

Processing Reflexive Pronouns in Sinhala

Shayne Sloggett & Malkini Chandrasiri (University of York)

shayne.sloggett@york.ac.uk

Reflexive pronouns hold an influential position in both theoretical and psycholinguistic research. In syntactic theory, constraints on their distribution are used to motivate important primitives like c-command and locality [1, i.m.a.], leading psycholinguists in turn to ask how these notions are applied in real-time language processing in interaction with general memory mechanisms [2–5]. However, research in psycholinguistics has largely focused on only a handful of languages (e.g. English). In this work, we expand the empirical coverage of the field by investigating the processing of reflexive pronouns in Sinhala.

Sinhala is an Indo-Aryan language spoken by approximately 18 million speakers, predominately in Sri Lanka. Despite its Indo-Aryan origins, it displays many features observed in Dravidian languages, likely due to areal language contact. Notably, the reflexive system of Sinhala is very similar to its Dravidian neighbors (e.g. Tamil, Telugu). In Sinhala, the reflexive form *taman* (lit. “self”) is understood to be human, and preferentially singular. In root clauses, it may be used in conjunction with the reflexivising verbal morpheme *-gatta* (a form of the verb “take”) to indicate a reflexive meaning (1). In embedded contexts, *taman* obligatorily refers to a non-local sentence subject, unless the embedded verb is reflexive-marked, in which case it must refer to the embedded subject (2) [6, 7].¹ Recent work on a similar system in Telugu suggests that despite the anti-local bias of their reflexive, Telugu speakers preferentially attend to local referents in initial stages of processing [8]. Here we extend this work to Sinhala by pitting referent number and verbal morphology against one another to determine (i) the degree to which verbal morphology determines local reflexivity in Sinhala (ii) the role of number morphology in resolving reflexive reference (iii) the strength of any (anti-)local bias in Sinhala reflexive processing.

Methods: 54 self-reporting native Sinhala speakers read 48 sentences like (3) in a moving-window self-paced reading paradigm. Sentences were manipulated such that the embedded verb either indicated local reference (*+gatta*), or non-local reference (*-gatta*); and whether the local and non-local subjects were singular or plural (*(Non)Local ±match*). Because the reflexive *taman* is normatively singular, plural referents are treated as “*-match*”. After reading the sentence, participants were shown a binary-choice comprehension question probing their interpretation of the reflexive pronoun, after which they also provided a six-point acceptability rating of the sentence. Results were analysed using Bayesian Mixed Effects Regression, wherein “credible effects” are those whose 95% credible intervals (CI) exclude 0.

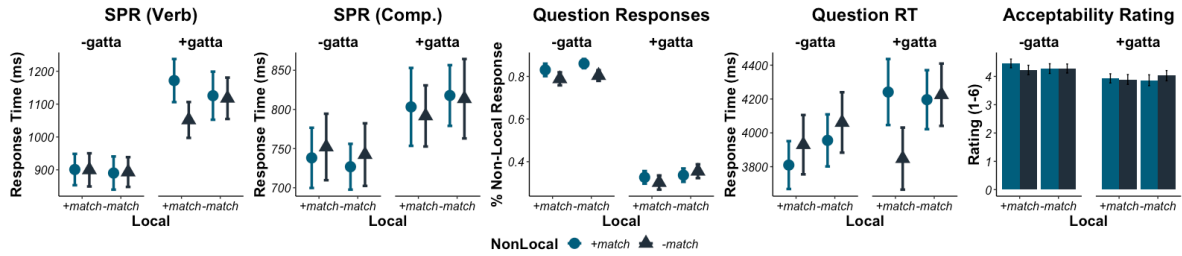
Results: Selected results are summarised in Figure/Table 1. In all measures, reflexive marking of the verb had the strongest influence on responses. In the self-paced reading portion of the task, the verb and sentence final complementiser were read more slowly if the verb was reflexive marked. Responses to questions probing the interpretation of the reflexive indicated a strong preference for non-local interpretations in the absence of the morpheme “*gatta*” (82%), and a preference for local binding when “*gatta*” was present (66%). Similarly, participants took longer to respond to locally-bound reflexives, and rated them as (very slightly) less natural (mean rating 4.3 vs. 3.9; $\hat{\beta}=0.35$, 95%CI[0.16, 0.54]). Counter to our predictions, the number marking of the two referents (singular vs. plural) did not exert a strong (or at least, statistically credible) influence on responses. This may be due to a lack of power, but nevertheless suggests that verbal marking provides a stronger cue than referent number.

Conclusions: These data confirm the strong role of verbal reflexivity in governing local binding in colloquial Sinhala [6], and suggest that number marking plays a relatively small role in determining reference. Interestingly, our results do not comport with [8], and suggest that Sinhala speakers both disprefer local reference with *taman*, and find it more difficult to process, implicating a fairly strong anti-local bias in both offline, and online measures of comprehension. Given the surface similarity of Sinhala and Telugu pronominals, further investigation is needed to understand this divergence.

¹N.b. There is considerable disagreement in currently published grammars of the language. [6] report that verbal reflexivisation strictly constrains interpretation, where [7] indicate a weaker role. We report here the judgments of our native speaker informants, supplemented by our experimental work.

- (1) a. Laməya_i (taman-wə_{i/*j}) tuwaalə-kara-gatta.
child_i (self-ACC_{i/*j}) hurt-do-REFL.
*The child_i hurt himself_{i/*j}.*
- b. * Laməya_i taman-wə_i tuwaalə-karaa.
child_i self-acc_i hurt-do.
The child_i hurt himself_i.
- (2) a. Piyal_i kiwwa Sriya_j taman-wə_{i/*j/*k} tuuwalə-karaa kiyəla.
Piyal_i said Sriya_j self-acc_{i/*j/*k} hurt-do that.
*Piyal_i said that Sriya_j hurt himself_{i/*j/*k}.*
- b. Piyal_i kiwwa Sriya_j taman-wə_{i/*j/*k} tuuwalə-kara-gatta kiyəla.
Piyal_i said Sriya_j self-acc_{i/*j/*k} hurt-do-REFL that
*Piyal_i said that Sriya_j hurt himself_{i/*j/*k}.*
- (3) a. Teechər_i kiwwaa laməyaa_j purawesi adyaapanəya paadəma mædədi
Teacher_i tell.PST child_j civics education lesson in-the-middle-of
thaman-wə_{i/j} pahath lesə wiwechanəya kara(-gatta) kiyəla.
self-ACC_{i/j} mean way criticise do(-REFL) that.
"The teacher said that the child criticized her(self) meanly in the middle of the civics lesson."
- b. **Question:** Who was criticised? (The teacher/The child)

Figure 1: Mean by-participant response (times). Error bars represent standard error ($n = 54$).



	SPR (Verb)		SPR (Comp.)		% NonLocal Resp.		Question RT	
	$\hat{\beta}$	95%CI	$\hat{\beta}$	95%CI	$\hat{\beta}$	95%CI	$\hat{\beta}$	95%CI
NonLocal	0.01	[-.01, .03]	0.01	[-.02, .04]	0.12	[-.02, .27]	0.00	[-.02, .03]
Local	0.01	[-.01, .03]	-0.01	[-.03, .02]	-0.13	[-.30, .03]	-0.01	[-.03, .01]
Gatta	-0.10	[-.13, -.08]	-0.04	[-.07, -.01]	1.53	[1.22, 1.85]	-0.04	[-.06, -.01]
NLoc:Loc	0.01	[-.01, .04]	-0.01	[-.03, .02]	0.00	[-.16, .17]	0.01	[-.01, .03]
NLoc:Gatta	-0.01	[-.03, .02]	-0.01	[-.04, .02]	0.09	[-.05, .23]	-0.02	[-.04, .00]
Loc:Gatta	0.00	[-.02, .03]	0.01	[-.01, .03]	0.03	[-.12, .18]	0.00	[-.02, .03]
NLoc:Loc:Gatta	-0.01	[-.04, .01]	0.00	[-.03, .02]	-0.06	[-.22, .10]	-0.02	[-.04, .01]

Table 1: BRMS estimates for response time (shifted lognormal) and %response (logistic) data.

- [1] Chomsky, N. (1986). MIT Press (MA).
- [2] Sturt, P. (2003). *Journal of Memory and Language*, 48(3).
- [3] Lewis, R. L., & Vasishth, S. (2005). *Cognitive science*, 29(3).
- [4] Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). *Journal of Memory and Language*, 69(2).
- [5] Parker, D., & Phillips, C. (2017). *Journal of Memory and Language*, 94.
- [6] Henadeerage, K., Butt, M., & King, T. H. (1998). *Proceedings of the LFG98 Conference. The University of Queensland, Brisbane: CSLI Publications.*
- [7] Gair, J. W., & Karunatilake, W. (1999). *Empirical Approaches to Language Typology.*
- [8] Arvindam, V., & Wagers, M. In: Conference on Human Sentence Processing. 2022, March.