

## Semantically Driven Tagalog Morphology

Linda Walton & Antti Arppe  
University of Alberta

Tagalog verbs are marked for focus or voicing. ‘Focus’ refers to which noun in the sentential context is the main argument of verb. Most root words can be inflected for multiple foci, meaning that for a given verb, the actor, object, location, or noun phrase, may be the focus, manifested with a verbal affix. For instance, in a sentence like *The boy ate rice*, either ‘boy’ or ‘rice’ may be the main argument of the verb depending on whether the verb is marked as being actor focus (1) or object focus (2).

- (1) K-um-ain                      ang lalaki                      ng kanin  
eat-ACTFOC                      FOC-boy                      rice  
*The boy ate rice* (actor focus)
- (2) K-in-ain                      ng lalaki                      ang kanin  
eat-OBJFOC                      boy                      FOC-rice  
*The boy ate rice* or *Rice was eaten by the boy* (object focus)

While there has been much morphosyntactic description of the phenomenon (Schachter & Otones, 1972; Ramos & Bautista, 1986), there has been little data to show which foci are more frequently used with given root words or the motivation for selecting particular affixes for different roots. Nevertheless, Arppe (2006), extending on work by Karlsson (1986), has shown in the case of Finnish with a rich verbal morphological system that the co-occurrence of verbs with various morphological features are semantically driven, evident at all levels of granularity of semantic categories.

The purpose of this study was to create Tagalog verb profiles based on corpus data and to study whether focus preferences of Tagalog verbs might also be semantically driven. First, a corpus of Tagalog usage consisting of 2.47 million words was compiled from online Tagalog texts, representing blogs, Wikipedia articles, and religious texts. Using a general classification of verbs into five semantic categories (adapting for Tagalog those used by Arppe, 2006), i.e. psychological states and processes, motion, action, speech act, and possession, eight of the most frequent roots for each semantic category were selected from the corpus, amounting to 40 verbs in all. Then, frequency data for all the potential verbal derivations and inflections of each root were collected from the corpus. Finally, the distributions of focus marking across the five semantic categories and individual verbal roots were scrutinized by various statistical methods in the R statistical computing environment (2012), e.g. the Pearson’s chi-squared test and its posthoc analyses using the *polytomous* package (Arppe, 2012) and visualizations using correspondence analysis (Lebart, 1998).

The statistical analysis of the data showed that the selected verbs in the five semantic categories differ significantly ( $p < 2.2e-16$ ) in terms of their focus marking. Based on the posthoc analysis of divergences of the observed co-occurrences of focus and semantic category from a homogeneous distribution, action verbs prefer object or location focus, while they disprefer actor and direction focus; psychological and speech verbs also prefer object focus and disprefer actor; motion verbs are the only verbs which prefer an actor focus; and possession verbs prefer direction, location, and benefactive focuses, while disprefer both actor and object.

Focus/Semantic class	action	psychological	speech	motion	possession
<b>Object</b>	+	+	+	-	-
<b>Actor</b>	-	-	-	+	-
<b>Direction</b>	-	+	-	+	+
<b>Locative</b>	+	-	-	-	+
<b>Benefactive</b>	-	-	-	-	+
<b>Instrumental</b>	0	0	0	0	+

Moreover, a visual inspection of a correspondence analysis plot of the roots within the same semantic groups shows that they prefer similar affixes and focuses. In conclusion, the results demonstrate that in terms of actual usage Tagalog verbal morphology is clearly semantically motivated.