Communication

ISSUES IN SINHALA SYNTAX: SENTENCE PROCESSING AND WORD ORDER

1. Introduction

Sinhala [also referred to as Sinhalese, Singhala and Singhalese, (Englebretson, et al., 2005)], is one of the major languages spoken in Sri Lanka (others being Tamil and English) a member of the Indo-Aryan language group which also consists of a number of other East Asian languages such as Bengali, Marathi, Punjabi, Maldivian, Hindi, and others. It has a history of more than two thousand years (Dissanayaka, 2007; A. Herath et al 1994), and since Pali and Sanskrit are said to be the root languages of Sinhala, there are a number of transferred and borrowed items from Pali and Sanskrit both in lexical and grammatical terms.

Although Sinhala belongs to the Indo-Aryan family, it has many unique features which cannot be seen in other members of that family. Sinhala is composed of two different arrays in spoken and written forms. According to previous studies (Dissanayaka, 2007; Noguchi, 1984; Miyagishi, 2003, 2005), these two arrays are different both in lexical and grammatical terms. This paper focuses only on the spoken form in relation to sentences processing.

1.1 Aim of this paper

Despite the fact that the Sinhala language has a long history of more than 2000 years, linguistic research on its syntactic issues such as word order, scrambling, sentence processing are considered to be limited. Therefore, this paper aims to analyze two linguistic features of the Sinhala language; the information cues available for Sinhala sentence processing ('sentences processing' is a phenomenon which accounts for speakers or listeners processing any given language), and the attributes of free word order of the Sinhala language.

2. Sinhala language and its structure

Words in any language have a category called "parts-of-speech" which provide syntactic information. These categories provide necessary information in order to construct grammatical sentences. For example, although a speaker knows the meaning of the words '*hit*, *the table*, *I*' and they belong to English language, s/he must possesses the syntactic information of these words in order to construct a grammatical sentence; *I hit the table*

(SVO). Native speakers of any given language intuitively know these syntactic categories in order to make syntactically and semantically sensitive sentences for the purpose of communication (transform information). For example, consider the sentence below in 1).

1) අමර නිමල ගැහුව

amara nimala-ta gehuwa

Amara (φNOM) Nimala (DAT) hit (V+PAST)

Amara hit Nimala.

Example 1) illustrates a Sinhala sentence with a transitive verb. Native speaker of the Sinhala language without profound linguistic knowledge would know that Amara did the action of hitting Nimala. Yet, identifying the elements in a sentence (for example, actor, action, and receiver) by a native speaker would require something more than simply being able to speak the language. They must acquire some kind of information to identify the lexical elements, and moreover, referring to the example 1), would have to know that in a canonical sentence of the Sinhala language, the subject (*amara*) comes before the object (*nimala-ta*) and then the verb (*gehuwa*) is placed at the end. There must be some sort of information cues in the sentences providing the necessary information for native speakers to identify the structure and its word order. Otherwise, processing such a sentence would be difficult for any native speaker.

2.1 The information cues

From a linguistic perspective, there are several kinds of attributes of languages that provide information for identifying different lexicons. For example, phonological information tells the speaker how a word is pronounced; morphological information makes the speaker aware of the consistency of meaningful elements of a word; semantic information provides knowledge to understand the meaning of a word; and finally, syntactic information provides knowledge to use the words in a smaller or larger context as such in phrase level or sentence level (Nakayama, 1999). This paper will basically be focussing on syntactic and semantic information.

3. Syntactic information in Sinhala Sentences - Case particles and grammatical functions

Syntactic information tells us about the structure of a language. For example, consider figure 1 which is built based on example 1). The syntactic structure of *nimala amara-ta gehuwa* can be illustrated as either in bracketed diagrams [$_{NP}$ φNOM [$_{VP}$ ACC [V+PAST]]] or in tree structures as depicted in figure 1. 'Case particles' in syntactic structures provide information to identify the roles of noun cases such as nominative, dative, accusative, etcetera. The Sinhala language has seven cases, namely nominative (NOM), accusative (ACC), dative (DAT), locative (LOC), genitive (GEN), instrumental (INST) and ablative (ABL). Although the NOM is unmarked in Sinhala (Englebretson, et al., 2005; Henadeerage, 2002), other cases have a case marker or an inflectional marker to illustrate the properties. Due to this case marking, this paper hypothesizes that these case particles provide information for native speakers on sentence processing to understand the noun cases.

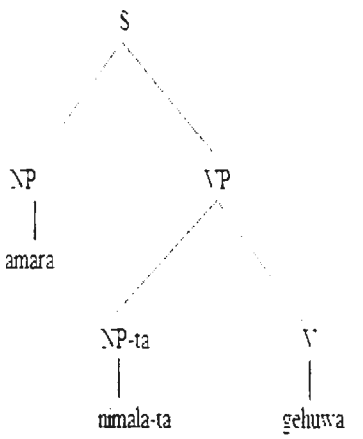

FIG.1 *An active sentence consisting of a transitive verb*
Note: NP-ta refers to the dative case-marked noun phrase

Grammatical functions, on the other hand, provide information on subject (S) object (O) verb (V). Sinhala is a subject-object-verb (SOV) language. When a native speaker encounters a sentence such as *amara nimala-ta gehuwa*, they know intuitively that a grammatical sentence in the Sinhala language should arrange the words in a way that the subject precedes the object and the object precedes the verb. Thus, the sentence can be interpreted as *Amara* did an action of *hitting Nimala*. Hence, grammatical functions can also be a significant candidate for information cues in the Sinhala language sentence processing.

3.1 Semantic information in Sinhala sentences - Thematic roles

Arguments which correspond to the verb in a sentence have different kinds of thematic relations. These thematic relations (i.e., thematic roles) include different elements such as agent, patient, instrument, experiencer, theme, goal, etcetera. With the relation of thematic roles, example 1) again can be explained. For instance, *amara nimala-ta gehuwa*, has an agent (*amara*), theme/experiencer (*nimala*) and goal (*gehuwa*). In languages exhibiting the SOV word order, it is known that the thematic relations have an order of agent-theme-goal (for example see Tamaoka, et al., 2003; 2005 for Japanese language SOV order with thematic relations). Even though a native speaker does not know about these technical terms, when s/he processes a sentence, it is assumed that s/he searches for the information such as the person who took action, what the main action was, and who was the receiver etcetera. Thus, the intuitive knowledge that a native speaker possesses of thematic roles can also be another factor which provides information for sentence processing.

3.2 Priority information in Sinhala sentences - case particles, grammatical function and thematic roles

As mentioned above, there are three main information cues available for the Sinhala language sentence processing; case particles, grammatical functions and thematic roles. The present analysis examined active sentences consisting of transitive verbs to confirm the availability of these three information cues. However, other sentence types (such as ditransitive sentences, passive sentences etc.) should also be examined to reconfirm the availability of these information cues. This part aims at examining the accountability of these information cues with three other sentence types.

Example 2) illustrates an active sentence consisting of a ditransitive verb. The sentence *amara nimala-ta pota dunna* has the structure of [$_{NP}$ φNOM [$_{VP}$ DAT [φACC [V+PAST]]]] meaning 'Amara gave a book to Nimala'. On one hand, case particles indicate that, in an active sentence containing a ditransitive verb, an NP without a nominative case marker (marked with φ) precedes the dative case */ta/* marked NP which in return precedes the accusative NP without the ACC case marker */wa/* and the verb is placed at the end. On the other hand, grammatical functions indicate that the subject (*amara*) with null case marker comes before the indirect-object (*nimala-ta*), and then the direct-object (*pota*) is placed before the verb (*dunna*). Finally, thematic roles indicate that, an agent (*amara*) precedes the theme/ experiencer (*nimala-ta*) and the source (*pota*) precedes the goal (*dunna*).

2) Active sentence with a ditransitive verb

අමර නිමල පොත දුන්න

amara nimala-ta pota dunna

Amara (φNOM) Nimala (DAT) book (φACC) give (V+PAST)

Amara gave Nimala a book.

3) Passive sentence with a transitive verb

නලින්ද විසින් අමල්ට බැනිනු ලැබුවා

nalinda wisin amalta baninu lebuwa

Nalinda (NOM) Amal (ACC) scold (V+PSS+PAST)

Amal was scolded by Nalinda.

4) Potential sentence

අමරට පිහිනන්න පුළුවන්

amarata piinanna puluwan

Amara (DAT) swimming (φNOM) can (V+PRE)

Amara can swim.

Example 3) illustrates a passive sentence containing a transitive verb. The sentence *nalinda-wisin amal-ta baninu lebuwa*, has the structure of [$_{NP}$ NOM [$_{VP}$ ACC [V+PSS+PAST]]] meaning 'Amal was scolded by Nalinda'. On the one hand, case markers indicate that in a passive sentence containing a transitive verb, an NP accompanied by */wisin/* precedes an NP accompanied by the dative case marker */ta/* (the verb should also be in passive voice). On the other hand, grammatical functions indicate that, the subject (*Nalinda-wisin*) comes before the object (*amal-ta*) and then the verb (*baninu lebuwa*) is placed at the end. Finally, thematic roles indicate that, an agent (*nalinda*) precedes the theme/ experiencer (*amal*) and the goal (*baninu lebuwa*) comes at the end.

Finally, example 4) illustrates a potential sentence. The sentence *amara-ta piinanna puluwan*, has the structure of [$_{NP}$ DAT [$_{VP}$ φNOM [V+PRE]]][1] meaning 'Amara can swim'. On the one hand, case markers indicate that, a dative NP accompanied by */ta/* precedes the nominative NP with no case markers. On the other hand, grammatical functions indicate that an object (*amara-ta*) comes before the subject (*piinanna*) and then the verb (*puluwan*) is placed at the end. Finally, thematic roles indicate that, an agent (*amara-ta*) precedes the theme (piinanna) and the goal (*puluwan*) comes at the end. Thus, it is clear that other sentences of the Sinhala language can also be explained using case particles, grammatical functions and thematic roles. However, a further question remains as to which information cue is being used by the native speakers in the processing of these sentences. This will be argued in the discussion part with respect to the word order and scrambling phenomena.

---

[1] It should be noted that *puluwan (*can) in Sinhala is rather an adverb. However, in this paper, it will be glossed as a V (verb).

4. Word order and 'scrambling' phenomena

Previous sections examined different sentences to confirm the availability of information cues in Sinhala language. However, we should also examine the word order and its attributes in order to consider their usage. This part of the paper will examine the word order of the Sinhala language and its attributes. Although Sinhala has a favored SOV order (Dissanayaka, 2007; Herath A, et al., 1994; Noguchi, 1983), some studies (e.g., Gair, 1998; Herath A, et al., 1994) suggest that, the word order of spoken Sinhala is quite flexible. For example, consider example 5).

5) [illegible]

gunapala aliyekwa dekka

Gunapala (φNOM) elephant (ACC) see (V+PAST)

Gunapala saw an elephant.

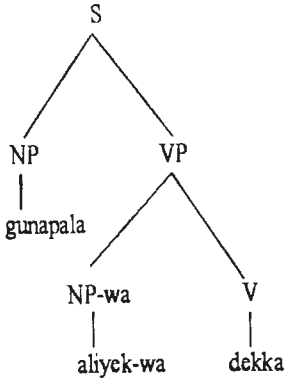

FIG. 2.

Note : NP-wa refers to the accusative case-marked noun phrase.

Example 5) illustrates a canonical word order of spoken Sinhala. The subject (S) *gunapala* precedes the object (O) *aliyekwa* and then the Verb (V) *dekka* is placed at the end. The sentence can be interpreted as 'Gunapala saw an elephant'. However, as mentioned above, previous studies (Gair, 1998) suggest that, the same sentence can have another five different word orders as depicted below (Example retrieved from Gair, 1998, p51).

5.1) අම්මාගේ ඇතෙක්ව දැක්ක. OSV

5.2) අම්මාගේ දැක්ක ඇතෙක්. OVS

5.3) දැක්ක ඇතෙක්ව අම්මාගේ. VSO

5.4) දැක්ක අම්මාගේ ඇතෙක්. VOS

5.5) ඇතෙක්ව දැක්ක අම්මාගේ. SVO

The sentence illustrated in example 5) with SOV order, can be re-ordered into five different word orders (ex., OSV, OVS, VSO, VOS and SVO) as depicted from 5.1) to 5.5). It should be noted that, all these sentences carry the same fundamental meaning of example 5) 'Gunapala saw an elephant'. While the sentences with original word order of a language (for example, SOV in Sinhalese) is called 'canonical ordered sentences', linguistic studies (e.g., Mazuka, et al., 2002; Miyamoto & Takahashi, 2004; Nemoto, 1999; Tamaoka, et al., 2005) refer to the re-ordered sentences as 'scrambled ordered sentences' associated with 'scrambling' phenomena[2]. Since Sinhala language has two types of word orders (canonical and scrambling), two essential questions arise. First, can SOV word order still be considered the canonical order? Second, to what extent Sinhala native speakers uses scramble ordered sentences? Since the present study only examine the availability of information cues with relation to word order, a further study has to be carried out to clarify these questions. However, in the next part, the relation between the information cues and the word order of the Sinhala language is further evaluated with relation to scrambling phenomenon.

5. Discussion

This paper concludes that, the Sinhala language has three kinds of information cues available for sentence processing. Case particles, grammatical functions and thematic roles

[2] "Scrambling" is a term used in linguistic studies for observable facts with free word order, which originally proposed by (Ross, 1967).

evidently provide necessary information to identify the canonically ordered (SOV) sentences as shown in table 1. The canonical order according to case particles predict that the noun phrase cases in Sinhala language should be ordered as NOM - DAT - ACC (-V), where as the canonical order predicted by grammatical functions could possibly have an order of S - IO - DO (-V). Finally, the canonical order according to thematic roles predicts that an agent precedes theme and goal.

Table I. Information Cuse and Predicted Canonical Word Orders

Information Cases	Predicted Canonical Word Orders
Case Particles	Nominative > Dative > Accusative
Grammatical Functions	Subject > Indirect Object > Direct Object
Thematic Roles	Agent > Theme > Goal

However, according to previous studies (Gair, 1998) the Sinhala language also has other five different scramble ordered sentences. Therefore, this paper assumes that, the information is available in a different order in different settings. Moreover, as previous studies suggest that, the Sinhala language has quite a flexible word order, implying that it could possibly have a non-configurational syntactic structure[3] (a flat structure) as depicted in figure 3.2.

[3] For a better explanation of 'non-configurational structure' see (Farmer, 1984; Hale, 1980, 1981).

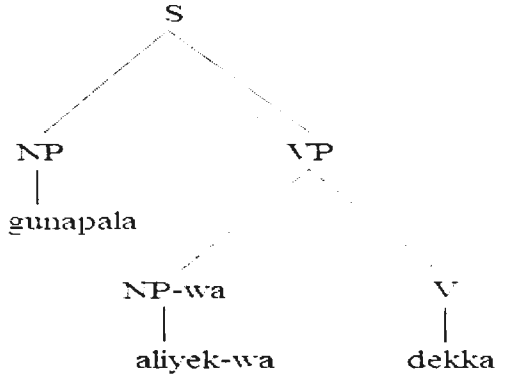

FIG. 3.1 *A configurational structure*

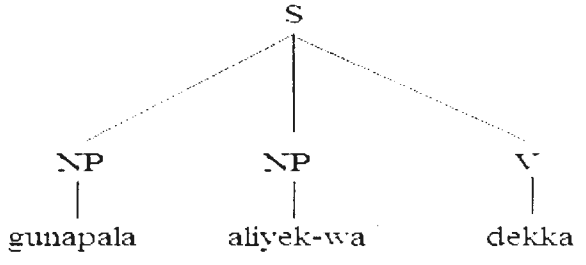

FIG. 3.2 *A non-configurational structure*

This flat structure predicts that there will be no difference in sentence processing between the canonical ordered (SOV) sentences and the scramble ordered (other than SOV) sentences. In other words, this structure predicts that the processing speeds (i.e., reaction

Hale, K. 1980. Remarks on Japanese Phrase Structure: Comments on the papers on Japanese Syntax. *MIT Working Papers in Linguistics*, 2, 185-203.

Hale, K. 1981. *On the Position of Warlpiri in a Typology of the Base*. Bloomington, IN: IULC.

Henadeerage, D. K. 2002. Topics in Sinhala syntax. Ph.D. dissertation, The Australian National University, Australia.

Herath A, Hyodo Y, Kawada Y, Ikeda T, 1994. A Practical Machine Translation System from Japanese to Modern Sinhalese, Gifu University, pp153-162.

Mazuka, R., Itoh, K., & Kondo, T. 2002. Costs of scrambling in Japanese sentence processing. In M. Nakayama (Ed.), *Sentence Processing in East Asian Languages* (pp.131-166). CA: CSLI Publications, Stanford.

Miyagishi, T. 2003. A Comparison of Word Order between Japanese and Sinhalese, *Bulletin of Japanese Language and Literature*, Yasuda Women's University, 33, Japan.

Miyagishi, T. 2005. Accusative Subject of Subordinate Clause in Literary Sinhala, *Bulletin of Japanese Language and Literature*, Yasuda Women's University, 33, Japan.

Miyamoto, E. T., & Takahashi, S. 2004. Filler-gap dependencies in the processing of scrambling in Japanese. *Language and Linguistics*, **5**, 153-166.

Nakayama, M. 1999. Sentence processing. In M. Tsujimura (Ed.), *The Handbook of Japanese Linguistics* (pp. 398-424). Malden, Massachusetts: Blackwell Publishers.

Nemoto, N. 1999. Scrambling. In M. Tsujimura (Ed.), *The Handbook of Japanese Linguistics* (pp. 121-153). Malden, Massachusetts: Blackwell Publishers.

Noguchi, T. 1984. *Shinhara-go nyuumon [Introductory to the Sinhalese language]*. Tokyo: Daigaku Shorin.

Ross, J.-R. 1967. *Constraints on Variables in Syntax*. Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge.

Tamaoka, K., Sakai, H., Kawahara, J., & Miyaoka, Y. 2003. The effects of phrase-length order and scrambling in the processing of visually-presented Japanese sentences. *Journal of Psycholinguistic Research*, 32, 431-454.

Tamaoka, K., Sakai, H., Kawahara, J., Miyaoka, Y., Lim, H. & Koizumi, M. 2005. Priority information used for the processing of Japanese sentences: Thematic roles, case particles or grammatical functions. *Journal of Psycholinguistic Research*, 34, 273-324.

A. B. PRABATH KANDUBODA

times) of the canonical (SOV) ordered *gunapala aliyekwa dekka* [$_{NP}$ φNOM [$_{VP}$ ACC [V+PAST]]] and the other five different orders do not differ in sentence processing. However, despite carrying the same fundamental meaning between canonical ordered sentences and scramble ordered sentences, the syntactic structure is assumed to be intricate in scramble ordered sentences, as NP properties and VP properties appear in different locations compared to the canonical ordered (SOV) sentences.

Furthermore, when the word order is changed, the information provided by the cues also appears in a different formation from that mentioned in table 1. Processing of *gunapala aliyekwa dekka* [$_{NP}$ φNOM [$_{VP}$ ACC [V+PAST]]] and *aliyekwa gunapala dekka* [$_{VP}$ ACC [$_{NP}$ φNOM [$_{VP}$ *gap* [V+PAST]]]] has two different syntactic structures, and therefore, the information given by the cues also has two different forms. For example, on one hand, the information provided by the case particles indicates that an NP with null case marker precedes an NP accompanied by an accusative case marker */wa/* and the verb is placed at the end, as in *gunapala aliyekwa dekka* (SOV). On the other hand, scramble ordered (OSV) sentence *aliyekwa gunapala dekka*, case particles tell that an NP with accusative case marker */wa/* precedes an NP with null case marker and the verb is placed at the end. This paper assumes that, the situation is complex when a native speaker processes a sentence. Although previous studies (Gair, 1998; Noguchi, 1984) have provided descriptive data on the complex usage of word order of Sinhala language, it is very important to carry out an experimental investigation in order to obtain raw data from native speakers.

References

Dissanayaka, J. B. 2007. *Say it in Sinhala*. Colombo, Sri Lanka: Stamford Lake Publication.

Englebretson, R. and C. Genetti, eds. 2005. *Santa Barbara papers in linguistics: proceeding from the workshop on Sinhala linguistics*. Santa Barbara, CA: Department of Linguistics at the University of California, Santa Barbara.

Farmar, A. 1984. *Modularity in Syntax: A Study of Japanese and English*. Cambridge, MA: MIT Press.

Gair, W. James 1998. Syntax: Configuration, Order, and Grammatical Function. In Lust, C. Barbara (Ed.), *Studies in South Asian Linguistics*; Sinhala and Other South Asian Languages(pp47–110), Oxford University Press, New York.