English (SAE) variety. In the present paper, the focus will be on structural nativisation in relation to linguistic acculturation in three SAEs: Indian English (IndE), Pakistani English (PakE), and Sri Lankan English (SLE). As our database, we will make use of the South Asian Varieties of English (SAVE) Corpus, a newspaper corpus with six national components from South Asia, including India, Pakistan and Sri Lanka (cf. Bernaisch et al. 2011). It has been shown in previous corpus-based studies that SAEs display many common features (i.e. ‘pan-South Asian structures’) as well as structural differences. Such differences between varieties are often quantitative in nature and show themselves, for example, in different frequencies of – and preferences for – verb-complementational patterns across SAEs (cf. Schilk et al. 2012).

It should not go unmentioned that the concepts of structural nativisation (and linguistic acculturation, for that matter,) refer to essentially diachronic developments. In corpus-based research into New Englishes, one would ideally need comparable historical data in addition to synchronic corpora (such as the SAVE corpus) in order to be able to directly trace and compare structural changes across time in individual varieties of English. The lack of historical corpora for most of the colonial and postcolonial settings world-wide is one of the most severe challenges for research into New Englishes, including SAEs. As we have argued elsewhere (cf. Mukherjee & Schilk 2012), there are various ways of tackling this problem with the help of ‘indirect evidence’, such as by establishing an ‘apparent time construct’ (cf. Bailey et al. 1991) with individual varieties representing individual stages of a basically uniform process of variety-formation as posited by Schneider (2003, 2007), or by construing a diachronic British English (BrE) database representing the historical input variety at the time of transplantation to South Asia and triangulating the findings with observations from present-day corpora of IndE and BrE (cf. Hoffmann & Mukherjee 2007). In the present project, we have opted for yet another option, namely to restrict ourselves to present-day data of

SAEes (and BrE) because we assume that the shared features and differences we find between the varieties under scrutiny are manifestations of diachronic developments in the three individual SAEes after the end of the British Raj in South Asia in 1947/48. In fact, it is reasonable to assume that all the three SAEes originate in a largely uniform proto-South Asian variety of English given the shared sociolinguistic characteristics in the pre-Independence periods of the countries concerned. The English language was used by a relatively small number of English-educated speakers and modelled on BrE standards, which were systematically disseminated to the respective elite minorities in what are India and Pakistan today following Macaulay's Minute in 1835 and in Sri Lanka after the suggestions of the Colebrooke-Cameron Commission in 1831/1832. Given these and other pre-Independence sociolinguistic similarities (e.g. the adoption of English as a vehicle for creative writing, the formal teaching of English in missionary schools, etc.), it is likely that the three varieties at hand displayed a high level of homogeneity on the brink of Independence. Amongst others, Trudgill (1986: 145) and Schneider (2007: 51) stress the importance of the social and regional homogeneity in early acrolectal forms of New Englishes. In the context of Schneider's (2003, 2007) model based on group-interaction and identity-construction, one could easily argue that the fight against English colonial rule in the decades before Independence made English-educated South Asian speakers emphasise the uniformity of their English usage.¹ While it is true that the homogeneity in English across South Asia of the pre-Independence period might have been more or less present at different linguistic levels, it makes sense to assume that for the object of inquiry of the present study, namely the routinised patterns in which socio-politically relevant high-frequency cultural keywords are used, it is the three fundamentally different post-Independence habitats of three independent countries which have shaped the major routines in the use of the culture-related lexical items under scrutiny. Thus, for the purpose of the present study of linguistic acculturation of cultural keywords we view the synchronic distance between SAEes today as an immediate result of post-Independence processes of diachronic divergence. At a later stage, it should be insightful to reassess our findings on grounds of historical data representing English in South Asia in, say, the 1930s and 1940s.

As in previous SAVE-based studies (e.g. Schilk et al. 2012), our interest lies in aspects of unity and diversity across SAEes. However, our focus of interest has shifted to an area of lexicogrammar which has been neglected in research into many

1. Consider as anecdotal evidence, for example, the largely indistinguishable linguistic styles of Muhammad Ali Jinnah, the first Governor-General of Pakistan, and of Jawaharlal Nehru, the first Prime Minister of India, in their speeches on Independence Day on 14 and 15 August 1947, respectively.

postcolonial Englishes so far, although it clearly lends itself to the description of differences in structural nativisation across SAEs, namely the lexicogrammatical routines associated with ‘cultural keywords’. ‘Cultural keyword’ is a concept that goes back to Williams (1976), and that was taken up by Stubbs (1996, 2002), who propagates the use of large corpora to analyse the habitual patterns associated with frequent lexical items that are ‘key’ in a specific Anglophone culture. Given that nativisation in its entirety, including structural nativisation, is a manifestation of the acculturation of the English language in a new habitat (see above), it is high time that corpus resources now available for SAEs were used to detect differences between neighbouring, but culturally vastly different, Anglophone cultures in South Asia with regard to the habitual patternings associated with high-frequency cultural keywords. This is certainly one of the areas in which the link between the linguistic acculturation of English and its structural nativisation in a given socio-cultural context is most immediate. The pilot study reported in the present chapter suggests how to approach this area of linguistic acculturation with the help of corpus resources.

Against this background, the aim of the present paper is two-fold. Firstly, we will analyse the use of a selection of cultural keywords in three SAEs and make use of the SAVE Corpus as a new environment of comparable corpus components. Secondly, and specifically, we will identify lexicogrammatical differences in the use of cultural keywords that are shared by all three SAEs in order to identify differences in structural nativisation between SAEs that are linked to differences between the underlying cultural habitats of India, Pakistan and Sri Lanka.

The paper is structured as follows. In Section 2 we will briefly compare the three South Asian cultures under scrutiny and re-assess the concept of cultural keywords against this background. In Section 3 the corpus data and the methodological steps of our analysis will be sketched out. In Section 4 we will present and discuss the findings of the corpus analysis. Section 5 will offer some concluding remarks.

2. South Asian cultures and cultural keywords in South Asian contexts

In the context of discussing aspects of unity and diversity in SAEs, it has already been sketched out elsewhere that the processes of nativising the English language in individual South Asian countries are different for various functional and historical reasons (cf. Schilk et al. 2012: 139f.). These include the influence of other European languages, different constellations of indigenous languages and English, different language policies with different statuses granted to the English language, and the differing needs for English as a ‘link language’ alongside other national languages. Above and beyond these genuinely linguistic factors, it is of course particularly relevant for the present study to take note of the underlying socio-cultural and socio-political

differences between India, Pakistan and Sri Lanka – differences that could be seen as partly responsible for the transformation of the colonial territories, all of which once formed parts of the British Raj, into three politically independent nations in 1947/48.

At the risk of gross oversimplification, let us overview some of the major socio-cultural and socio-political differences between India, Pakistan and Sri Lanka (cf. Laporte 2002; Mishra 2002; Holt 2011). India is a secular country with no state religion, although more than 70% of the population follows the Hindu faith. The parliamentary system was designed after the British model, and the democratic state of law has been relatively stable since Independence (except for a period of emergency rule in the mid-1970s). There have been changes in government due to free elections but no military coups, there is a range of opposition parties, and there is a free and critical press. In the history of India, various political movements have tried to fight for Independence of specific regions (e.g. Punjab in the 1980s) and terrorist groups have been active in various parts of the country (at present the 'Naxalite' movement in particular). Of all South Asian countries, India, the second most populous country in the world, displays the highest degree of ethnic, linguistic and religious diversity as well as the highest degree of political stability and economic dynamism.

Pakistan was created in 1947 as a homeland for the Muslims in former British India. In fact, Islam has been the state religion ever since. The Islamic orientation is clearly visible in the constitution, for example with regard to the role of the *Sharia*. Although the Constitution defines Pakistan as a parliamentary republic, the country has been ruled by military and authoritarian governments for the major part of its post-Independence history. The most traumatic event in the history of Pakistan certainly is the loss of its Eastern part, which became independent, with military support from India, as Bangladesh in 1971. The conflict with India (especially over Kashmir, which is claimed both by India and Pakistan and has been divided between the two countries ever since the late 1940s) as well as the fight against Islamist terrorists affiliated with Al-Qaeda who use Afghanistan and Pakistan as their base stations, characterise Pakistani politics today. In the light of severe political controversies, the great number of terrorist attacks and the deteriorating standard of living, some have referred to Pakistan as a failed or failing state.

The history of Sri Lanka, the official name of which was Ceylon until 1972, has been marked by a grave ethnic conflict between the Sinhalese majority population, most of whom are Buddhists, and the Tamil minority living in the Northern and Eastern provinces, most of whom are Hindus. This ethnic conflict has been caused by various factors, including the privileged status of Tamils under British rule and the dominance of Sinhalese Buddhist nationalism after Independence. This ethnic conflict also had an enormous impact on the language policy in the 1950s, when the then government implemented a Sinhala-only agenda. Much of the history of Sri Lanka was marked by a full-scale (civil) war between the Sri Lankan Army and the Liberation

Tigers of Tamil Eelam (LTTE) fighting for a separate state for the Tamils. This war also affected neighbouring India: for example, India was engaged in Sri Lanka with peace-keeping troops in the 1980s, and the then Prime Minister Rajiv Gandhi was later assassinated by an LTTE suicide bomber in 1991. Today, after the end of the civil war, there is a clear attempt to create a new pan-ethnic national identity and to rebuild and develop Sri Lanka with the help of foreign investors.

Given the marked differences between the socio-cultural and socio-political contexts of India, Pakistan and Sri Lanka, all of which represent historically related and geographically adjacent South Asian cultures, a comparative analysis of cultural keywords of the three South Asian cultures should be a promising endeavour.

As pointed out by Stubbs (1993), it was Firth (1935:40), the founding father of British contextualism, who proposed “research into the distribution of sociologically important words”. Such words were referred to by Stubbs (1996, 2002) as ‘cultural keywords’, which he defined as the small set of words in in different languages “whose meanings give insight into the culture of the speakers of those languages” (Stubbs 2002:145). In this context, English is a particularly interesting case because it is used as a communicative vehicle in a wide range of vastly different cultures. That is, across individual varieties of English, one and the same lexical item may follow different routines of usage, providing insights into its cultural associations. Note that the culture-specificity of the meaning(s) of a cultural keyword often derives from the typical contexts in which it is habitually used. Stubbs (1995) refers to the meaning components of a cultural keyword in its typical contexts of usage as ‘cultural connotations’, which may entail ‘evaluative connotations’ as well as ‘semantic prosodies’ (cf. Louw 1993; Rocci & Wariss Monteiro 2009:71).

Linguistic corpora are particularly helpful in unveiling cultural connotations of this kind as they include a large number of natural contexts of language use in a given speech community. These natural contexts provide direct access to the socio-cultural habitat in which a cultural keyword tends to be used. From a corpus-linguistic perspective, one way of defining cultural keywords is by categorising them as the high-frequency content words included in a large and representative corpus of a language or a language variety (cf. Mukherjee 2009:69f.). Given their cultural significance and their high frequency, the typical use of cultural keywords in a given speech community provides important insights into the socio-cultural setting of that community and, thus, the community-specific linguistic acculturation of the English language.

In the present pilot study, our focus is on a selection of cultural keywords in three neighbouring postcolonial Englishes in South Asia. More specifically, we seek to capture differences in the acculturation of the English language in India, Pakistan and Sri Lanka by looking at typical collocates in the contexts in which high-frequency keywords are used. We thus view the present paper as complementing the focus on structural nativisation of corpus-based research into World Englishes. Our interest

lies in how differences between the extralinguistic realities of the three South Asian countries are reflected in the lexicogrammatical patternings in the varieties of English that have emerged in the respective localities. In this context, it makes sense to view cultural keywords as “words that are revealing of a culture’s beliefs or values” (Rocci & Wariss Monteiro 2009: 66).

3. Methodology

As observed by Wierzbicka (1997), there is no single and objective way of identifying keywords in a culture. Therefore, she proposes a number of possible means of identification, involving examining words in terms of their (a) frequency of occurrence, (b) frequency of occurrence in particular domains, (c) frequency of occurrence in book titles, songs, proverbs, sayings, etc., and (d) richness of phraseological patterns. Rocci & Wariss Monteiro (2009: 67) add that “in order to decide whether a certain word is indeed a cultural keyword one should look at how exactly this word is used in arguments in a corpus of texts representative of the cultural community under consideration”. In this study we assume that it is reasonable to identify cultural keywords on the basis of quantitative analyses of large collections of authentic text, and furthermore that attention has to be paid not only to each keyword as such, but also to its surrounding pattern(s) and context(s).

The SAVE Corpus (cf. Bernaisch et al. 2011) lends itself ideally to the identification of cultural keywords and the linguistic patterns in which they are used in IndE, PakE and SLE.² The SAVE Corpus features a total of six national components, each of which is comprised of approximately three million words representing local acrolectal newspaper English from two leading national English-medium newspapers. The newspaper data were obtained from online archives including texts from the years 2000 to 2008. From a structural perspective, newspaper English to a large extent fulfils a standardising function in South Asia given the absence of reference works such as full-fledged dictionaries or grammars for many SAEs (cf. Schilk 2012: 47). It is a central asset of the SAVE Corpus that articles from international news agencies (not representing local varieties of English) as well as the large amount of duplicates that are typical of online archives have been systematically removed (cf. Bernaisch et al. 2011: 3). In light of the implicit meanings and connotations that may be associated with cultural keywords in particular, newspapers in the South Asian countries represent the most relevant corpus-linguistic resource since they address and potentially

2. The SAVE Corpus was compiled in the context of the project ‘Verb complementation in South Asian Englishes: A study of ditransitive verbs in web-derived corpora’ funded by the German Research Foundation (*Deutsche Forschungsgemeinschaft* MU 1683/3–1, 2008–2011).

influence nation-wide audiences.³ For this reason, newspapers may be considered ‘cultural loudspeakers’ with a nationwide range. They grant access to important local as well as international issues and events, provide interpretation schemata for – and relevant opinions on – these issues and events, and disseminate these interpretation schemata and opinions among millions of readers, thus possibly shaping their readers’ world views and, on a larger scale, cultural connotations in the speech community. In order to identify cultural keywords, their associated linguistic structures and their potential cultural connotations in IndE, PakE and SLE, we proceeded in the manner sketched out in Figure 1.

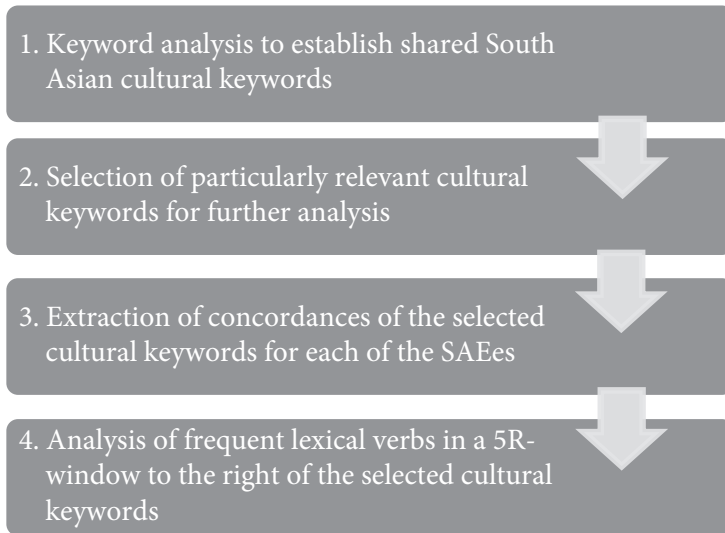


Figure 1. Describing cultural keywords and their cultural connotations in SAEs

Based on comparable data representative of British newspaper English in the form of the daily news section in the British National Corpus (BNC news), keyword analyses (in which the British texts served as reference data) were conducted for the Indian, Pakistani and Sri Lankan components of SAVE (SAVE-IND, SAVE-PAK and SAVE-SL, respectively) with WordSmith Tools 5.0 (Scott 2008).⁴ These analyses resulted in

3. The term ‘connotation’ has been defined in different ways by various scholars (e.g. Lyons 1968; Beardsley 1975). For the present study, we regard connotations as “*expressive* components of meaning, most obviously in the case of terms which carry ‘favourable’ or ‘unfavourable’ connotations. Many lexical units serve to express the attitudes or feelings of the speaker towards what they describe.” (Bright 1992: 297).

4. The exact words counts for the individual datasets derived from the WordList statistics available in WordSmith Tools 5.0 are as follows: BNC news comprises 8,972,033 words, SAVE-IND 3,104,430 words, SAVE-PAK 3,103,816 words and SAVE-SL 3,083,206 words. While it is

one keyword list for each SAE variety. We then opted for a socio-culturally motivated selection of nominal cultural keywords from different semantic fields shared by the three keyword lists (i.e. a selection of lexical items that occur frequently in all three SAEs). The recovery of potential cultural connotations associated with the selected cultural keywords necessitated a close look at their semantico-structural environment. Hence, for each of the cultural keywords chosen, concordances were drawn from part-of-speech tagged (POS-tagged) versions of SAVE-IND, SAVE-PAK and SAVE-SL.

In work on semantic prosody, which can be defined as “[a] consistent aura of meaning with which a form is imbued by its collocates” (Louw 1993: 157), it is usually collocating adjectives and their evaluative descriptions which are studied in order to establish the positive or negative nature of the prosody of a lexical item in question. However, in contexts where media freedom may be severely limited,⁵ it is necessary to reconsider the value of the insights we can expect from studying these overtly evaluative adjectives with politically sensitive lexical items such as *government* or *terrorism*. It is not unlikely that negative descriptive values assigned via adjectives to, say, *government* are sometimes deleted or replaced in the editing process of local newspapers, because of the negative evaluation of the current authorities. It is also for this reason that we chose not to focus on (overtly descriptive) adjectives, but on verbs, which have a higher potential to assign and convey meanings in a more subtle way which may escape local editors’ awareness and thus provide a more unbiased perspective. Consequently, in the present pilot study we have placed special emphasis on the verbs with which the selected cultural keywords are associated to the right, i.e. on noun-verb collocations. Thus, only those concordances in which the nominal cultural keywords were followed by a lexical verb (i.e. by a word form with a VV*tag in CLAWS C7 terms)⁶ in a window of five words to the right of the noun (i.e. in its 5R-window) were considered for further analysis as in (1),⁷ where the nominal cultural keyword is *government* and the first VV*-tagged verb in the 5R-window is *find*.

certainly true that larger synchronic as well as diachronic mega-corpora are available for some native varieties of English such as The Corpus of Contemporary American English (COCA; Davies 2008–) or The Corpus of Historical American English (COHA; Davies 2010–), this is not the case for SAEs. At present, the individual SAVE components represent the largest comparable corpus resources for comparative studies of SAEs. It is difficult to assess the degree of comparability of the South Asian components of the recently launched Corpus of Global Web-based English (GloWbE; (<http://corpus2.byu.edu/glowbe/>) (30 May 2014)).

5. According to the 2013 World Press Freedom index (<http://en.rsf.org/press-freedom-index-2013,1054.html>) (30 May 2014), which features a total of 170 countries, India ranks 140th, Pakistan 159th and Sri Lanka 162nd.

6. CLAWS stands for Constituent Likelihood Automatic Word-tagging System (<http://ucrel.lancs.ac.uk/claws/>) (30 May 2014); CLAWS C7 is the current standard tagset.

7. *Be*, *do* and *have*, which can function as auxiliary verbs and are often considered semantically ‘empty’, were not extracted from the data and were thus excluded from the analysis.

- (1) Similarly_RR our_APPGE *government_NN1* will_VM *find_VVI* the_AT means_NN to_TO ensure_VVI a_AT1 safe_JJ and_CC secure_JJ future_NN1 four_MC our_APPGE youth_NN1. (SAVE-SL-DN_2002-01-23)

The first VV*-tagged verb in the 5R-window of the nominal cultural keyword was extracted from each concordance line with the help of a PERL script and was then lemmatised.⁸ If the resulting verb lemma occurred at least five times,⁹ it was subjected to further analyses focusing on its influence on the semantic prosody of the cultural keyword concerned, and on the degree to which these verbal collocates of the nominal cultural keywords are variety-specific (as opposed to being shared by all SAE varieties, i.e. 'pan-South Asian').

The fact that New Englishes simultaneously display shared features and differences between varieties has recently spawned greater interest, also because it is now feasible to compare a wide range of varieties of English world-wide by utilising corpus resources such as the International Corpus of English (cf. Hundt & Gut 2012). In this context, we have already shown that neighbouring New Englishes in South Asia, too, are characterised by aspects of unity and diversity (cf. Schilk et al. 2012).

In order to quantify the degree to which a certain structural feature (e.g. the verbal collocates of a cultural keyword in a given variety) is marked by cross-varietal unity or diversity, we propose a 'diversity/unity (*d/u*) ratio'. The *d/u* ratio is calculated as shown in (2).

$$(2) \quad d/u \text{ ratio} = \frac{\text{variety - specific structures}}{\text{variety - specific structures} + \text{structures shared across all varieties}} \times 100$$

The *d/u* ratio takes the sum of the variety-specific structures of the object of investigation as the numerator and the sum of the total structures of the object of investigation as the denominator of the fraction which is then multiplied by 100.¹⁰ The *d/u* ratio

8. We would like to thank Benedikt Heller for the PERL script extracting the verbs in the 5R-window from the concordances of the nominal keywords selected. The lemmatisation of the verbs was performed on the basis of the slightly modified lemma list available at (http://www.lexically.net/downloads/BNC_wordlists/e_lemma.txt) (28 July 2013).

9. The cut-off point of a minimum frequency of five occurrences per keyword-related verb lemma has been used as an operationalisation of the concept of 'cultural salience' (cf. Wierzbicka 1997:12).

10. As the present paper examines three varieties, there can be: (a) variety-specific structures, (b) structures shared by two varieties, and (c) structures shared across all varieties. The *d/u* ratio as calculated in the present paper only takes into account variety-specific structures and structures shared across all three SAE varieties, since structures shared by two varieties can be regarded both as semi-variety-specific and as semi-shared at the same time.

ranges from 0 to 100 and shows whether and to what degree a given linguistic object or area of variation is marked by cross-varietal unity or diversity since d/u ratios smaller than 50 indicate a dominance of cross-varietally shared structures and d/u ratios larger than 50 stand for a dominance of variety-specific structures. In other words, the more variety-specific the object of investigation is, the closer the d/u ratio is to 100; the more cross-varietally stable it is, the closer the d/u ratio is to 0.

It needs to be stressed here that, apart from the analysis of the influence of the verbal collocates on the semantic prosody of the cultural keywords at hand, the procedure depicted in Figure 1 is a fully automated approach. This was necessary in order to come to grips with the huge amounts of corpus data, but it does not come without its downsides in recall and precision. Note that only verbs occurring in a 5R-window to the right of the cultural keyword were extracted and verbs to the left were not considered because they would have negatively affected the precision of the automatic data extraction. This means, for example, that constructions in which a cultural keyword occurs as a passive *by*-agent in sentence-final position are not included in the present study.¹¹ Also, false positives may have entered the data because we did not check whether a given nominal cultural keyword was also the head of the noun phrase governing the verb extracted for further analysis. While the results of our analysis thus have to be taken with a measure of caution, it goes without saying that it is the high degree of automatism that has enabled us to analyse the right-hand contexts of nominal cultural keywords in three national components of the SAVE corpus in their entirety.

4. Results

For SAVE-IND, SAVE-PAK and SAVE-SL, we conducted keyword analyses with the BNC news section as a source of reference data. Given that nouns may provide the most immediate insights into central and recurrent topics in the varieties covered, only nominal cultural keywords were considered for further analysis. In order to assess the range of semantic fields that are covered by cultural keywords shared by the three

11. It should be noted, however, that in general 'long' passives with an explicit *by*-agent are comparatively rare in actual usage, depending on the genre. Biber et al. (1999) give the following figures: long passives occur around 750 times per million words (pmw) in news and academic writing while short passives with dynamic verbs (without *by*-agent) occur around 2,500 times pmw in news and 5,000 times pmw in academic writing. Against this background, we seem to have neglected a small minority of all passives with our automatic procedure.

SAE varieties, the shared nominal cultural keywords were semantically categorised using Wmatrix3 (Rayson 2008): see Table 1.¹²

Table 1. Shared nominal South Asian English cultural keywords arranged according to their semantic classification in Wmatrix 3

Semantic class	Shared nominal South Asian English cultural keywords
Business general	<i>commerce, infrastructure</i>
Crime	<i>terrorism, violation</i>
Education general	<i>education, students, university</i>
General actions/making	<i>activities, commission, implementation, project, projects</i>
Giving	<i>irrigation, issue, supply</i>
Government	<i>constitution, country, governance, government, governments, govt, minister, ministry, president, revenue, state</i>
In power	<i>administration, chief, coordination, order, power, raja</i>
Information technology and computing	<i>internet, website</i>
Medicines and medical treatment	<i>Dr, rehabilitation</i>
Money general	<i>finance, rupee, rupees</i>
Other proper names	<i>alliance, district, process</i>
Participating	<i>participation, parties</i>
People	<i>human, persons, populations</i>
Places	<i>areas, countries, districts, region</i>
Various	<i>agencies, construction, coordination, corporation, corruption, demand, department, deputy, development, due, elections, employees, funds, industries, institutions, issues, kg, media, ms, NGOs, peace, petitioner, present, prime, purpose, release, religion, rice, rs, secretariat, sector, security, seminar, situation, sources, steps, tariff, technology, water</i>

A total of 90 shared nominal South Asian cultural keywords could be identified. What renders these keywords *cultural* is their specificity to the South Asian Sprachraum. True, the list in Table 1 contains lexical items such as *administration, finance* or *country*, the South Asian nature of which is not immediately obvious. However, there are other

12. Wmatrix3 was used to semantically tag the list of shared nominal SAE cultural keywords in order to (a) gain an overview of the range of semantic classes from which the keywords stem, and (b) ensure that the keywords to be studied more closely cover different semantic fields and are not restricted to one semantic class only. The key semantic domains module in Wmatrix3 was not used in the present study.

items like *rupee(s)* and *raja*, for example, which can immediately be identified as integral parts of South Asian culture because the terms denote local currencies and rulers respectively. Further, *security* and *irrigation* are recurrent issues in India, Pakistan and Sri Lanka, and thus also markedly South Asian. When it comes to the lexicon of SAEs, Meyler (2007:xiv) attests that:

SLE includes a number of features characteristic of British colonial language which has fallen out of fashion in contemporary British English, including “Anglo-Indian” words which date from colonial times and are common to the whole Subcontinent. These words are considered archaic in British English, or are restricted to more formal contexts.

This argument is in line with descriptions of the vocabulary of New Englishes in general, which deem their respective lexical repertoires as to some extent ‘bookish’ (Mesthrie & Bhatt 2008: 114ff.), and may thus also apply to IndE and PakE. Against this background, the fact that *persons*, a lexical item which is “restricted to more formal contexts” (Meyler 2007:xiv) in BrE, occurs as a shared nominal South Asian cultural keyword, indicates that this lexical item is indeed used significantly more frequently in SAEs than in BrE and is thus a less overtly marked South Asian cultural keyword than *rupee* or *raja*.¹³

The lemmata *government*, *terror* and *religion* suggest themselves for a more fine-grained analysis of their semantico-structural environment because the concepts they express are highly relevant in each of the countries, but may be associated with different connotations in the three South Asian countries, given the different extralinguistic political, historical and societal contexts in which they are used in India, Pakistan and Sri Lanka.¹⁴

For *government*, the extraction of the verbs in its 5R-window produced three groups of verbs: (a) verbs that occur exclusively in the context of *government* in one variety, (b) verbs shared by two varieties, and (c) verbs shared across all varieties. Each of the three groups is certainly suitable for the recovery of possibly implicit meanings associated with *government*. However, in terms of shared pan-South Asian – as

13. The reason why media-related terms such as *Internet* and *website* entered the list of shared nominal South Asian cultural keywords probably lies in the fact that BNC news, the British reference data on the basis of which the keyword lists were created, is at least 10–20 years older than SAVE and is therefore less marked by the Internet age.

14. *Government*, *religion* and *terror* were structurally (and *terror* also semantically) lemmatised in that *government* and *religion* were searched for with tokens of the respective singular and plural forms while the concordances for *terror* cover the search terms *terror(s)*, *terrorist(s)* and *terrorism(s)*. For *government*, the cultural keyword *govt* as shown in Table 1 was not used because it is only marginally frequent in that it constitutes 1.19% of all word forms referring to the concept of government in SAVE-IND, 2.49% in SAVE-PAK and 2.87% in SAVE-SL.

opposed to variety-specific – structures and meanings associated with *government*, the first group covering variety-exclusive verbs used alongside *government*, and the third group covering verbs used in all three varieties in the 5R-window, are the most relevant ones. Figure 2 depicts the lexical verbs which occurred in the 5R-window of *government* in SAVE-IND, SAVE-PAK and SAVE-SL.

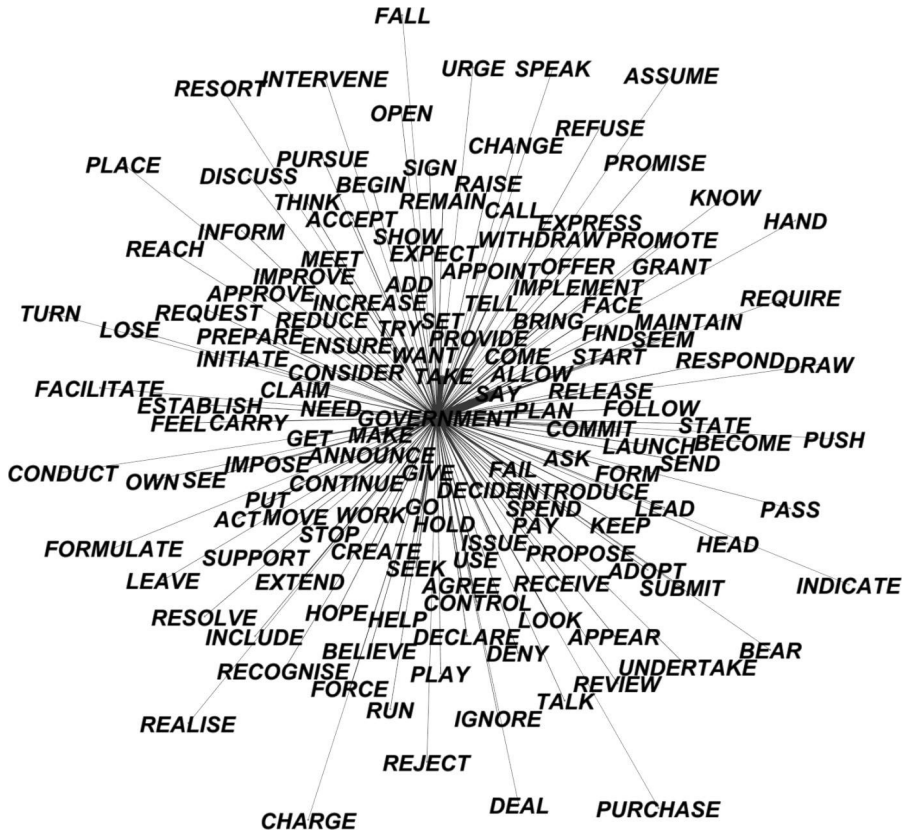


Figure 2. Verbal collocates in the 5R-window of *government* shared across SAVE-IND, SAVE-PAK and SAVE-SL*

* The network visualisations of this paper have been created with the open-source graph visualisation and manipulation software Gephi (<<https://gephi.org/>>, accessed 29 August 2013).

The cultural keyword *government* is at the centre of Figure 2. The 145 lemmatised verbs surrounding *government* are lexical verbs which occurred at least five times in a window of 5 words to the right of *government* in SAVE-IND, SAVE-PAK and SAVE-SL. These verbs thus constitute the pool of lexical verbs with which *government* is habitually associated in South Asia. The distance between *government* and the respective verbs is meaningful in the sense that the closer a lexical verb is to

government, the more frequent their co-occurrence. The frequency of co-occurrence of a given verb with *government* is also mirrored in the thickness of the connecting line. Thick lines denote high frequencies of co-occurrence and thin lines low(er) frequencies. The verbal lemma closest to *government* is *say* since it is the verb which, on average, occurs most frequently with *government* in SAVE-IND, SAVE-PAK and SAVE-SL.¹⁵ *Purchase*, in contrast, co-occurs least frequently with *government*, which is why it is found at the very periphery of Figure 2.

Figure 3 depicts these variety-specific verbal collocates of *government* in three separate networks representing IndE, PakE and SLE.

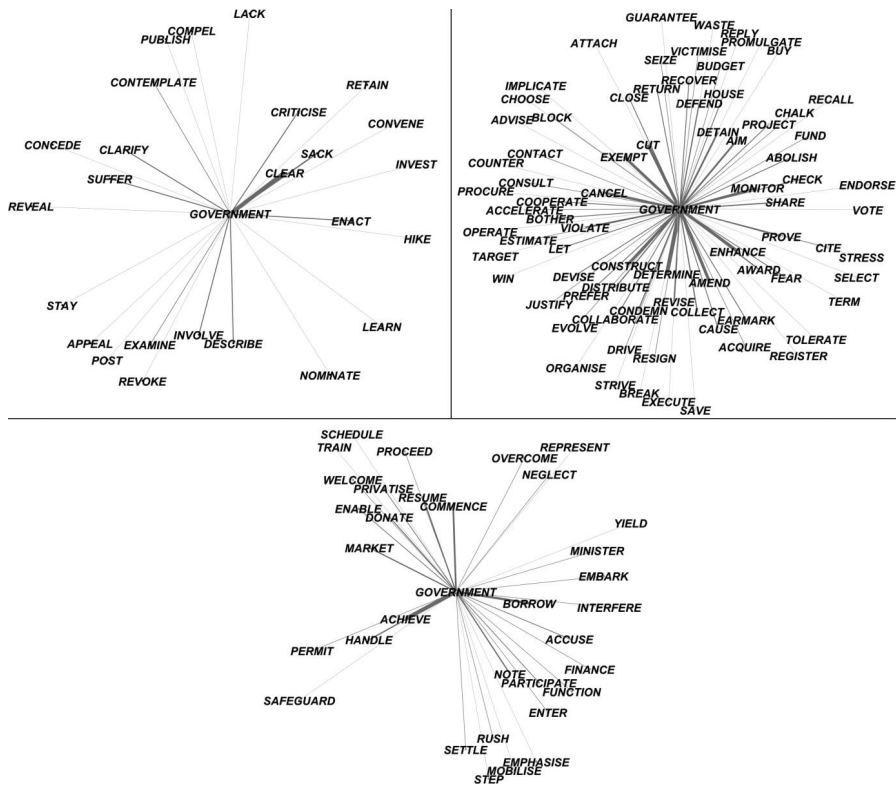


Figure 3. Variety-specific verbal collocates in the 5R-window of *government* in SAVE-IND (top left), SAVE-PAK (top right) and SAVE-SL (bottom)

15. To be more precise, *say* occurs in the 5R-window of *government* with a normalised per-million-word (pmw) frequency of 102.55 in SAVE-IND. The respective normalised (pmw) frequency for SAVE-PAK is 138.68 and for SAVE-SL 62.30. Consequently, the average normalised (pmw) frequency for *say* across the three varieties is 101.18. It is this cross-varietal average normalised (pmw) frequency which forms the basis for the position of each of the verbal lemmas in relation to *government* in Figure 2.

For SAVE-IND, which features 25 variety-specific verbal collocates of *government*, *clear* is the most frequent variety-exclusive verb used with *government*. Combinations of *government* followed by *clear* are exemplified in (3) and (4).

- (3) The *government* today *cleared* the first batch of 500 skilled and unskilled workers for post-war reconstruction work in Iraq.
(SAVE-IND-SM_2003-06-18)
- (4) In addition, the central *government* *has cleared* the joint proposal of PCMC and Marhatta Chamber of Commerce, Industries and Agriculture (MCCIA) for establishing an “autocluster” in PCMC so as to promote automobile and rubber industry.
(SAVE-IND-TI_38406)

One of the most frequent verbs used variety-exclusively in the context of *government* in SAVE-PAK is *amend*.¹⁶ Relevant examples are provided in (5) and (6).

- (5) Opposing the bill, Punjab Minister for Law and Parliamentary Affairs Muhammad Basharat Raja said the *government* *had* already *amended* the rules to improve “thana” culture.
(SAVE-PAK-DT_2003-10-29)
- (6) The Punjab *government* *has amended* the Local Government Ordinance 2001, deleting the method of the election of naib nazims of district governments and tehsil councils through secret ballot.
(SAVE-PAK-DA_2005-12-05)

Among the 33 lexical verbs which occur exclusively in the Sri Lankan data, *achieve* is the most frequent right-hand verbal collocate of *government*. This typical noun-verb collocation is exemplified in (7) and (8).

- (7) Following a period during which there was political instability and a serious economic downturn in 2001, the *Government*, in our opinion, *has achieved* considerable progress by signing a MOU on a ceasefire with the LTTE, which has continued despite problems and violations for the longest period when compared to previous ceasefires between governments and LTTE.
(SAVE-SL-DM_2003-11-19)
- (8) “Government has made progress in curbing inflation, which is expected to be below 6.5% in 2003. The *government* *achieved* some success in its efforts to consolidate the fiscal situation, although actual performance was below expectations for the first four months of 2003,” Mr. Zveglic pointed out.
(SAVE-SL-DM_2003-08-23)

16. The automated analysis produced *determine* as the most frequent variety-exclusive verb occurring with *government*. However, a closer look at the respective concordance lines reveals that many of the combinations of *determine* and *government* are combinations in which *determined* is used not as a verb, but as an adjective such as in “the government is determined to initiate [...]” and that the POS-tagging produced many false positives in this context.

Examples (3) to (8) provide a first insight into what structures and meanings surround *government* in the three varieties. The meanings associated with *clear* in (3) and (4) and *achieve* in (7) and (8) are certainly of a positive nature, while *amend* in (5) and (6) is comparatively neutral. In order to recover the connotations *government* is associated with on a larger scale, each verb has been semantically categorised according to whether it generally has a negative meaning (e.g. *suspend*, *arrest*), a neutral meaning (e.g. *examine*, *transfer*), or a positive meaning (e.g. *solve*, *cooperate*). The results based on this categorisation of the variety-specific verbal collocates of *government* are shown in Table 2.¹⁷

Table 2. Connotations of the variety-specific verbal collocates of *government*

Meaning	SAVE-IND		SAVE-PAK		SAVE-SL	
	abs. freq.	%	abs. freq.	%	abs. freq.	%
Positive	52	34.21%	206	37.80%	122	52.14%
Neutral	58	38.16%	185	33.94%	74	31.62%
Negative	42	27.63%	154	28.26%	38	16.24%
Total	152	100.00%	545	100.00%	234	100.0%

The connotations of the variety-specific verbs in the context of *government* yield statistically highly significant cross-varietal differences ($\chi^2 \approx 21.21$, $df = 4$, $p < 0.001$, Cramer's $V \approx 0.11$) with a relatively weak correlation.¹⁸ In the Indian data, the majority of the meanings of the verbs at hand is neutral (38.16%) followed by positive (34.21%) and negative meanings (27.63%). Positive meanings (37.80%) constitute the majority of meanings in the Pakistani data; neutral (33.94%) and negative meanings (28.26%) figure less prominently. In the Sri Lankan data, more than half of the meanings (52.14%) are of a positive nature with neutral (31.62%) and negative meanings (16.24%) occurring less frequently.

These findings have implications for the semantic prosody of (or the implicit meaning that is conveyed by the use of certain verbs with) *government*. First, in none of the varieties does *government* have a predominantly negative semantic prosody (i.e. positive and neutral verbs generally accompany this concept in the majority of cases).

17. The nature of the connotation of the variety-specific verbal collocates shown in Table 2 was established manually.

18. In pairwise comparisons, SAVE-IND and SAVE-PAK are not significantly different ($p > 0.05$) from each other in terms of the meanings associated with *government*. The pairwise comparison between SAVE-IND and SAVE-SL is statistically significant ($p < 0.01$), and so is the comparison between SAVE-PAK and SAVE-SL ($p < 0.001$).

Second, in the Sri Lankan data, it is particularly noticeable that positive meanings figure much more prominently than in SAVE-IND and SAVE-PAK and that negative meanings are markedly less frequent. Consequently, there seems to be a cross-varietally stable positive semantic prosody of *government* which is particularly pronounced in Sri Lanka.

Still, given the long-lasting Sri Lankan civil war which the local government was able to end after decades in 2009, it is to some extent surprising that *government* should have a more positive cultural connotation in the Sri Lankan than in the Indian and Pakistani data. What should probably not be overlooked is that the Sri Lankan data have been drawn from two local newspapers (the *Daily Mirror* and the *Daily News*), the latter of which is a government-owned daily newspaper. It is not too far-fetched to assume that what we see in our results might be the impact of a generally positive evaluation of the government in the discourse of one of the newspapers:

Unfortunately the end of the war in Sri Lanka three years ago does not mean that the media can be called properly free. I would echo a column in the *Wall Street Journal* in saying that the Sri Lankan media are possibly the least free in the South Asian region [...]. The lack of real media freedom is of course most marked when it comes to the state media. Although some official outlets like the *Daily News* sometimes cover mild criticism of various government practices, others, like state TV and radio broadcasters, are more or less propaganda channels of a type more akin to Communist Cold War outlets. That means that “news” is dripping in nationalistic ideology, talk of “traitors” and “terrorist-lovers” and full of factual inaccuracies. (Haviland as quoted in Natarajan 2012: 12)

This means that analyses of semantic prosodies associated with organs of government in general have to pay particular attention to the sources from which the data to be examined are drawn. In the Sri Lankan context, for example, one may otherwise fall into the trap of positing a process of linguistic acculturation which is based on ideologically biased data.

From a diachronic perspective, it is not only interesting to observe that cultural keywords such as *government* have developed variety-specific profiles of linguistic acculturation. It is also relevant to observe to what extent cultural keywords display different degrees of unity and diversity that have emerged in terms of the structures in which these cultural keywords are used. The *d/u* ratio as shown in (2) can be used to establish the degree to which a given cultural keyword is marked by a tendency towards unity or diversity. In (9) this measure is applied to *government* and its verbal collocates as the related structures under scrutiny.

$$(9) \quad d/u \text{ ratio} = \frac{303.65}{303.65 + 1098.59} \times 100 = 21.6647$$

The value for the variety-specific structures (i.e. the numerator in the multiplier) is the sum of the normalised frequencies per million words (pmw) of the verbal collocates of *government* which occurred exclusively in SAVE-IND (49.48), SAVE-PAK (177.84) and SAVE-SL (76.33). The value in the denominator is the sum of the variety-specific structures (i.e. the same value as in the numerator), and the structures shared across all the varieties (i.e. in this case the sum of the average occurrence of each verbal collocate of *government* shared by SAVE-IND, SAVE-PAK and SAVE-SL). For instance, *make* is a frequent example of a cross-varietally shared verbal collocate of *government*. *Make* occurs 20.84 times pmw in SAVE-IND, 44.71 times pmw in SAVE-PAK and 30.99 times pmw in SAVE-SL. The resulting mean value across the three datasets is 32.18 pmw and thus one of the 145 summands that constitute the sum of the cross-varietally shared structures (1098.58). For the verbal collocates of *government*, the *d/u* ratio is 21.6647, which is indicative of the dominance of a pan-South Asian usage pattern as far as noun-verb collocations are concerned. Although there clearly are a large number of variety-specific verbal collocates of *government* in each of the three SAE varieties, they are not as frequent as the verbal collocates of *government* shared by all the SAE varieties.

In comparison to *government*, there are fewer verbs which occur at least five times in the 5R-window of *terror*. To be more precise, there are only three verbal collocates of *terror* shared across SAVE-IND, SAVE-PAK and SAVE-SL, namely *say* with a frequency of 15.76 pmw, *make* with a frequency of 2.83 pmw and *attack* with a frequency of 1.85 pmw. SAVE-IND and SAVE-PAK also display a number of variety-specific verbal collocates of *terror*: see Figure 4.

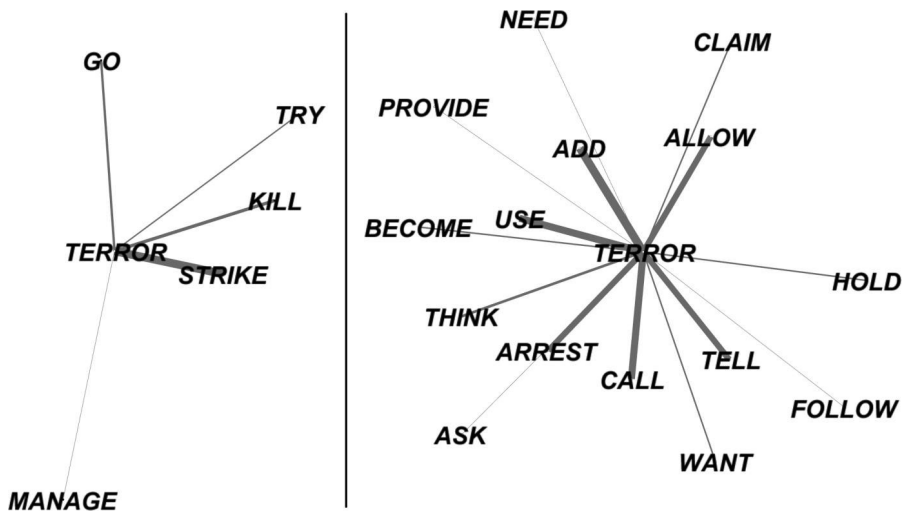


Figure 4. Variety-specific verbal collocates in the 5R-window of *terror* in SAVE-IND (left) and SAVE-PAK (right)

Although the data at hand are too limited to make any generalisations about the semantic prosodies of *terror* in any of the varieties concerned, it is nevertheless noteworthy that negatively connotated verbs such as *strike* in (10) or *kill* in (11) are most frequent in the Indian data while more neutral terms such as *add* or *call* make the top of the list in the Pakistani data.

- (10) Now that *terrorists have struck* again, the chief minister needs to answer as to whether his opposition to Poto holds water.
(SAVE-IND-SM_2002-01-23)
- (11) [...], the *terrorist who was killed* in Millennium Park near Nizamuddin, was reportedly inspired by a recorded speech of Masood Azhar.
(SAVE-IND-TI_37865)

The *d/u* ratio for *terror* is 71.09. In contrast to *government*, *terror* is thus strongly marked by variety-specific usage patterns as far as noun-verb collocations are concerned. We can thus draw the conclusion that *terror* has undergone vastly different processes of linguistic acculturation in the three South Asian socio-cultural settings. In more general terms, one could argue that terror and terrorism are concepts with connotations that reflect the national individualities of the histories of terror and terrorism rather than a shared South Asian experience.

For the cultural keyword *religion*, there are no verbal collocates shared by SAVE-IND, SAVE-PAK and SAVE-SL and only two verbs occur variety-specifically in the 5R-window of *religion*. These two variety-specific verbs are *make* in SAVE-PAK with a frequency of 1.63 pmw, as exemplified in (12), and *practise* in SAVE-SL with a frequency of 1.63 pmw, as exemplified in (13). Neither of the verbs contributes to a positive or negative semantic prosody.

- (12) Fasting in this environment also provides me with a private space where my *religion makes* sense to me and has meaning. (SAVE-PAK-DT_2004-10-23)
- (13) These values are there in all the *religions practised* in our country and are taught in our schools. (SAVE-SL-DN_2003-10-25)

As no verbal collocate of *religion* is shared by all the three SAE varieties and as *make* occurs exclusively in SAVE-PAK and *practise* in SAVE-SL, the *d/u* ratio for *religion* is 100. In principle, then, *religion* is marked by a very high degree of variety-specificity in its linguistic acculturation. To some extent, this is intuitively appealing, as the religious contexts of the three South Asian countries are markedly different, with India representing a secular country with a Hindu majority, Pakistan an Islamic nation with hardly any religious minorities, and Sri Lanka a country with a Buddhist majority. Against this background, it seems plausible that religion as a South Asian cultural keyword has undergone three markedly different processes of linguistic acculturation,

manifesting itself in variety-specific profiles of noun-verb collocations. But obviously, this conclusion is based on a limited database and therefore has to be taken with a measure of caution.

5. Concluding remarks

In the present study, we have examined aspects of the linguistic acculturation of nominal South Asian cultural keywords in IndE, PakE and SLE. Specifically, we have looked at the verbal collocates to their right in order to determine whether the profiles of noun-verb collocations are determined more by pan-South Asian patternings shared by all three SAE varieties or whether they are more variety-specific in nature. To this end, we have utilised the SAVE corpus and have introduced the *d/u* ratio as a measure capturing the degree of diversity vs unity in the lexicogrammatical routines associated with three cultural keywords: *government*, *terror* and *religion*. We have shown that the use of *government* is marked by a high degree of cross-varietally shared verbal collocates, with a more positive semantic prosody of *government* in the Sri Lankan data than in the other components. The latter finding would warrant further investigation, however, given the potential ideological bias of one of the Sri Lankan newspapers included in SAVE. In contrast, the verbal collocates as well as the semantic prosodies of *terror* seem to be largely variety-specific. *Religion* appeared to be marked by the highest degree of variety-specificity in its linguistic acculturation, but the database in the SAVE corpus is limited.

The present paper should be regarded as a pilot study in that we have tried to illustrate with selected data how corpus resources may be used to detect the effect of processes of linguistic acculturation on varieties of English. Lexicogrammatical routines such as noun-verb collocations associated with the use of cultural keywords are a particularly interesting area in which structural nativisation is strongly linked to the overarching cultural appropriation of the English language to a new postcolonial habitat. Additionally, by introducing the *d/u* ratio we have shown how the degree of unity and diversity in linguistic patternings (e.g. at the level of lexicogrammatical routines associated with cultural keywords) can be measured and compared across varieties of English. Again, although our focus has been on three neighbouring Englishes in South Asia, this approach and method can be applied to the analysis and comparison of New Englishes world-wide. It goes without saying that the description and analysis of processes of variety-specific routinisation would profit from the availability of historical data. The lack of comparable diachronic corpora remains a challenge for research into the formation of most New Englishes world-wide, including SAEs.

As for the three cultural keywords that we have looked at in the present paper, future research will certainly have to go into further detail about the way they are

routinely used in the three SAEs. For example, we need to find out more about the reasons for the comparatively positive semantic prosody of *government* in noun-verb collocations in SLE: is this really part of the linguistic acculturation of *government* in Sri Lanka, or is this an artefact of the corpus data? Also, we will have to look at lexicogrammatical routines above and beyond noun-verb collocations to the right in order to sketch a more comprehensive picture of the ways in which *government*, *terror* and *religion* are habitually used in SAEs (and, thus, of their linguistic acculturation). It certainly is high time that Stubbs' (1995, 1996, 2002) suggestions as to how to analyse the use of 'words in culture' from a corpus-informed cultural perspective are systematically applied to postcolonial Englishes. It is in these contexts that the transplantation of the English language system to an entirely new socio-cultural setting can be observed, and it is this relocation of English that leads to a new Anglophone speech community that cannot but reassess and remodel the world around it with a new linguistic vehicle.

References

- Achebe, Chinua. 1975/1993. The African writer and the English language. In *Colonial Discourse and Post-Colonial Theory: A Reader*, Patrick Williams & Laura Chrisman (eds), 428–434. London: Harvester Wheatsheaf.
- Bailey, Guy, Wikle, Tom, Tillery, Jan & Sand, Lori. 1991. The apparent time construct. *Language Variation and Change* 3: 241–264.
- Basso, Keith H. 1967. Semantic aspects of linguistic acculturation. *American Anthropologist* 69(5): 471–477.
- Beardsley, Monroe C. 1975. Figurative language. In *Introductory Readings on Language*, Wallace L. Anderson & Norman C. Stageberg (eds), 148–157. New York NY: Holt, Rinehart and Winston.
- Bernaisch, Tobias, Koch, Christopher, Mukherjee, Joybrato & Schilk, Marco. 2011. *Manual for the South Asian Varieties of English (SAVE) Corpus: Compilation, Cleanup Process, and Details on the Individual Components*. Giessen: Justus Liebig University.
- Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan & Finegan, Edward. (eds). 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Bright, William. (ed). 1992. *International Encyclopedia of Linguistics*, Vol. 1. Oxford: OUP.
- Davies, Mark. 2008–. *The Corpus of Contemporary American English: 450 Million Words, 1990–Present*. (<http://corpus.byu.edu/coca/>) (3 December 2013).
- Davies, Mark. 2010–. *The Corpus of Historical American English: 400 Million Words, 1810–2009*. (<http://corpus.byu.edu/coha/>) (3 December 2013).
- Firth, John R. 1935. The technique of semantics. *Transactions of the Philological Society* 34(1): 36–72.
- Giles, Howard. 1984. *The Dynamics of Speech Accommodation*. Berlin: Mouton de Gruyter.
- Hoffmann, Sebastian & Mukherjee, Joybrato. 2007. Ditransitive verbs in Indian English and British English: A corpus-linguistic study. *Arbeiten aus Anglistik und Amerikanistik* 32: 5–24.

- Holt, John Clifford. (ed.). 2011. *The Sri Lanka Reader: History, Culture, Politics*. Durham, NC: Duke University Press.
- Hundt, Marianne & Gut, Ulrike. (eds). 2012. *Mapping Unity and Diversity World-Wide: Corpus-Based Studies of New Englishes* [Varieties of English around the World G43]. Amsterdam: John Benjamins. DOI: 10.1075/veaw.g43
- Johnson, Jean Bassett. 1943. A clear case of linguistic acculturation. *American Anthropologist* 45(3): 427–434.
- Kachru, Braj B. 1985. Institutionalised second-language varieties. In *The English Language Today*, Sidney Greenbaum (ed), 211–226. Oxford: Pergamon Press.
- Laporte, Robert. 2002. Pakistan – political system. In *Encyclopedia of Modern Asia – 4. Malaysia to Portuguese in Southeast Asia*, David Levinson & Karen Christensen (eds), 436–440. New York, NY: Scribner.
- Louw, William E. 1993. Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In *Text and Technology: In Honour of John Sinclair*, Mona Baker, Gill Francis & Elena Tognini-Bonelli (eds), 157–176. Amsterdam: John Benjamins. DOI: 10.1075/z.64.11lou
- Lyons, John. 1968. *Introduction to Theoretical Linguistics*. Cambridge: CUP.
- Mesthrie, Rajend & Bhatt, Rakesh M. 2008. *World Englishes: The Study of New Linguistic Varieties*. Cambridge: CUP.
- Meyler, Michael. 2007. *A Dictionary of Sri Lankan English*. Colombo: Mirisgala.
- Mishra, Patit Paban. 2002. India – profile. In *Encyclopedia of Modern Asia – 3. Iaido to Malay-Indonesian Language*, David Levinson & Karen Christensen (eds), 9–13. New York, NY: Scribner.
- Mukherjee, Joybrato. 2007. Structural nativisation in Indian English: Exploring the lexicogrammar interface. In *Rainbow of Linguistics: Volume I*, Niladri Sekhar Dash, Probal Dasgupta & Pabitra Sarkar (eds), 98–116. Calcutta: T. Media Publication.
- Mukherjee, Joybrato. 2009. *Anglistische Korpuslinguistik: Eine Einführung*. Berlin: Erich Schmidt.
- Mukherjee, Joybrato & Schilk, Marco. 2012. Exploring variation and change in New Englishes: Looking into the International Corpus of English (ICE) and beyond. In *The Oxford Handbook of the History of English*, Terttu Nevalainen & Elizabeth Closs Traugott (eds), 189–199. Oxford: OUP.
- Natarajan, Swaminathan. 2012. Media freedom in post war Sri Lanka and its impact on the reconciliation process. *Reuters Institute for the Study of Journalism Fellowship Papers*. <https://reutersinstitute.politics.ox.ac.uk/fileadmin/documents/Publications/fellows_papers/20112012/Media%20Freedom%20in%20Post%20War%20Sri%20Lanka.pdf> (23 September 2013).
- Rayson, Paul. 2008. From key words to key semantic domains. *International Journal of Corpus Linguistics* 13(4): 519–549.
- Rocci, Andrea & Wariss Monteiro, Márcio. 2009. Cultural keywords in arguments: The case of interactivity. *Cogency* 1(2): 65–100.
- Schilk, Marco. 2012. *Structural Nativization in Indian English Lexicogrammar* [Studies in Corpus Linguistics 46]. Amsterdam: John Benjamins. DOI: 10.1075/scl.46
- Schilk, Marco, Bernaisch, Tobias & Mukherjee, Joybrato. 2012. Mapping unity and diversity in South Asian English lexicogrammar: Verb-complementational preferences across varieties. In *Mapping Unity and Diversity World-Wide: Corpus-Based Studies of New Englishes* [Varieties of English around the World G43], Marianne Hundt & Ulrike Gut (eds), 137–165. Amsterdam: John Benjamins. DOI: 10.1075/veaw.g43.06sch

- Schneider, Edgar W. 2003. The dynamics of New Englishes: From identity construction to dialect birth. *Language* 79(2): 233–281. DOI: 10.1353/lan.2003.0136
- Schneider, Edgar W. 2007. *Postcolonial English: Varieties Around the World*. Cambridge: CUP. DOI: 10.1017/CBO9780511618901
- Scott, Mike. 2008. *WordSmith Tools*. Version 5. Liverpool: Lexical Analysis Software.
- Stubbs, Michael. 1993. British traditions in text analysis: From Firth to Sinclair. In *Text and Technology: In Honour of John Sinclair*, Mona Baker, Gill Francis & Elena Tognini Bonelli (eds), 1–33. Amsterdam: John Benjamins. DOI: 10.1075/z.64.02stu
- Stubbs, Michael. 1995. Collocations and cultural connotations of common words. *Linguistics and Education* 7(4): 379–390. DOI: 10.1016/0898-5898(95)90011-X
- Stubbs, Michael. 1996. *Text and Corpus Analysis: Computer-Assisted Studies of Language and Culture*. Oxford: Blackwell.
- Stubbs, Michael. 2002. *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.
- Thomason, Sarah G. 2001. *Language Contact: An Introduction*. Washington DC: Georgetown University Press.
- Trudgill, Peter. 1986. *Dialects in Contact*. Oxford: Blackwell.
- Wierzbicka, Anna. 1997. *Understanding Cultures Through their Keywords*. Oxford: OUP.
- Williams, Raymond. 1976. *Keywords*. London: Fontana.

