Filipino Grammar

Stephen Borja, Justin Ching, Zhean Ganituen

February 22, 2025

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 "es"-words and "is"-words

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

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.

Example 1. *Here are examples for English and Spanish loan words in Filipino:*

1. Ako ay papasok sa **eskwelahan**. (correct)
   Ako ay papasok sa *iskwelahan*. (incorrect)
   eskwelahan ($\mathcal{F}$) from escuela ($\mathcal{S}$)

2. Aba! Malaki pala ang **espasyo** rito. (correct)
   Aba! Malaki pala ang *ispasyo* rito. (incorrect)
   espasyo ($\mathcal{F}$) from espacio ($\mathcal{S}$)

3. Marami kaming **estudyante** sa Computer Science. (correct)
   Marami kaming *istudyante* sa Computer Science. (incorrect)
   estudyante ($\mathcal{F}$) from estudiante ($\mathcal{S}$)

4. Mahilig sila Turing at Sipser maglaro ng **eskrima**. (correct)
   Mahilig sila Turing at Sipser maglaro ng *iskrima*. (incorrect)
   eskrima ($\mathcal{F}$) from esgrima ($\mathcal{S}$)

5. Marami raw **espiritu** rito. (correct)
   Marami raw *ispiritu* rito. (incorrect)
   espiritu ($\mathcal{F}$) from espiritu ($\mathcal{S}$)

6. Kinausap mo na ba iyung **ispiker**. (correct)
   Kinausap mo na ba iyung *espiker*. (incorrect)
   ispiker ($\mathcal{F}$) from speaker ($\mathcal{E}$)

7. Marami naman daw **isports** na puwedeng pagpilian. (correct)
   Marami naman daw *esports* na puwedeng pagpilian. (incorrect)
   isports ($\mathcal{F}$) from sports ($\mathcal{E}$)

8. Si Dijkstra ay hindi raw **iskolar**. (correct)
   Si Dijkstra ay hindi raw *eskolar*. (incorrect)
   iskolar ($\mathcal{F}$) from scholar ($\mathcal{E}$)

Example 2. *Here are some examples of literature and webpages where the "es-" and "is-" prefixes are misused:*

1. **TODO**

1.2.2 "kump"-words and "kumb"-words

If $s \in \mathcal{S}$ 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:

- $s \models (\text{conf})\mathbb{L}* \rightarrow \mathcal{F}(s) \models (\text{kump})\mathbb{L}*$

- $s \models (\text{conv})\mathbb{L}* \rightarrow \mathcal{F}(s) \models (\text{kumb})\mathbb{L}*$

Example 3. *Here are examples for Spanish words in Filipino with the "kump-" and "kumb-" prefixes:*

1. Pumasok si Cormen sa **kumbento**. (correct)
   Pumasok si Cormen sa *konbento*. (incorrect)
   kumbento ($\mathcal{F}$) from convento ($\mathcal{S}$)

2. Tinanong ko si Leiserson bilang **kumpirmasyon**. (correct)
   Tinanong ko si Leiserson bilang *kunpirmasyon*. (incorrect)
   kumpirmasyon ($\mathcal{F}$) from confirmacion ($\mathcal{S}$)

3. Ang **kumpetisyon** sa agham ay ginanap sa paaralan. (correct)
   Ang *kunpetisyon* sa agham ay ginanap sa paaralan. (incorrect)
   kumpetisyon ($\mathcal{F}$) from competición ($\mathcal{S}$)

4. Nagbigay siya ng **kumpisal** sa pari noong Linggo. (correct)
   Nagbigay siya ng *kunpisal* sa pari noong Linggo. (incorrect)
   kumpisal ($\mathcal{F}$) from confesión ($\mathcal{S}$)

5. Sumali ako sa isang **kumbento** sa simbahan natin. (correct)
   Sumali ako sa isang *kunbento* sa simbahan natin. (incorrect)
   kumbento ($\mathcal{F}$) from convento ($\mathcal{S}$)

6. Hindi ko alam ang **kumbinasyon** sa Graph Coloring. (correct)
   Hindi ko alam ang *kunbinasyon* sa Graph Coloring. (correct)
   kumbinasyon ($\mathcal{F}$) from combinación ($\mathcal{S}$)

7. Ipinahayag niya ang kanyang **kumbersasyon** sa kanila. (correct)
   Ipinahayag niya ang kanyang *kunbersasyon* sa kanila. (incorrect)
   kumbersasyon ($\mathcal{F}$) from conversación ($\mathcal{S}$)

Example 4. *Here are some examples of literature and webpages where the "kump-" and "kumb-" prefixes are misused:*

1. **TODO**

1.2.3 Morphophonemic Alteration with Suffixes

For a $k \in \mathcal{F}$ that is constructed with the concatenation $k = \alpha \cdot \text{root}_{\text{allomorph}} \cdot \omega$ for α is the prefix of k, ω is the suffix of k, and $\text{root}_{\text{allomorph}}$ is the root allomorph word of k.

The morphophonemic alteration of the root word is given by:

1. $\text{root} \models \mathbb{L} + e \rightarrow \text{root}_{\text{allomorph}} \models \mathbb{L} + i$
2. $\text{root} \models \mathbb{L} + o \rightarrow \text{root}_{\text{allomorph}} \models \mathbb{L} + u$

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" [3].

Here are some examples of literature and webpages where that have the incorrect root allomorph:

1. **TODO**

1.2.4 Hyphenated Reduplications

For a $k \in \mathcal{F}$such that$k \models b\backslash\text{-}r$ for b is the base word of k and r is the reduplicant of b. The reduplicant r is given by:

1. $b \models \mathbb{L} + e \rightarrow r \models \mathbb{L} + i$
2. $b \models \mathbb{L} + o \rightarrow r \models \mathbb{L} + u$

Here are some examples of literature and webpages where the wrong reduplicant is used:

1. **TODO**

1.2.5 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 α is any noun, adjective, verb, or adverb. The proper usage of the enclitic particles are given by:

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

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".

Here are some examples of literature and webpages where the enclitic particles, and the adverbs "raw" and "daw" are misused:

1. **TODO**

1.2.6 Ng v. Nang

1.2.7 *Gitling* Usage