

Acculturation in Sri Lankan English

M.D.S.S. Kumara

Abstract

Because of the worldwide spread of the English language in the past centuries, it has become necessary for linguists to refer to several 'Englishes' instead of one English language. Such new Englishes have emerged in cultural and linguistic contexts different from those in which English originated, and English has been in contact with diverse kinds of other languages in this process. Because the sociocultural and linguistic contexts of new varieties of English such as Sri Lankan English (SLE) are thus different, they inevitably reflect the sociocultural identities of the people who use the language in the new context, a phenomenon referred to as 'acculturation' of a language.

Acculturation of a language is overtly shown in the vocabulary through 'borrowings' or 'loans'. The present paper illustrates four 'loans' in SLE in which the words are English, but their meanings are either partly or completely borrowed from local source languages: Sinhala and Tamil. Employing a corpus-based methodology, the paper shows how the four lexical items have acquired new semantics in the new cultural and linguistic contexts in Sri Lanka.

Keywords – Acculturation, SLE, semantic extension, corpus

INTRODUCTION

According to the Oxford dictionary, the term 'English' is used to refer to 'the West Germanic language of England, now widely used in many varieties throughout the world' (OED Online, 2017). For centuries, the term was used as a singular uncountable noun in English with no plural form. However, because of the widespread usage of this language 'in many varieties throughout the world', it has become necessary for linguists and scholars in the field to use the term in the plural as 'Englishes', as exemplified by phrases such as *World Englishes*, *Global Englishes*, *International Englishes*, *New Englishes*, *Nativized Englishes*, *Indegenized Englishes*, *Non-native Englishes*, *Transplanted Englishes*, and *Postcolonial Englishes*. Sri Lankan English (SLE) is one such variety of 'New Englishes', in the sense that its linguistic and literary institutionalization has been recently recognized (cf., Kachru, 2001, p.520).

One reason for referring to several *Englishes* instead of one *English* thus is that there is some considerable language change happening in the evolution of these new *Englishes*. As Winford (2003, p.243) states, there are two central processes of such language change - 'nativization' and 'acculturation'. According to Kachru (1992, p.235), 'nativization is the linguistic readjustment a language undergoes when it is used by members of another speech community in distinctive sociocultural contexts and language contact situations', whereas 'acculturation focuses on the people learning a transplanted language'. The latter process, which is applicable to the present paper more, refers to the reflection of the sociocultural identities of such people in a nativized language. As Sridhar (2012) outlines, 'language acculturation consists of a variety of processes of language change by which a language, not one's own, is modified to serve as an

effective vehicle for the expression of one's own sociocultural and cognitive experiences'.

Linguistic acculturation in New Englishes is overtly shown in the vocabulary through lexical borrowings (among others). According to Winford (2003, p.43), lexical borrowings, which involve imitation of some aspect of source language model, are of two types: Loan words and Loan shifts. In loan words, all or part of the morphemic composition of the loan derives from the source language. For example, the italicized parts of the following SLE lexical items from Meyler (2007) are from the source language, Sinhala: *Cadjunut*, *rasthiyadufy*, *kuppi* class, *kitul* toddy, *gajeying*. In loan shifts (loan meanings), the morphemic composition of the item is entirely from the recipient language, though its meaning derives at least in part from the source language. The four lexical items analyzed in the present paper exemplify loan shifts. Loan shifts are of two subcategories: semantic loans or extensions, which Winford (2003, p.45) defines as 'shifts in the semantics of a native word under influence from a foreign word', and loan translations or calques, which are 'combination(s) of native morphemes in imitation of foreign pattern' (ibid). The present paper argues that three out of the four SLE lexical items analyzed are semantic extensions and the other is a calque.

Out of the four lexical items analysed in the paper, the first, *up country/up-country*, was recorded in several sources including Passé (1955), Halverson (1966), Fernando (2003), and Meyler (2007). Two others, viz., *low country/low-country* and *hill country/hill-country*, were culled from Meyler (2007). However, the meanings illustrated in the present investigation for these three lexical items have not been recorded. The fourth *mid country/mid-*

country, which is not known to be recorded previously, is a similar lexical item found in the course of the present investigation. All four of these lexical items occur both as a separate word compound and as a hyphenated compound.

Although three out of the four lexical items analysed in the present paper have been recorded in SLE literature, much of the writings on SLE have been based on random examples and personal experience, as Parakrama (1995, p.34) observes. Therefore, 'the findings of linguists remain more impressionistic than necessary, and even the acceptability of the few cited examples are contested' (ibid). Baker (2010, p. 94) also points out that 'humans do not always make accurate introspective judgments regarding language, instead relying on cognitive and social biases'. Thus, the purpose of the present study is to empirically study whether the said lexical items occur significantly in SLE, and whether their meanings are distinct in SLE. It is hypothesised that these lexical items are distinctively SLE specific, showing *acculturation* because the sociocultural and linguistic contexts in which English is used in Sri Lanka are different from those in Great Britain.

Methodology And Experimental Design

In contrast to the impressionistic human observations about language mentioned in the previous section, *corpus methods* rely on 'real-world instances of language use in order to derive rules or explore trends about the ways in which people actually produce language' (Baker, 2010, p. 94). Mukherjee (2012, p.202) also views that large-scale corpora of SLE provide a systematic way of describing SLE which is not only 'empirically sound' and 'linguistically plausible' but also 'socio-culturally and politically acceptable', in light of 'the current debate on language standards and norms in Sri

Lanka'. Hence, the present investigation is primarily *corpus-based*. A *corpus*, which is a computer-readable collection of texts or transcribed speech representative of a language, can basically be used to get information on frequencies of words, or other structures. The present study uses all three central corpus-linguistic methods used to exploit frequency information in a corpus that are outlined in Gries (2009, p.11): frequency (lists), collocation, and concordances. Corpus access software, or concordancers, are used in these methods.

Primarily, quantitative comparison of corpus frequencies of the selected lexical items in three varieties of English, viz., SLE, British English (BE), and Indian English (IndE), is made in order to trace their comparative frequency of occurrence in SLE. The corpus environment of the study consists of three types of written language corpora representing the three varieties of English. The first type of written language corpora is from written parts of the respective components of International Corpus of English (ICE), namely Sri Lankan component (ICE-SL), Great Britain component (ICE-GB), and Indian Component (ICE-Ind). ICE corpora provide comparable language data from each variety with a high level of representativeness covering a wide range of genres. Each written component of ICE consists of approximately 400,000 words (cf., Greenbaum, 1996).

As ICE corpora are comparatively small in size, a second type of corpora, representing news genres, is also used in the analysis. For data for SLE and IndE, web-derived newspaper corpora from the South Asian Varieties of English (SAVE) project compiled at Justus Liebig University of Giessen, Germany, in the context of the research project on "Verb-complementational profiles in South Asian Englishes" are used. Each of these components, SAVE-SL and SAVE-Ind, is approximately 3 million words in size. To

represent BE, news genres of the British National Corpus (BNC-News) are used. They compose approximately 8.9 million words. To make the comparisons fair, all results are normalized to 1 million words. Newspaper data are norm-providing in outer-circle varieties, and 'relatively unlikely to be what are often considered learner mistakes' (Schilk, 2011, p.47).

The third type of written corpora used in the analysis is a large online database recently made available. It is the corpus of Global Web-based English - GloWbE (Davies & Fuchs, 2015), which is composed of 1.9 billion words in 1.8 million web pages from 340,000 websites in 20 different English-speaking countries. The relevant components to the present study contain 46,583,115 words of SLE (GloWbE-SL), 96,430,888 words of IndE (GloWbE-Ind), and 387,615,074 words of BE (GloWbE-Gb). This large database is useful in detecting those innovative features of SLE which are low-frequency phenomena.

In addition to the comparative frequencies of the selected lexical items, concordance lines or the Key Words in Context (KWIC) lines are used for qualitative analysis of these items in their larger contexts. KWIC lines are manually analysed in order to recognize the semantics of the selected lexical items. The third central corpus-linguistic method, collocation, is also used for this purpose. Collocation is probabilistic co-occurrence of words. For example, the most frequent preposition collocate of the verb lemma [*participate*] in the entire GloWbE corpus at the first position to the Right of the verb (R1) is 'in'. Concordancing software extract collocates based either on their frequency of occurrence or on another measure of collocation strength such as Mutual Information (MI). ICE and News corpora data were analyzed using the concordancer (software), Wordsmith Tools-Version 5

(Scott, 2008). For GloWbE data, the corpus web interface was used as the concordancer.

Wordsmith Tools-Version 5 is a lexical analysis software, which provides three main tools: Concord, KeyWords, and WordLists. For the purposes of the present study, only the 'Concord' tool was used. After uploading the corpus files (.txt format) to the program, just by typing the search term in the concord window, concordances are generated. In order to find patterns, concordances can be sorted using various parameters. For high frequency search terms, different sub-windows of the 'Concord' tool provide additional information such as collocates, patterns, filenames, and clusters. For example, 'patterns' sub-window for the search item 'look' in ICE-SL written component shows that the R1 collocates of 'look' are *at, forward, for, after, into* and *around*. Online web interface of GloWbE also functions as a concordancer. Just when the search term is typed, the interface outputs its frequency in all 20 varieties. Collocates of the search term can also be obtained either by typing '*' for any collocate, or by specifying the collocate or its Part Of Speech (POS). It is also possible to compare the results for two search terms, or two varieties specified.

Results

A. Dictionary meanings

According to the Oxford English Dictionary (OED) (Simpson & Weiner, 1989), the term *low country* refers to 'a region or district whose level is lower than that of the surrounding country'. However, the term is used in SLE in a different, wider context as will be revealed in the present investigation. Similarly, the compound *hill country* has a usage in SLE which is different from that given in the OED as 'a district composed of hills or elevated

ground'. Likewise, the term *up-country* is defined in OED as 'an uplying or inland district, the inland part of a country ', while in SLE it acquires different semantics. For the fourth item, *mid country/mid-country*, neither the OED nor the over 900 online dictionaries linked to 'OneLook Dictionary Search' (OneLook, 2017) provide a definition.

B. *Frequencies of occurrence*

Figure 1 below shows absolute frequencies of the four lexical items analyzed as they appear in the three ICE corpora. As can be seen from the table, none of the four lexical items appear either in ICE-GB or in ICE-Ind. *Mid country/mid-country* is not attested even in ICE-SL.

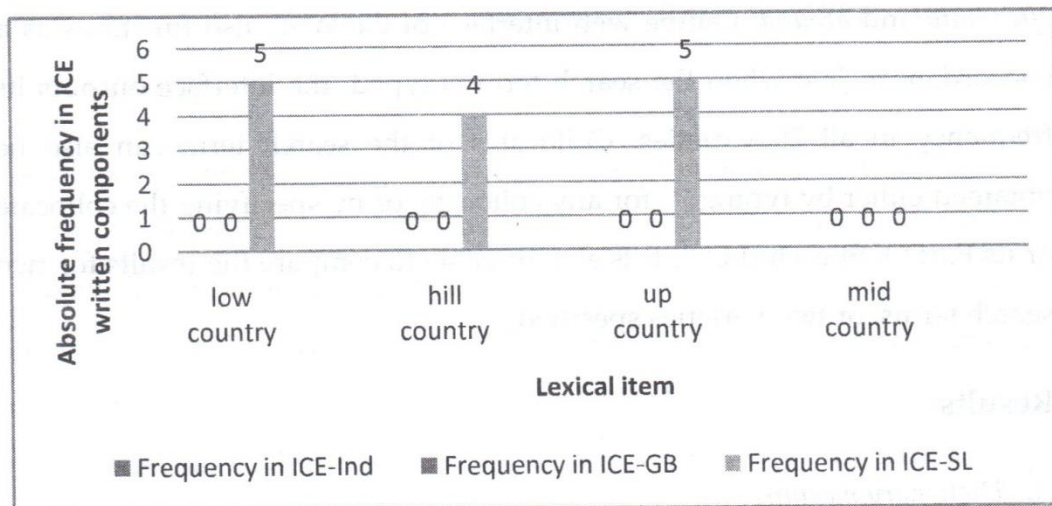


Figure 1. Absolute frequencies of *low country*, *hill country*, *up country*, and *mid country* in ICE-Ind, ICE-GB, and ICE-SL

Normalized frequencies of the four lexical items in the three corpora of news genre representing the three varieties under scrutiny also reveal the

comparative significance of occurrence of *low country*, *hill country*, and *up country* in SLE (Figure 2). Additionally, the data show the presence of *mid country/mid-country* in SLE and the absence of it in IndE and BE data.

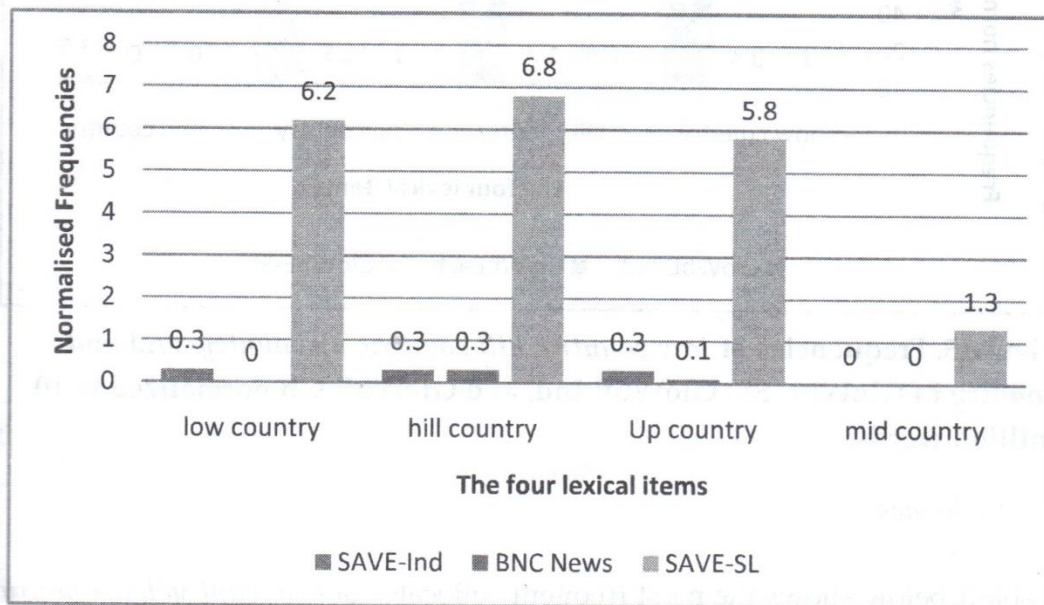


Figure 2. Frequencies of *low country*, *hill country*, *up country* and *mid country* in SAVE-SL, SAVE-Ind, and BNC News normalized to 1 million words.

Figure 3 below shows frequencies of the four lexical items under investigation as they occur in the three relevant components of the GloWbE corpus. Normalization has been done to ten million words because of the large corpus sizes and the ease in comparison. As the figure shows, the compounds *low country*, *hill country*, and *up country* appear in all three components of the GloWbE corpus, but with a high comparative frequency in the Sri Lankan component. The compound *mid country*, however, does not appear in IndE and BE data even within this large data base.

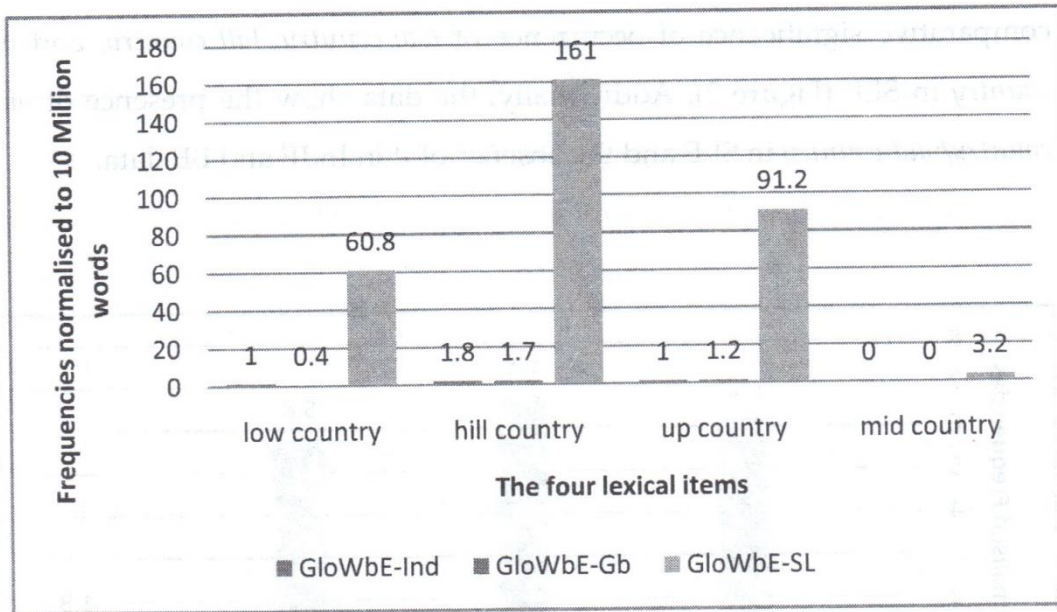


Figure 3. Frequencies of *low country*, *hill country*, *up country*, and *mid country* in GloWbE-SL, GloWbE-Ind, and GloWbE-Gb normalized to 10 million words.

C. Collocates

Table 1 below shows the most frequent collocates of *low country/low-country* in the three components of the GloWbE corpus as extracted by the corpus web interface, along with their frequencies and MI values. Among the most frequent collocates of *low country/low-country*; words like *Sinhalese*, *Kandyan (s)*, *castes*, *headmen*, *Karava* (a caste), and *dancing* suggest a context of use of this term which is different from the geographical context suggested in its dictionary meaning. None of these collocates appear as frequent collocates for the term in IndE and BE data. Even the frequent collocates with a geo-climatic reference, such as *zone*, *province*, *wet*, *dry*, and *hill*, do not have the association in IndE and BE data. Instead, the main association, especially in IndE data, is with *south* and *Carolina* which, upon observation of the concordance lines, show a reference to a region in America.

Table 1. Top ten most frequent collocates for *low country* (left) and *low-country* (right) in GloWbE-SL, GloWbE-Ind, and GloWbE-Gb within a collocation span of L4 and R4, reproduced from the GloWbE web interface.

| Collocates <i>low country</i> | Total | GB | Ind | SL | MI | Collocates <i>Low-country</i> | Total | GB | Ind | SL | MI |
|----------------------------------|-----------|----|-----|-----------|-------|----------------------------------|-----------|----------|-----------|-----------|-------|
| <u>SINHALESE</u> | <u>53</u> | | | <u>53</u> | 10.89 | <u>SINHALESE</u> | <u>42</u> | | | <u>42</u> | 9.33 |
| <u>ZONE</u> | <u>10</u> | | | <u>10</u> | 8.43 | <u>KANDYAN</u> | <u>20</u> | | | <u>20</u> | 11.85 |
| <u>KANDYAN</u> | <u>8</u> | | | <u>8</u> | 11.76 | <u>ZONE</u> | <u>19</u> | | | <u>19</u> | 8.12 |
| <u>CASTES</u> | <u>8</u> | | | <u>8</u> | 11.28 | <u>WET</u> | <u>15</u> | | | <u>15</u> | 8.44 |
| <u>PROVINCE</u> | <u>7</u> | | | <u>7</u> | 8.61 | <u>COUNTRY</u> | <u>13</u> | | <u>2</u> | <u>11</u> | 3.89 |
| <u>WET</u> | <u>6</u> | | | <u>6</u> | 8.35 | <u>CAROLINA</u> | <u>11</u> | <u>1</u> | <u>10</u> | | 9.97 |
| <u>DRY</u> | <u>6</u> | | | <u>6</u> | 7.50 | <u>SOUTH</u> | <u>11</u> | <u>1</u> | <u>10</u> | | 5.03 |
| <u>HEADMEN</u> | <u>4</u> | | | <u>4</u> | 15.17 | <u>DANCING</u> | <u>9</u> | | | <u>9</u> | 7.76 |
| <u>LOW-COUNTRY</u> | <u>4</u> | | | <u>4</u> | 14.94 | <u>HILL</u> | <u>8</u> | | | <u>8</u> | 6.72 |
| <u>KARAVA</u> | <u>4</u> | | | <u>4</u> | 14.76 | <u>KANDYANS</u> | <u>7</u> | | | <u>7</u> | 13.58 |

The most frequent collocates for *hill country/ hill-country* in the three components of the GloWbE corpus as extracted by the corpus web interface shown in Table 2 below, except *park*, refer to specific socio-cultural nuances associated with the people of Sri Lanka, especially with the hill country Tamils. The collocate *park* in BE data represents specific 'non-generic' references such as 'Robin Hill country park'. The other collocates are present significantly in SLE data and marginally in BE data. Even the marginal presence of them in BE data has references to Sri Lanka (except for *central*). Therefore, these data suggest that the meaning of *hill country* is distinctive in SLE, expanding the semantics of the term 'continued' from the historical input variety, BE.

Table 2. Top ten most frequent collocates for *hill country* (left) and *hill-country* (right) in GloWbE-SL, GloWbE-Ind, and GloWbE-Gb within a collocation span of L4 and R4, reproduced from the GloWbE web interface.

| Collocates <i>Hill country</i> | Total | GB | Ind | SL | MI | Collocates <i>hill-country</i> | Total | GB | Ind | SL | MI |
|-----------------------------------|-----------|-----------|----------|-----------|------|-----------------------------------|-----------|----|-----|-----------|-------|
| <u>TAMILS</u> | 94 | <u>2</u> | | <u>92</u> | 7.81 | <u>TAMILS</u> | <u>31</u> | | | <u>31</u> | 9.78 |
| <u>SRI</u> | <u>59</u> | <u>1</u> | | <u>58</u> | 4.88 | <u>YOUTH</u> | <u>5</u> | | | <u>5</u> | 7.11 |
| <u>LANKA</u> | <u>54</u> | <u>1</u> | | <u>53</u> | 5.16 | <u>SINHALESE</u> | <u>4</u> | | | <u>4</u> | 7.67 |
| <u>KANDY</u> | <u>25</u> | <u>1</u> | | <u>24</u> | 8.38 | <u>PLANTATIONS</u> | <u>3</u> | | | <u>3</u> | 10.29 |
| <u>TEA</u> | <u>25</u> | <u>2</u> | | <u>23</u> | 5.86 | <u>CITIZENSHIP</u> | <u>3</u> | | | <u>3</u> | 8.87 |
| <u>CENTRAL</u> | <u>24</u> | <u>2</u> | <u>1</u> | <u>21</u> | 4.81 | <u>LANKA</u> | <u>3</u> | | | <u>3</u> | 4.56 |
| <u>PARK</u> | <u>21</u> | <u>16</u> | | <u>5</u> | 4.98 | <u>SRI</u> | <u>3</u> | | | <u>3</u> | 4.15 |
| <u>TAMIL</u> | <u>16</u> | <u>1</u> | | <u>15</u> | 4.38 | <u>STATELESS</u> | <u>2</u> | | | <u>2</u> | 11.98 |
| <u>SINHALESE</u> | <u>15</u> | | | <u>15</u> | 6.01 | <u>CWC</u> | <u>2</u> | | | <u>2</u> | 11.93 |
| <u>PLANTATIONS</u> | <u>13</u> | <u>1</u> | | <u>12</u> | 8.84 | <u>SLAUGHTERED</u> | <u>2</u> | | | <u>2</u> | 10.01 |

Table 3 reproduced below shows the most frequent collocates for *up country/ up-country* in the three components of GloWbE corpus as extracted by the corpus web interface. Just as with *low country* and *hill country*, the most frequent collocates; like *Tamil/s*, *Sinhalese*, *Muslim*, *leader*, *violence* etc.; top the SL list, and they are associated with socio-cultural nuances in the Sri Lankan society rather than with a context of geo-climatic reference as suggested in dictionary definitions.

Table 3. Top ten most frequent collocates for *up country* (left) and *up-country* (right) in the three GloWbE components within a collocation span of L4 and R4, reproduced from the GloWbE web interface.

| Collocates up country | Total | GB | IN | SL | MI | Collocates up -country | Total | GB | IN | SL | MI |
|--------------------------|-------|----------|----|-----------|-------|---------------------------|-------|----|----|-----------|-------|
| <u>TAMILS</u> | 41 | | | <u>41</u> | 9.33 | <u>TAMIL</u> | 94 | | | <u>94</u> | 10.29 |
| <u>MESSED</u> | 19 | <u>4</u> | | <u>1</u> | 10.33 | <u>TAMILS</u> | 66 | | | <u>66</u> | 9.09 |
| <u>FUCKED</u> | 15 | <u>4</u> | | <u>1</u> | 10.32 | <u>WOMAN</u> | 19 | | | <u>19</u> | 3.09 |
| <u>TAMIL</u> | 13 | | | <u>13</u> | 6.75 | <u>LEADER</u> | 14 | | | <u>14</u> | 4.15 |
| <u>SINHALESE</u> | 11 | | | <u>11</u> | 8.29 | <u>VIOLENCE</u> | 13 | | | <u>11</u> | 5.45 |
| <u>SCREWED</u> | 10 | <u>2</u> | | | 8.83 | <u>CLUB</u> | 8 | | | <u>8</u> | 3.69 |
| <u>MOVING</u> | 6 | <u>6</u> | | | 4.20 | <u>ESTATE</u> | 7 | | | <u>7</u> | 5.22 |
| <u>TRAVEL</u> | 6 | <u>2</u> | | <u>1</u> | 3.84 | <u>ESPECIALLY</u> | 7 | | | <u>4</u> | 3.06 |
| <u>CLUB</u> | 6 | | | <u>1</u> | 3.72 | <u>MUSLIM</u> | 7 | | | <u>6</u> | 4.81 |
| <u>AREAS</u> | 6 | | | <u>4</u> | 3.32 | <u>SPORT</u> | 7 | | | <u>7</u> | 3.95 |

The results of the collocates for *mid country/mid-country* are not reproduced here because their frequencies are low. They are discussed in the following section.

Discussion And Conclusion

Results of the corpus-linguistic methods of analysis presented in the previous section revealed that the four lexical items analysed in the paper show a very high comparative frequency of occurrence in SLE in all 03 types of written language corpora used for the analysis. The most frequent collocates, which have a high co-occurrence probability with these lexical items, show a context of usage which is different from the context suggested in their dictionary definitions. In light of these findings, this section discusses how this semantic extension in SLE has occurred.

Figure 4 below reproduces the concordance lines from ICE-SL written component for *low country*. The use of the term in concordance lines 3-5 in the figure refers to the low-lying geo-climatic region of Sri Lanka, as opposed to the central hills. This usage corresponds to the usage given in the OED. The usage in concordance lines 1 and 2, however, refers to a kind of social stratification that prevails in the Sinhala community of Sri Lanka,

which divides the community into two social classes: *low-country* and *up-country* or *Kandyan*. The latter type of Sinhalese considered themselves descendants of the kingdom and aristocratically superior over the former type. This distinction infiltrates into most aspects of a Sinhala person's life including marriages and even matrimonial dresses, in which case only Kandyans wear the traditional royal attire. It surpasses geographical boundaries and relates to the caste system. Frequent collocates described in section 3C, such as *Kandyan (s)*, *castes*, *headman*, *Karava* (a caste) too corroborate this cultural nuance. Additionally, this compound also refers to a kind of traditional dancing system in Sri Lanka, which is, again, different from up-country or Kandyan dance. Attesting of *dancing* as a frequent collocate of *low-country* (section 3C) testifies this. Thus, this usage of the term to refer to specific cultural nuances of the Sri Lankan society is clearly different from the usage given in the OED.

| 18 Concordance | |
|----------------|--|
| 1 | to Kanda-Uda-Paa-Rate, the British colonial rulers referred to the Sinhalese toiling under their yoke as low-country Sinhalese and those Sinhalese who lived in the Kingdom of the King of Kandy |
| 2 | their decisive empire policy, did their best to drive a wedge between the so-called Kandyan and the low-country Sinhalese by bestowing official recognition to the division by various court and d |
| 3 | Sindurita, Lobophyton and Sarcocofyton spp) are also present. The National Park is located within the low-country dry zone with an annual temperature of around 27.0-27.5 degrees C and annual rain |
| 4 | workers during the 35 plus years the writer of this article worked as a Tea Planter in the High, Medium and low-country districts of Sri Lanka - of course you need to speak their language fluently to re |
| 5 | these leaves on branched hard heavy twigs (Jayaweera, 1982). It is very common everywhere in the low-country of Sri Lanka. The wide distribution in Sri Lanka has led to the extensive use of th |

Figure 4. Concordances for *low country* extracted from ICE-SL written component.

As exemplified in the above concordance lines, the lexical item *low country* occurs both as an adjectival compound (lines 1-4) and a nominal compound (line 5). In the source language, Sinhala, a similar compound, functioning both adjectivally and nominally, is found. The Sinhala compound is /pahaṭṭa ratə/, in which /pahaṭṭə/ means 'low', and /ratə/ means 'country'. Thus, in both languages, attributive modification of a noun to form a compound is present. Just as for English 'low', Sinhala /pahaṭṭə/ can refer both to 'a lower elevation' and to 'a lower status'. The use of /ratə/ to refer to 'a politically

autonomous area or a state' descends from pre-colonial times of Sri Lanka, as exemplified in the division of the country to three 'states': /ruhunu ratə/, /ma:ja: ratə/ and /pihiti ratə/. Thus, the term for 'low country Sinhalese' in Sinhala, in the specific cultural reference of the term, is /pahaṭṭə ratə sinhələjə:/. What has happened here is that a new meaning with a socio-cultural reference, derived from the source language, has attached to the 'continued' meaning of the compound *low country* given in the OED, having only a geographical reference, through language contact. This new meaning can be viewed as a semantic extension as it is in line with Winford's (2003, p.45) definition of semantic extension (see section I).

Sample concordances for *hill-country* extracted from GloWbE-SL shown in Figure 5 below depict a special use of this compound in SLE with reference to the Tamil community of Sri Lanka. This usage is relevant to the two main types of Tamils in Sri Lanka: Sri Lankan Tamils and Indian Tamils. *Hill country Tamils* means Indian Tamils, who descend from plantation workers sent from South India to Sri Lanka, and they live mostly in the central hills of Sri Lanka. Attesting of collocates such as *citizenship*, *stateless*, *CWC* (Ceylon Workers Congress-a political party) etc. which refer to the affairs of plantation Tamils shown in section 3C above also illustrates this use of the term. Thus, in SLE, this compound has a geo-climatic reference to the central hills of Sri Lanka, a cultural reference to *hill country Sinhalese* (as opposed to *low-country Sinhalese*), and a cultural reference to *hill country Tamils*, among other references. This may explain the very high comparative frequency of this compound in GloWbE-SL data as shown in Figure 3. The semantics of this compound in SLE is, thus, different from the meaning given in the OED.

language and land rights, disenfranchisement of the hill-country Tamils in 1948, and the introduction of quota system for Sinhalese and the Muslims or the Sinhalese and the hill-country Tamils in several localities in the South. Some of them are according to a ministry directive that 50% should be Hill-Country Tamils, 25% Sinhalese and 25% Ceylon Tamils and Muslims (they prefer to be identified as a separate ethnicity), hill-country Tamils (migrants from Tamilnadu state in India during the British period). The government passed laws in 2009 to grant citizenship to hill-country Tamils living among other Sri Lankan ethnic Tamils in refu

Figure 5. Sample concordance lines for *hill-country* from GloWbE-SL

The corresponding term for *hill country* in Sinhala is /kaḍũ ratə/, in which /kaḍũ/ means 'hills'. The corresponding singular term for 'hill' is /kandə/, which is etymologically related to Sanskrit 'skandha' (Geiger, 1941, p.36). The term for 'hill country Tamil' in Sinhala is /kaḍũ ratə demələ/. Because of the reference to Tamil, we searched the Tamil equivalent of *hill country* as well. In Tamil, the corresponding term is [malai na:tu], and in the adjectival form it is [malai naṭṭil] as in [malai nāṭṭil tamil], meaning 'hill country Tamils'. Thus, both Sinhala and Tamil have corresponding models for this compound. Therefore, it can be posited that the distinctive meaning of the compound *hill country* in SLE referring to the cultural nuances of Sri Lanka is a semantic loan or an extension based on the parallel terms used in the local languages of Sri Lanka, through language contact phenomena.

Extension of the semantics of the English term to describe socio-cultural nuances of the Sri Lankan society is even more obviously seen in *up country/up-country*. As the concordance lines from GloWbE-SL in Figure 6 below show, the term is used to refer to *aristocratic Sinhalese, plantation Tamils, dancing traditions* etc. Collocates of the term attested in section 3C, such as *woman, leader, and violence*, too refer to the cultural issues of the up-country communities, particularly of plantation Tamils. This linguistic context has little to do with the term's geographical reference. The source

language model for this extended meaning of the term seems to be from Sinhala, for Tamil speakers use the equivalent for *hill country* ([malai na:tu]) for *up country* as well. The Sinhala equivalent for *up country* is /udə ratə/, in which /udə/ means ‘up,’ referring both to ‘an upper elevation’ and ‘an upper social class’. Hence, in SLE, the term *up country/up-country* too is a semantic loan modelled on the semantics of the Sinhala equivalent, /udə ratə/.

entity only by combining their numbers with that of the Up Country Tamils or Tamils of recent Indian origin high ranking low country Maha mudaliyar nobility and up country radala aristocracy. # Many of those attached among the Sinhalese, we had low country Sinhalese, up country Sinhalese identities. So identities have been number there were several local cultural traditions- the Up country tradition of Walauwe society, the low-country movements. This has been developed according to the up country dancing traditions. Uduwela, Paranagar

Figure 6. Sample concordance lines for *up country* from GloWbE-SL

The last term analysed in the present paper, *mid country/mid-country*, was neither known to be previously recorded nor was a dictionary meaning for it found. This lexical item refers to *the geo-climatic region in Sri Lanka in between the up country and the low country*, and it seems exclusively specific to Sri Lankan settings because even a dictionary entry is not found for it. As the concordance lines in Figure 7 below depict, this, too, occurs both as an adjectival compound (line 1) and as a nominal compound (lines 2-4). The corresponding Sinhala compound is /mædɔ ratə/, in which /mædɔ/ means ‘mid’. The Tamil equivalents are either [Mattiya nātu] or [Mattiya pakuti] in which [Mattiya] means *mid/central*, [nātu] means *country*, and [pakuti] means *region*. The term’s collocates in SLE (though not very frequent) are words like *cultivation, estates, biodiversity, forests, and garden*, which are more

to do with geo-climatic contexts than to socio-cultural contexts. As this meaning was not found as a 'continued' form in BE altogether, we posit that *mid country* in SLE is a loan translation (a combination of native morphemes in imitation of foreign pattern) modelled based on the corresponding local language equivalents through contact phenomena.

| N | Concordance |
|---|--|
| 1 | worries given up animal husbandry" by P. Rajaratnam, Newera Eliya Group Correspondent The hill and the mid-country zones which were the best milk producers in Sri Lanka a few years |
| 2 | National Milk Board collected 205,000 pints of milk of which 200,000 pints came from the Hill Country and mid-country. An appreciable quantity was from the Estate sector. Taking into |
| 3 | in the aftermath of the nationalization of plantation, being more pronounced in the up-country than in the mid-country, while tea grown extent expanded significantly. The total tea cultivation |
| 4 | produces. We have long way to go from the present 5th position. Uneconomical Tea lands in the low and mid-country are also potential areas for rubber planting. Skill Development, Tea |

Figure 7. Concordances for *mid-country* extracted from SAVE-SL

When compared with the other three lexical items, however, the frequency of *mid country* in SLE data is significantly low, both in SAVE and GloWbE corpus data: For example, in GloWbE-SL *mid country* produces only 15 hits, whereas, *low country*, *hill country* and *up country* produce 283, 750, and 425 hits respectively. One reason for this can be that *mid country* is used in SLE mainly with reference to a geo-climatic region, while the other three terms have additional socio-cultural references. This also confirms our previous observation that the other three compounds are used in SLE with a distinctive meaning referring to socio-cultural nuances.

Thus, the word associations and contexts of usage of the four lexical items - *low country*, *hill country*, *up country*, and *mid country* - show that they can be described as loan shifts (loan meanings) in SLE, in which the words are English, but their meanings are either partly or completely borrowed from local source languages. This has become necessary because English functions as a *transplanted language* in contexts such as those in Sri Lanka where 'it (English) is used by a significant number of speakers in social,

cultural and geographical contexts different from the contexts in which it was originally used' (cf. Fernando, 2003).

The context provides "meaning", and as the cultural and linguistic contexts change the language acquires new meanings. In new contexts, therefore, new *uses* and *users* of English have developed appropriate linguistic tools. (Kachru, 1986, p.30)

The four lexical items analysed in the present paper, thus, reflect the sociocultural identities of the people who use a transplanted language, which Kachru (1992, p.235) refers to as *acculturation*. Therefore, as hypothesised, these lexical items are distinctively SLE specific. Elsewhere, even when they are used, they do not refer to peoples' ethnicities, caste systems, and other similar cultural nuances exemplified in the paper. Since English has become a language of Sri Lanka (at least as a 'link language' officially), inevitably, it has to readjust itself to represent and reflect such localized phenomena. This is why English has borrowed the discussed meanings from local languages into these English lexical items. Such distinctive uses have made the English used in Sri Lanka different from the English used in, Great Britain, for instance. This may be why linguists in the field of 'World Englishes' tend to refer to different 'Varieties of English' or different 'Englishes', rather than to just one English at present.

References

- Baker, P. (2010) *Corpus Methods in Linguistics*. In Litosseliti, L (ed): *Research Methods in Linguistics*. London: Continuum, pp. 93-113.
- Davies, M., and Fuchs, R. (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, pp.1-28.
- Fernando, S. (2003) *The Vocabulary of Sri Lankan English: Words and Phrases That Transform a Foreign Language into Their Own*. Full Paper No. 26. 9th International conference on Sri Lanka Studies, 28th - 30th November 2003, Matara, Sri Lanka.
- Geiger, W. (1941) *An etymological glossary of the Sinhalese language*. Colombo, Royal Asiatic Society, Ceylon Branch.
- Greenbaum, S. (1996) *Comparing English worldwide: the International Corpus of English*. Oxford, Clarendon Press
- Gries, S. T. (2009) *Quantitative Corpus Linguistics with R: a Practical Introduction*. New York, Routledge.
- Halverson, J. (1966) Prologomena to the study of Ceylon English. *University of Ceylon Review* 24, pp. 61-75.
- Kachru, B. B. (1986) *The alchemy of English: the spread, functions and models of non-native Englishes*. Oxford, Pergamon Institute of English.
- Kachru, B. B. (1992) The Second Diaspora of English. In: T.W. Machan & C. T. Scott, (eds), *English in its social contexts: essays in historical sociolinguistics*, New York: Oxford University Press, pp. 230-252.

- Kachru, B., B. (2001) New Englishes. In: R. Mesthrie & R. E. Asher (eds). *Concise encyclopaedia of sociolinguistics*. Amsterdam: Elsevier, pp. 519-524
- Meyler, M. (2007) *A dictionary of Sri Lankan English*. Colombo: Michael Meyler.
- Mukherjee, J. (2012) English in South Asia - Ambinormative orientations and the role of corpora: the state of the debate in Sri Lanka. In: A. Kirkpatrick and R. Sussex. (eds). *English as an International Language in Asia: Implications for Language Education*, New York, NY: Springer, pp. 191-208.
- OED Online. (2017) Oxford University Press. Available from: <https://en.oxforddictionaries.com/definition/us/English/> (accessed July 09, 2017)
- OneLook. (2017) Datamuse. Available from: <http://www.onelook.com/> (accessed July 13, 2017).
- Parakrama, A. (1995) *De-Hegemonizing Language Standards: Learning from (Post) Colonial Englishes about 'English'*. London: MacMillan.
- Passé, H. A. (1955) *The use and abuse of English; common errors in Ceylon English*. London: Oxford University Press.
- Schilk, M. (2011) *Structural nativization in Indian English lexicogrammar*. Amsterdam, John Benjamins Publishing Company.
- Scott, M., (2008) *WordSmith Tools version 5*, Liverpool: Lexical Analysis Software.

Simpson, J.A. and Weiner, E. (1989) *The Oxford English Dictionary. Second edition*: Oxford: Oxford University Press.

Sridhar, K. K. (2012) Acculturation in World Englishes. In Chapelle, C. (ed). *The encyclopedia of applied linguistics*. Oxford, Wiley-Blackwell, pp. 12-15

Winford, D. (2003) *An introduction to contact linguistics*. Malden, Mass, Blackwell Pub.