Filipino Grammar

Stephen Borja, Justin Ching, Zhean Ganituen

February 22, 2025

Contents

1 Preliminaries

1.1 The Filipino Alphabet

Let $\mathcal{F}$ be the Alphabet for the Filipino language, this alphabet is composed of 56 scripts and 11 punctuation marks [2]. The 56 scripts are divided into two, the first half being the capital letters of the modern Latin script with the addition of "$\tilde{\text{N}}$" and "Ng"; while the other half is the lower case variants of each letter.

The 11 punctuation marks in the Filipino language are the: *tuldok* (.), *tandang pananong* (?), *tandang padamdam* (!), *kuwit* (,), *kudlit* (’), *gitling* (-), *tutuldok* (:), *tuldok-kuwit* (;), *panipi* (”), *pambukas na panaklong* ((), *pampasarang panaklong* ()), at ang *tutuldok-tuldok* (...)

In mathematical notation, we can represent $\mathcal{F}$ as the set:

$$\mathcal{F} = \{\text{a}, \text{b}, \ldots, \text{z}, \tilde{\text{n}}, \text{ng},\ \text{A}, \text{B}, \text{C}, \ldots, \text{Z}, \tilde{\text{N}}, \text{Ng}\}\ \cup \{., ?, !, ,, ’, -, :,\ ;, ”, (,), \ldots\}$$

and the size of $\mathcal{F}$, $|\mathcal{F}| = 67$.

We can also introduce subsets the following which are subsets of $\mathcal{F}$.

1. $\mathbb{M} = \{.,?,!,,,’,-,:,;,”,(,),...\}$, the set of punctuation marks
2. $\mathbb{V} = \{\text{a,e,i,o,u,A,E,I,O,I}\}$, the set of upper and lower case vowels
3. $\mathbb{C} = \mathcal{F} - (\mathbb{M} \cup \mathbb{V})$, the set of upper and lower case consonants
4. $\mathbb{V}_{\text{upper}}$ is the set of upper case vowels
5. $\mathbb{V}_{\text{lower}}$ is the set of lower case vowels
6. $\mathbb{C}_{\text{upper}}$ is the set of upper case consonants
7. $\mathbb{C}_{\text{lower}}$ is the set of lower case consonants
8. $\mathbb{L} = \mathcal{F} - \mathbb{M}$, the set of consonants and vowels

1.1.1 Remarks on the Digraph: ng/Ng or "*en dyi*"

Although the letter "Ng" or "ng" is a concatenation of two separate graphemes or symbols in $\mathcal{F}$ (since $\text{Ng} = \text{N} \cdot \text{g}$ and $\text{ng} = \text{n} \cdot \text{g}$), the letter "Ng" is officially recognized as a symbol in $\mathcal{F}$ since it represents a distinct Filipino sound. In particular, it represents the voiced velar nasal sound, or in the International Phonetic Alphabet (IPA), the ŋ sound [1].

For instance, the word "hangin" has 5 letters namely: "h", "a", "ng", "i", "n", since "ng" is pronounced as a velar nasal sound, not as two separate sounds n-g. So, "hangin" is pronounced as haŋin ("ha-ngin"). Take for instance the English word "manger" where "ng" is a substring but is not pronounced as the velar nasal sound. Instead, its pronunciation is ˈmeɪ ndʒ ər ("meyn-jer"); not ˈmæŋ ʒ ər ("mang-jer"), ˈmæŋ ər ("mang-er") or ˈmæŋər ("manger").

1.2 Common Errors

1.2.1 Loan Words, U/O, and I/E

Given the Spanish and English roots of Filipino, some *loan* words have rules for Filipino spelling. Let s be any string, the English language $\mathcal{E}$, the Spanish language $\mathcal{S}$, and $\mathcal{F}(s)$ is the translation of s in $\mathcal{F}$

1. $s \in \mathcal{S} \rightarrow \mathcal{F}(s) :\ \text{es} \cdot (\mathbb{M}|\mathbb{C})*$
2. $s \in \mathcal{E} \rightarrow \mathcal{F}(s) :\ \text{is} \cdot (\mathbb{M}|\mathbb{C})*$

Rule (1) denotes that if s is a Spanish word, translating s to a Filipino word would use "es" as the prefix to the word to denote that $\mathcal{F}(s)$ is a word of Spanish origin. On the other hand, for rule (2), if s is an English word, then $\mathcal{F}(s)$ would use "is" as the prefix to the word to denote that it is of English origin.

Another rule for Spanish loan words is the "o" to "u" change. If $s \in \mathcal{S}$ was translated to $\mathcal{F}$, and the the prefix of s is given by the regular expression (C|c)on(f|v). Then, $\mathcal{F}(s)$ is prefixed with (C|c)um(p|b). In particular, if the $s \models \text{conf} \rightarrow \mathcal{F}(s) \models \text{kump}$, otherwise $\mathcal{F}(s) \models$ kumb.

In addition to this, let α be the prefix, ω be the suffix, and $s \in \mathcal{F}$ be any Filipino word and $s \models (\mathbb{M}|\mathbb{C}) + (\text{e}|\text{o})$. If another word $k \in \mathcal{F}$ and $k = \alpha \cdot s \cdot \omega$, then $k \models \alpha(\mathbb{M}|\mathbb{C}) + (\text{i}|\text{u})\omega$. That is to say, that if a Filipino word is the

concatenation of a root word that ends in "e" or "o" and a suffix. Then, "e" will change to "i" and "o" will change to "u".

Finally, "e" changes to "i" and "o" changes to "u" if the word is a concatenation of a root word with itself without a *gitling*. Let $s \in \mathcal{F}$ and $s \models s'\backslash\text{-}s$ for s' is the word s but if s ends in "e" s' ends in "i", or if s ends in "o" then s' ends in "u".

1.2.2 Raw v. Daw (Enclitic Particles); Rin v. Din

Let $p =$ " " the blank symbol or the symbol containing space. And, we have the sentence structure $S = \alpha + p + EP$. The Enclitic Particles ($\mathbb{EP}$) is the set $\mathbb{EP} = \{\text{"raw"}, \text{"daw"}\}$ and n is any α is any noun, adjective, verb, or adverb. The usage of "raw" and "daw" is given by:

$$\left(\alpha \models \left[\text{a-z}\tilde{\text{n}}\text{(ng)A-Z}\tilde{\text{N}}\text{(Ng)}\right] + [\text{aeiou}] \equiv (\mathbb{V} \cup \mathbb{C}) + \mathbb{V}_{\text{lower}}\right) \implies EP = \text{"raw"}$$

Otherwise,

$$EP = \text{"daw"}$$

In other words, if the preceeding word to the enclitic particle ends in vowel, then the enclitic particle is "raw". If it is a consonant, then it is "daw".

This is the same idea with the adverbs "rin" and "din". If the sentence structure $S = \omega + p + (\text{"rin"}|\text{"din"})$ and ω ends in a vowel then the adverb used is "rin". Otherwise, the adverb is "din".

1.2.3 Ng v. Nang

1.2.4 *Gitling* Usage

References

[1] Valeria Malabonga. "Heritage Voices: Language - Tagalog". In: *Heritage Voices Collection* (Jan. 2009). DOI: https://www.cal.org/heritage/pdfs/heritage-voice-tagalog.pdf.

[2] Komisyon sa Wikang Filipino. *Ortograpiyang Pambansa*. Komisyon sa Wikang Filipino, 2014. ISBN: 978-971-0197-33-0.