SINHALA AND TAMIL INFLUENCE ON SRI LANKAN ENGLISH PARTICLE USE: A CORPUS-BASED STUDY ON THE CASE OF 'FOR'.

MDSS Kumara[12]

[1]DELT, Rajarata University of Sri Lanka, Mihintale, Sri Lanka

[2]Justus Liebig University of Giessen, Germany

\# skmenikpure@gmail.com

Abstract - Colonial expansion brought English into contact with different languages. In Sri Lanka, English has been in contact with Sinhala and Tamil for over 200 years. Today, all users of Sri Lankan English (SLE) are either bilingual or multilingual. In bilingual language processing, linguistic habits of one language exert influence on the other causing the latter to restructure. Although restructuring of a language in a language contact situation is overtly seen in phonology and lexis, syntactic and grammatical innovations are subtle and they take time to establish. When grammatical innovations do occur, they tend to begin at the intersection of grammar and lexis.

The present paper analyses the lexis- grammar interface of Particle Verbs (PVs) in SLE with special reference to the particle/preposition 'for'. The study uses a corpus-based methodology, and data are from the Sri Lankan, Indian, and Great Britain components of the International Corpus of English and Corpus of Global Web-based English. Results are presented of two PVs, 'sit for' and 'contest for', which are innovations having significant frequency of occurrence in SLE data.

When these structures are compared with their corresponding equivalents in Sinhala and Tamil, it is revealed that the case environments of the relevant verbs in Sinhala and Tamil make it compulsory for those equivalents to have a counterpart for 'for'. This may be why SLE prefers these PVs, while its historical input variety, British English, prefers simplex verbs without the particle, i.e. 'sit' and 'contest', to convey the same idea.

Keywords - SLE, Particle Verbs, language contact, corpora

I. INTRODUCTION

Although English was originally the language of the people of England, it has been displaying the feature of widespread expansion right from its origins. In the 16th and 17th centuries, it spread to Wales and other British Isles such as Scotland; then, with the movement of English speaking populations, it spread to other 'Inner Circle' countries like America, Canada, Australia, and New Zealand (cf. Kachru, 1992). Later introduction of English to Asian, African, and Latin American countries with the British and American colonial expansion, however, brought it into contact with a large number of other genetically and culturally unrelated languages (cf. Bhatt, 2001). As a result of this widespread movement of English, a large number of International Varieties of English (Crystal, 1995) have emerged. Sri Lankan English (SLE) is one such variety of English.

The English language reached Sri Lanka with the arrival of the British East India Company in 1796. Since then, it has been in contact with the two prominent local languages of the country- Sinhala and Tamil. Additionally, as Mendis & Rambukwella (2010) observe, all speakers of SLE today are bilingual and some are trilingual. In bilingual language processing, one language exerting influence on the other (and vice versa) is inevitable, and it results in restructuring of the systems of the contact languages. Restructuring of languages in a contact situation can affect all systems of a language. Of these, phonology and lexis are the most affected. In contrast, syntax and grammar tend to be much more stable and resistant to change comparatively. As Schneider (2007) points out, when grammatical

innovations do emerge, they start out at the intersection of grammar and lexis. World Englishes (WE) literature has already recorded that innovations in the lexis-grammar interface such as different complementation patterns of verbs, co-occurrence and collocational tendencies of words in phrases, and patterns of word formation can be found especially with regard to WEs (Schneider, 2004). Particle/preposition use, which is the object of investigation in the present study, is also common among such lexis-grammar innovations.

The object of investigation in the present study is termed 'particles' following the tradition of the WE field as a neutral cover term for prepositions and spatial adverbs, as was done by Schneider (2004). Since the focus of the present paper is only on the particle 'for', the term 'prepositions' could also have been used. The term 'particles' suits better, nevertheless, because references are made in the paper to Sinhala and Tamil as well. Although Sinhala is an Indo-Aryan language and Tamil a Dravidian language, both have 'postpositions' rather than 'prepositions'. Additionally, both in Sinhala and Tamil, not only postpositions but also case inflections play a role in performing the functions of prepositions/particles in English. Thus, certain meanings expressed by the English particle for are represented in Sinhala by the Sinhala dative suffix /tə/ and in Tamil by the Tamil dative suffix /kku/. In addition, certain meanings expressed by the English particle for are represented by the Sinhala postposition /saḏə̃ha:/ and the Tamil bound postposition /a:ka/ (Lehmann, 1989) .

The lexis-grammar interface of the particle 'for' (or any other particle for that matter) can be studied when it goes into combinations with verbs, adjectives, nouns etc. The focus of the present paper is verb-particle combinations, which is, perhaps, the most frequently discussed combination with particles in WE literature. The term 'Particle Verb (PV)' is used to refer to such combinations in the present paper following Zipp (2014), who adapted the term from Quirk et al (1985) as a cover term for all combinations of verb and particle, subsuming the three main categories defined in Quirk et al (1985): phrasal verbs, prepositional verbs, and phrasal-prepositional verbs. In fact, the PVs with 'for' discussed in the present paper fit prepositional verbs category out of these three categories. Only those PVs with 'for' having objects that answer the questions 'who' and 'what' are used for analysis (as opposed to those answer questions 'when', 'where' etc., which are free adverbial combinations).

II. METHODOLOGY AND EXPERIMENTAL DESIGN

The present study employs a synchronic, corpus-based methodology as its principal analytical procedure. Performance data for the linguistic analysis are extracted from 2 types of written language corpora representing SLE, which are compared with data from parallel corpora representing British English (BE), the historical input variety of SLE; as well as Indian English (IndE), in order to effectively evaluate the significance of occurrence of selected structures in SLE. The term corpora (singular corpus) is used in the present paper to refer to collections of texts or transcribed speech that are representative of a language, and are stored and accessed electronically.

The first type of written language corpora of the present study's corpus environment is from the written parts of the respective components of International Corpus of English (ICE) (Greenbaum, 1996), 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 the ICE consists of approximately 400,000 words. The second 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 (including online newspapers and blogs) 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.

Selection of only written language data in the present study is motivated both by practical and theoretical reasons. Practically, as the spoken part of ICE-SL is still in the compilation stage, there are no standard corpora of SLE representing spoken data. Compiling a sufficiently sized representative corpus of spoken SLE is also beyond the scope of a study like the present study. Theoretically, although linguistic innovations are considered to be accumulated in genres of spoken language at first, their infiltration into written genres signals the advancement of a variety through its evolutionary stages. According to Schneider (2003, p.252), interaction between spoken

behavior and written norms in a society takes place in the phase of endonormative stabilization, with codification. Thus, it is hypothesized that the appearance of distinctive features of SLE in standard written language corpora, covering a wide range of both formal and informal genres, signals that SLE is reaching the stage of endonormative stabilization.

ICE corpora data were analyzed using the concordancer (corpus access software), Wordsmith Tools-Version 5 (Scott, 2008), with 'for' as the search word and verbs occurring with it at positions 01, 02, and 03 to the left of it (L1, L2, and L3) as collocates in all three written components. In order to exclude adverbial combinations, all concordance lines were manually read. The frequencies of the selected concordances were normalized to 1 million words because the word counts of each ICE component are not exactly the same. Selected PVs in the three ICE components were, then, categorized as recorded, partly-recorded, and unrecorded based on the combinations' extent of availability in the online dictionaries indexed at Onelook dictionary search <http://www.onelook.com/>. If the combinations are not recorded in any of the online dictionaries indexed at Onelook dictionary search, they were categorised under unrecorded. If the combinations are available in a wide range of dictionary sites indexed at Onelook dictionary search, including Oxford, Cambridge, Collins, or Macmillan, they were categorised under recorded. If they are not recorded in any of the above dictionary sites, but recorded only in the online The Free Dictionary, they were categorised under partly-recorded. From the unrecorded PVs, those that occur 05 or more times in ICE-SL were considered for evaluating contact language influence.

Additionally, as ICE corpora are comparatively small in size, unrecorded and partly-recorded PV types selected through the analysis of ICE corpora data were searched in GloWbE corpora using its online interface as the concordance under individual scholar license. With GloWbE, the verb (lemma) was used as the search term, and 'for' as the first collocate to the right of the verb lemma for brevity's sake (without going to the second and third position to the right). The frequencies of these PV types in GloWbE corpora were normalized to 100 million words, and those that hit over 100 normalized counts of the raw frequencies in SL component were manually read in order to exclude adverbial combinations. In addition to the significantly frequent PVs selected through the analysis of ICE and GloWbE corpora data, PVs already recorded in WE literature were also searched in the two corpus environments.

III. RESULTS

The page length restrictions of the present paper do not warrant reporting the complete results of the study. Therefore, the results of only two PVs that show evidence of local language influence are discussed here: The first, 'sit for', is found through the analysis of ICE corpora data, and the other, 'contest for', is described through the analysis of GloWbE corpora data.

Among the unrecorded PVs that are 05 or more times frequent in ICE-SL, 'sit for' meaning 'take an exam' triggered particular interest of the present study because Gunesekera (2005) outlines this as a case of 'overuse of prepositions', which, according to her, is a characteristic of SLE syntax. She explains that the preposition 'for' in the sentence 'they did not sit for the exam' is not required because the structure 'sit the exam' alone 'works'. This PV occurs 14 times in ICE-SL data, whereas both ICE-GB and ICE-Ind show no occurrence of it. Figure 1 below shows normalised frequencies of the PV 'sit for' occurring in the three components of ICE and GloWbE. Data in the figure show that the presence of this PV is highly significant in SLE, whereas in IndE and BE, its presence is marginal, more so in the latter.
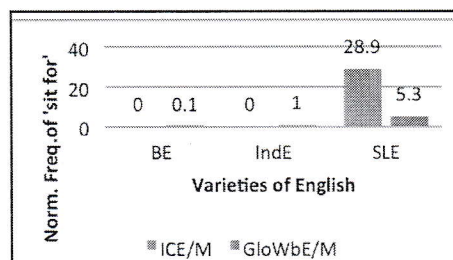

Figure 1. Frequencies of 'sit for' in the three components of ICE and GloWbE normalised to 1 million words.

As this PV is an instance of an 'addition of a particle' (to the simplex verb 'sit'), the occurrence of the simplex verb 'sit' to convey the same meaning was examined. Figure 2 below shows normalised frequencies of the PV 'sit for (an exam/test)' and its single-verb alternative 'sit (an exam/test)' in the three ICE-components.
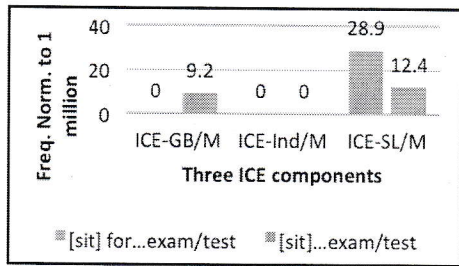
Figure 2. Normalised frequencies of 'sit for' and 'sit' in the three ICE-components

As depicted in Figure 2, written component of ICE-Ind records neither the PV nor the simplex verb, while ICE-GB attests only the simplex verb. ICE-SL shows the occurrence of both the PV and the simplex verb, but with a clear preference for the former. Cross-checking with 1000 random hits of the lemma [sit] in GloWbE components too produced similar results (Figure 3).
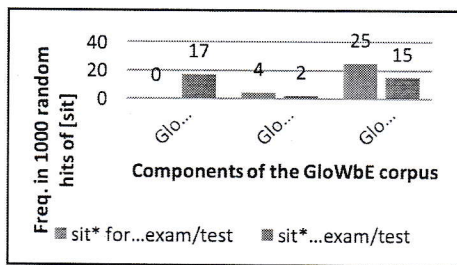

Figure 3. Frequencies of 'sit for' and 'sit' in 1000 random hits of the lemma [sit] in the three GloWbE components

Data depicted in Figure 3 show statistically significant differences at p<0.001, df=2 (x2≈ 19.90) and a 'redundant' correlation (Cramer's V≈0.56). Further, as elaborated by Figures 1 through 3, the two 'new' varieties of English (SLE and IndE) tend to prefer the innovative combination 'sit for' over the simplex verb 'sit', continued from the historical input variety (BE). For BE, the preference is still the simplex verb.

Key Word In Context (KWIC) lines (concordances) for 'sat for' extracted from the written component of ICE-SL, which are depicted in Table 1 below, show that six out of the nine concordance lines are from the text category W2A, which is Academic Writing (Printed), and the others are also from printed categories; Instructional Writing (W2D) and Creative Writing (W2F).

Table 1. Concordances for 'sat for' from ICESL written component and their text category names

Concordance line	Text category
after we left the school we sat for the public service exam	W2F-006.txt
estern Music privately and sat for the examination throug	W2D-018.txt
gic AlianceAlliance "After I sat for this course, II knew wh	W2D-011.txt
the students who have not sat for or passed GCE (AL), n	W2A-008.txt
MCQs of all trainees who sat for the pre and post tests	W2A-006.txt
listrict (NTH) all 9 trainees sat for both pre and post tests	W2A-006.txt
9 participants respectively sat for the pre-test and particip	W2A-006.txt
ly the marks of those who sat for the post-test were usec	W2A-006.txt
e A total of 3477 students sat for the Grading Tests in 19	W2A-005.txt

The concordances for 'sit* for' from ICE-SL given below show that this PV appears in Non-printed text categories W1A (Student Writing) and W1B (Letters) as well (Table 2). In GloWbE-SL, the PV occurs in online written genres including blogs. Thus, it is quite obvious that this PV is used in a wide range of written genres in SLE, including very formal genres such as Academic Writing.

Table 2. A section of concordances for 'sit for' from ICE-SL written component and their text category names

Concordance line	Text category
when i come back i have to sit for 4 papers! NOT fun! :(bu	W1B-006.txt
s of all those who could not sit for the post-test have been	W2A-006.txt
be chosen even though they sit for the exam the next time	W1A-012.txt
Thirteenthirteen years they sit for A/L's which desidedeci	W1A-013.txt

The second PV analysed in the present paper, 'contest for', was culled from Nihalani (2004) with regard to IndE. This PV was found to be significant in SLE through the analysis of GloWbE corpus data. The absolute frequencies of this combination are higher in the two new varieties of English than in BE (Figure 5). With regard to frequencies normalised to 100 million words, this variation is significantly pronounced specially in SLE, in which the figure even exceeds 100 (120), whereas BE records only 11 occurrences. The written components of the relevant ICE corpora, however, do not record this combination.
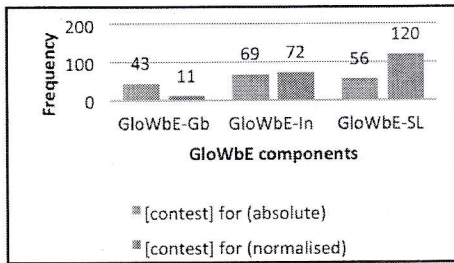
Figure 5: Absolute and normalised frequencies of 'contest for' in the three GloWbE components

This PV too appears to be a case of an 'addition of a particle' (to the simplex verb 'contest'). Therefore, total occurrences of the lemma [contest] were sought in GloWbE corpora. As the occurrence of the verb lemma itself is higher in the two new varieties, frequency counts of the PV were worked out as a percentage of the counts for the entire lemma. As can be seen in Figure 6 below, the PV's occurrence is significant in the two new varieties of English as a percentage as well (more so in IndE though).
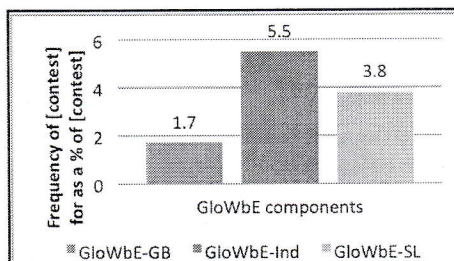

Figure 6: Frequency of '[contest] for' as a percentage of the frequency of [contest] in the three GloWbE corpora

IV. DISCUSSION AND CONCLUSION

The results of the corpus data analysis presented in the previous section show that the frequency of occurrence of both of the analysed PVs stands out in the two new varieties of English. Previous studies in the field also reveal that lexico-grammatical innovations in new varieties of English don't necessarily stand out categorically from the more established varieties (such as BE), but only show frequency-related differences with them (cf. Mukherjee 2007, p.175; Bernaisch, 2015, p.137).

In light of the significant frequency of occurrence and widespread distribution of the PV 'sit for' in SLE, it is important to trace the diagnostic processes that may have led to the creation of this innovation. An analogy based explanation is difficult to trace because, for one, the verb 'sit' in this combination (in the context of exam/test) is used in a rather figurative sense (meaning 'take an exam' or 'appear for an exam'). Further, most of the other verbs that collocate with exam/test, such as take, pass, write, and fail do not provide a model (such as *take for an exam or *pass for an exam) for this combination with the particle 'for'. Therefore, it is worth tracing the possible contact effects from Sri Lankan local languages, Sinhala and Tamil.

The figurative use of 'sit for' in the context of 'exam/test' finds its equivalent in less formal Sinhala. In more formal Sinhala, an equivalent of 'appear' (for), /peni:sitin əvʌ/, is used in this context. The equivalent of 'for' in this PV, /t ə/, manifests as a dative suffix for the prepositional object of the construction as shown below:

Sinhala: /viba:getə   va:divenəvʌ /
Gloss: the exam for   sit
SLE: sit for the exam

The case environment of the Sinhala verb, / va:divenə,vʌ/, is such that the preceding noun taking the suffix,/tə/, is mandatory. Thus, the following construction is unacceptable in Sinhala:

Sinhala: */viba:ge va:divenəvʌ /
Gloss: the exam sit
SLE: sit the exam

The other Sri Lankan local language, Tamil, too has an equivalent for 'sit for...exam/test'. In Tamil too, the equivalent of 'for' is a dative suffix, /kku/, as illustrated below:

Tamil: [avarkaḷ aṭuttamuṟai parītcaikku amarvārkaḷ]
Gloss: they   next time exam for   sit
SLE: They sit for the exam the next time

Just as in Sinhala, this case marking is mandatory in Tamil, making the following construction unacceptable in Tamil:

Tamil:*[parītcai amarvārkaḷ]
Gloss: the exam   sit
SLE: sit the exam

Thus, the equivalent of 'sit the exam' being absent in Sinhala and Tamil, and only the equivalent of 'sit for the exam' being present in the two local languages may have contributed to the significant, widespread occurrence of 'sit for the exam' in SLE. This corroborates Schneider's (2007, p.107) position on 'contact' as a linguistic process which triggers or accelerates various types of innovation. Examination of the ICE-SL metadata reveals that in seven

out of the ten texts from which the 14 occurrences of 'sit for…exam/test' come, the informant has Sinhala first language background, while in two texts the informant has Tamil first language background. In the remaining text, the metadata information is not sufficient to trace the first language background of the informant. Hence, the possible transfer effect for the said structure seems to be from both Sinhala and Tamil. Perhaps, fewer instances of the PV have been recorded from informants with Tamil first language background because of the lesser proportion of such informants in the SLE speech community.

As with 'sit for', in the corresponding Sri Lankan local language structures of 'contest for', an equivalent for the particle 'for' is mandatory. There are two main equivalents in Sinhala for the particle 'for' in 'contest for'. One is a postpositive particle, /sʌd̪əha:/, which is more formal, and the other less formal equivalent is the dative suffix /tə/. The following structures illustrate these uses:

Sinhala: /ohu mæt̪ɪvʌrənəyə sʌd̪əha: t̪ʌrəğəkərai/
Gloss: he   the election   for   contests
SLE: He contests for the election.

Sinhala: /ohu mæt̪ɪvʌrənəyətə t̪ʌrəğəkərənəvʌ/
Gloss: he   the election for   contests
SLE: He contests for the election.

Since the case environment of the Sinhala verb, /t̪ʌrəğəkərənəvʌ (kərai)/ makes the structure to have the suffix, /tə/ (less formal versions), or the postpositive particle, /sʌd̪əha: / (more formal versions), mandatorily, the constructions below are less acceptable in Sinhala (though not totally unacceptable):

* Sinhala: /ohu mæt̪ɪvʌrənəyə t̪ʌrəğəkərai/
Gloss: he   the election   contests
SLE: He contests the election.

*Sinhala: /ohu mæt̪ɪvʌrənəyə t̪ʌrəğəkərənəvʌ/
Gloss: he   the election contests
SLE: He contests the election.

As the following structure illustrates, in Tamil too an equivalent for the particle is mandatory in this combination:

Tamil: [Aṇṇātturai pārāḷumaṉrattirku pōṭṭiyiṭṭār]
Gloss: Annadurai the parliament for contested
SLE: Annadurai contested for the parliament.

For this reason, the structure below, which is without the particle, is less acceptable in Tamil as well. Thus, the mandatory particle use in the Sri Lankan local languages, Sinhala & Tamil, in the equivalent of this PV may have contributed to its significant occurrence in SLE data.

*Tamil:[Aṇṇātturai Pārāḷumaṉram pōṭṭiyiṭṭār]
Gloss: Annadurai the parliament contested
SLE: Annadurai contested the parliament.

Because the compulsory requirement of having an equivalent for 'for' in this PV is there both in Sinhala and Tamil, it is possible that the contact effect is from both Sinhala and Tamil. However, unlike with 'sit for', it is difficult to trace the first language backgrounds of the informants of the concordances of 'contest for' because it is only attested in GloWbE corpora data. As GloWbE corpora are based on webpage data, comprehensive metadata about the sociolinguistic backgrounds of the informants are absent, which is one disadvantage of GloWbE corpora in comparison with ICE corpora. Nevertheless, 'contest for' itself shows a strength of GloWbE corpora over ICE corpora; that the PV was found in none of the relevant ICE corpora, but was found over 100 instances in some GloWbE corpora. GloWbE corpora thus have the advantage of attesting low frequency structures. Additionally, it should also be mentioned that the PVs, especially 'contest for' was found to be in a significant frequency in IndE data as well. This does not negate a possible contact language influence in SLE, however, because major local languages in India are also from Indo-Aryan and Dravidian families just as the major local languages in Sri Lanka.

In conclusion, the foregoing discussion makes it evident that the PVs 'sit for' and 'contest for' exemplify lexico-grammatical innovations in new varieties of English, particularly in SLE. Infiltration of these PVs into genres of written language (including Academic Writing) testifies that they are being accepted in terms of variety internal norms. Whereas the historical input variety of SLE, i.e., BE, is found to prefer the simplex verbs without the particle, SLE prefers these PVs with the particle. Thus, the two PVs are instances of 'addition of a particle' to a simplex verb to form new PVs, which is said to be a feature of New Englishes. The analysis of corresponding Sinhala and Tamil equivalents revealed that for both PVs an equivalent for the particle 'for' is mandatory in Sinhala and Tamil structures, and structures without the particle are unacceptable or less acceptable in Sinhala and Tamil. Hence, it can be argued that in the process of bilingual processing of language, SLE users bring their habit of using a particle in the Sinhala and Tamil equivalents of these structures into the English structures, thus preferring the PVs over the simplex verbs. Therefore, it can be posited that these PVs show influence of Sinhala and Tamil on SLE particle use. However, these findings have to be substantiated in a future study