

## Efficient marking of argument focus: A trade-off between focus particles and word order in Sinhala

Shigeki Yoshida\*

**Abstract.** Colloquial Sinhala has focus concord constructions in which finite verbs with the suffix *-e* mark argument focus. In such constructions, the focal constituent is optionally marked via a focus particle, via movement to a postverbal (rather than preverbal) position, or via a combination of a focus particle and postverbal position. In this paper, I quantitatively analyze the preferred position of focal constituents marked by the three focus particles *=yi*, *tamaa* and *tamai*, as well as those not marked by any particle. Using data from a news website, I show that marking via a particle and the movement to the postverbal position are not randomly used. Rather, focal constituents marked by a focus particle are less likely to occur preverbally, while those not marked by a particle are more likely to occur postverbally. Moreover, among constituents marked by a focus particle, those marked by the particle *=yi* are more likely to occur postverbally. Based on these findings, I argue that the positional tendencies of focal constituents can be regarded as efficient marking patterns. When a constituent is marked by a focus particle, it is less likely to occur preverbally because the focal constituent is already explicit and the marking by the position would be redundant. The preference of *=yi* for the postverbal position does not seem to follow this general tendency until one takes into account that it requires less effort to produce *=yi* due to its phonological status. With this in mind, the distribution of constituents marked by *=yi* can also be regarded as an efficient pattern. The discussion in this paper contributes to both the typology of efficiency and the typology of optional focus marking.

**Keywords.** Sinhala; Indo-Aryan; focus concord constructions; word order; optionality; redundancy; efficiency

**1. Introduction.** Recent literature on typology has focused on efficiency as a factor which shapes the structures of languages (Gibson et al. 2019, Levshina & Moran 2021, Haspelmath 2021). It has been argued that languages are shaped to meet speakers' tendency to use language effectively. One of the most-frequently discussed topics in this line of research centers on how grammatical relations like subject or object are efficiently marked. The trade-off between word order and inflectional morphology has been discussed in previous typological studies (McFadden 2003, Koplenig et al. 2017, Levshina 2021). For example, Koplenig et al. (2017) show that languages with inflectional morphology, such as Quechua, tend to allow relatively free word order, while languages with no morphological marking, such as Mandarin Chinese, tend to show

---

\*I am grateful to Mai Hayashi, Yuko Morokuma, Naonori Nagaya, and Yui Suzuki, for their helpful and insightful comments on the earlier drafts of this paper. I would also like to thank the audiences at the LSA meeting, especially John Bundschuh, Ettien Koffi, and Teigo Onishi for their questions and suggestions. Thanks also to Aidan Aannestad for giving typological information and Elisabeth Kerr for asking valuable questions about focus types. Needless to say, any remaining errors are my own responsibility. This work was partially supported by LIXIL Ushioda East Asian Humanities Initiative Open Research Project A (PI: Naonori Nagaya) and JSPS KAKENHI Grant Number JP19H01264 (PI: Yo Matsumoto). Author: Shigeki Yoshida, The University of Tokyo ([shige.mountain.linguistics@gmail.com](mailto:shige.mountain.linguistics@gmail.com)).

rigid word order. The findings of the aforementioned studies show that languages efficiently mark grammatical relations: a language tends to rely on one of the above strategies, but not both.

In this paper, I examine the use of two marking strategies, which are similar to those mentioned above, in focus marking in Colloquial Sinhala (Indo-European, Sri Lanka; SOV). Colloquial Sinhala has focus concord constructions (Sumangala 1992, Henadeerage 2002, Hagstrom 2004, Kishimoto 2005, 2018, Paolillo 1994, Slade 2011, 2018 among others) in which finite verbs with the suffix *-e* (E-form; glossed as E) are used for marking argument focus. In such constructions, the focal constituent is optionally marked by two strategies: a focus particle (glossed as FOC), e.g., *=yi*, *tamaa*, and *tamai*, or the rightward movement of the constituent to the postverbal position. The possible four combinations of these two optional strategies are shown in Table 1.

	PreV Word Order	PostV Word Order
Zero-Marking	(1)	(3)
Marked by a Focus Particle	(2)	(4)

Table 1. Possible combinations of two optional marking strategies

Each combination is represented by the examples in (1) to (4). The constituent in bold in each example is the focal constituent.

- (1) **Zero-marking & preV word order**  
 mamə    **gamə-ʈə**    yann-e.  
 1SG.NOM   village-DAT   go.NPST-E  
 ‘It is to the village that I go.’ (Slade 2011: 49)
- (2) **Focus particles & preV word order**  
 mamə    **gamə-ʈə**{=yi<sup>1</sup>/tamaa/tamai}    yann-e.  
 1SG.NOM   village-DAT {=FOC/FOC/FOC}    go.NPST-E  
 ‘It is to the village that I go.’ (Slade 2011: 49)
- (3) **Zero-marking & postV word order**  
 mamə    yann-e    **gamə-ʈə**.  
 1SG.NOM   go.NPST-E    village-DAT  
 ‘It is to the village that I go.’ (Slade 2011: 49)
- (4) **Focus particles & postV word order**  
 mamə    yann-e    **gamə-ʈə**{=yi/tamaa/tamai}.  
 1SG.NOM   go.NPST-E    village-DAT {=FOC/FOC/FOC}  
 ‘It is to the village that I go.’ (Slade 2011: 49)

In the examples in (2) and (3), the focal constituent is marked by one of the two strategies, specifically, a focus particle is used in (2) and rightward movement in (3). In the example in (1), the focal constituent is not marked by either of the two optional strategies. Lastly, in the example in (4), the focal constituent is redundantly marked by both strategies.

The focus particles examined in this paper are *=yi*, *tamaa*, and *tamai*, which have similar meanings and functions as zero-marking. These three focus particles and zero-marking are interchangeable (Henadeerage 2002, Slade 2011), as indicated in the pairs in (1) and (2) or (3)

<sup>1</sup>Slade (2011, 2018) analyzes *yi* as a suffix and it is written as *-yi* in his papers. However, I analyze it as a clitic because it can take various word classes as its host. Throughout this paper, I cite examples from Slade (2011, 2018) with *-yi* modified to *=yi*.

and (4). These three particles can be considered optional because of this interchangeability with zero-marking. In addition to the three focus particles, there are other particles which need the E-form in Colloquial Sinhala (Henadeerage 2002: 205, Chandralal 2010: 217). These particles not only indicate the constituent in focus, but they also entail additional semantics such as exhaustiveness or negation. Because of these additional semantics, these other particles are not interchangeable with zero-marking and cannot be optional (see section 2.3). The present study only deals with the three optional focus particles that are interchangeable with zero-marking.

Two word orders are observed in (1) to (4): preV word order (where the focal constituent occurs preverbally) and postV word order (where it occurs postverbally). The postV position of the focal constituent is the result of rightward movement of the constituent.<sup>2</sup> The postposed focal constituent has been referred to by various terms, including “dislocation” (Slade 2011, 2018) and “pseudo-clefting” (Kishimoto 2005, 2018). The focal constituent is explicitly marked in a sentence with the postV word order, as in (3) and (4), as the result of rightward movement.

Given that both the focus particles and the rightward movement to the postV position explicitly mark the focal constituent, the question arises as to whether the two strategies of focus marking are used efficiently. This study investigates the usage of optional marking strategies in focus concord constructions in Sinhala. Until now, little has been known about the behavior of these two strategies of focus marking, despite the many studies on focus concord constructions in Sinhala (Sumangala 1992; Herring & Paolillo 1995; Henadeerage 2002; Hagstrom 2004; Kishimoto 2005; Kishimoto 2018; Paolillo 1994; Slade 2011; Slade 2018). Most previous studies have focused on syntax (Henadeerage 2002, Hagstrom 2004, Kishimoto 2005, 2018) or diachronic development (Paolillo 1994, Slade 2011, 2018). It remains to be investigated whether the optional focus marking strategies are used effectively.

To answer the question of whether focus marking in Sinhala is efficient, this study explores the hypothesis in (5) using data from a news website.

(5) Hypothesis on the usage of optional focus marking strategies:

The focus marking patterns in (2) and (3), where the focus constituent is only marked once, are more efficient than the patterns in (1) and (4), where the constituent is marked redundantly or marked by neither strategy.

If focus marking in Sinhala reflects speakers’ tendency toward communicative efficiency, the usage of the optional focus marking strategies should represent an effective pattern like that observed in the cross-linguistic marking patterns of grammatical relations. I expect the trade-off to be observed between the postverbal position and the use of the focus particles. I test the hypothesis in the form of the prediction that more efficient patterns are more frequent than less efficient patterns. Table 2 shows the possible combinations of the optional marking strategies, which have been exemplified in (1) through (4), and the predictions of their usage. I predict that the patterns marked in gray, in which constituents marked by a focus particle appear preverbally and zero-marked constituents appear postverbally, are more frequent than the other two patterns.<sup>3</sup>

---

<sup>2</sup>Sumangala (1992) categorically distinguished focus concord constructions with the preV and postV word orders as monoclausal and biclausal structures; however, the diagnostic tests they used were later critiqued by Henadeerage (2002) and Slade (2011). The later studies argued that there is no such structural difference between the two word orders.

<sup>3</sup>Ettien Koffi reminded me of the importance of acoustic studies of focus at the LSA meeting. Here, however, I do not consider acoustic realizations of focus here because we know little about them. Weerasooriya (2013) is the only study I have found which describes the prosodic realizations of Sinhala focus concord constructions beyond using the ambiguous term “intonational prominence”. It represents boundary tone on focal constituents and E-form verbs

	PreV Word Order	PostV Word Order
Zero-Marking	(1) Focus is not explicit → Less frequent	(3) Focus is disambiguated → More frequent
Marked by a Focus Particle	(2) Focus is disambiguated → More frequent	(4) Focus is disambiguated but redundant → Less frequent

Table 2. Possible marking patterns and prediction of their usage

This paper is organized as follows. In section 2, I introduce the corpus used in this study, the methodology used to extract the data from the corpus, and the analysis of the extracted data. In section 3, I present the results of the corpus research. I show that focal constituents with the particles *tamaa* and *tamai* almost always appear preverbally, while zero-marked focal constituents often appear postverbally. I also show that focal constituents marked by =*yi* tend to appear postverbally. In section 4, I examine how the observed frequency differences between the optional focus particles and the position of constituents can be accounted for. First, in section 4.1, I argue that the positional tendencies of constituents marked by *tamaa*, *tamai*, and zero-marked constituents can be regarded as efficient marking patterns. Then, in section 4.2, I propose that the constituent marked by the particle =*yi* can also be considered an efficient pattern when taking into account that it requires less effort to produce =*yi* due to its phonological dependency. Lastly, in section 5, I conclude the paper.

**2. Methodology.** In this section, I introduce the methodology used in this study. I firstly delimit the scope of the study to Colloquial Sinhala (section 2.1). Then, I introduce the overview of the corpus used (section 2.2), the way relevant data were extracted from the corpus (section 2.3), and how I analyzed the data (section 2.4).

2.1. COLLOQUIAL SINHALA AND LITERARY SINHALA. Modern Sinhala exhibits diglossia: Colloquial Sinhala and Literary Sinhala are used under different conditions (Paolillo 1992, 1997). Colloquial Sinhala is the variety used in everyday conversation or conversational parts of novels, while Literary Sinhala is used in written communication or formal speech (Paolillo 1997). These two varieties significantly differ both grammatically and lexically. Importantly, focus concord constructions exhibit different structures in the two varieties in some respects, as can be seen in (6) and (7).

- (6) Colloquial Sinhala (Slade 2018: 3)  
 mamə kiyewwe ee potə(=yi).  
 1SG.NOM read.PST.E that book(=FOC)  
 ‘It was that book that I read.’
- (7) Literary Sinhala (Slade 2018: 3)  
 maa kiyevuvec ema pota ya.  
 1SG.ACC read.PST.3SG that book COP.3SG  
 ‘It was that book that I read.’

---

by the denotation of L and H, where focal constituents realize as LH and E-form verbs end with LL%. However, no other phonetic information is described. Moreover, it remains unclear how the phonetic realizations of focus constructions differ depending on the two strategies, i.e., the focus particles and the different position of the constituent. Thus, more experimental phonetic research is needed to better understand acoustic realizations of focus in Sinhala.

Literary Sinhala differs from Colloquial Sinhala in that it exhibits obligatory postV word order and requires a copula after the focal constituent (Paolillo 1994, Slade 2011, 2018). The focal constituent *ee potā* ‘that book’ in (6) can appear both preverbally and postverbally, while the focal constituent *ema potā* ‘that book’ in (7) always appears postverbally. Furthermore, the focus particle =*yi* in (6) is optional, while the copula *ya* in (7) is obligatory.

Since this study is concerned with the optional uses of marking strategies, Literary Sinhala is out of the scope of the present study.

2.2. OVERVIEW OF THE CORPUS. The corpus used in this study was constructed by using data from the news website BBC Sinhala (<https://www.bbc.com/sinhala>). I used this resource because the Colloquial Sinhala data available on BBC Sinhala are transcribed and the contents can be extracted as HTML files. In this news site, interview spoken in Colloquial Sinhala are transcribed, while the other parts are written in Literary Sinhala.

Since this study is concerned with the optional focus marking strategies in Colloquial Sinhala, I distinguished the two varieties in the data by tagging the varieties using TEI (Text Encoding Initiative) XML markup. Data in Colloquial Sinhala are enclosed in double quotes in the original data. I manually tagged them with <said>, as in (8).

(8) <p><said> “එහෙම රෝගී වුණු සුළු දෙනා ඉක්මනින්ම සුවය ලබා තියෙනවා.” </said> ඔහු පැවසීය.</p>

I collected 30,493 words in Colloquial Sinhala from the interview parts of this site and 103,508 words in Literary Sinhala from the non-interview parts. Again, only Colloquial Sinhala was considered in this paper.

2.3. EXTRACTION OF RELEVANT SENTENCES. From the constructed corpus, I extracted relevant data, i.e., Colloquial Sinhala clauses where the main verbs were non-past E-form and one of the constituents was in focus. The extraction was conducted as follows.

I extracted word forms containing the string *nne* (න්නේ), which every nonpast E-form contains, as in *yanne* in (1). This string mostly occurred in nonpast E-form in the data, but there were words which contained this string but did not use the nonpast E-form (e.g., *dunne* ‘give.PST.E’). I manually excluded such cases, except for those that used the nonpast E-form. E-forms can be in the past tense as well as the nonpast tense. Past tense E-forms were, however, not included in the present study because it is difficult to extract a past E-form only by looking at a string. For example, the past E-forms *dunne* ‘give.PST.E’ above and *binde* ‘break.PST.E’ in (12) only share the suffix *-e*. Thus, the decision to exclude past tense E-forms was based on the difficulty of extracting data of this type.

Out of the extracted Colloquial Sinhala sentences with the E-form, I examined sentences with a focal constituent marked by =*yi*, *tamaa*, or *tamai*, as well as those not marked by any particle. Examples of such sentences are shown in (9) and (10).

(9) mamə yann-e gamə-ʔə.  
 1SG.NOM go.NPST-E village-DAT  
 ‘It is to the village that I go.’ (Slade 2011: 49)

(10) mamə gamə-ʔə tamai yann-e.  
 1SG.NOM village-DAT FOC go.NPST-E  
 ‘It is to the village that I go.’ (Slade 2011: 49)

I excluded clauses which contained non-past E-form verbs but were out of the scope of the present study. In addition to =*yi*, *tamaa*, and *tamai*, there are other particles which need the finite

verb to be E-form. These particles include negative, exhaustive, confirmation, hearsay, conditional, and interrogative particles (Henadeerage 2002: 205, Chandralal 2010: 217). An example of the exhaustive particle is shown in (11).

- (11) mamə yann-e gamə-ʈə witərai.  
 1SG.NOM go.NPST-E village-DAT EXH  
 ‘It is only to the village that I go.’

The exhaustive particle *witərai* requires the verb to be E-form. However, because of the exhaustive meaning ‘only’ of *witərai*, the examples in (9) and (11) have different meanings. Marking via the particles with additional meanings is not interchangeable with the zero-marking. I did not take the particles with additional semantics into account and focused only on =*yi*, *tamaa*, and *tamai* because this study centers on the optionality only allowed for by these three focus particles. I also excluded a type of negative sentence with E-form followed by the negative particle *nææ* as in (12). This negation does not necessarily indicate focal negation.

- (12) ranjit wiiduruə bind-e nææ.  
 Ranjit glass break.PST-E NEG  
 ‘Ranjit did not break the glass’ (Chandralal 2010: 223)

Fixed expressions with E-form were also excluded. For instance, the expression X *kiyanne* Y ‘X means Y’ as in (13) was excluded. There were 58 instances of this expression in the data.

- (13) uubar kiyann-e jaatyəntərə aayətəny-ak.  
 Uber say.NPST-E international institution-INDF  
 ‘Uber means an international institution.’

2.4. ANNOTATIONS AND ANALYSES OF THE EXTRACTED DATA. I annotated and analyzed the extracted data as follows. First, I annotated the position of a focal constituent in each sentence as preV or postV. Then, I annotated the presence and type of any focus particles. I classified the type of marking on focal constituents into four types: zero-marking, =*yi*, *tamaa*, and *tamai*. Based on these classifications, I investigated the relationships between the position and focus particles to test the hypothesis in (5).

**3. Results.** Using the methodology described in section 2.3, I collected a total of 303 clauses. Overall, focal constituents appeared in the postV position in 207 clauses (68.3% of the sample), while they appeared in preV positions in 96 clauses (31.7%). This result is similar to what Herring & Paolillo (1995) report: “In the Sinhala narrative corpus, 28.8% of emphatics [focus concord constructions] involve an initial or immediately preverbal focused element [focal constituent], while the remaining 71.2% are postposed” (p. 174). The results of this study have replicated Herring & Paolillo’s (1995) study in this respect.

The position of focal constituents differs depending on the focus particles (Table 3, Figure 1). Table 3 shows the raw frequencies of instances of each focus particle and the position of said particles. It also shows the ratio of postV focal constituents marked by each focus particle (postV ratio = postV/ (preV + postV)).

As shown in Figure 1, most of the zero-marked focal constituents (87.8%) appeared postverbally. However, focal constituents with the different focus particles tended to appear in different positions. On the one hand, focal constituents with the particles *tamaa* and *tamai* almost always appeared in preV positions. On the other hand, focal constituents with the particle =*yi*

tended to appear in the postV position. Note that the postV ratio of constituents marked by =yi was lower than that of zero-marked focal constituents.

Formal Marking	PreV	PostV	PostV Ratio
Zero-Marking	24	173	87.8%
=Yi	15	30	66.7%
Tamai	51	4	7.3%
Tamaa	6	0	0.0%
<b>Total</b>	<b>96</b>	<b>207</b>	<b>68.3%</b>

Table 3. Focus particles and position of focal constituents

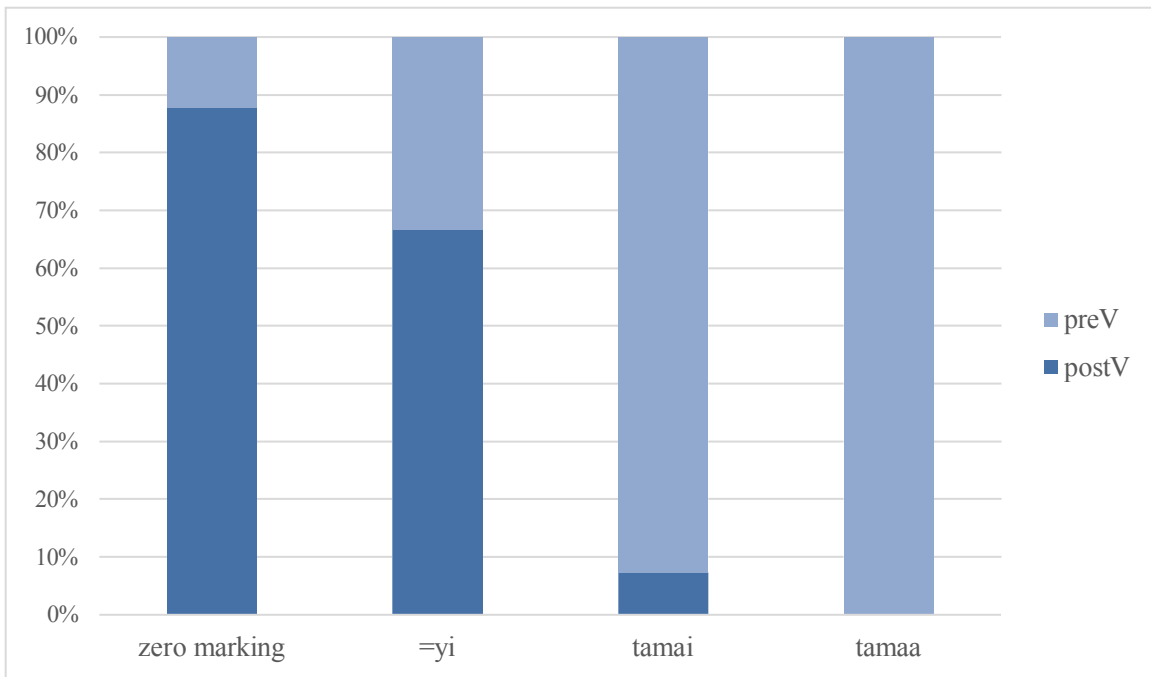


Figure 1. Focus particles and position of focal constituents

As seen in the examples below, the similar focus meaning was expressed by the different combinations of the strategies.

- (14) [...] wædipuramə kərənn-e geem gahanə ekə.  
 [...] most do.NPST-E game play.NPST.VA NMLZ  
 ‘[Many kids who use laptops to study do other things.] It is playing games that most of them do.’ (<https://www.bbc.com/sinhala/sri-lanka-57779832>)
- (15) [...] anit ayə-ʈə læbenn-e sottʃamə=yi.  
 [...] other people-DAT receive.NPST-E little=FOC  
 ‘[Some people receive more than they need.] Others receive little.’  
 (<https://www.bbc.com/sinhala/world-57474174>)
- (16) [...] sallikaarəyo tamai meewaa kərə-wə-nn-e.  
 [...] rich\_people FOC these do-CAUS-NPST-E  
 ‘[It is politicians who do these things. We cannot do it.] It is the rich who make them do these things.’ (<https://www.bbc.com/sinhala/sri-lanka-57401533>)

The examples in 0 to 0 represent the more frequent position of focal constituents marked by each focus particle. Zero-marked focal constituents mostly occurred in the postV position, as in *geem gahanə ekə* ‘playing games’ in 0. Focal constituents with =*yi* tended to occur in the postV position, as in *sottfamə=yi* ‘little’ in 0. Focal constituents with *tamai* and *tamaa* mostly occurred in the preV position, as in *sallikaarayo* ‘rich people’ in 0. The variation in focus marking is observed in the actual corpus data.

A logistic regression model showed that the position of focal constituents tends to differ depending on the focus particles. I computed a logistic regression model in which the position of the focal constituents was the dependent variable and the presence and kind of focus particle were the independent variables (Table 4).

	Estimate	SE	Z Value	Pr ( z )	
(Intercept)	1.71	0.20	8.78	<0.001	***
=Yi	-1.02	0.37	-2.75	<0.001	***
Tamaa & Tamai	-4.33	0.55	-7.83	<0.001	***

Table 4. Results of the logistic regression modeling

Zero-marking served as the reference level. The model evaluates to what extent the position of focal constituents marked by a focus particle differs from those not marked by a particle. *Tamaa* and *tamai* were combined into the value “tamai & tamaa” here because the uneven distribution of the focal constituents with *tamaa* failed to make the model converge.<sup>4</sup> In this model, each independent variable, i.e., the focus particles, was shown to significantly affect the dependent variable, i.e., the position of focal constituents.

**4. Discussion.** In this section, I propose factors that could have affected the usage frequency distributions observed in this study. First, in section 4.1, I argue that efficient marking patterns are observed in optional focus marking in Sinhala. The distributions of zero-marked focal constituents and focal constituents marked by the two focus particles *tamaa* and *tamai* tend to be efficient in that focus is marked only once, either by the focus particle or by the postV position. However, the distribution of =*yi* cannot immediately be considered to be efficient marking because its positional preference is different from that of the other focus particles. Then, in section 4.2, I suggest that the particle =*yi* can be analyzed as requiring less effort than the other focus particles due to its phonological status. Accordingly, constituents marked by =*yi* are more likely to occur postverbally.

4.1. EFFICIENT PATTERNS OF FOCUS MARKING. In this section, I discuss how the observed frequency patterns confirm the prediction of the efficient marking pattern. Then, I argue that the usage patterns of focus marking reflect speakers’ tendency to use language effectively. To begin with, I compare the prediction of the usage of the optional focus marking strategies in section 1 and the frequency distributions observed in section 3.

As discussed in section 1, I predicted that a focal constituent would tend to be only marked once by one of the optional focus marking strategies: the focus particle or the postV position. More specifically, focal constituents marked by one of the focus particles, i.e., =*yi*, *tamaa*, or *tamai*, were predicted to frequently appear preverbally. A focus particle can make the focus

<sup>4</sup>The decision to put the two particles together is also justified by their etymological similarity. Etymologically, *tamai* is formed by *tamaa* ‘self’ plus the focus particle =*yi*, and thus both *tamaa* and *tamai* have developed from the reflexive pronoun meaning ‘self’ (Turner 1962–1966: #5983, Paolillo 1994: 168).

interpretation of the constituent explicit. I predicted that the postV word order would not be needed because the focal constituent is already made explicit by the particle. In contrast, I predicted that zero-marked focal constituents would often appear in the postV focus position. I predicted that the postV word order would be needed to make the focus interpretation explicit and unambiguous.

These predictions were mostly verified by the results shown in section 3. The two focus particles *tamaa* and *tamai* and zero-marking appeared where they were predicted to. There was a clear difference between the focus particles *tamaa* and *tamai* and zero-marking. As shown in section 3, focal constituents with *tamaa* and *tamai* were far less likely to appear postverbally (0.0% and 7.3%, respectively): they almost exclusively appeared preverbally. In contrast, most of the zero-marked focal constituents appeared postverbally (87.8%). =*Yi* seems to be an exception; in the sample, focal constituents with =*yi* were more likely to occur postverbally (66.7%) than the other two focus particles. In the next section, however, I suggest that =*yi* may also be used efficiently.

The distributions of zero-marked constituents and constituents marked by the focus particles *tamaa* and *tamai* can be regarded as efficient patterns. On the one hand, when a focal constituent is marked by the particles *tamaa* and *tamai*, it is obvious which constituent is in focus and thus there is no need for the constituent to redundantly appear in the postV position. Consider the example in (17) (=16)) and the modified sentence in (18).

(17) *sallikaarəyo tamai meewaa kərə-wə-nn-e.*  
 rich\_people FOC these do-CAUS-NPST-E  
 ‘It is the rich who make them do these things.’  
<https://www.bbc.com/sinhala/sri-lanka-57401533>

(18) *meewaa kərə-wə-nn-e sallikaarəyo tamai.*

The only difference between the sentences in (17) and (18) is the position of the focal constituent *sallikaarəyo* ‘the rich’. Focal constituents marked by *tamai* were more frequently observed in preV positions, as in (17), than in the postV position, as in (18). In (17), the presence of the focus particle *tamai* makes it clear that the constituent *sallikaarəyo* is in focus. There is no need for the constituent to appear postverbally, as in (18), for the purpose of explicitly indicating the focal constituent. Moreover, postverbal constituents marked by the focus particle are not efficient because it is redundant to use two strategies with a similar focus-marking function at the same time. The redundant marking pattern seen in (18) requires more effort than the pattern seen in (17).

On the other hand, when a focal constituent is not marked by a particle, it is unclear which constituent is in focus and thus the explicit focus marking strategy of the postV position is preferred in this case. Consider the example in (19) and the modified sentence in (20).

(19) [...] *eewaa godak liyə-we-nn-e ingrisi-yen.*  
 [...] those many write-ANTIC-NPST-E English-INS  
 ‘[There are always posts about mistakes made by drivers on Facebook.] Most of them are written in English.’ <https://www.bbc.com/sinhala/sri-lanka-57423692>

(20) *ingrisi-yen eewaa godak liyə-we-nn-e.*

The only difference between the sentences in (19) and (20) is the position of the focal constituent *ingrisi-yen* ‘in English’. Focal constituents not marked by a focus particle were more frequently observed in the postV position, as in (19), than in the preV position, as in (20). In (20), where

the focal constituent *ingrishi-yen* ‘in English’ appears preverbally without any focus particle, it is unclear whether *ingrishi-yen* ‘in English’ or the other NP *eewaa* ‘these’ is in focus. In contrast, in (19), it is clear that the postverbal constituent *ingrishi-yen* is in focus even though it is not marked by any focus particle. Here, the postV position of the constituent helps to disambiguate which constituent is in focus.

The preferred patterns of focal constituents marked by *tamaa* and *tamai* and zero-marked focal constituents discussed here can be seen as efficient marking patterns: the focal constituent is marked only once by one of the two optional strategies.

4.2. THE PARTICLE =*yi* REQUIRES LESS EFFORT DUE TO ITS PHONOLOGICAL STATUS. In the preceding section, I showed that the preferred patterns of focus marking with *tamaa* and *tamai* and focus marking not requiring a focus particle can be considered as efficient, but constituents marked by the particle =*yi* do not seem to follow this tendency. In this section, I propose that the constituency marking via the particle =*yi* can also be seen as efficient if the effect of the phonological status of the particle =*yi* is considered.

The focus particle =*yi* is different from the focus particles *tamaa* and *tamai* in its phonological dependency. Phonologically, =*yi* is not an independent word and has fewer segments than *tamaa* and *tamai*. Compare the elicited examples in (21). Both A1 and A2 are answers to the focused yes/no question.

- (21) Q: oyaa mee pot də bæluw-e?  
 2SG these book.PL Q search.PST-E  
 ‘Was it these books that you were searching for?’  
 A1:ou mamə mee pot=ui bæluw-e.  
 yes 1SG these book.PL=FOC search.PST-E  
 ‘Yes, I was searching for these books.’  
 A2:ou mamə mee pot tamai bæluw-e.  
 yes 1SG these book.PL FOC search.PST-E  
 ‘Yes, I was searching for these books.’

When =*yi* is attached to a word ending with a consonant, as in A1, the vowel *u* is inserted between the host word and =*yi*. This insertion is conditioned by the final segment of the host word. Since the form of =*yi* is phonologically conditioned by the host word, =*yi* is phonologically dependent on the host. In contrast, there is no such vowel insertion between a host word and *tamaa* or *tamai*, as in 0. In addition, =*yi* has only one syllable, while *tamaa* and *tamai* each have more than one.

Due to the phonological dependency described above, the focus particle =*yi* requires less effort for speakers to produce than *tamaa* and *tamai*. Haspelmath (2021) proposes that, crosslinguistically, the more frequent construction in an asymmetric coding pair tends to be left uncoded or coded with fewer segments. The reason one construction in an opposition tends to be shorter is because it requires less effort to produce shorter forms and thus fulfill a speakers’ communicative needs. Of greatest importance to the present study is the proposal that shorter forms require less effort. I propose that the focus particle =*yi*, which has fewer segments and is more dependent, requires less effort to produce than the particles *tamaa* and *tamai*.

If it requires less effort to produce =*yi*, then it can be more easily used with the other optional strategy of rightward movement. In the previous section, I argued that the focal constituents marked by *tamaa* or *tamai* are more likely to appear preverbally because the combination of the particle and the postV word order requires much more effort and is not an

efficient marking pattern. In contrast to *tamaa* and *tamai*, it requires less effort to produce =*yi* due to its phonological status. This characteristic of =*yi* may enable a constituent it marks to appear in the postV position more easily.

The continuum of effort required to use or not use a particle allows for a better understanding of the frequency distribution described in section 3. Focus particles can be organized on a scale of increasing postV ratio, as shown in Figure 2.

less effort is required	zero-marking (postV: 87.8%)	= <i>yi</i> (postV: 66.7%)	<i>tamaa</i> , <i>tamai</i> (postV: 6.6%)	more effort is required
----------------------------	--------------------------------	-------------------------------	--	----------------------------

Figure 2. The scale of effort required to use each particle and postV ratio

The means of marking a constituent can be aligned in the following order: zero-marking, the particle =*yi*, and the particles *tamaa* and *tamai*. I have proposed that =*yi* requires less effort than *tamaa* and *tamai*. In terms of the number of phonological segments, zero-marking, obviously, requires less effort than =*yi*. As shown in Figure 2, constituent-marking patterns requiring less effort are more likely to be used with postV word order. Focal constituents not marked by a particle are the most likely to appear postverbally, those marked by *tamaa* and *tamai* are the least likely to appear postverbally, and those marked by =*yi* are in between.

4.3. SUMMARY. In this section, I have argued that focal constituents tend to be marked efficiently in focus concord constructions in Sinhala: overall, more effective patterns are used more frequently than less effective patterns. When a constituent is marked by the focus particles *tamaa* and *tamai*, it is likely to appear preverbally because the focal constituent is already explicitly marked; when a constituent is not marked by a particle, it is more likely to appear postverbally to mark the focal constituent explicitly. Moreover, I have proposed that constituents marked by the particle =*yi* are more likely to appear postverbally because =*yi* requires less effort due to its phonological status.

**5. Conclusion.** In this study, I quantitatively investigated the usage of the two optional strategies of focus marking in focus concord constructions in Colloquial Sinhala. I showed that the two strategies are not used randomly, and that focal constituents tend to occur in different positions depending on the particles by which they are marked. I also showed that focal constituents with the particles *tamaa* and *tamai* mostly appeared preverbally, while zero-marked focal constituents mostly appeared postverbally. Furthermore, focal constituents marked by =*yi* tended to appear postverbally. Based on these findings, I argued that the positional tendencies of constituents marked by each focus particle can be regarded as efficient marking patterns. When a constituent is marked by the focus particles *tamaa* and *tamai*, it is likely to appear preverbally because the focal constituent is already explicit; when a constituent is not marked by a particle, it is more likely to appear postverbally to indicate the focal constituent explicitly. Moreover, I proposed that constituents marked by the particle =*yi* are more likely to appear postverbally because it requires less effort to produce =*yi* due to its phonological status.

This study has implications for the literature on typology and efficiency. Efficient patterns of focus marking can be considered to be a reflection of speakers' and hearers' communicative needs. This view unites the phenomenon of optional focus marking with the phenomenon of optional case marking or, more generally, with the marking of grammatical arguments, which has been more widely discussed. Specifically, I showed that optional focus marking in Sinhala is similar to the marking of grammatical arguments in the following two ways. First, I showed that

the trade-off between word order and marking on the constituent is observed in focus marking. This resembles the trade-off between word order and inflectional morphology for argument marking, which has been discussed by McFadden (2003), Koplenig et al. (2017), and Levshina (2021). Second, I showed that optional focus particles are minimally used to the extent the focus interpretation is not hard to recover. The same motivation is suggested in studies on optional case marking (Kittilä 2005, Kurumada & Jaeger 2015). The discussion presented in this paper contributes to the literature on typological patterns of efficiency (Gibson et al. 2019, Levshina & Moran 2021, Haspelmath 2021) by extending findings on the marking of grammatical relations to a different area of focus.

Furthermore, the quantitative study of the multiple focus marking strategies in Sinhala presented here also contributes to the literature on typology and optional focus marking. Recent studies on optional focus marking have used experimental or corpus data to investigate the differences in the use of optional focus marking in, for example, Hausa (Hartmann & Zimmermann 2007) and Italian (Bianchi & Bocci 2012). In both Hausa and Italian, a focal constituent can optionally occur both *in situ* and *ex situ*. These studies revealed that there are statistical tendencies for specific contexts to be marked by a specific strategy. Optional focus marking in Sinhala is more complicated in that Sinhala has both focus particles and the movement of the focal constituent as optional marking strategies, while Hausa and Italian only exhibit leftward movement of the focal constituent. Thus, the present investigation of optional marking strategies contributes to the literature on typology, as well as enriching previous descriptions of Sinhala.

### Abbreviations

1	first person	EXH	exhaustive	NPST	nonpast
3	third person	FOC	focus	PL	plural
ACC	accusative	GEN	genitive	PP	past participle
ANTIC	anticausative	INDF	indefinite	PST	past
CAUS	causative	INS	instrumental	Q	question
COP	copula	NEG	negation	SG	singular
DAT	dative	NMLZ	nominalizer	VA	verbal adjective
E	E-form	NOM	nominative		

### References

- Bianchi, Valentina & Giuliano Bocci. 2012. Should I stay or should I go? Optional focus movement in Italian. In Christopher Piñón (ed.), *Empirical issues in syntax and semantics 9*, 1–18. <http://www.cssp.cnrs.fr/eiss9/>.
- Chandralal, Dileep. 2010. *Sinhala*. Amsterdam: John Benjamins. <https://doi.org/10.1075/loall.15>.
- Gibson, Edward, Richard Futrell, Steven P. Piantadosi, Isabelle Dautriche, Kyle Mahowald, Leon Bergen & Roger Levy. 2019. How efficiency shapes human language. *Trends in Cognitive Sciences* 23(5). 389–407. <https://doi.org/10.1016/j.tics.2019.02.003>.
- Hagstrom, Paul. 2004. Particle movement in Sinhala and Japanese. In Veneeta Dayal & Anoop Mahajan (eds.), *Clause structure in South Asian languages*, 227–52. Dordrecht: Kluwer Academic Publishers. [https://doi.org/10.1007/978-1-4020-2719-2\\_8](https://doi.org/10.1007/978-1-4020-2719-2_8).
- Hartmann, Katharina & Malte Zimmermann. 2007. In place — out of place? Focus strategies in Hausa. In Kerstin Schwabe & Susanne Winkler (eds.), *On information structure: Meaning and form*, 365–403. Amsterdam: John Benjamins. <https://doi.org/10.1075/la.100.20har>.

- Haspelmath, Martin. 2021. Explaining grammatical coding asymmetries: Form–frequency correspondences and predictability. *Journal of Linguistics* 57(3). 605–633. <https://doi.org/10.1017/S0022226720000535>.
- Henadeerage, Deepthi Kumara. 2002. *Topics in Sinhala syntax*. Canberra: The Australian National University dissertation.
- Herring, Susan C. & John C. Paolillo. 1995. Focus position in SOV languages. In Pamela Downing & Michael Noonan (eds.), *Word order in discourse*, 163–198. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/tsl.30.08her>.
- Kishimoto, Hideki. 2005. Wh-in-situ and movement in Sinhala questions. *Natural Language & Linguistic Theory* 23(1). 1–51. <https://doi.org/10.1007/s11049-004-6574-0>.
- Kishimoto, Hideki. 2018. Sinhala focus concord constructions from a discourse-syntactic perspective. *Glossa: A Journal of General Linguistics* 3(1). 9. <https://doi.org/10.5334/gjgl.260>.
- Kittilä, Seppo. 2005. Optional marking of arguments. *Language Sciences* 27(5). 483–514. <https://doi.org/10.1016/j.langsci.2004.10.005>.
- Koplenig, Alexander, Peter Meyer, Sascha Wolfer & Carolin Müller-Spitzer. 2017. The statistical trade-off between word order and word structure – Large-scale evidence for the principle of least effort. *PLoS ONE* 12(3). e0173614. <https://doi.org/10.1371/journal.pone.0173614>.
- Kurumada, Chigusa & T. Florian Jaeger. 2015. Communicative efficiency in language production: Optional case-marking in Japanese. *Journal of Memory and Language*. 83. 152–178. <https://doi.org/10.1016/j.jml.2015.03.003>.
- Levshina, Natalia. 2021. Cross-linguistic trade-offs and causal relationships between cues to grammatical subject and object, and the problem of efficiency-related explanations. *Frontiers in Psychology* 12. 648200. <https://doi.org/10.3389/fpsyg.2021.648200>.
- Levshina, Natalia & Steven Moran. 2021. Efficiency in human languages: Corpus evidence for universal principles. *Linguistics Vanguard* 7(s3). 20200081. <https://doi.org/doi:10.1515/lingvan-2020-0081>.
- McFadden, Thomas. 2003. On morphological case and word-order freedom. *Berkeley Linguistics Society (BLS)* 29. 295–306. <https://doi.org/10.3765/bls.v29i1.997>.
- Paolillo, John C. 1992. *Functional articulation in diglossia: A case study of grammatical and social correspondences in Sinhala*. Stanford, CA: Stanford University dissertation.
- Paolillo, John C. 1994. The co-development of finiteness and focus in Sinhala. In William Pagliuca (ed.) *Perspectives on grammaticalization*, 151–170. Amsterdam: John Benjamins. <https://doi.org/10.1075/cilt.109>.
- Paolillo, John C. 1997. Sinhala diglossia: Discrete or continuous variation? *Language in Society* 26(2). 269–296. <https://doi.org/10.1017/S0047404500020935>.
- Slade, Benjamin. 2011. *Formal and philological inquiries into the nature of interrogatives, indefinites, disjunction, and focus in Sinhala and other languages*. Urbana, IL: University of Illinois dissertation. <http://hdl.handle.net/2142/26068>.
- Slade, Benjamin. 2018. History of focus-concord constructions and focus-associated particles in Sinhala, with comparison to Dravidian and Japanese. *Glossa: A Journal of General Linguistics* 3(1). 2. <https://doi.org/10.5334/gjgl.241>.
- Sumangala, Lelwala. 1992. *Long-distance dependencies in Sinhala: The syntax of focus and WH questions*. Ithaca, NY: Cornell University dissertation.

Turner, Ralph Lilley. 1962–1966. *A comparative dictionary of Indo-Aryan languages*. London: Oxford University Press.

Weerasooriya, Tharanga. 2013. Wh-in situ questions in Sinhala: A prosodic account. *McGill Working Papers in Linguistics* 23.

<http://people.linguistics.mcgill.ca/~mcgwpl/McGWPL/2013v23n01/Weerasooriya2013.pdf>.