UCLA Working Papers in Linguistics, no. 12, September 2005
Proceedings of AFLA XII, Heinz & Ntelitheos (eds.)

AN ASSOCIATIVE SEMANTICS FOR BASIC SENTENCES IN MALAGASY

BENJAMIN KEIL
University of California, Los Angeles
bkeil@ucla.edu

Referent systems, originally introduced in Vermeulen (1995) as a tool for text semantics, can also be used as a tool for modeling sentential semantics. As formulated by Vermeulen, they provide a reasonable means to model the basic sentences of languages like English (where the verb sits in a medial position and can identify its arguments based on their direction) on the one hand, and languages like Korean and Tagalog on there other hand, where the verb is in a peripheral position, but each argument has distinctive morphology on or near it. They do not, however, provide a convenient means to model a language like Malagasy, where the verb is in a peripheral position and the arguments are not reliably marked with distinctive morphology. Referent systems can be modified in a straightforward way which enables them to handle not only basic Malagasy sentences, but also various other constructions, like the English double object construction.

1. INTRODUCTION

Vermeulen (1995) introduced a formalism called "referent systems" for calculating text semantics. Kracht (1999) demonstrated that referent systems could also be applied to doing sentential semantics. Kracht as well as Keil (to appear) showed simple referent system models for the basic sentence types in English and Korean. Keil also demonstrated the inability of Vermeulen's referent systems to provide a model of the basic sentences of Malagasy. This paper provides a brief introduction to referent systems and shows how they can be straightforwardly modified to handle the basic sentences of Malagasy.

The property that English and Korean (as well as, e.g., Tagalog) have in common is this: in their basic sentence types, the relationships between the verb and its arguments are determined either by the side of a verb on which the argument appears (as in English, and other SVO languages) or by the morphology on or near the arguments. In basic Malagasy sentences, on the other hand, the verb is peripheral and the arguments are not reliably distinguished morphologically.

The problematic configuration in Malagasy is also present in slightly more complicated sentences in many languages. English double object constructions, for example, have two post verbal objects without distinctive morphology; sentences with topicalized objects have two preverbal arguments. Tagalog double object sentences present the same configuration when they have a directional focus. Extending referent systems to handle basic Malagasy sentences, then, provides a means of modeling many more constructions from various languages.

2. Referent Systems, Informally

Referent systems were created to incrementally calculate the *meaning* of a text from the meanings of its parts without referencing its *structure*. Applied to sentential semantics, referent systems ability to calculate meaning without structure has several attractive qualities. First, they provide a hypothesis about how hearers are able to interpret sentences as they perceive them, starting with the first word and building up an interpretation from there. Second, they provide a hypothesis about the closely related problem of how hearers can start interpreting the sentence before its structure can be known (i.e., they *cannot* reference the structure of the sentence during the interpretation). Third, they have the enticing property that an incomplete sentence will still have an interpretation, albeit an incomplete one, and—more generally—, they assign a meaning to any continuous segment of a sentence, even if the segment is not syntactically well-formed.

2.1. Morphology, Referent Systems' Lifeblood

A referent system model of a language is constructed in the following way. First, each lexical item is associated with a set of semantic truth conditions. The Tagalog word *bata*, for example, might be associated with the truth condition *x is a boy*. Second, each lexical item must also be associated with a referent system, which associates the semantic variables in the conditions with "tags." These tags determine the referent systems' combinatoric properties. The rightmost construct in Figure 1, associates the variable from *bata* with the tag "*". The rest of Figure 1 shows a model of the Tagalog sentence in (1).

Some of the other constructs in Figure 1 have the following properties. The verb *naghugas* 'wash' sets the semantic conditions that there is a washing event with a theme and an agent. The referent system connected to the verb associates the variable for the event's agent with a tag called "ANG" and the variable for the event's theme with a tag "NG". The nominal *pinggan* 'plate', similar to *bata* 'boy', has the semantic condition that it is talking about a plate and also has a referent system that connects the variable in that condition to a generic tag called "*". The most interesting feature are the two morphemes *ng* and *ang*. These morphemes set no truth conditions, but they associate their case-specific tag ("NG" or "ANG") with whatever their generic tag "*" is associated with, in effect "re-tagging" the variable.

(1) Naghugas ng pinggan ang bata. (Tagalog)
PAST.AF.wash NG plate ANG child
"The child washed a plate."
(Schachter and Otanes 1972)

To calculate the meaning of (1) using the constructs in Figure 1, one merges the referent systems (along with their associated semantic conditions and lexical items) together into larger constituents

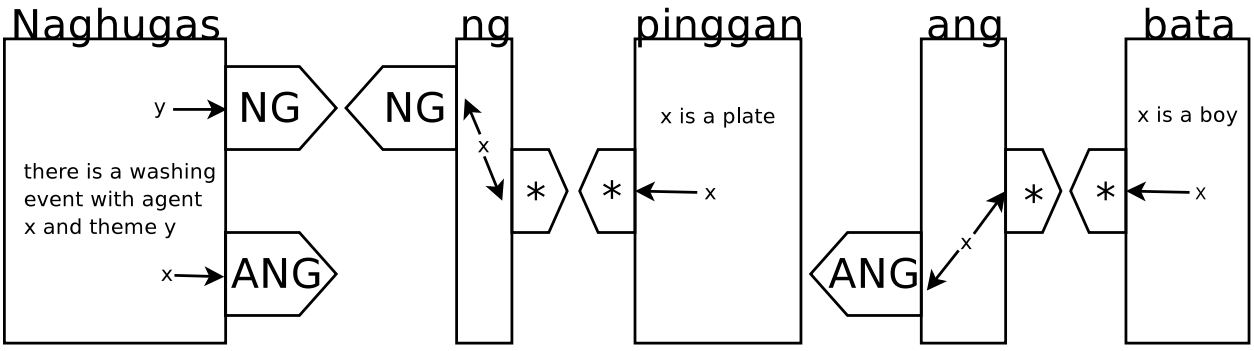

Figure 1: A referent system model of a Tagalog sentence

until there is but one remaining. Because referent systems are designed to calculate meaning independent from structure, any two adjacent referent systems may be combined in each step of a derivation. Having associated the words with referent systems and conditions (as in Figure 1), the individual words can then be merged. Two possible derived referent systems are shown in Figure 2.

Figure 2: Different merges of referent systems

In Figure 1, both *ng* and *pinggan* have a variable called x. These variables are associated with matching tags (i.e., the "*" tags) facing each other, and so in the merged referent system in Figure 2a, the variables have been unified; the variable conditioned by *pinggan* to be a plate is associated with *ng*'s tag "NG." Looking back to Figure 1, *pinggan* and *ang* also have a variable x in common, but these variables are not associated with matching tags, and so in the merged construct in Figure 2 they are not unified. In order to clearly state which variable in the merged construct corresponds to which variable in the original pair of referent systems, the one from the left system has been suffixed with an 'l' and the one from the right has been suffixed with an 'r'.

The pair of structures in Figure 3 show a possible penultimate step in a derivation of (1). It should be clear that the final merge produces something with the appropriate meeting. Although the two nominals both had their variable associated with the same tag "*", they are associated with

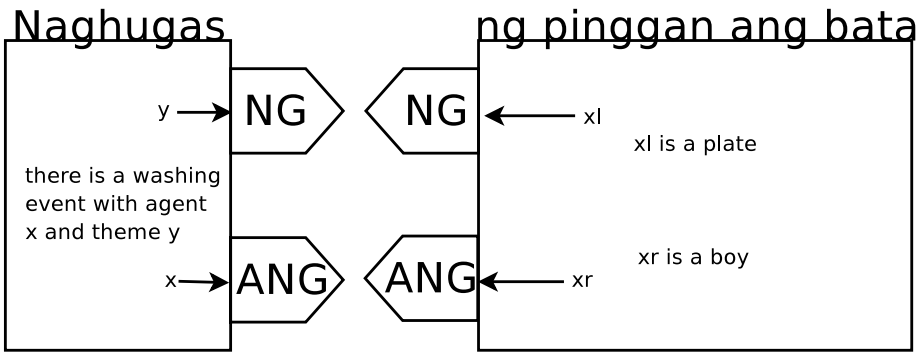

Figure 3: Before the final merge of the Tagalog sentence

the correct role in the event represented by the verb because of the re-tagging action of the articles. From a referent systems point of view, then, morphology plays the role of a broker, negotiating relations between arguments and those that need them.

2.2. When Morphology Fails

As stated in the introduction, there are two kinds of patterns that are very easily modeled with referent systems. In the previous section, it was shown that in sentences which have sufficiently rich voice and case morphology (e.g., Tagalog, Korean, and Latin) arguments can be identified by their case markers (as in Figure 1). When there is not sufficient morphology, directionality can sometimes fill its gap. Basic sentences in SVO and OVS languages like English and Guarijio can be modeled in this way; the verb can identify its arguments according to which side of the verb they are on. Figure 4 shows a set of referent systems associated with the lexical items in the Swahili sentence (2).

(2) Simba a-na-fahamu Ki-ingereza.
lion 1-Pres-understand 7-English
“The/a lion understands English.”

3. REFERENT SYSTEMS, FORMALLY

Mathematically, referent systems are fairly simple objects. They are composed of the set of tag names (N, or “Names”), the set of referents (R), an partial injection I (“Import”) from N into R, and a partial injection E (“Export”) from R into N. The function I represents the left-pointing tags

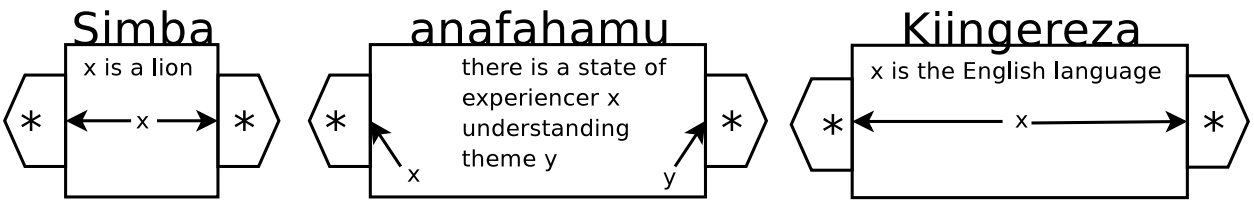

Figure 4: A referent system model of a Swahili sentence

in the figures from the previous section, while E represents the right-pointing tags, so the referent system for Tagalog *ng* corresponds sets and function in (3):

(3) a. $N = \{\mathrm{NG}, *\}$
b. $R = \{x\}$
c. $I = \{\langle \mathrm{NG}, x\rangle\}$
d. $E = \{\langle x, *\rangle\}$

The sole operation defined on referent systems is the binary operation merge. If $\langle N_1, R_1, I_1, E_1\rangle$ represents the left operand of merge and $\langle N_2, R_2, I_2, E_2\rangle$ represents the right operand, then result of the merge, $\langle N_3, R_3, I_3, E_3\rangle$ is calculated in the following manner:

First, the tag name set (i.e., the potential names of tags) of the merged system is union of the tag names of the individual , that is $N_3 = N_1 \cup N_2$.

The second step is to calculate the set of referents in the merged system; this is itself a process with three steps. The first step is to take each of the referents in R_1 and add a distinctive suffix, conventionally a superscript ^{1}. The set of all the elements in R_1 with the suffix added is referred to as R_1^1. In the second step, a similar operation is performed on the members of R_2; they are suffixed with a different suffix, conventionally a superscript ^{2}, and the resulting set if referred to as R_2^2. The referents of the merged system are a subset of $R_1^1 \cup R_2^2$. The third step is to determine which members of $R_1^1 \cup R_2^2$ are not in R_3. A member of R_2^2 is not in R_3 if it is "supervened" by a member of R_1^1, and a member x^1 of R_1^1 supervenes a member y^2 of R_2^2 just in case there is some η such that $\langle x, \eta\rangle \in E_1$ and $\langle \eta, y\rangle \in I_2$, or, graphically, just in case x is associated with a right-pointing tag and y is associated with a matching left-pointing tag.

The third step is to calculated the new import and export functions. I find it more natural to give the definition of the inverse of I_3, so for each u in R_3:

(4) $$I_3^{-1}(u) =_{def} \begin{cases} \psi & \text{iff } u = x^1 \text{ and } \langle \psi, x\rangle \in I_1 \\ \phi & \text{iff } u = y^2,\ \langle \phi, y\rangle \in I_2, \text{ and } \neg(\exists x)(\langle \phi, x\rangle \in I_1) \\ \text{not defined} & \text{otherwise} \end{cases}$$

(5) $$E_3(u) =_{def} \begin{cases} E_2(I_2(E_1(x))) & \text{if } u = x^1 \text{ and } E_2(I_2(E_1(x))) \text{ is defined.} \\ E_1(x) & \text{if } u = x^1 \text{ and } E_1(x) \text{ is defined, } I_2(E_1(x)) \text{ is undefined,} \\ & \text{and there is no } v \text{ such that } E_2(v) = E_1(u). \\ E_2(y) & \text{if } u = y^2 \text{ and } E_2(y) \text{ is defined.} \\ \text{not defined} & \text{otherwise} \end{cases}$$

This defines the merged referent system, but there is another step that must be performed on the accompanying semantic conditions. The merge of two referent systems produces also two translation functions, τ_1 and τ_2. The first, τ_1, is very simple; it merely appends the distinctive suffix from the left side to its argument. The second, τ_2 is defined as follows:

(6) $$\tau_2(y) =_{def} \begin{cases} x^1 & \text{if } x^1 \text{ supervenes } y^2 \\ y^2 & \text{otherwise} \end{cases}$$

The (string of) lexical items associated with the merged referent system is the concatenation of the strings associated with the left and right operands, and the semantic conditions of the merged referent system are the union of the semantic conditions of each referent system, with τ_1 having been applied to the variables of the semantic conditions associated with the left operand and τ_2 having been applied to the variables of the semantic conditions associated with the right operand.

4. The Malagasy Configuration

The verb in basic Malagasy sentences—as well as other languages which lack rich case morphology and which have SOV, OSV, VSO, or VOS as their basic word order—cannot distinguish between the two arguments based on their direction, nor can it reliably identify which argument is which by the case morphology on or near the arguments. In this configuration, Vermeulen's referent systems fail to provide a reasonable model for sentential semantics.

The Malagasy sentence in (7) demonstrates the problem. Both arguments appear to the right of the verb, like in the Tagalog sentence (1), but in (7) both arguments have the same determiner, and thus they both associate their nominal with the same tag. In Vermeulen's system, this causes two problems. The first problem is that of "tag clash." When the two nominals are merged together, the second clause of (4) causes the nominal further from the verb to lose its tag. With no tag, it cannot be interpreted as an argument of the verb. The second problem, visualized in Figure 5, is that the verb itself cannot present two tags to the right with the same name, so even without the tag clash, it can only take one argument per tag name.

(7) Nanasa ny vilia ny zazalahy.
PAST.AF.wash DET plate DET boy
"The boy washed the plate."

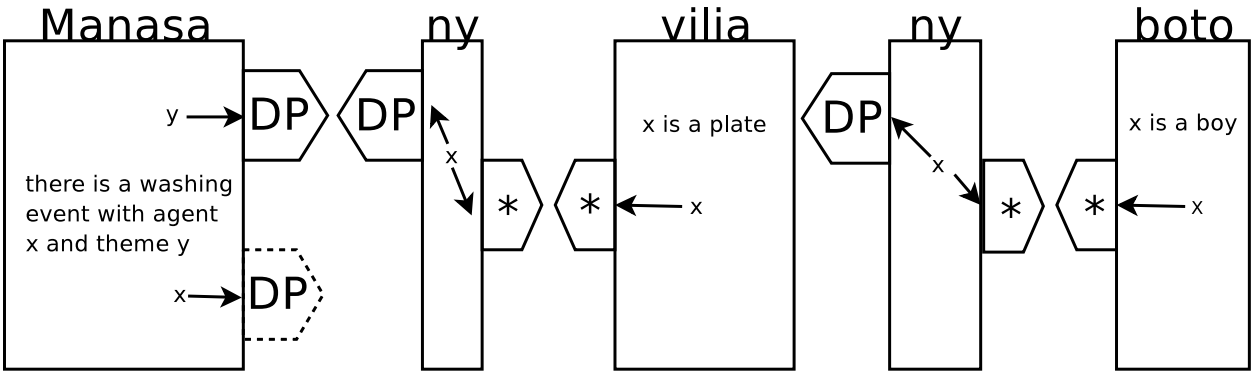

Figure 5: A referent system model of the problematic Malagasy sentence

To provide a reasonable model of Malagasy, then, referent systems must changed to do at least two things. First, they must allow two arguments with the same tag to coexist on the same side of the verb. Second, they must be able to resolve "tag clash" without losing essential information, such as the linear order in which the tags originally appeared. Both of the above requirements are met when tags dissimilate whenever they come into clash. The dissimilation must be done in a way that certain tags (like the arguments tags on a Malagasy verb) can be "pre-dissimilated."

A possible way to perform the dissimilation would be to use numerical indices. The nouns and determiners from (7) would then appear as in Figure 6. Mathematically, the import function for a referent system would then be a partial injection from $N \times \mathbb{N}$ into R, and the export function would be a partial injection from R into $N \times \mathbb{N}$. Keil (to appear) gives a full mathematical treatment of this approach.

Merging two of these indexed referent systems is slightly more complicated than before. Tags still match on strict identity, but the dissimilation must be integrated into the merge and it works differently on each side. For leftward tags, the merged system gives precedence to the tags of the left argument; for rightward tags, the merged system gives precedence to the right argument of the merge. The graphic in Figure 7 shows how this works; in the rightmost figure, the black tags are the tags that have come from the right operand and the white tags are those that have come from the left operand. In particular, it should be noted that the black tags are indexed with a 2 on the left and with a 1 on the right and that the situation is exactly the reverse for the white tags. This peculiar merging process ensures that the interaction of a tag with its neighbors to the left is not altered by anything that it might merge with on the right and that its interaction with its neighbors to its right is similarly unaffected by any merging it may do on the left. The merge is thus fully associative and can be calculated independent of constituent structure.

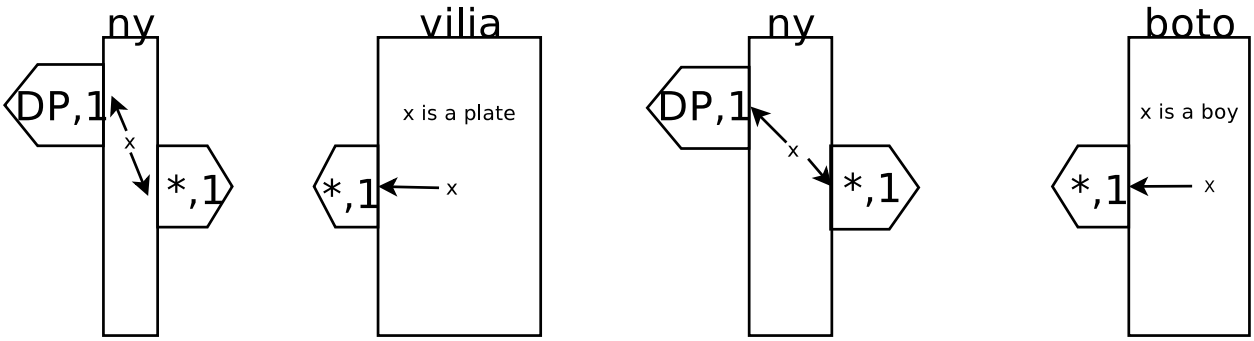

Figure 6: The arguments from the problematic Malagasy sentence (with indices)

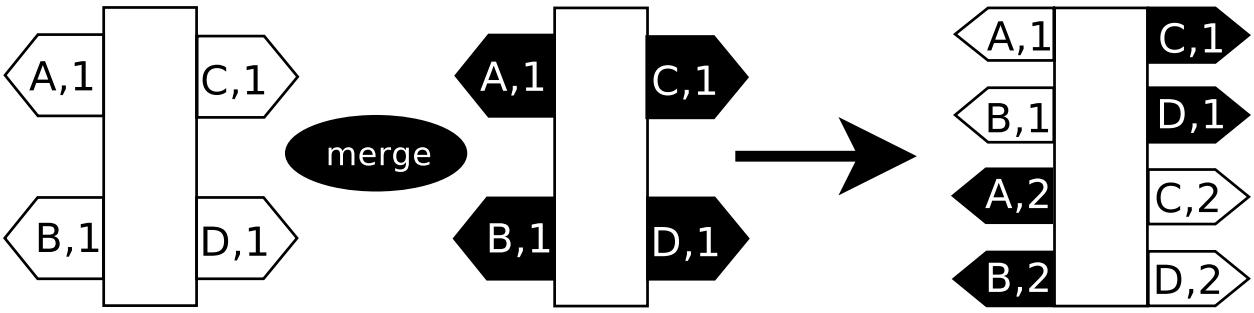

Figure 7: Merging two indexed referent systems

Using this new, indexed merge procedure, the elements in Figure 6 merge together to form the structure on the right in Figure 8. Figure 8 also shows the "pre-dissimilated" referent system associated with the verb. The tags on the verb match up with the appropriate tags from the arguments, and the final merge will produce the desired results.

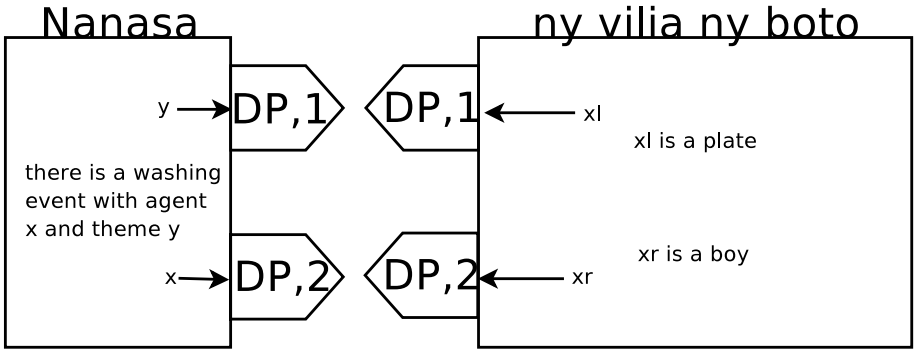

Figure 8: A final merge of the Malagasy sentence (with indices)

Although everything in Figure 8 appears to be well and good, there is another situation that must be considered. Because any two adjacent referent systems may be merged, it is possible that the verb will merge with the closer argument before that argument has merged with the other argument. In this case, no dissimilation will occur and the "DP,2" tag on the verb will remain unmatched. For this reason, tags must be updated after a merge, advancing the index of any tag whose higher-ranking namesake has been eliminated. Figure 9 shows how the tag on the verb would be advanced if the first step of the derivation was to merge *nanasa* with *ny*.

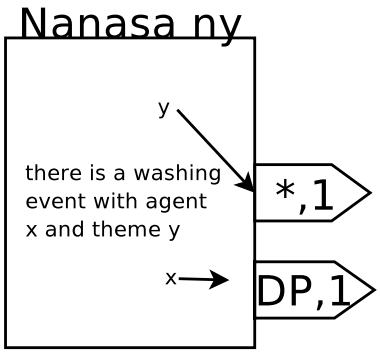

Figure 9: An "advanced" tag

5. Conclusion

Adding indices to the tags and manipulating these indices during the merging process allows for Vermeulen's insights to be maintained while still giving a treatment of basic Malagasy sentences. More generally, the formalism of referent systems demonstrates that the semantics of many sentences can be calculated without reference to their syntactic structures, and that semantics can be computed incrementally. Because a the meaning of a sentence can be calculated without reference to its structure, the semantics for a sentence that is not perfectly well formed syntactically may be calculated in just the same way as any other sentence. All of these attributes make referent systems an attractive formalism for further study. Future work should focus on the structures and semantics of coordinate constructions and quantification patterns, as these will likely be the most difficult challenges for the formalism.

References

Keil, Benjamin. *Associative Structures for Sentencial Semantics*. Master's thesis, University of California Los Angeles, to appear.

Kracht, Marcus. 1999. Agreement Morphology, Argument Structure, and Syntax. In *ESSLLI*.

Schachter, Paul and Fe T. Otanes. 1972. *Tagalog Reference Grammar*. University of California Press.

Vermeulen, C. F. M. 1994. Incremental Semantics for Propositional Texts. *Notre Dame Journal of Formal Logic* 35(2), 243–271.

———. 1995. Merging without Mystery or: Variables in Dynamic Semantics. *Journal of Philosophical Logic* 24, 405–450.