

# A Not-So-Gentle Morphological Analyzer of Tagalog

Resty M. Cena

Barring performance issues, the listener succeeds in parsing words 100% of the time. This paper describes the development of an automated parser for Tagalog words that aims for as high a success rate. To achieve this goal, the parser is forced to employ methods that are heavy-handed and not pretty. Nonetheless, the results have many uses. The output parse tree can be input to the next level of text analysis (parts of speech and phrase tagging). The parser is packaged as a browser-based program. A learner can consult it for the morphological composition of words. The parser feeds a frequency counter, which provides a count of Tagalog words, roots, affixes, and affix combinations. The words and affixes and their frequency counts and morphological analyses are stored in a database. The database serves as an online dictionary of Tagalog words and their morphological analyses. The parser can be easily modified for use with other Philippine languages.

## CONTENTS

How might listeners do it?  
What's in a derived Tagalog word?  
The challenges of a Tagalog word parser  
When morphemes collide  
A pile of prefixes, an array of affixes  
Orthography is sloppy  
Doubling the complication: reduplication  
A hopping infix  
Words with double infixes  
Variants  
Compound words are simple  
Implementing the parser  
Spelling regularization rules  
When everything else fails: lookup lists  
Left-to-right, right-to-left, or inside-out scan?  
Organization and presentation of results