

The interface between cultures and corpora: Tracing reflections and manifestations

Edgar W. Schneider, University of Regensburg

1 Introduction: Traces of cultures in corpora?

The ICAME 37 conference, held at the Chinese University of Hong Kong in May 2016, had the conference theme of “Corpus Linguistics across Cultures”. Clearly, this was a most timely decision – cultural studies has grown to become a major sub-discipline in the humanities, and there is also an increasing number of publications and activities that somehow deal with culture and language. But on closer inspection much of this remains fairly abstract and theory-oriented, not primarily interested in looking into language forms and structures as such – which can be studied with corpus-linguistic methods. The topic of this paper, originally a plenary at the Hong Kong conference, was thus inspired and intrigued by the conference topic. The questions I wish to ask address the empirical, formal representations of ‘culture’: Do differences between cultures find systematic manifestations in language forms, and specifically in text collections as we find them in modern electronic corpora? How and to what extent can traces of cultural impact be detected in corpora, using a corpus-linguistic methodology?

Surprisingly little research has been conducted on these questions so far, on the interface between culture and not only language but corpora. Different cultures are typically associated with different regions and nations; so are languages and emerging varieties of English, and the latter constitute a main topic in corpus study, via projects and collections such as the ‘International Corpus of English’ (ICE; Greenbaum 1996), GloWbE or NOW (corpus.byu.edu). While these corpora have become central tools for major branches of World Englishes research, this discipline has not paid much attention to the notion of culture(s) as such (Schneider 2007, 2011) – again, an interface that needs to be developed further. The goal of this paper is to address the issue of the relationship between cultures and corpora in a systematic, principled manner, and to sketch out a framework for research. Since the topic is broad and multi-faceted, my approach will necessarily have to remain somewhat exploratory and exemplary, a pilot study which later research could elaborate on.

I start by looking into definitions of the notion of ‘culture(s)’ and by briefly outlining some earlier research on language and culture, with a special eye on studies of language forms. A section on methodological issues will ask how corpora can be employed for searching for traces of culture, i.e. search terms and strategies. In the main, results, part I will then subsequently document three layers of reflections of culture(s) in corpora: cultural objects, dimensions of cross-cultural analysis, and syntactic constructions. In each of these three parts and in several sub-sections I will first outline some theoretical background, will develop a specific hypothesis, and will then describe and interpret the data and results accordingly. In conclusion, a summary evaluation will be offered.

2 Culture and language

2.1 Defining ‘culture(s)’

The discipline of cultural studies has come to be a highly popular, successful and influential branch of the humanities over the last few decades (Assmann 2012). On the other hand, the notion of culture itself is characterized by a high degree of fuzziness, versatility, and underspecified semantics (Moran 2001; Hua 2013: ch. 11). It encompasses aspects of time and space (as covered, for example, by related disciplines such as history and geography), identity and memory, gender and social roles (i.e. sociology, anthropology, and psychology), arts and the media, and many more approaches and facets. Minkov (2013: 2, 9) explicitly states that ‘culturology’ encompasses a range of varying conceptualizations. Customarily, approaches and definitions map onto a cline from relatively concrete to abstract modes of understanding the notion. At the concrete end, there is a straightforwardly material ‘culture-as-content approach’ which looks into objects, customs, social hierarchies, etc. (Hua 2013: 4). Relatively more intermediate is an understanding of sociopsychological elements of culture (Minkov 2013: ch. 3), including meanings, rituals, taboos, institutions, values, norms, beliefs, attitudes, behaviors, and the like. At the abstract end culture has been defined as “the collective programming of the mind that distinguishes the members of one group or category of people from another” (Hofstede 2001: 9).

Clearly, this continuum also reflects a narrow vs. wider understanding of ‘culture’ and the objects of cultural studies. The relatively narrow notion entails a focus on objects, artefacts, food, clothing, religious and social terms, and so on – a rather encyclopedic, concrete set of terms. The wider notion, in contrast, highlights knowledge systems, beliefs, values, assumptions, and behavior as core elements, much more abstract objects of investigation (Hua 2013: 86-87; cf. Moran 2001: 24–25). This range of definitions and topics corresponds to the

three layers, or domains, which will be investigated below: terms and objects, dimensions, and schematic constructions, respectively.

2.2 Language and culture

The interrelationship between language and culture is not a core topic in linguistics, although the issue has been addressed from various perspectives. However, all of these discussions remain on a fairly abstract, generic or theoretical level; reflections of cultures are commonly identified in cognition systems and pragmatic conventions but not in linguistic forms.

Perhaps the best-known line of thinking closely related to this topic is the theory of ‘linguistic relativity’, widely known also as the ‘Sapir-Whorf-hypothesis’ after its main early proponents, which argues that the structure of a language shapes patterns of cognition in the community of speakers of that language (cf. Salzmann et al. 2012: 225–256; Hua 2013: 173–176; Leavitt 2015; Sharifian 2017: ch. 12). It has been debated widely, and in various formats. A strong version of ‘linguistic determinism’ held that linguistic structures actually determine frames of thought (so that, for instance, the community of speakers of a language without past tense morphology was projected not to conceptualize pastness); this is no longer seriously assumed today. In contrast, the weak version of ‘linguistic relativity’ suggests that language influences cognitive categories. This hypothesis is still under investigation, and it seems to be valid in certain semantic domains, though only weakly. The theory has triggered many debates, essentially still ongoing and related to cognitive linguistics; no real consensus has been reached so far. It seems clear that an interrelationship of some kind is evident, but there is no strict monocausal relationship.

There is a recent, growing branch in linguistics known as ‘cultural linguistics’ (see a recent book with that title, Sharifian 2017; or the *International Journal of Language and Culture*, founded in 2016). Its orientation is fairly different from the corpus-based, empirical approach advocated here, however – Sharifian (2017: 47) has a one-paragraph section on “corpus-based analysis”, in which a single, programmatic paper (Jensen 2017) is summarized. It considers “cultural cognition” and its linguistic manifestations via “cultural conceptualizations” (Sharifian 2011). Its focus is on “cultural key words” (cf. Hua 2013: 180–181; Levisen and Waters 2017), schemas, “conceptual metaphors” (Wolf and Polzenhagen 2009), culture-specific conceptualizations and intercultural pragmatics, or embodiment, generally arguing on a fairly abstract level and explaining concepts and notions peculiar to specific cultures. The *Routledge handbook of language and culture* (Sharifian, ed. 2015) covers anthropology, cultural psychology, politeness, cognitive structures, and the like.

From a broadly pragmatic angle there has been some work on intercultural communication (e.g. Hua 2013), characterized by applied approaches and with a focus on conversational routines and politeness issues, notably for business encounters. This is closely related to the teaching of intercultural skills as an important component of teaching foreign languages in general (cf. Byram 1997; Byram, Nichols and Stevens 2001).

In contrast, empirical analyses on structural preferences as reflections of cultural differences have been extremely rare. Occasionally some speculative hints have been offered, as for example in Olavarria de Errson and Shaw (2003). This example, and some similar suggestions and speculative hints, will be briefly discussed in section 4.3.1. An interesting investigation along such lines can be found in a paper on “linguistic acculturation” (Mukherjee and Bernaisch 2015), based on the South Asian Varieties of English (SAVE) corpus. The authors investigate “lexicogrammatical routines associated with ‘cultural keywords’” (2015: 415); i.e. they carry out a verb co-occurrence analysis of the noun lemmata *government*, *terror*, and *religion* and show in a network visualization which co-occurrence relationships (specific verbs going together with specific nouns) are peculiar to or shared across the varieties under investigation.

3 *Tracing cultures in corpora: Methodology*

The present study searches for and quantitatively assesses formal surface manifestations of specific cultural traits and orientations, with both components (the forms and cultural categories) identified and decided on beforehand, based on earlier writings on the subject and some hypothesizing of my own. Like many others, it is based on a comparative analysis of selected components of the ICE corpus project. The ICE corpora clearly have advantages and have turned out to be quite suitable for such a comparative investigation: they are well balanced and almost fully comparable in structure; they represent both speech and writing as well as various genres and styles, and the proportion of 60 per cent of spoken texts guarantees a strong presence of localized usage (since speaking is less strongly subject to the norming tendencies found in writing). The magnitude of the forms searched for, mostly mid-frequency expressions and structures, has yielded largely reasonable numbers for comparison and analysis. I have chosen five ICE corpora, representing English as used in these countries, as typical of major cultural orientations: Great Britain (ICE-GB), representing a western culture and also the baseline, input variety in most postcolonial contexts; Hong Kong (ICE-HK), Singapore (ICE-Sing) and India (ICE-Ind) as three important Asian cultures, which may be taken to share an Asian collectivist orientation but

also vary along important parameters: Singapore and Hong Kong represent two Chinese-dominated, Buddhist-influenced cultures, which have undergone varying degrees of modernization and westernization; and India is a more conservative, non-Confucian, Hindu country. In contrast, Nigeria (ICE-Nig) was selected as a prototypical representative of (West) African cultures. In a single case study, in 4.1., I have also supplemented this design by data from New Zealand (ICE-NZ). And occasionally, largely on an ad-hoc basis determined by the issue in question, I add select data from the 'Global Web-Based English' (GloWbE) corpus (Davies 2013; Davies and Fuchs 2015), which, as is well known, offers a substantially larger magnitude (with a corpus size of 1.9 billion words overall) but is less balanced and thus perhaps less representative of a variety overall, since it consists predominantly of blogs and online newspapers.¹

As stated earlier, I investigate three different object domains (layers) of manifestations of cultures (objects, dimensions, and constructions, respectively). In each case (the sub-sections of section 4), I offer a brief outline of the theoretical background, followed by the formation of a hypothesis as to possible manifestations of the layer in question in corpora, and then proceed to the documentation and interpretation of the data. The database consists of records of what I call 'manifestation forms' of cultural objects, specific sets of forms and structures selected beforehand and assumed to represent and formally express characteristics of local cultures. These forms have been searched in the study corpora, and I offer frequency-based documentation and comparisons, statistically tested for possible significance of distribution differences.

Obviously, then, these manifestation forms, the search terms, constitute a central methodological tool in this investigation. Some are rather straightforward and self-explanatory, while others may need some justification and discussion in individual instances. For the first layer of investigation clearly the lexemes denoting cultural terms and objects, typically words borrowed from indigenous languages, serve as natural manifestation forms in the sense defined above. I identified and compiled such terms from both linguistic publications (e.g. many of the Maori terms addressed in section 4.1.1 from Schneider 2007: 129) and other sources, including websites, on local cultures. For the second layer, cultural dimensions, I decided on a set of 'indicator terms' in each case – words or phrases assumed to indicate, reflect or be associated with the cultural domain or issue in question, culled from the literature on (cross-)cultural theory, from socio-psychological writings, or similar sources. Obviously, these sets of indicator terms call for discussion and justification; they have been decided on and picked for the present study but in principle they could be expanded or alternated. Thirdly, for schematic constructions I defined a tightly circumscribed set

of ‘indicator structures’, linguistic surface patterns assumed to reflect or possibly to have been strengthened by cultural schemes. In order to tease out any impact of alternative cultural orientations, in this case the strategy was often to set up two functionally equivalent but formally different structural options putatively representing alternative cultural orientations. Thus, I obtain a controlled ‘envelope of variation’ in the sense of Labov’s (1972) variation theory, controlling for lexical influences and designed to maximize precision and recall by mostly employing high-frequency verbs, personal pronouns, and similarly constrained choices. Frequencies are then compared, and differences interpreted.

Clearly, in the practical analyses in the corpus investigations a number of decisions in detail had to be taken, and a few considerations and caveats need to be considered. Corpus sizes did not turn out to be a problem: On the whole, searches in the ICE corpora yielded reasonably many hits for a cross-variety comparison. Comparisons in GloWbE were carried out mainly for a few low-frequency constructions. Lexical search terms were mostly lemmatized; i.e. I searched for inflectional variants of nouns and verbs as well, and sometimes (indicated in the specific instances below) I also pooled derivation forms and cognates belonging to different word classes but denoting the same concept (e.g. *Jain* / *Jainism*; *promiscuous* / *promiscuity*). Frequencies were counted in Ant-Conc, using the WordList and Concordance tools (the latter when context mattered for disambiguation or excluding unwanted instances). As far as was reasonably possible non-target homograph forms were excluded, e.g. when homonyms (unwanted formally identical items) were returned (for example, a search for *waka*, a New Zealand cultural object, yielded the same form in ICE-Nig, which turned out to be a Pidgin form of *walk*), or when polysemic uses (alternative meanings) different from the target concept came up (as in the collocation *naked truth* when the target word *naked* was searched as an indicator term for sociosexuality). Hence, idiosyncratic inclusion or exclusion decisions in specific instances may entail a small fringe of uncertainty as to the precision of the frequencies reported, but on the whole the figures are robust. In principle, of course, the formal identifiability of cultural notions and dimensions, i.e. the choice of manifestation forms and indicator terms assumed to represent cultural phenomena, remains a basic issue, obviously. For example, when Minkov (2013: 426) states that “pride is almost a sin” in Asian cultures the assumption is that this cultural concept can be tested by the presence or relative frequency of the lexemes *proud* and *pride* – but clearly this assumption is open to debate.

The quantitative distributions of the token frequencies of indicator terms across different corpora were statistically assessed by comparing that number relative to the respective corpus size (which by subtraction yields the number of

non-occurrences) in any given corpus to all other frequencies. Basically this follows a procedure suggested and documented by Paul Rayson (n.d.), though the Fisher's Exact test (suitable for small token numbers) was employed instead of a log likelihood test, with Bonferroni correction and automatic adjustment of benchmark p values when several pairwise comparisons are conducted (Gries 2008: 243–244). An R script written by Thomas Brunner (for which I am grateful to him) was consistently used for these calculations. Significance levels are shown in the conventional fashion by asterisks: $p < .001 = ***$; $p < .01 = **$; and $p < .05 = *$. All pairwise distribution comparisons between corpora were assessed, but for reasons of space and clarity in the tables reproduced below significance symbols relate to the pairwise comparison between GB (the donor and benchmark variety, as it were) and Asian/African Englishes, respectively.

The search for indicator terms in all varieties and for all layers of investigation turned out to be fairly time-consuming and laborious, and the results are rich and versatile overall. Especially for the middle layer, the study of dimensions of culture, the breadth of possible issues and data was remarkable. As many as 15 different dimensions, some sub-divided into more specific manifestation types, were tested. Given that amount of data and details, in the present context the presentation of results has to remain selective by necessity. I therefore focus on mainly results on the forms and dimensions which have yielded statistically significant or otherwise interesting distributions, and refrain from reporting a few others – which will be mentioned briefly but not documented in detail.

4 Cultural reflections in linguistic form

4.1 Cultural terms and objects

The first, most simplistic layer analyzes cultural terms and objects, the most straightforward and directly evident manifestations of cultures, in line with 'naive' assumptions of what constitutes a culture. Hua (2013: 4) calls this the "culture-as-content approach", which highlights the "four F's", i.e. "food, fairs, folklore, facts": typical objects and artefacts; food and clothes; social roles, values, and the like. In language contact settings, and consequently in many localized varieties of English, such terms are typically not translated but rather denoted by terms borrowed from indigenous languages.

Since these terms relate to regionally restricted cultural manifestations and habits, it is likely that they are also regionally constrained. The obvious hypothesis on their distribution is that localized use is likely; i.e. they can be expected to be largely restricted to corpora from their origin countries. Exceptions are

possible, of course, and even likely in small numbers: in times of globalization some outstanding local culture notions from various regions have come to be transnationally known and talked about. Hence, a small number of terms which have become internationalisms, signaling local cultures from an outside perspective, may be expected to spill over to other corpora as well.

4.1.1 *Objects of New Zealand / Maori culture*

The culture of the Maori in New Zealand is characterized by a large number of distinctive objects, notions and terms; and the fact that this culture is generally recognized as an important component of the country's heritage, more so than in historically comparable nations, prompted the question to what extent this is reflected in New Zealand English texts (or in other varieties). Since I considered this a particularly interesting, special case, I added it to the studies of the core corpora for comparison. Table 1 shows the distribution of seven terms which are typical of New Zealand, all but the last one borrowed from the Maori language. Four more (*kea* 'kind of bird', *tuatara* 'kind of reptile', *hangi* 'earth oven', *hongi* 'traditional Maori greeting pressing one's nose') yielded lower figures and are not reported here.

Table 1: New Zealand cultural terms

	GB	NZ	Hong Kong	Singapore	India	Nigeria
<i>kiwi</i> 'kind of bird'	0	129***	4	1	0	1
<i>kauri</i> 'kind of tree'	0	16**	0	0	0	0
<i>waka</i> 'canoe'	0	24**	0	0	0	0
<i>whare</i> 'house'	0	26**	0	0	0	0
<i>mana</i> 'prestige, standing'	0	39***	0	0	0	1
<i>pakeha</i> 'person of European descent'	0	183***	0	0	0	0
<i>all black/s</i> 'New Zealand's rugby team'	0	120***	0	0	0	0

The results strongly confirm the hypothesis: the Maori and New Zealand English terms are all very frequent in New Zealand English but almost completely restricted to that variety, not used anywhere else. The only exception is *kiwi*, which has become an internationalism and comes up in most of the other corpora as well. This distribution builds a strong case for the special role of cultural objects as culture indicators.

4.1.2 Objects of Hong Kong culture

Table 2: Hong Kong cultural terms

	GB	Hong Kong	Singapore	India	Nigeria
<i>dim sum</i> ‘Cantonese dish’	0	10*	0	0	0
<i>triad</i> ‘organized crime syndicate’	0	13**	7	0	0
<i>majhong</i> ‘Chinese game’	1	42**	3	0	0

The distribution of terms associated with Hong Kong culture is similar, though less straightforward, partly due to lower token numbers. Three terms, the statistically significantly distributed ones, are shown in Table 2; three more (*dragon boat*, *red packet* ‘monetary gift’, *canto-pop*) show low figures but similar distributions; *spirit money* ‘bank notes on incense paper for offerings’, *lucky money* ‘monetary gift’ and *mooncake* ‘Chinese bakery product’ were not found at all.

All these expressions are common and focussed in Hong Kong itself. Some spill-over effects to and similarities with Singapore can be observed, presumably through the shared Chinese culture input. There are hardly any occurrences elsewhere (except *mahjong*, once in ICE-GB).

4.1.3 Objects of Indian culture

Table 3: Indian cultural terms

	GB	Hong Kong	Singapore	India	Nigeria
Jain/ism	2	0	0	15*	0
<i>lakh</i> ‘one hundred thousand’	0	0	0	66**	0
<i>chutney</i> ‘Indian sauce’	0	0	1	9*	0
<i>sari</i> ‘female Indian garment’	0	0	3	5	0

As before, Indian cultural terms are significantly focussed in India and not found elsewhere. With lower figures, but also in India only, the words *Namaste*, *tikka* ‘type of Indian food’, *Ayurveda*, and *Mughal* have been recorded. *Dhoti* ‘male garment’ and *lunggi* ‘male garment’ were not found in the corpora. *Jain* and *sari*, with highly limited outside occurrences in GB and Singapore, are on their way towards becoming internationalisms.

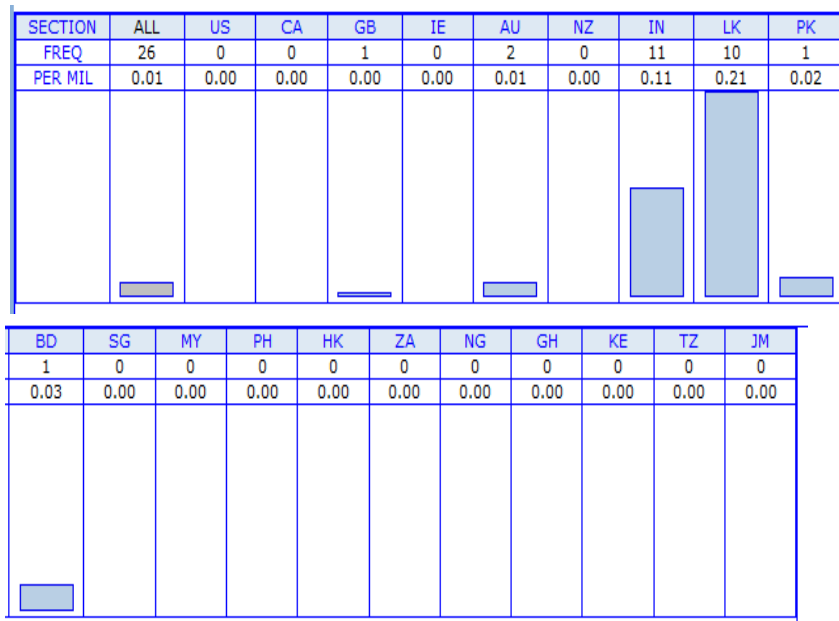


Figure 1: bed tea in GloWbE

Figure 1 shows *bed tea* ‘early morning serving of tea’, a distinctly Indian and Lankan expression, in GloWbE. In addition, it occurs only occasionally in the two other South Asian countries and in countries where substantial numbers of South Asian immigrants live, GB and Australia.

4.1.4 ‘Deculturation’: Objects of British culture

Kirkpatrick (2015: 460) argues that some varieties show signs of ‘deculturation’, the opposite of acculturation, “where the new variety of English divests itself of cultural references to older varieties, such as British English.” To test this, I checked for nine terms considered typical of British culture. Four of them are displayed in Table 4. *Tea time*, *scones*, and *cheddar* were found in relatively low numbers, but, interestingly enough, not in GB but in Singapore, Hong Kong and India. *Rice pudding* and *plum pudding* did not occur in any of the corpora.

Table 4: Objects of British culture

	GB	Hong Kong	Singapore	India	Nigeria
duke	20	8*	1***	3**	4*
pub	25	11*	12	2**	3**
pie	11	6	19	15	7
afternoon tea	4	1	0	1	0

The results confirm Kirkpatrick's hypothesis in parts. Some terms (*duke*, *pub*, *afternoon tea*) show comparatively high usage figures in GB as opposed to lower ones in former colonies, a fact which can be interpreted in the light of deculturation. In other cases we find clear evidence of the retention of elements of the British colonial culture (cf. *pie*, *tea time*, *scones*, *cheddar*).

4.1.5 Summary of results: Cultural objects

In line with the hypothesis developed in 4.1., the distributions found are extremely straightforward, focused, and in line with expectations. Terms for cultural objects, notions and artefacts are used almost exclusively locally, with a few exceptions: some 'indicator terms' have spilled over across their original confines, are found in other corpora as well, and show transnational diffusion. They have become internationalisms with a pointer function to local cultures. Cases in point include *kiwi*, *sari*, *triad*, and others. Some traces of both British culture leftovers and deculturation in former colonies have also been worked out. On the whole, thus, and not surprisingly, cultural terms and objects turn out to be solid reflectors of cultures.

4.2 Cross-cultural analysis: Dimensions of cultures

The second layer investigates reflections of 'dimensions of cultures' as posited in the socio-psychological discipline of 'cross-cultural analysis'. The central idea of this approach is that differences between cultures can be captured along specific 'dimensions', which reflect consistent thematically ordered attitudes and values in a society. Studies in this discipline are usually based upon psychological questionnaire data which systematically collect reactions to stimulus statements given by subjects from many different regions and cultures, sometimes also on the systematic observation of behavior in communities. Typically, the results of such studies report broad regional, sociocultural tendencies, often as scales or index values of a dimension relative to a society. The orientation of this discipline has tended to have a strongly applied character, with an overall

focus on organizational culture, managerial and leadership styles, and business-directed applications in cross-cultural behavior.

The American anthropologist Edward Hall, with work between the 1950s and 1970s (cf. Hall and Hall 1990), counts as the founder of the study of intercultural communication as a branch of anthropology. He posited ‘high-context’ vs. ‘low-context’ cultures and other dimensions (including proxemics and monochronic vs. polychronic time). The discipline’s key figure and best known, most influential representative is the Dutch sociologist Geert Hofstede, who explicitly posited and studied “dimensions of cultures” (for a useful summary, see Minkov 2013: 201–216). Hofstede (2001; originally 1980) has been immensely influential in the social sciences; the book counts as a milestone and core reference of cross-cultural psychology to the present day, and has been very often and widely referred to. Hofstede’s ideas have been tested in and have inspired very many follow-up projects and a very large number of publications (more than 1400 alone on the individualism – collectivism dimension, following House et al. 2004: 437). A huge follow-up project, involving 170 researchers and covering 61 societies, was the ‘GLOBE’ project (House et al. 2004; described in Minkov 2013: 310–329). But while cross-cultural psychology has grown to be a strong research tradition in sociology and business studies, it has had surprisingly little impact in the humanities and in linguistics. There is not a single reference to this school in the *Handbook of Language and Culture* (Sharifian 2015).

4.2.1 Cultural dimensions: Collectivism vs. individualism

Kashima and Kashima (2003: 125) state that “[i]ndividualism and collectivism ... are without a doubt the most significant constructs in research on culture and psychology”. This is Hofstede’s (2001) most central, most uncontroversial dimension, one which has “attained the status of paradigm in cross-cultural psychology” (House et al. 2004: 437). It is conventionally associated with a regional division, with western cultures valuing individualism as opposed to eastern and Asian cultures in which a collectivist community orientation is called for (Fang 2012: 28). There is a putative reason behind these differences, namely different economic types (Minkov 2013: 428–433): European-rooted cultures can be traced back to hunter-gatherers and herders who had to compete for scarce land, and hence the strength of individuals constituted a decisive advantage for a community’s survival. In contrast, pastoralism as practiced in Asia, especially the cultivation of rice involving complex communal irrigation systems, is a labor-intensive, necessarily collective activity, so the reconciling and downplaying of possible differences, the blending in of individuals to com-

munity interests and structures, is of utmost importance. East Asian languages and cultures are proclaimed to more or less avoid an ‘I’ (Minkov 2013: 428). Of course, this is essentially a graded phenomenon: the results by Hofstede and others typically yield index values rather than absolute or qualitative distinctions.

My hypothesis is thus that there will be a correlation between sociocultural indices of collectivism vs. individualism and the frequency of indicator terms representing these dimensions.

There are very few extant linguistic applications of this line of thinking and of Hofstede’s concepts in general. An unpublished paper which investigates singular and plural pronoun usage as indicators of individualism (Fuchs 2012) finds a correlation between the individualism dimension and uses of 1st person singular pronouns in ICE corpora. Uz (2014) documents a weak correlation between the frequencies of first person singular and plural pronoun use in Google’s Ngram database with Hofstede’s individualism scores. Yu et al. (2016) also compare pronoun frequency data from Google’s Ngram database. Across nine languages, they find British and American English, respectively, showing the highest rates of first-person singular pronoun usage, and Chinese the lowest in 1949, but one rising substantially until 2008.

Hypothesizing that individualism will entail a higher usage of the first person singular pronouns and collectivism will result in more frequent usage of the first person plural pronoun, Figure 2 shows the proportion between both in the ICE corpora investigated.

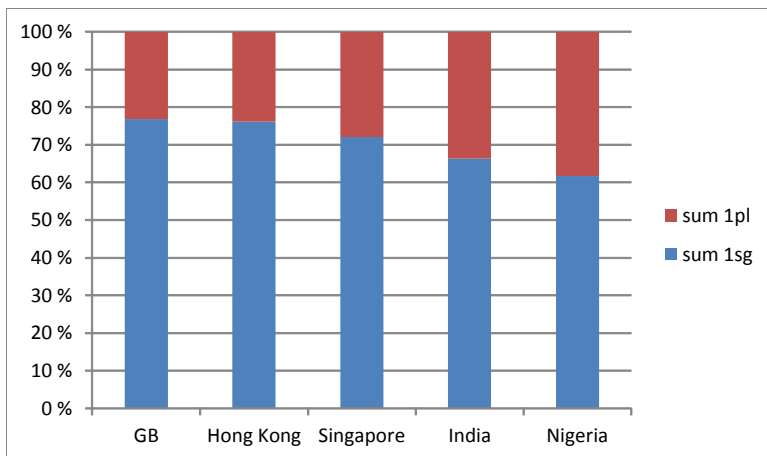


Figure 2: First person singular vs. plural usage

The results show a highly significant distribution overall ($p < .000$) (and also for all pairwise Fisher's Exact comparisons except HK:GB), largely in line with the hypothesis: GB displays the strongest proportion of 1st person singular usage (77% of all first person personal pronouns; $n = 29,906$), as opposed to Singapore (72%), India (66%) and Nigeria (62%), which thus have relatively higher proportions of first person plural pronouns, taken to be indicative of a more collectivist orientation.

Table 5: Collectivism-oriented indicator terms

	GB	Hong Kong	Singapore	India	Nigeria
take care of	9	70***	54***	79***	93***
protect*	118	215*	186**	147	136
loyal*	19	27	13	28	28
harmon*	11	57***	43***	36**	27
share	49	77	131***	42	55
together	270	471***	350	259	260
concerned about	19	26	22	13	15
sensitive	3	45***	29***	51***	25***
sum	498	988***	828***	655**	639***

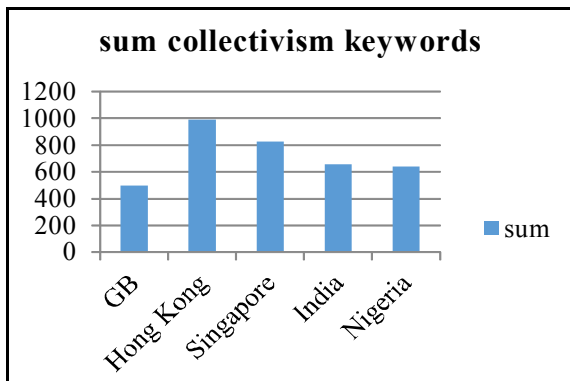


Figure 3: Sum total of collectivism-oriented indicator terms

Table 5 reports frequencies of collectivism-oriented indicator terms, and Figure 3 visualizes their overall distribution. The hypothesis is strongly confirmed: all keywords appear substantially more frequently in the Outer Circle than in GB

(especially in Hong Kong and Singapore), some several times as often, with very few exceptions; and many of the comparisons are statistically highly significant.

The robustness of these distributions and relationships and the validity of the hypothesis can also be demonstrated by calculating correlations. Hofstede (2001: 215) reports an “individualism index value” for many countries; it is 89 for GB, 25 for Hong Kong, 20 for Singapore, 48 for India, and 20 for Nigeria. Calculating the Pearson correlation coefficient between these index values and the percentages of 1sg pronouns yields a positive correlation ($r=.44$); the same coefficient between Hofstede’s index and the frequency of collectivism indicators shows a strong inverse correlation of $r = -.71$.

Figure 4 (partly) shows a piece of evidence from GloWbE which is largely in line with that of ICE: as with ICE, the GloWbE evidence is inconclusive for *concerned about* but strongly in line with the hypothesis for *sensitive towards*, which is very strong in India and across Asia but low in GB and the ‘Inner Circle’.

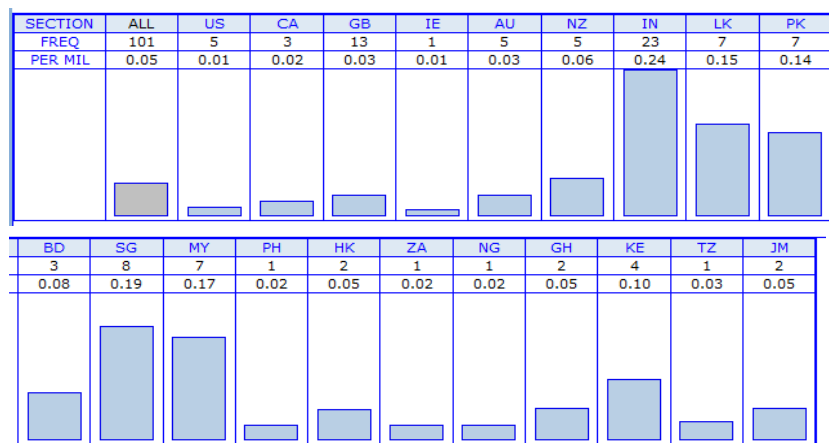


Figure 4: sensitive towards in GloWbE

Thus, the overall results for the important individualism vs. collectivism dimension yield evidence which is quite strongly in line with the hypothesis: indicators of collectivism predominate across Asia (when compared to GB), and are also strong in Nigeria.

4.2.2 Cultural dimensions: 'Long-term vs. short-term orientation'

The dimension of long-term orientation captures “the extent to which a culture programs its members to accept delayed gratification of their material, social, and emotional needs” (Hofstede 2001: xix–xx; similarly 29). Hofstede’s results can be translated almost directly into my distributional hypothesis: “East Asian countries scored highest, Western countries on the low side, and some Third World countries lowest.” (351) The indicator terms chosen to test this dimension were *persist*, *continuity*, *long-term*, *postpone*, *delay*, *later*, and *lifetime*.

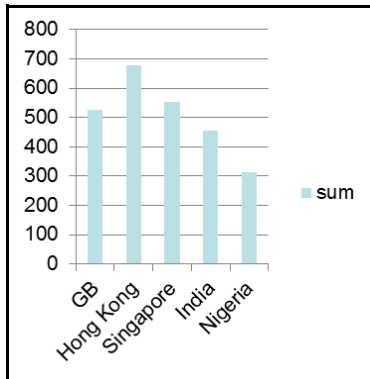


Figure 5: Sum total of indicator terms for long-term orientation

The overall results, summarized in Figure 5, are largely in line with the hypothesis. Long-term orientation terms appear most frequently in Hong Kong, and in Singapore also more commonly than in GB. Their frequency is lower in India and lowest by far in Nigeria.

In contrast, the distribution of a set of short-term orientation terms turned out to be inconclusive.

4.2.3 Cultural dimensions: Power distance

This dimension reflects “the extent to which the less powerful members of organizations and institutions accept and expect that power is distributed unequally” (Hofstede 2001: xix), reflecting the basic social fact of human inequality and attitudes towards it. Hofstede measures power distance indices; within an overall range of 104 to 11, India scores 77, West Africa 77, Singapore 74, Hong Kong 68, and GB 35 (2001: 87). This dimension is strongly inversely correlated with individualism (House et al. 2004: 441).

Amongst various forms of expression which may be associated with this dimension I choose three: the use of address terms, indicators of status differences, and the use of polite requests.

Table 6: Honorifics and terms of address

	GB	Hong Kong	Singapore	India	Nigeria
sir	194***	144***	355***	997***	878***
madame	9	180***	31*	90***	39***
respected PN/Title	0	0	0	22***	0
respected	9	11	9	34**	18

The results for honorifics and formal, distancing address terms, shown in Table 6, are strongly in line with the hypothesis: The traditional, respectful address terms *sir* and *madam* are used much more commonly across Asia and Africa (except for *sir* in Hong Kong), with significant distributions throughout. Interestingly, *respected* is predominantly an Indian English term (cf. Kachru and Smith 2008: 49, 136), and ‘*respected* + personal name/title’ appears exclusively in India in ICE. Similarly, in GloWbE, *dear respected* occurs mainly in Malaysia and is also strong in India, Pakistan and Bangladesh; *respected* + *PN* is strongest in Pakistan, Sri Lanka, and India.

As to indicator terms for status differences, the distribution of low status difference indicators (*equal*, *at leisure*, *consult*) turned out inconclusive, but some high status difference indicators (out of 16 terms tested), presented in Table 7, show significant or interesting patterns.

Table 7: Indicator terms for high status differences

	GB	Hong Kong	Singapore	India	Nigeria
respect	122	155	147	204**	176**
authority	156	292**	64***	66***	105*
embarrass*	54	48	21***	19***	23**
beg	23	12	41	30	63***
disgrace	3	0	1	0	12
humble/-y	16	7	7	9	24

The term *respect* occurs more frequently in ESL countries than in GB, mostly in India and also Nigeria. *Authority* appears strongly in Hong Kong but is comparatively weak in Singapore, India, and Nigeria. Interestingly, *embarrassing* and *embarrassment* are strongly British notions. The strong distance indicators *beg*, *disgrace*, and *humble* are typically Nigerian expressions.

Table 8: Polite requests as power distance indicators

	GB	Hong Kong	Singapore	India	Nigeria
would you	124	208	90	110	50***
why don't you	19	39	45*	35	10
would like to	79	243***	126*	193***	62
ask you to	8	39*	11	3	11
appreciate it if	2	3	7	0	0
sum	232	532***	279	341**	133***

Finally, as Table 8 shows, polite requests are a strong means of signaling power distance across Asia (in line with the hypothesis) but, in contrast, are very rare in Nigeria. Overall, these terms are very frequent, especially in Hong Kong, and also in India. In particular, *why don't you* and *would like to* are strongly Asian phraseologisms.

4.2.4. Cultural dimensions: 'Social relations'

Table 9: Kinship terms as indicators of social relations

	GB	Hong Kong	Singapore	India	Nigeria
uncle	13	39	53***	14	28
aunt/ie	31	76*	42	26	17
brother	59	195***	121***	152***	216***
cousin	35	39	21	60	29
sum	138	349***	237***	252***	290***

Table 9 shows distributions of kinship terms, a mainstay of signaling culture-specific social relations (cf. Kronenfeld 2015; Kachru and Smith 2008: 49–50, 108), and there are indeed interesting differences, indicative of cultures. Kinship terms are generally highly significantly less frequently used in Great Britain –

which is remarkable and may be taken to reflect the reduced importance of family bonds in western cultures. The values for *uncle* and *aunt* are high in Singapore and Hong Kong, but female *aunt* is low in Nigeria. The frequency of *brother* is high across Asia and even more so in Nigeria. *Cousin* appears most commonly in India.

Another notable manifestation of social relations are indicators of politeness, known to vary across cultures. The most common expression of politeness, the word *please*, occurs significantly less commonly in Great Britain than in all ESL varieties, as is shown in Table 10 and also, based on GloWbE data, in Figure 6. Table 10 also shows that *courtesy* appears remarkably frequently in Nigeria. The adjectives *polite* and *impolite* themselves are not distributed in any noteworthy fashion; *arrogant* appears more frequently in Great Britain but the difference is not statistically significant.

Table 10: Indicator terms for politeness as a manifestation of social relations

	GB	Hong Kong	Singapore	India	Nigeria
please	209	410***	410***	342***	685***
courtesy	3	2	6	6	16*

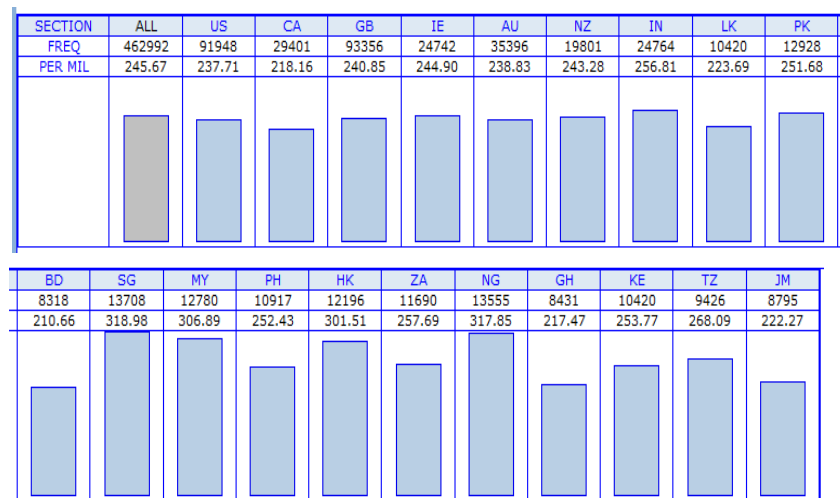


Figure 6: please in GloWbE

Overall, it can thus be stated that in line with the hypothesis politeness expressions occur more frequently in Asia and in Nigeria.

4.2.5 Cultural dimensions: Emotions

Expressing emotions in public clearly is also a culturally sensitive domain. Two main observations on characteristics of cultures are possible: (a) the strength of expressions of emotions, and (b) their predominating directionality (positive or negative) (cf. Minkov 2013: 345–349, after work by Kuppens et al. 2006). Minkov (2013: 347–348) claims that Asian cultures score relatively low on positive and high on negative emotions.

Table 11: Summary of indicator terms for emotions

sum	GB	Hong Kong	Singapore	India	Nigeria
positive	296 (56%)	484 (55%)	337 (55%)	372 (67%)	446*** (68%)
negative	233 (44%)	392* (45%)	276 (45%)	187 (33%)	210 (32%)
sum	529	876	613	559	656

Terms tested and counted as indicators of this dimension are: *happy/iness, cheerful**, *grateful/itude, pleasant, content/ment; unhappy/iness, grief/ve, guilt/y, jealous/y, anger/gry, sad/ness, unpleasant, and worry*. Table 11 summarizes their overall frequencies. It documents surprisingly many emotion terms in use in Hong Kong but relatively few in Great Britain (and in addition, this counts especially for terms for positive emotions, e.g. *happy*). India and Nigeria are strong in the tally of positive terms, with comparatively few negative ones. So in general the table offers evidence for relatively ‘constrained’ (Great Britain) vs. more ‘emotion-friendly’ (esp. Hong Kong) cultures, and identifies Nigeria and India as more strongly positively oriented ones.

4.2.6 Cultural dimensions: Sociosexuality

The final dimension to be discussed here goes back to work by Schmitt (2005; cf. Minkov 2013: 341–343), who developed scores of ‘sociosexuality’, based on mating strategies, display of courtship and romantic closeness, emotional investment, attitudes to monogamy versus promiscuity, and so on (cf. Carbaugh 2005: 61–68). Sub-Saharan Africa in particular is reported to be characterized by strong mating competition and high violence rates in this context (Minkov 2013: 430–433).

Thirteen indicator terms were tested for this dimension. Of these, eight show significant distributions (*sex*, *sexual*, *girlfriend*, *boyfriend*, *pregnant/cy*, *contraception/ives*, *naked*, *abortion*), but five others do not and have low token figures (*sexy*, *premarital*, *promiscuous/ity*, *topless*, *nude/ity*). Interesting differences in the overall frequencies of these lexical items can be observed. Sociosexuality seems a strong phenomenon in Hong Kong in particular (421 occurrences of indicator terms overall) and also in Nigeria (273; the difference between both figures and the one for GB is highly significant). In contrast, and not surprisingly given cultural stereotypes, British people address sexuality only very reluctantly (only 172 uses of all terms), with Singapore and India in between (213 and 209, respectively) but also on the relatively low side.

Table 12: Select indicator terms for sociosexuality

	GB	Hong Kong	Singapore	India	Nigeria
boyfriend	17	128***	46*	16	9
girlfriend	27	43	26	8**	36
contraception	4	1	3	51***	7
pregnant/cy	14	44*	23	38*	76***
abortion	9	8	17	12	44***

Table 12 points out the distribution of some select concepts in this domain. Being or having a *boyfriend(s)* is a strong discourse topic in Hong Kong and also, not equally predominantly, in Singapore. On the other hand, *girlfriend(s)* is more a topic of discourse than *boyfriends* in Great Britain and Nigeria but one almost not addressed at all in India. In contrast, interestingly, *contraception* is a strong topic in India, and so are *pregnancy* and *abortion* in Nigeria. Overall, fairly clear cultural differences have manifested themselves with respect to this dimension.

A few more cultural dimensions based on Hofstede's approach have been investigated but the results will not be reported here, partly due to space constraints and partly because they have not yielded clear and interesting results. These are: high/low context (the amount of detail provided in a society on the background of a statement); time orientation (mono- or polychronic, with pref-

erences for doing things either subsequently or simultaneously); proxemics (one's personal space negotiation); and a mixed bag of 'Chinese Culture Connection' terms.

4.2.7 Summary of results: Cultural dimensions

Overall, a substantial proportion of the indicator terms show significant distributions across the corpora, and very often these have been found in line with culturally-based hypotheses. So yes, cultural dimensions as posited in cross-cultural anthropology do show manifestations and substantial traces in corpora and language forms, defined as sets of indicator terms. Yet this is not the whole story: manifestations have not been identified consistently. They are strong for some dimensions and many terms, but for other distributions they are random or show token frequencies too low to be meaningful. So the basic hypothesis that cultural dimensions would show reflections in corpora is confirmed largely but with some limitations. Given the nature of this paper as a pilot study one question is whether more or other indicator terms would have produced different results.

The various dimensions clearly behave differently with respect to their manifestations in corpora. Table 13 offers a summary assessment of varying degrees of sensitivity by symbolizing whether the impact of the respective dimension (including the ones not reported in detail here) on linguistic forms is strong, medium or weak. Dimensions where linguistic manifestations are readily found are collectivism, address terms, polite requests, kinship terms, and expressions of emotions. An intermediate degree of impact can be found for individualism, long-term orientation, status indicators, politeness, Chinese culture terms, and sociosexuality.

Table 13: Strength of linguistic manifestations by sociocultural dimensions

	strong	medium	low
collectivism	X		
individualism		X	
long-term orientation		X	
short-term orientation			X
high/low context			X
mono-/polychronic time			X
proxemics			X
power distance / address terms	X		
power distance / status indicators		X	
power distance / polite requests	X		
social relations / kinship terms	X		
social relations / politeness		X	
social relations / small talk			X
Chinese culture notions		X	
emotions	X		
sociosexuality		X	

4.3 Constructions and culture?

4.3.1 Theoretical background

Adopting a Construction Grammar perspective (Hoffmann and Trousdale 2013), I have been moving up one more level in a hierarchy of constructions, from the lexical-material (looking at cultural objects) to the lexical-conceptual (i.e., cultural dimensions) level; and here we proceed one more step towards the schematic level, investigating the possibility of culturally-inspired syntax. The question to be asked is this: Are schematic constructions (or is the choice between alternative constructions) possibly motivated by culturally-based principles? Clearly, this implies adopting an interesting, more abstract perspective, and it would be important if any positive evidence could be identified.

Unlike the study of cultural concepts and keywords, possible relationships between constructions and culture have not been a strong topic in linguistics. However, this is also not completely uncharted territory – a few suggestions, hypotheses, attempts have been brought forward. One of the earliest hypotheses concerning culturally motivated variety-specific constructions in World Englishes research can be found in Olavarria de Errson and Shaw (2003). The authors investigate frequency differences of verb complementation patterns between British English and Indian English and interpret these as possibly reflecting cultural differences, “different ways of perceiving the world”, claiming that they found “an extraordinarily ... direct connection between grammar and ‘culture’” (159). With certain verbs with two complements, e.g. *provide*, British English consistently prefers a construction type *provide somebody with something*, while Indian English prefers *provide something to somebody*. The British pattern, they argue, reflects European subjectivism, seeing the individual at the center, as opposed to a South Asian reluctance to profile the recipient and a tendency to focus on the object of conversation.²

Mukherjee and Hoffmann (2006) have a similar hunch on a preference of Indian English for monotransitive recipient-less constructions (154), but they remain essentially skeptical of such cultural explanations of syntactic preferences.

Gladkova (2015) published a paper on ‘ethnosyntax’, where she posits “grammatical constructions [which] are not semantically arbitrary and ... related to broader cultural understandings” (33) – without offering in-depth exemplification.

An interesting, rather deeply reflected claim on the growth of culturally motivated constructions can be found in Burridge (2015). She posits a dependency relationship: “Cultural preoccupations give rise to ways of thinking and [...] can then end up embodied in the grammar; habitual conversation practices

[...] solidify into specific morphosyntactic constructions” (72). Her examples are the loss of the ‘impersonal construction’ (e.g. *it worries me* vs. *I worry*) and in general the disappearance of dative experiencers in Early Modern English, which, she thinks, is partly to be explained by “the rise of the modern self-determining individual” (70).

No assumptions of culturally-determined constructions are made in Construction Grammar (Hoffmann and Trousdale 2013; Hoffmann, personal communication) or in Cognitive Sociolinguistics (Hollmann 2013), two approaches where such lines of thinking would not seem too far-fetched. Both emphasize the basis of constructions in social context and cognition, as well as the relevance of varieties for construction choices – but there are no suggestions on possible cultural motivations.

So in sum, some hunches and hypotheses on possible influences of cultures on syntax have been presented, but there is no strong evidence so far, and this is clearly not a big topic in linguistics. Methodologically, extrapolating from the suggestion made by Olavarria de Errson and Shaw (2003), my attempts at practical applications and translating the notion of culture-sensitive constructions to possible corpus-based searches largely boils down to manifestations of the individualism vs. collectivism dimension after Hofstede. The basic hypothesis thus is the assumption that collectivist cultures may prefer constructions which syntactically downplay the agent, the individual as a persona, and conversely may highlight collectivity as opposed to individuality grammatically. All examples in this section pin this idea down to specific constructional alternatives, defined as choices which are otherwise functionally equivalent and can thus be directly quantified and compared. Five different construction types will be investigated.

4.3.2 Constructions and culture: Constructions with possible end-focus on the individual

Ditransitive verbs in English allow a choice between two sequences of the two objects: either the indirect object, mostly with human referents, comes first (Oind_[human]-Odir) or it is placed after the direct object and governed by *to* (Odir *to* Oind/PrepC_[human]) – e.g. *give me something* versus *give something to me*. In a functional sentence perspective analysis, sentences are assumed to begin with a theme, a known topic, followed by a rheme, new information about the theme, so due to the principle of end-focus the clause-final position receives special prominence and information value. For such alternative, functionally equivalent constructions, then, the hypothesis is that individualistic cultures might display a higher proportion of a ‘*V NP to me*’ pattern, highlighting an individual in the

end position, as opposed to ‘*V me NP*’, with the individual in the less prominent medial position.

Methodologically, I selected and tested a small number of highly frequent ditransitive verbs (*give*, *send*, *promise*, and *show*, in all their finite present and past tense forms), and I also reduced the search objects to a finite set of pertinent, highly frequent, and easily searchable forms, namely the singular human personal pronouns only (*me/you/him/her*). This guarantees a tightly controlled range of variation, a classic Labovian ‘envelope of variation’. It implies that in a corpus-linguistic perspective the recall of the pattern in question is low (certainly there are many more ditransitive constructions in the corpora where the same issue could be investigated), but the precision of the retrieved instances is extremely high. The restriction to small sets of verbs and objects guarantees a concentration on clearly defined, objectively measurable syntactic alternatives, with all other potentially interfering factors (e.g. lexical preferences for either construction) eliminated.

Table 14: Construction alternatives with possible end-focus on singular pronouns

sum	GB	Hong Kong	Singapore	India	Nigeria
V PRON (sth.)	153 (86.4%)	216 (85.4%)	199 (91.3%)	125 (89.9%)	142 (88.7%)
V (sth.) to PRON	24 (13.6%)	37 (14.6%)	19 (8.7%)	14 (10.1%)	18 (11.3%)

While in principle the indirect object position directly after the verb clearly represents the default choice, the proportion of end-focus individualizing pronouns is consistently higher in Great Britain than in Singapore, India and (less so but also clearly) Nigeria. This is largely in line with the hypothesis. However, the exception is Hong Kong, with a very high proportion of end-focus pronouns (reminiscent of the remarkably high proportion of first person singular versus plural pronouns observed there as well; cf. Figure 1 in section 4.2.1 above).

4.3.3 *Constructions and culture: Passives*

In active constructions the agent, who performs the verbal activity expressed in the predicate, fills the subject role, while passive constructions, especially without a *by*+NP complement, camouflage agency. Assuming that individualistic agency may be dispreferred in collectivistic cultures, a possible hypothesis is the

assumption that a higher proportion of passives might turn up in more collectivistic cultures.

As before, in order to achieve high precision and a controlled envelope of variation I adopted the same methodological procedure as in the previous section. I chose eight prototypically monotransitive verbs (*tell, offer, support, help, beat, inform, admire, encourage*, again in all present and past finite verb forms) with human personal pronouns only as direct objects.

Table 15: Construction alternatives active - passive

sum	GB	Hong Kong	Singapore	India	Nigeria
active	171 (87%)	400 (90%)	372 (94%)	319 (93%)	314 (91%)
passive	26 (13%)	45 (10%)	24 (6%)	24 (6%)	33 (9%)

The results, shown in Table 15, clearly contradict the hypothesis: there are consistently fewer passives in the ESL varieties than in Great Britain. The data offer no explanation, but a plausible assumption might be syntactic complexity: perhaps due to their less advanced developmental stage (cf. Schneider 2007) or the more persistent impact of simplification as a feature of second-language acquisition it may be the case that ESL varieties tend to avoid the passive as the syntactically slightly more complex construction.

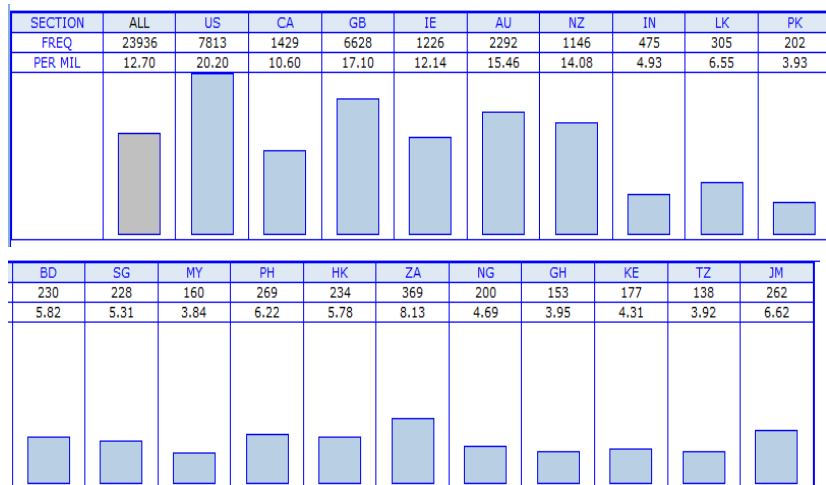
4.3.4 Constructions and culture: Impersonal constructions

In a similar vein, and inspired by Burridge's (2015) thoughts, I looked into a few impersonal constructions, assuming that they also downplay agency by moving the human 'agent' of a mental process from an (active) subject to a (recipient) object or complement position. Thus, the hypothesis here is that relatively more impersonal constructions should be found in collectivistic cultures. Pertinent patterns are identified in Table 16. The first two lines present a structural alternative, as in the previous cases; the other two document token frequencies of impersonal constructions. As before, the pronoun set after *occur to* comprises all singular personal pronouns (so *It occurred to him* would be included) while *seems to me* is assumed to be a firm collocation with first singular pronouns only. Again, both tenses were looked at.

Table 16: Impersonal constructions

	GB	Hong Kong	Singapore	India	Nigeria
PRON WORRY	4	18	4	5	4
(sth.) WORRY PRON	8	6	7	1	1
seems to me	48	21***	4***	4***	3*
occur to PRON	10	11	2	7	4

The results are largely contrary to the hypothesis. All impersonal constructions are strong (and in most cases strongest) in Great Britain, as opposed to the ESL varieties, where they appear to have been adopted only to a limited extent. The pattern *something worries someone* predominates in Britain and Singapore, while in all other ESL varieties *someone worries* is used more frequently. The construction *occur to someone* seems characteristic of Great Britain as well (and also Hong Kong in this case) and is used rarely or less commonly in second-language varieties. The clearest case is *seems to me*, which is statistically significantly typical of British English over all ESL varieties (and among these it is relatively common in Hong Kong but rare elsewhere). A similar distribution in GloWbE (Figure 7) confirms that *seems to me* is very much an L1 construction, most common in the US, GB, Australia and New Zealand.

Figure 7: *seems to me* in GloWbE

Hence, this construction type shows no effect of cultural impact in the expected direction. As in the previous case, it may be presumed that the distributions observed reflect an effect of syntactic complexity, which may be reduced in second-language and less advanced varieties.

4.3.5 Constructions and culture: Recipient-less constructions

In the construction type ‘verb + experiencer-object + finite *that*-clause’ the recipient individual can be downplayed in World Englishes by omitting the person in the object / experiencer position (so that, for example, *inform [somebody] that ...* becomes *inform that ...*). Thus, it may be hypothesized that there should be more recipient-less constructions in collectivistic cultures (cf. Mukherjee and Hoffmann 2006). The hypothesis was tested by searching for two high-frequency verbs (*assure*, *inform*) directly followed by a *that*-clause (in both tenses; other functions of *that* were deleted).

Table 17: Recipient-less constructions

sum	GB	Hong Kong	Singapore	India	Nigeria
assure that	0	4	2	5	12**
inform that	0	0	2	11*	3
sum	0	4	4	16***	15***

The results, shown in Table 17, are strongly in line with the hypothesis. Recipient-less constructions with these two verbs are found in all ESL varieties, representing more collectivist cultures, and especially so in India (where the pattern is strong with *inform*) and Nigeria (strong with *assure*) – but not at all in Great Britain. This is fully confirmed by GloWbE data as well, where recipient-less constructions are rare in L1 but widespread in L2 varieties – the distributions for *assure that* and *inform that* are shown in Figures 8 and 9, respectively. These preferences may be caused by a possible impact of culture – but alternative interpretations are conceivable as well: omitting the object could also be an outcome of syntactic simplification and innovation.

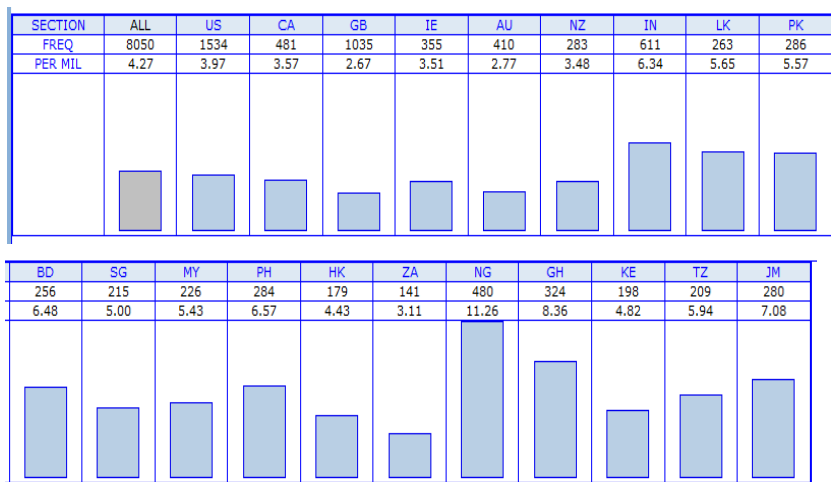


Figure 8: assure that in GloWbE

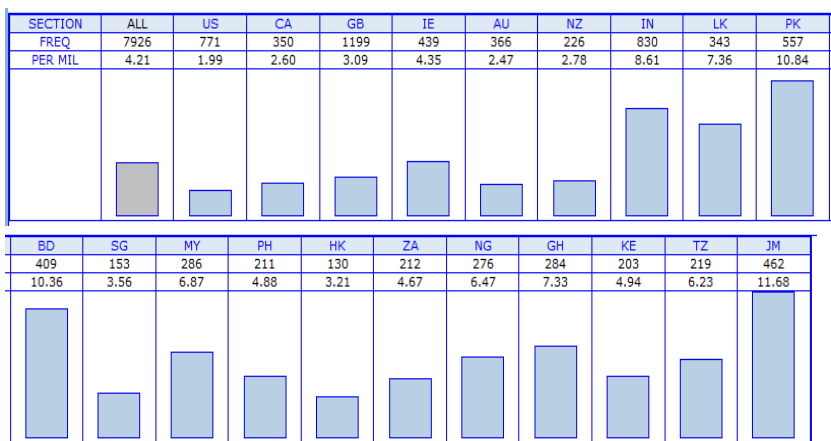


Figure 9: inform that in GloWbE

4.3.6 Constructions and culture: Verb concord with collective nouns

Collective nouns allow a choice between singular and plural verb forms (e.g. *my family is/are happy*). The choice of verb form reflects a difference in the conceptualization of the respective nouns as either a holistic group or a multitude of individuals – a difference which obviously parallels the one between collectivist and individualist orientations. The hypothesis is, thus, that collectivist cultures prefer holistic (= singular; e.g. *N is*) rather than individualistic (= plural; e.g. *N are*) conceptualizations and their corresponding forms. The methodological procedure, along the same lines as before, was to choose a set of appropriate nouns (*team / military / council / government / company / family / committee*) and a set of controlled verb forms which consistently show the morphological difference in number marking, i.e. the primary auxiliaries *is/are* and *has/have*.

Table 18: Construction alternatives: verb concord with collective nouns

sum	GB	Hong Kong	Singapore	India	Nigeria
N Vsg	95	327	131	233	229
N Vpl	25	34	16	18	34
% sing.	79%	91%	89%	93%	87%

The results are strongly in line with the hypothesis: Asian varieties, and also (though less so) Nigeria, show substantially more singular concord (around and above 90 %) than Great Britain (79%), thus treating the subject nouns as a single, collective entity. This may be an indicator of a possible impact of culture on conceptualization and, ultimately, grammatical form. However, a possible alternative explanation is also available and cannot be ruled out: It is well known that singular concord is more common in American than in British English (e.g. Hundt 2006: 209), so this distribution may also be caused or influenced by American impact on the new World Englishes.

4.3.7 Constructions and culture: Summary

An overall look at the results for constructions calls for a somewhat reluctant, balanced assessment. Out of five construction types selected to test predictions derived from the culture hypothesis, two (recipient-less constructions and collective concord) have yielded results which are confirming, strongly in line with the hypotheses; one more distribution (personal end-focus constructions) is also largely in line with the hypothesis and basically confirming; but two others (passives, impersonal constructions) have produced results which contradict the hypotheses.

So, in principle, there is some partial support for the hypothesis that cultures may influence construction choices, but the evidence remains somewhat limited, and causal relationships are not wholly clear. Whether the constructions selected as possible manifestations of cultural principles were chosen and interpreted adequately remains open to discussion, of course (it is conceivable, for instances, that the assumption that passives or impersonal constructions might reflect collective orientations is misguided). In a few cases alternative explanations for the distributions observed are also conceivable – such as syntactic complexity reduction, varying developmental stages, or American impact. Still, the basic idea remains a viable hypothesis, with some evidence in favor of it – an observation which may be regarded as remarkable, given the scarcity of relevant documentation so far. The issue remains open to discussion.

5 Reflections of culture in corpora: Summary evaluation

The research question that I started out from was: are there any systematic reflections of cultural differences in corpora? The results from three different ‘layers’ of evidence show that this is indeed the case – but to varying degrees and with limitations. Fundamentally, it can be observed that some positive evidence has been identified on all levels. There is a systematic relationship with the increasingly abstract levels of language organization, however: the more concrete the perspective, the stronger and more directly evident are the cultural reflections; conversely, the more abstract and general the perspective, the more elusive and indirect are the cultural traces.

For concrete objects, artefacts, and notions the relationship is straightforward, positive, and quite direct – terms for these items appear in regional corpora, mainly locally but not exclusively so (especially for internationally known concepts).

As to cultural dimensions (as posited in cross-cultural anthropological analysis after Hofstede), the answer is affirmative, too, but with relatively stronger limitations: yes, there are clearly recognizable influences, but they vary from one dimension to another and by indicator terms, and they tend to be graded rather than absolute.

When looking at schematic constructions, results are fuzzier but still not overly elusive: yes, possibly such influences exist and have an impact, but if so they are clearly more indirect and somewhat abstract.

I have emphasized the claim that, while this is a systematic study of the issue of formal, corpus-based reflections of cultural differences, due to limitations in scope it is a pilot study. Given the results, I do think the topic deserves more in-depth investigation.

Notes

1. Cf. Hundt and Leech (2012) on the issue of corpus size and composition.
2. The line of thinking is interesting – but the specific interpretation seems to imply a priority of the direct object immediately after the verb (which is the constituent they focus on), while it could also be argued that the clause-final position highlights a constituent most effectively.

References

- Assmann, Aleida. 2012. *Introduction to cultural studies: Topics, concepts, issues*. Berlin: Erich Schmidt-Verlag.
- Burridge, Kate. 2015. The body, the universe, society and language: Germanic in the grip of the unknown. In R. La Polla and R. de Busser (eds.). *The shaping of language: The relationship between the structures of languages and their social, cultural, historical, and natural environments*, 45–76. Amsterdam, Philadelphia: Benjamins.
- Byram, Michael. 1997. *Teaching and assessing intercultural communicative competence*. Clevedon: Multilingual Matters.
- Byram, Michael, Adam Nichols and David Stevens (eds.). 2001. *Developing intercultural competence in practice*. Clevedon: Multilingual Matters.
- Carbaugh, Donal. 2005. *Cultures in conversation*. Mahwah, New Jersey and London: Lawrence Erlbaum Associates.
- Davies, Mark. 2013. *Corpus of global web-based English: 1.9 billion words from speakers in 20 countries*. Available online at <http://corpus.byu.edu/glowbe/>.
- Davies, Mark and Robert Fuchs. 2015. Expanding horizons in the study of World Englishes with the 1.9 billion word Global Web-based English Corpus (GloWbE). *English World-Wide* 36: 1–28.
- Fang, Tony. 2012. Yin Yang: A new perspective on culture. *Management and Organization Review* 8: 25–50.
- Fuchs, Robert 2012. ‘I, me and myself’ – individualism and collectivism in World Englishes. Paper presented to the International Association for World Englishes (IAWE), Hong Kong / Guangzhou, China.
- Gladkova, Anna. 2015. Ethnosyntax. In Sharifian (ed.). 2015: 33–50.
- Greenbaum, Sidney (ed.). 1996. *Comparing English worldwide. The International Corpus of English*. Oxford: Clarendon.

- Gries, Stefan Th. 2008. *Statistik für Sprachwissenschaftler*. Göttingen: Vandenhoeck & Ruprecht.
- Hall, Edward T. and Mildred R. Hall. 1990. *Understanding cultural differences*. Yarmouth: Intercultural Press.
- Hoffman, Thomas and Graeme Trousdale (eds.). 2013. *The Oxford handbook of construction grammar*. Oxford: Oxford University Press.
- Hofstede, Geert. 2001. *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. 2nd ed. Thousand Oaks, CA: Sage.
- Hollmann, Willem B. 2013. Constructions in cognitive sociolinguistics. In Hoffman and Trousdale (eds.), 491–509.
- House, Robert J., Paul J. Hanges, Mansour Javidan, Peter W. Dorfman and Vipin Gupta. 2004. *Culture, leadership, and organizations: The GLOBE study of 62 societies*. Thousand Oaks: Sage.
- Hua, Zhu. 2013. *Exploring intercultural communication. Language in action*. London, New York: Routledge.
- Hundt, Marianne. 2006. *The committee has/have decided ...* On concord patterns with collective nouns in inner- and outer-circle varieties of English. *Journal of English Linguistics* 34: 206–232.
- Hundt, Marianne and Geoffrey Leech. 2012. ‘Small is beautiful’: On the value of standard reference corpora for observing recent grammatical change. In T. Nevalainen and E. Closs Traugott (eds.). *The Oxford handbook of the history of English*, 175–188. Oxford: Oxford University Press.
- Jensen, Kim Ebensgaard. 2017. Corpora and cultural cognition: How corpus-linguistic methodology can contribute to Cultural Linguistics. In F. Sharifian (ed.). *Advances in Cultural Linguistics*, 477–505. Singapore: Springer Nature.
- Kachru, Yamuna and Larry E. Smith. 2008. *Cultures, contexts, and World Englishes*. London, New York: Routledge.
- Kashima, Yoshihisa and Emiko S. Kashima. 2003. Individualism, GNP, climate, and pronoun drop. Is individualism determined by affluence and climate, or does language use play a role? *Journal of Cross-Cultural Psychology* 34: 125–134.
- Kirkpatrick, Andy. 2015. World Englishes and local cultures. In Sharifian (ed.). 2015: 460–469.

- Kronenfeld, David B. 2015. Culture and kinship language. In Sharifian (ed.). 2015: 154–169.
- Kuppens, Peter, Eva Ceulemans, Marieke E. Timmerman, Ed Diener and Chu Kim-Prieto. 2006. Universal intracultural and intercultural dimensions of the recalled frequency of emotional experience. *Journal of Cross-Cultural Psychology* 37 (5): 491–515.
- Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Leavitt, John. 2015. Linguistic relativity: precursors and transformations. In Sharifian (ed.). 2015: 18–30.
- Levisen, Carsten and Sophia Waters (eds.). 2017. *Cultural keywords in discourse*. Amsterdam, Philadelphia: Benjamins.
- Minkov, Michael. 2013. *Cross-cultural analysis. The science and art of comparing the world's modern societies and their cultures*. Thousand Oaks, CA: Sage.
- Moran, Patrick R. 2001. *Teaching culture: Perspectives in practice*. Boston, MA: Heinle & Heinle.
- Mukherjee, Joybrato and Tobias Bernaisch. 2015. Cultural keywords in context: A pilot study of linguistic acculturation in South Asian Englishes. In P. Collins (ed.). *Grammatical change in English World-Wide*, 411–435. Amsterdam, Philadelphia: Benjamins.
- Mukherjee, Joybrato and Sebastian Hoffmann. 2006. Describing verb-complementational profiles of New Englishes: A pilot study of Indian English. *English World-Wide* 27: 147–173.
- Olavarria de Ersson, Eugenia and Philip Shaw. 2003. Verb complementation patterns in Indian Standard English. *English World-Wide* 24: 137–161.
- Rayson, Paul. n.d. Log-likelihood and effect size calculator. <http://ucrel.lancs.ac.uk/llwizard.html>. (06_01_2018).
- Salzmann, Zdenek, James Stanlaw and Nobuko Adachi. 2012. *Language, culture and society. An introduction to linguistic anthropology*. Boulder, CO: Westview Press.
- Schmitt, David P. 2005. Sociosexuality from Argentina to Zimbabwe: A 48-nation study of sex, culture, and strategies of human mating. *Behavioral and Brain Sciences* 28: 247–275.
- Schneider, Edgar W. 2007. *Postcolonial English. Varieties around the world*. Cambridge: Cambridge University Press.
- Schneider, Edgar W. 2011. *English around the world. An introduction*. Cambridge: Cambridge University Press.

- Sharifian, Farzad. 2011. *Cultural conceptualizations and language. Theoretical framework and applications*. Amsterdam, Philadelphia: Benjamins.
- Sharifian, Farzad. 2015. Language and culture: overview. In Sharifian (ed.). 2015: 3–17.
- Sharifian, Farzad (ed.). 2015. *The Routledge handbook of language and culture*. London, New York: Routledge.
- Sharifian, Farzad. 2017. *Cultural linguistics*. Amsterdam, Philadelphia: Benjamins.
- Uz, Irem. 2014. Individualism and first person pronoun use in written texts across languages. *Journal of Cross-Cultural Psychology* 45 (10): 1671–1678.
- Wolf, Hans-Georg and Frank Polzenhagen. 2009. *World Englishes: A cognitive sociolinguistic approach*. Berlin: Mouton de Gruyter.
- Yu, Feng, Theodore Peng, Kaiping Peng, Shi Tang, Chuan Shi Chen, Xiaojun Qian, Pei Sun, Tingting Han and Fangyuan Chai. 2016. Cultural value shifting in pronoun use. *Journal of Cross-Cultural Psychology* 47 (2): 310–316.