A Metalexicographic Analysis of the Diksyunaryong Monolingwal sa Filipino (DMF): The Two Challenges Facing Filipino Lexicography

Janus Ruel Cabazares

ABSTRACT

Filipino lexicography deals with two problems related to the linguistic features of the language: the lexicon and the grammatical features including orthography. This paper analyzes how the Diksyunaryong Monolingwal sa Filipino (DMF) copes with these issues. With regards to the first, the DMF dealt with the problem of defining the Filipino lexicon by creating a corpus using several types of written texts. Its definition of Filipino, however, is still tied to the old form which was primarily based on Tagalog, an outdated formulation of the national language. Nevertheless, the DMF separates itself from previous dictionaries through its use of a clearly defined mechanism in building a corpus. In its use of the corpus to describe the meaning of the lemma, the DMF displays several faults. This includes definitions which refer to senses and have sexual connotations, direct translation, vague definition of some verbal derivations, and use of technical definition. In the lexicographic description of grammar, the DMF contains some irregularities in the morphological marking, inconsistencies in orthography and spelling of borrowed words. The morphological features of the language also lead to some challenges that the DMF was not able to consistently address. Overall, the DMF's use of lexicographic principles is an achievement that contributes to the development of the Filipino Language. It can pave the way for better monolingual dictionaries in the future.

1. Introduction

Of the few monolingual dictionaries of Filipino, only the Diksyunaryong Monolingwal sa Filipino (DMF) so far has benefited from the advantages of using lexicographic principles. This aspect of the DMF is primarily evident in the methods it employed to build a corpus of Filipino as well as its use of concordance to provide the definitions. This feat is further highlighted when looking at the previous significant monolingual dictionaries: Diksyunaryo ng Wikang Filipino (DWF) (1989, 1998) and the UP Diksyonaryong Filipino (UPDF) (2001, 2010). The former was made by the Komisyon ng Wikang Filipino (KWF) while the UPDF was published by the Sentro ng Wikang Filipino (SWF).

For the 1989 version of the DWF, Lee (2010) asserts that the corpus did not reflect the language of the time period when it was made leading to outdated or possibly nonexistent words in the entries. For the UPDF, he points out that the presence of several musical and zoological terms without explaining the mechanism and justification for the choice of such words are the effects of a questionable corpus. Another criticism was levelled on the representation of the features of the language either orthographic, phonetic, morphological, or syntactic. According to Lee, the 1989 version of the DWF did not include the letter F while the centennial version added eight letters. Also, the centennial edition of the same dictionary did not provide sample sentences which provides the crucial linguistic context that informs users how a particular word is used. This is particularly necessary for loanwords such as those from English (Lee, 2010).

Guillermo, Cajote, and Logronio (2015) alludes to the fundamental role of a properly collected corpus. The mistakes that they found in the UPDF is a consequence of the problematic selection and creation of a corpus. Numerous entries were defined through translation using various Philippine languages. This, according to the authors, resulted in circular definitions. Another problem pointed out is the inclusion of words with no attested occurrence or use in the language as entries. Its only appearance is in the UPDF itself. This self-referential feature of the entries

makes it difficult to claim that the UPDF is describing a language used by a group of speakers.

Underlying these critical observations on both the DWF and UPDF are two themes: the identification of the Filipino lexicon and the lexicographic description of grammatical features of the language. It can be said that the challenges that confront the creation of every Filipino monolingual dictionary involve these two issues. The first is related to the problem of establishing the national language while the second is concerned with the progress of linguistic analyses on the grammar of Filipino and other Philippine languages.

Filipino, the national language, is based not on one but on many languages of the Philippines (Flores, 2015). Prior to this the national language was called Pilipino and was primarily based on the Tagalog language. The drive towards the establishment of Filipino as a truly national language necessitates the identification of linguistic features from various local languages, including a lexicon that reflects this purpose. This circumstance contributes to the lexicographic problem of building a corpus of Filipino.

Linguistic analysis of Philippine languages have yet to achieve consensus on certain aspects of grammar. Since lexicographic description also needs to take into account certain grammatical features of the language, this condition has contributed to the problems found in presenting important linguistic features in monolingual dictionaries of Philippine languages. For example, several lexical items (e.g. *ay*) are categorized differently depending on the theory being subscribed by the linguist. Similarly, there is also disagreement on the syntactic categorization of roots (eg. *kain, tulog*). Furthermore, different linguists adhere to different theories on the syntactic status of the noun often marked by *ang* in Filipino sentences. This noun is co-indexed with a verbal affix often called the focus. This grammatical feature affects the semantic relationship between the verb and the noun, a crucial requirement in explaining the semantic based aspect of verb-noun collocations. Disagreements also arise in orthographic representation of phonemes. This includes the inconsistent

representation of the glottal stop and stress, both of which play an important role as phonemes in Filipino. The standardization of orthography for loanwords, particularly those borrowed from English, is another problem.

An analysis of the achievements and weaknesses of the DMF involves looking at how it grapples with the two problems of Filipino lexicography discussed above. This article provides an analysis of whether the DMF has achieved its aims as a monolingual dictionary of Filipino. Specifically, it examines how the DMF has defined the Filipino lexicon and how it employed a grammatical analysis in presenting linguistic features relevant to lexicographic descriptions.

2. The DMF and the Filipino Lexicon

The Diksyunaryong Monolingwal sa Filipino (DMF) was created by the KWF as an improvement of its previous dictionaries. As explained in its foreword, it utilized the study of McFarland (1989) which made a frequency count of the language. The result of the study which identified 2,053 most frequently occurring words guided the selection of headwords for the dictionary. This density makes it smaller than the average pocket dictionary. With the help of the corpus, sample sentences were chosen to identify the meaning and the linguistic context of the headwords. This method helps provide the most important and most frequently used senses of the lemma (Sinclair, n.d.). The 422-page dictionary has only two parts: the lemma list which comprises essentially what the dictionary offers and a very short outside matter, a foreword in Filipino, and its translation in English.

The DMF is intended for native and second language speakers rather than those who are still learning the language (KWF, 2005). Since it is assumed that users are already familiar with the language, the dictionary can serve as reference for identifying the meaning of words for both production and reception. In schools where Filipino is being taught, it can help students writing Filipino essays or reports as well as those reading and studying Filipino literature. The dictionary is considered a prototype,[1] a preparatory work

intended as a starting point for the creation of a comprehensive monolingual dictionary using lexicographic methods. The authors themselves expressed their hope to continue its development and expansion.

To develop the DMF, the lexicography division of the KWF trained under Curtis McFarland during the drafting of the dictionary (KWF, 2005). McFarland was also responsible for the selection and analysis of the corpus as well as the criteria used to define the Filipino language as conceived in the DMF. In the Frequency Count of Filipino, McFarland (1989) observes that the language being called either Tagalog, Pilipino, or Filipino is "historically and geographically continuous" and that the three cannot be easily differentiated into discrete languages (p. 5). He subscribes to the term Pilipino since the materials used for the corpus were made at a time where the language was referred to as Pilipino. Also, he explains that "Filipino as a combination of Philippine languages does not exist" (1989, p. 5).

According to McFarland (1989), the corpus was also composed entirely of written materials due to technical problems that spoken materials bring about. Another drawback, according to the author, was the shortage of Filipino materials. This limited the type of language use that the corpus can represent. The materials were composed of short Pilipino mini novels, short stories from Liwayway, post war novels, play and radio scripts, older novels, non-fiction, proverbs, riddles, poetry, jokes, literature translated from other languages, and instructional materials (McFarland, 1989).

This definition of Filipino by McFarland (1989) is outdated. Currently, Filipino as a national language is being developed based on the various Philippine languages. But this problem of the DMF is understandable owing to the period when its corpus was produced. Filipino as it is now constituted was at that time still in its early stages. Future editions should redefine and expand the corpus to better reflect the Filipino language. This challenge includes the gathering of spoken materials since Philippine languages, including Filipino, are primarily used orally.

On the other hand, the DMF's use of lexicographically defined corpus is still a major development compared to the previous Filipino monolingual dictionaries. The principles used to choose the entries supports this point about the DMF. It employed the formula by Francis & Kucera (1982) for adjusted frequency rather than the raw frequency. This way, McFarland avoided the skewed effect of bias caused by the content of the text (e.g. proper nouns which can appear numerous times in fiction). For convenience sake, McFarland analyzed only those forms with at least five occurrences. He reasoned that this amount can provide sufficient information regarding the form. This also reduced the number of forms for analysis from 58,000 to a manageable 14,000. Newell (1995), in contrast, believes that a frequency rate of two hundred instances can serve as an excellent standard for lexicographic descriptions.

2.1 Meaning Description

The concordance served as the primary means used to provide definitions of the lemmas. The primary type found in the DMF is the lexicographic definition. In Figure 1, the meaning of the root *inis*, which denotes a type of emotion, is explained. The definition is composed of a superordinate which is *damdamin* and a distinctive feature, *pagkagalit o pagkayamot*.

There were also other types used. The second entry, *dapa*, uses a COBUILD type of definition. The meaning description is a sentence explaining the meaning by using the lemma. Another type of definition found is the synonym, as exemplified by the first sense of the entry *balak* (Figure 1). Four synonyms are given. All contribute to providing a paraphrase of the meaning of the lemma since none of them exhibit complete synonymy. There are lemmas that use a combination of synonyms and lexicographic definitions. These two types are the most frequently used in the DMF.

The inconsistency of the format used may be partly due to the approach chosen by the authors. Also, the difficulty in providing full lexicographic definitions may stem from the need for further development of the register of the language. Some lemmas can be

balak png. **1.** Plano, layon, panukala, proyekto. *Maganda ang mga balak niya sa darating na taon.* **2.** Kalkulasyon, tantya; akala; sapantaha. *Mali pala ang kanyang balak.* **balakin** *pd.* (-in) PL Planuhin, tantyahin. *Balakin mong mabuti ang gagawin natin bukas. Binalak ang gagawin; binabalak ang gagawin.* **magbalak** pd. (mag-) PTg. Magplano; magpa-nukala. *Siya ay nagbabalak na magliwaliw sa buong mundo.*

dapa png. **1.** Paghiga nang pataob ang dapa sa kama. **2.** Sungaba; subasob. *Bigla ang dapa niya sa lupa.* **3.** Isang uri ng isda. *Malapad ang dapa.* **dumapa** pd. (-um-) PTg. Tumaob; lumapat ang dibdib at tiyan sa sahig, lupa, kama, atb. *Dumapa siya sa lupa para hindi tamaan ng ligaw na bala noong nagpuputukan na.* **madapa** pd. (ma-) PTg. Masungaba; masubasob. *Huwag kang pumunta sa putikan at baka ka madapa. Nadapa si Maria sa putikan.* **nakadapa** pr. Nakataob; nakasubasob. *Nakadapa si Ate sa putikan nang madatnan ko.*

inis png. Damdamin ng pagkagalit o pagkayamot. *Ang inis ko sa taong iyan ay hindi maalis-alis.* pr. Yamot; suya; buwisit. *Inis ako sa iyo.* **inisin** pd. (-in) PL Yamutin, galitin, buwisitin. *Huwag ninyong inisin ang aso. Iniinis na naman niya ang bata.* **mainis** pd. (ma-) PTg, Mayamot, magalit. *Mainis mo kaya si Ana?*

Figure 1. Examples of the types of meaning description used in the DMF (KWF, 2005, p. 39, 77, 133)

easily defined using synonyms rather than using a lexicographic definition. For others, it is necessary to explain their function in the sentence rather than an explanation of the meaning. This case is true for function words. Furthermore, the use of multiple types also helps provide additional explanation when one type is not enough. Nevertheless, the mishmash of types is at the very least a weakness and must be addressed in future editions or expansions of the dictionary.

This variety of methods also leads to entries which did not provide clear meaning descriptions. Some of the mistakes such as the use of directly translated, technical, and encyclopedic definitions can be corrected if the lexicographer adhere to the tenets of defining lemmas through the use of concordance. For some lemmas, the concordance can also negatively impact the definitions particularly for those which have low frequency count. This can produce definitions that do not reflect the appropriate meaning.

2.1.1 Translation

There were lemmas which were defined through translation or borrowing from a foreign language. This problem is often found for lemmas using synonym definitions. The lemma *bato* is defined as *piedra* and *roka*. These two synonyms are borrowed words from Spanish. Not only are these words borrowed, but they are also hardly used by the speakers of the object language. The same is true for the lemma *bituin* which is defined using the word *estrella*. Fortunately, it is followed by a more appropriate synonym *tala*. It is questionable why the rarely used Spanish word *estrella* comes before the more precise and popular synonym. This is also the problem with the synonym definition for the lemma *dangal.* The word *onor,* which is also a borrowed word, comes first in the list of synonyms and is followed by the more fitting synonyms: *puri, kabunyian, kamahalan.*

All these cases of translation or use of borrowed words are clearly problematic since they do not provide a clear description of the meaning of the lemma and assume that a synonym can be used as a substitute for a word in the same linguistic context (Newell, 1995). This is a misconception and can lead to unacceptable or ungrammatical sentences.

2.1.2 Technical and Encyclopedic Definitions

The definition for the color *dilaw* indicates the color spectrum and the placement of yellow in relation to other colors. This type of definition is not advisable for general purpose dictionaries and clearly does not provide a clear description of the meaning of the lemma. Definitions should reflect the culture of the language and should use words found in the dictionary (Newell, 1995). An example of a good definition for color is found in the entry for *pula* which has a more suitable definition: *kulay na tulad ng dugo.*

There are several definitions which can be classified as encyclopedic, providing extra and unnecessary information about the lemma. These do not contribute in explaining the common and general definition. The lemma Cebu is provided with historical information identifying it as the location where the first Catholic mass was held hundreds of years ago. The lemma *demonyo* includes

niyog png. Isang uri ng mahalagang palma na itinatanim nang maramihan sa mga pook na hindi matagal ang tag-araw. Ito ay tuwid at makinis. Ang katawan ay mataba sa puno at paliit sa dulo at may tila bilog na gilit sa pinagtatanggalan ng dahon. Ang mga dahon ay nagsisiksikan sa dulo ng katawan at ginagawang palamuti, pambalot at pandinding ng bahay. Ang bunga nito ay may pagkabiluhaba. *Ang niyog ay maraming gamit.*

Figure 2. A long and descriptive meaning for the lemma niyog
(KWF, 2005, p. 210)

the story of Lucifer the angel who was banished from heaven. Some definitions are also long and descriptive but are more relevant to the description of the meaning such as those for the lemma *niyog* (Figure 2), *buwan* and *buto*. Nevertheless, encyclopedic information is unnecessary for the purposes of the DMF.

2.1.3 Narrow Definitions and Inappropriate senses

Some definitions were narrow. This problem might have been due to the skewed effect of the use of certain sentences in the concordance. This is exemplified by the lemma *banal* which is defined as *taong palasimba.* This meaning is restricted to an adjective for a quality of a person. This is in contrast to its popular use as an adjective that modifies different types of nouns and not only human beings.

Associated to this is the use of words which are not neutral and connote inappropriate senses. The meaning of the lemma *danas* is stated as *ang tikim ng katawan, isip, diwa, loob, at ng anuman. Tikim* is used to mean experience. Although this is one of its senses, it is hard to disambiguate it from its sexual connotation particularly because it is also collocated with *katawan.* This definition is a double entendre. The sample sentence (*ang danas ng hirap sa panganganak*) does not imply a sexual situation. Similarly, the lemma *katabi* is defined as *sinumang kasiping o anumang bagay na kalapit ng iba pang bagay*. The word *kasiping* like *tikim* also has a sexually suggestive meaning. This connotation is further highlighted by the use of the noun *sinuman* which refers to a person. The lemma itself does not possess the sexual sense implied in the meaning and the sample sentence

(*ang katabi ni Tony ay si Joan*) does not necessarily imply this connotation. The descriptions given for these two lemmas are not the commonly used definitions.

2.1.4 Other Errors

Some lemmas were clearly not checked for errors. The sublemma *basta-basta* does not have a definition. It is directly followed by a sample sentence. Also, the sublemma *mabubuti* does not use the plural form of the adjective in the sample sentence. Thus, it fails to differentiate its meaning from the other lemma *mabuti*.

Some errors lead to contradictory definitions. An example of which is that of the lemma *dumi* defined as *anumang bagay na nakapagpapapangit, nakapagpapabaho o* **nakasisira sa kinalalagyan ng alikabok** (emphasis added).

3. The DMF and the Grammar of Filipino in Lexicography

3.1 Establishment of Lemmas

The Filipino language characteristically possesses several derivations which are formed through the use of affixation, forms of reduplication, and compounding. Syntactic function of words are often derived using these morphological processes. This linguistic feature affects the criteria for establishing lemmas. Since a root can have numerous derivations, the use of McFarland's (1989) study was very helpful. Citing only the major inflections helps make the dictionary more useful and relevant to speakers since these are the forms which the users will likely search in the dictionary (Newell, 1995). Most entries in the DMF use the root as lemmas. The most frequent derivations, on the other hand, are included in the entries as sublemmas. This arrangement is an efficient way to group the derivations. This also provides a convenient way to show the similarities (and differences) between derivations--an information that a good dictionary must provide (Newell, 1995).

In the DMF, the lemma is a root and is categorized frequently as a noun. A few are labelled as adjectives (e.g. *ganap, gipit, hamak, hayag*) and rarely as verbs (e.g. hali). Each sense is numbered and followed by its descriptive meaning and sample

gaan png. **1.** Kababaan ng timbang. *Ang gaan ng kahon ay napansin namin at nagduda ang may-ari.* **2.** Pagiging madali o maalwang gawin ng isang bagay. *Dahil sa gaan ng kanyang trabaho, nakakapanood pa siya ng TV tuwing tanghali,* **gumaan** pd. (-um-) PTg. **1.** Bumaba ang timbang. *Gumaan ang bata matapos magkasakit.* **2.** Dumali at umalwan ang isang bagay. *Gumaan ang responsibilidad ng mga nakatataas.* **magaan** pr. **1.** Mababa ang timbang. *Magaan ang laman ng bag.* **2.** Hindi mahirap; maalwan. *Magaan ang trabaho ko.* **mapagaan** pd. (mapa-) PTg. **1.** Mabawasan ang bigat, mapagaan ang dala-dalahan **2.** Madalian; maalwanan. *Mapagaan ang iuutos na gawain sa aming grupo.*

Figure 3. Sample lemma entry from the DMF (KWF, 2005, p.90)

sentence. Derivations are also included as sublemmas. In Figure 3, the derived verbs *gumaan, mapagaan* and the adjective *magaan* are sublemmas of the root *gaan.* There are also nominalized forms for the lemma such as *pag-ikot* under the lemma *ikot.* The full form of the stem and the affix is written as entry for the sublemma and followed by its Part of Speech (POS) designation and derivational affix enclosed in parentheses. Unfortunately, only the verb-forming affixes are indicated. In figure 2, the sublemma *magaan* does not have its adjective-forming affix separately indicated unlike the sublemma for the verbs. This presentation of derivational affixes is not only absent for adjectives but also for nouns, adverbs, and other syntactic categories. Lastly, a meaning description is provided along with a sample sentence in italic typeface.

For the DMF, roots are often identified as lemmas while derived forms such as verbs, adjectives, adverbs, and nouns are relegated as sublemmas. The identical root binds all these words together into the same entry. This lemma arrangement creates a complicated microstructure but makes it easy to find the various derivations. The semantic link between the derivations are easily identified in this arrangement. Newell (1995) also considers this as the most practical procedure.

3.1.1 Homophones as Different Lemmas

If the roots are homophones, having entirely separate and unrelated meanings but possessing identical forms, each word is

> **panabik, pananabik** png. Damdaming nagpapasigla, nagpapamithi o nagpapasigasig upang gawin ang isang bagay o makita ang sinuman. *Pananabik ng ina sa anak.* **kapanabikan** png. Kasiglahan; kasigasigan upang gawin ang isang bagay o makita ang sinuman. *Nakalimutan niya ang bag dahil sa kapanabikang masalubong ang kapatid.* **nananabik** pd. (ma-) PTg. Minimithi; nagsusumigasig. *Nananabik na akong makita ang aking matalik na kabigan.*
>
> **pananabik** png. Damdamin ng pagnanais, paghahanap o pag-iinteres. *Ang pananabik niyang makita ang anak ay nakapagpaligaya sa kanya.*

Figure 4. Two separate lemmas for pananabik
(KWF, 2005, p. 230 and p. 231)

represented as an individual lemma. This is true for the lemmas *ay* (*katagang nagpapakilala ng pagkagitla, pagdaing…*) and *ay* (*katagang gamit sa pag-angkop…*) as well as the lemmas *buhat* (*angat; alsa*) and *buhat* (*mula; magmula; sapul*). This separation is particularly effective for function words such the homophone *ni* which has two senses: genitive marker and conjunction. For lemmas with senses that are related, the DMF considers them polysemous and are categorized into one lemma. This is true for *bola* with the first sense referring to the round object (*ang mabilog na bagay…*) and its second sense referring to the verb deceive (*pang-uuto*). Although this criterion provides a clear basis for establishing lemmas, there are several instances in the DMF which are inconsistent to this rule.

The entries in Figure 4 are separate lemmas. The first is found in page 230 while the second in page 231. Upon close inspection both mean the same thing. The meanings are stated differently but imply the same sense. The sample sentences are also different. This redundancy is a clear issue that needs to be addressed.

3.1.2 Variations as Lemmas

There are also lemmas that are related in meaning and can be considered variations of the same word but have been separated into different lemmas. The terms for mother (i.e. *inang, inay, nanay*) and father (i.e. *itay, tatang, tatay*) have separate lemmas. This is also true for other gender inflected nouns such as *nobya* and *nobyo* as well as *Pilipino, Pilipina*, and *Pilipinas. Peso* and

piso are also separate. The distinction between the forms does not justify the establishment of separate lemmas since most entries have sublemmas which have different forms owing to word formation processes such as affixation, compounding, or reduplication.

3.1.3 Derivations as Lemmas

Another issue frequently found in several entries in the DMF are derivations which have been granted lemma status. The roots of these derivations are also found as lemmas in other parts of the dictionary. The P and N section of the DMF are replete with such derivations. N has several lemmas which are verbal derivations using the affix *naka-* (see Figure 5). For other words, their derivations using the same *naka-* affix are subsumed under the root as sublemmas. For example, the roots yakap, yuko, abang, akbay, aliw, angat, etc. are lemmas which have *naka-* verbal forms as sublemmas. P even has more of these misplaced lemmas (see Figure 5).

naka-	pa(g)-	pag + CVr	(-in, -an, ka-)
Nakaupo	pabiro	paglalakbay	tanawin
Nakatira	pabaya	pagmamadali	damdamin
Nakatitig	padala	pagmamalaki	gawain
Nakatingin	pagaari	pagmamahal	tanggapan
Nakatayo	paghanap	pagnanasa	tanghalan
Nakatawa	pagbili	pagsasama	tanghalian
Nakatatakot	pag-asa	pagsasapalaran	tindahan
Nakasulat	paghanga	pagsisikap	silangan
Nakangiti	pagkakataon	pagtuturo	pamahalaan
Nakakahiya	pagkaroon		tuwiran
Nakahiga	pagkatapos		usapin, usapan, and kausap

Figure 5. Derived forms for *naka-*, *pa(g)-*, *pag-* + CVr, *-in*, *-an*, and *ka-* granted lemma status (KWF, 2005)

biro png. Tukso; tudyo; kantyaw. *Masakit ang biro niya sa akin.* **biruan** png. **1.** Pagtutuksuhan; pagkakantyawan. *Masaya ang kanilang biruan.* **2.** Panahon ng tuksuhan. *Hoy, hindi ito ang oras ng pagbibiruan.* **3.** Laro. *Nagkapikunan sila sa kanilang biruan.* **biruin** pd. (-in) PL. tuksuhin; tudyuhin; kantyawan. *Huwag mo nga akong biruin nang ganyan. Binibiro lamang; bibiruin lamang.* **magbiro** pd. (mag-) PTg. Manukso; mangantyaw. *Magbiro ka na sa lasing, huwag lang sa bagong gising. Nagbiro sa lasing; nagbibiro sa lasing.* **mapagbiro** pr. Mahusay manukso; magaling mangantyaw. *Ang kapatid niya ay taong mapagbiro.* **pagbibiro** png. Panunukso; pangangantyaw. *Nagbunga ng hindi mabuti ang iyong pagbibiro.* **palabiro** pr. Mahusay manukso; makantyaw. *Mabiro siya talaga kaya parating masaya kapag naroon siya.*

pabiro pb. Sa paraang hindi mataimtim o hindi matapat; sa paraang nanunukso. *Ang pagtatapat ng pag-ibig ng binata sa babae ay tila palabiro lamang* **Sk.** Patukso

higa png. Anyo o kalagayan ng tao o hayop na namamahinga nang lapat ang buong katawan. *Ang higa niya ay patagilid kaya nanakit ang kanyang braso.* **Sk.** Himlay. **higan** pd. (-an) PD Himlayan; tulugan. *Hihigan na niya ang bagong kama.* **humiga** pd. (-um) PTg. Humimlay; mamahinga nang nakalapat ang buong katawan sa kama, sahig o anumang malalapatan ng katawan. *Humiga siya sa kama ni Lily. Hihiga na si Nanay.* **Sk.** Matulog. **ihiga** pd. (i-) PL Ilagay sa higaan; tulugan o alalayan s paghiga. *Ihiga mo sa papag ang bata. Ihihiga ang bata.* **mahiga** pd. (ma-) PTg. Lumatag sa kama o sahig. *Mahiga ka sa sahig. Nahiga sa kama.*

nakahiga pr. Nakabuwal, nakalapat ang likod sa sahig o kama. *Nakahiga siya nang tuwid dahil sa pagod.* **pagkakahiga** png. Pagkakalapat ng likod sa sahig, kama, lupa o anuman. *Komportableng pagkakahiga.*

Figure 6. Derived forms *pabiro* and *nakahiga* are granted lemma status
separate from their roots *biro* and *higa*
(KWF, 2005, p.56, 216, 116-117, 206 respectively)

sikap png. Pag-uukol ng panahon at masusing pag-aasikaso upang matamo ang nais. *Nagtagumpay siya dahil sa sikap at tiyaga.* **magsikap** pd. (mag-) PTg. Magkaroon ng malasakit sa anumang gawain, magsipag, magsinop. *Magsikap kayong mabuti. Nagsikap na makaahon sa hirap; nagsisikap na kumita.* **pagsikapan** pd. (pag-...-an) PGn Pagpilitan; pagsinupan; gawan ng paraan. *Pagsikapang makatapos ng pag-aaral.* **sikapin** pd. (-in) PL Pilitin o sinupin. *Sikapin ninyong matapos ang inyong ginagawa; sinikap makatapos sa gawain; sinisikap na kumita; sisikaping maging tanyag.*

pagsisikap png. Pag-uukol ng panahon at masusing pag-aasikaso upang matamo ang nais. *Pagsisikap sa pag-aaral* **Sk.** Pagsasakit, pagpipilit, pagtitiyaga.

kausap png. **1.** Sinuman sa dalawang kataong nagsasalitaan; kapalitang sabi. *Si Nene at ang kausap niya ay nagtatawanan.* **Sk.** Kasalitaan. **2.** Kahablahan sa hukuman; sinuman sa nagsasakdal at isinasakdal. *Si Joey at ang kausap niya ay may sariling abogado.* **Sk.** Kalaban, katalo. **kausapin** pd (-n) PL Makipagtanungan. *Kausapin mo nga si Bert tungkol sa negosyo niya.* **magkausap** pd (magka-) PTg. Magkapalitang-kuro. *Gawan mo ng paraan na magkausap kami, Nagkausap na silang magkapatid.* **makausap** pd. (maka-) PTg. Makapalitang-kuro. *Kailangang makausap ko siya. Nakausap ko ang ina mo. nakakausap lagi; makakausap na.*

usap png. Salitaan o pagpapalitan ng salita; pagpupulong. *Ang usap ng matanda ay di dapat pakialaman.* **mag-usap** pd. (mag-) PTg. Magsalitaan; magkasundo; magkayari. *Mabuti pa ay mag-usap kayong mabuti. Nag-usap na ang magkapatid. Mag-uusap ang magkapatid.* **makipag-usap** pd. (makipag-) PTg. Makipagpulong; makipagsalitaan; makipagkasundo. *Makipag-usap sa nakatataas; nakipag-usap sa nakatataas; nakikipag-usap sa magulang; makikipag-usap sa mga kamag-aral.* **mapag-usapan** pd. (mapag-...-an) PGn. Mapagpulungan; mapagkasunduan. *Sang-ayon ako kung anuman ang mapag-usapan. Sundin natin ang napag-usapan. Napag-uusapan naman ito nang mahusay.* **pag-usapan** pd. (pag-...-an) PGn Pagpulungan; pagkasunduan. *Pag-usapan nang mahusay; pinag-usapan ng samahan; pinag-uusapan sa pulong; pag-uusapan na namin.* **pag-uusap** png. Pagpupulong; pagkakasundo; pagsasalitaan. *Nasa magandang pag-uusap ang ikaaayos ng lahat.* **pakikipag-usap** png. Pakikipagpulong; pakikipagkasundo. *Ang pakikipag-usap mo lamang ang makapagpapalinaw ng lahat.* **usap-usapan** png. Bulung-bulungan; Salisalitaan. *Marami kang maririnig na usap-usapan tungkol sa pamilyang iyan.*

usapan png. **1.** Ang salitaan; ang bagay na tinatalakay; ang anumang sinasabi ng mga taong nagpupulong. *Ang usapan ng magkapatid ay tungkol sa taas ng bilihin.* **2.** Ang kasunduan; ang pinagkayarian. *Sundin natin ang usapan. Sa Jolibee ang usapan nilang magkikita.*

usapin png. Ang sumbong sa hukuman; ang demanda; ang kaso sa husgado. *Ang usapin nila ay nakarating na sa Mataas na Hukuman.*

Figure 7. Derived forms *pagsisikap* and *kausap, usapin,* and *usapan* are given
lemma status separate from their roots *sikap* and *usap*
(KWF, 2005, p276, 220, 157, 337-338, 338 respectively)

Other forms not derived from *naka-* and *pag-* affixes are also common. There are also verb or noun derivations using the suffix *–in* and *–an* that are included in the lemma list. Figure 4 provides examples of these. The lemma *damdam* has a sublemma for *damdamin* as a verb but the morphologically and semantically related noun *damdamin* has a separate lemma. It is important to take note that other verbs such as *sugal* and its derivation *sugalan* are found within a single lemma highlighting the irregularity that may have been caused by basing the establishment of lemmas primarily on frequency counts without considering the morphological and semantic relations.

The lemma *pabiro* in Figure 5 exemplifies this uneven granting of lemma status. Its root *biro* is also a lemma and includes a sublemma *pagbibiro* which is morphologically and semantically related to *pabiro.* It could have been more convenient to just include *pabiro* as a sublemma. The same problem is exemplified by *higa* and *nakahiga* again in Figure 6. This is the same for *sikap* and *pagsisikap* in Figure 7.

An extreme example is the root usap which has three derived forms given separate lemma status without cross referencing (see Figure 7). Another example is the root *dala*. Its reduplicated form *dala-dala* is a separate lemma. Worse, both lemmas consist of meaning descriptions that do not explain clearly the difference in meaning.

When establishing the lemma list using the frequency count as reference, the relationship between the lemmas must be considered otherwise it will lead to an uneven organization of the headwords. In several instances the DMF contradicts the standard of using roots as lemmas and relegating as sublemmas the derived forms of the same root. This is complicated further by the lack of cross referencing. For speakers of the language, identifying the root of derivations is not difficult. Thus, in establishing lemmas, it is more convenient for the dictionary user and the lexicographer to prioritize the root as lemma over the frequency count to provide a more comprehensive lemma list and organized macrostructure. To make the structure consistent, the derivations must be incorporated

under the root as sublemmas even if they rank higher in the frequency count. If some derivations are granted lemma status the motivations for this must be explained in the outside matter, cross referencing must be used between the root, and the derivation separated from the root.

3.2 Labelling of Syntactic Categories

Roots are often categorized as nouns even those roots that semantically refer to activities and are often used as verbs in the language (e.g. *tanggi, tigil*). Those that refer to quality or feature of an object are labeled adjectives. Numbers are also included in this category. However, the lemma *tatlumpo* is marked as both noun and adjective.

Another inconsistency is the POS classification of negation markers. The lemma *huwag* is classified as verb while *hinde* is listed as adverb. The shortcomings of the POS designation can be easily fixed. It just needs to be consistent. A review of the categorizations will surely be enough to correct these. The abbreviations used for the POS markings should also be explained in the outside matter to properly guide users of the dictionary.

The abbreviation for the POS marking is also not explained. Although from context, some of the abbreviations can be understood. The roots serving as lemmas are often labelled as png. (*pangngalan*). Verbal derivation are categorized as pd. (*pandiwa*) while adjective are pr. (*pang-uri*). For pronouns, the abbreviation used is ph., while pb. for pang-abay, and ptg. for pangatnig. Other abbreviations are harder to identify. The abbreviation PTg., Pl/PL. and PD. are often found after parentheses enclosing the verbal affix. Another is the marking Sk. which often follows synonyms.

3.3 Orthographic Representation
3.3.1 Glottal Stop and Stress

The most important orthographic issue is the disregard for the glottal stop and stress. Although native speakers will be able to distinguish minimal pairs upon reading the description, this failure shows an inconsistent representation of the phonemes of the language.

This neglect leads to lemmas appearing as homophones, exhibiting similarity in spelling when in fact they are different words differentiated by glottal stop or stress. There are several such lemmas in the DMF. The two lemmas represented by *baka* are minimal pairs different only in the placement of stress. One form is stressed in its penultimate syllable while the other in its ultima (see figure 8). This issue is also true for the entries *'mura* (*mababang halaga*) and *mu'ra* (*masama at masakit na salit sa kapwa*). Furthermore, there are several minimal pairs differentiated by a word-final glottal stop: *kaya, kita, mama, sala,* and *tayo.* Figure 8, shows the placement of the glottal stop and the difference in meanings. These then can be confused with real homonyms such as *tira* (*bagay na hinde nagamit*) and *tira* (*pamumuhay sa isang bahay o lugar*). The *tuldik* can be used to represent stress as well as word final glottal stop. This is already an established orthographic method in Filipino. For glottal stops in other positions the *gitling* can be used.

Minimal pairs for stress	
'baka (uri ng hayop)	ba'ka (pagpapahiwatig ng pag-agam-agam)
'mura (mababang halaga)	mu'ra (masama at masakit na salit sa kapwa).
Minimal pairs for word final glottal stop	
kaya (magagawa)	kayaʔ (nagpapahiwatig ng kadahilanan)
kita (ikaw at ako)	kitaʔ (sweldo)
mama (nanay)	mamaʔ (tawag sa isang lalaking hinde kilala o sa isang nakatatanda)
sala (silid tanggapan nga mga panauhin sa tahanan o opisina)	salaʔ (mali)
tayo (panghalip na panao)	tayoʔ (patindig na posisyon)

Figure 8. Stress and glottal stop minimal pairs (KWF, 2005)

Filipino spelling is generally straightforward. Each sound is represented by a single grapheme. In spite of the simplicity of the orthography, there were still some inconsistencies. First is capitalization. Proper names have their first letter capitalized. This is the only use of capitalization in the representation of the lemma.

However, some names of places did not have their first letters in capital. (e.g. *asya*). Moreover, some common nouns have lemmas with capitalization such as the lemma for bihis and the sublemmas for *baha* and *bantay.*

3.3.2 Loanwords

A second issue is the spelling of borrowed English words. Some of them retained their English spelling: attorney, beer, boss, business, mother, mommy, and date. Others have been adapted to Filipino such as otel (*hotel*), *istasyon, weyter, kostumer, treyler* (as a variation of trailer), *taksi* and *tinedyer.* This variation in spelling of borrowed words may be due to the frequency count since most were based on written materials where different orthographic forms for loanwords words are used.

Nevertheless, their representation must be consistent. For instance, all these can be adapted to Filipino pronunciation while the variation in spelling can be listed as part of the microstructure.

class {klas} png. **1.** Pangkat o grupo ng mga mag-aaral. *Ang class '93 ay magtitipun-tipon sa Sabado.* **2.** Klase; oras ng liksyon. *May class kami sa Math ngayon.* **Sk.** Pasok. **3.** Uri. *Ang bigas na wagwag ay first class.*
club {klab} png. **1.** Ang samahan; kapatiran. *Ang club ng mga kababaihan.* **2.** Bulwagang pinagsasayawan at pinagkakainan; panggabing restawran at pinagsasayawan. *Nagsaya sila sa night club.* **3.** Bahay-samahan; bahay kapisanan. *Ang club na pinuntahan nila ay sa Quezon City.*
College {kaĺedz} png **1.** Kolehiyo; paaralan; dalubhasaan. *Malayo ang college sa aming lugar.* **2.** Antas ng pag-aaral na nauukol sa mga kursong nais kunin ng isang mag-aaral. *Ang batang ito ay nasa college.*

Figure 9. Only these three entries were given phonetic transcription (KWF, 2005, p. 68)

3.3.3 Pronunciation Guide

Since it is a monolingual dictionary, a pronunciation guide is not a necessary feature. However, there are several issues worth mentioning. In the lemma list for C, some had transcriptions that served as guide for enunciation. The lemma for Cebu (*isla sa Visayas*) however, was not supplied with one even though the other

lemmas (class, club, and college) which has a 'c' that is enunciated using the /k/ sound had transcriptions. These are borrowed words from English. Only these three have been transcribed phonetically while the rest of the borrowed words in the dictionary were not supplied with a transcription.

Given that the three words in figure 9 are borrowed words already used commonly, it is unnecessary to provide a pronunciation guide. The other option would be to provide a pronunciation guide for all borrowed words from English to make the microstructure consistent. This is also particularly useful since, unlike Filipino, English orthography is complicated.

3.3.4 Macrostructure

The DMF is a semasiological dictionary hence, it uses an alphabetical macrostructure. The sections follow the order of the English alphabet. However, the letters F, J, V, and Z are not included and the letter NG is added after N. For the lemma and its derived forms, they follow the access alphabet and conform to a strict alphabet organization. This structure is contrary to the orthography set by the KWF. It does not follow the traditional orthography of Balarilang Pilipino, nor the orthography reform of 1987 (Yap, 2010). The authors of the DMF did not provide the reasons behind the chosen macrostructure. This issue must be looked into in future editions of the dictionary.

4. Conclusion

Using McFarland's corpus, the DMF's definition of Filipino is out of date. It no longer reflects the national language. Its use of materials prior to the current definition of the Filipino language weakens its claim as a truly monolingual dictionary of Filipino. However, the use of a corpus has afforded the authors several advantages. One of the problems with previous dictionaries had been the criteria for choosing the lemmas. Most often the authors used secondary materials filled with words that did not reflect the usage of the language at that period. The corpus provided the authors a way to solve this problem. The most frequently occurring words were chosen as lemmas. As these are the words

that speakers often encounter, users are more likely to look them up in the dictionary. The corpus through the concordance sets also facilitated the method used to describe the meaning of the lemmas. The different senses of the words can be conveniently identified as well as ranked according to frequency. The sentences in the corpus also provided a good way of disambiguating the multiple senses of polysemous words. These advantages clearly help make the dictionary more useful for the users. A few cases discussed above lead to meaning descriptions that do not provide the commonly used senses of the words. This can be easily corrected by double checking the meaning descriptions to see if they refer to the most frequently used senses. Reviewing the definitions would also help improve some of the problematic definitions for some lemmas. The other disadvantages mostly involve the inconsistent use of the structure chosen for the dictionary.

In its lexicographic description of grammar, the DMF reflects the current state of linguistic analysis of Philippine languages. Its vacillation between granting lemma status to the root or the derived form is symbolic of the lack of consensus on the morphological analysis of Filipino words and affixes. This is also shown by the inconsistent labelling of the syntactic categories of the roots and the semantic relationship between the focus affix and the co-indexed noun. The same point can be said of the discrepancies in orthography. There are, after all, orthographic guidelines for glottal stops and stresses but these were not followed. For loanwords, the rules for standardization has not been settled.

The study highlights the progress of the Filipino language through lexicography. The DMF has undoubtedly shown the benefits of the lexicographic approach in making an excellent dictionary that provides precise information about the language. As a whole, the DMF is an accomplishment and certainly warrants an expansion and new editions as long as it addresses the fundamental criticisms that the discussion above has pointed out. A revised monolingual dictionary of this kind would significantly contribute to the development of Filipino as the national language.

Janus Ruel Cabazares teaches at the Social Sciences Department of the University of the Philippines in Mindanao where he specializes in Anthropological Linguistics, Semantics, and Lexicography. Correspondences may be sent to januscabazares@gmail.com.

End Notes

[1] Aside from publishing a digital version of the DMF, there has not been any news regarding the publication of a revised or expanded edition.

References

Almario, V. S. (Ed.). (2001). *UP Diksiyonaryong Filipino.* Quezon City: UP-SWF.

Almario, V. S. (Ed.). (2010). *UP Diksiyonaryong Filipino.* Binagong edisyon. Diliman: UP Sentro ng Wikang Filipino.

Flores, M. L. (2015). Nahuhuli at panimulang pagtatangka: Ang pilosopiya ng wikang pambansa/Filipino. *Daluyan: Journal ng Wikang Filipino, Espesyal na Isyu* (2), 10-39. Retrieved from http://journals.upd.edu.ph/index.php/djwf/article/view/4500/4063

Francis, W.N. & H. Kucera. (1982). *Frequency analysis of English usage: Lexicon and grammar.* Boston: Houghton Mifflin.

Guillermo R.G., Cajote, R.D., and Logronio, A. (2015). UP Diksiyunaryong Filipino: Sinonismong Walang Hangganan. *Daluyan: Journal ng Wikang Filipino, Espesyal na Isyu,* (2), 54-71. Retrieved from http://journals.upd.edu.ph/index.php/djwf/article/view/4503/4066

John Sinclair: Editor Information. (n.d.). In *Collins.* Retrieved from http://www.collins.co.uk/page/John+Sinclair

Komisyon sa Wikang Filipino. (1998). *Diksyunaryo ng Wikang Filipino* (Sentinyal Edisyon). Quezon City: Merylvin Publishing House.

Komisyon sa Wikang Filipino. (2005). *Diksyunaryong Monolingwal sa Filipino.* Retrieved from https://store.kobobooks.com/en-us/ebook/diksyunaryong-monolingual-sa-filipino

Lee, A. P. (2010). The Filipino monolingual dictionaries and the development of Filipino lexicography. *Philippine Social Sciences Review, 62,* (2), 369-401. Retrieved from http://journals.upd.edu.ph/index.php/pssr/article/viewFile/2109/2032

Linangan ng mga Wika sa Pilipinas. (1989). *Diksyunaryo ng mga Wika sa Pilipinas.* Mandaluyong, MM: Cacho Hermanos, Inc.

McFarland, C. D. (1989). *A frequency count of pilipino.* Manila: Linguistic Society of the Philippines.

Newell, L. E. (1995). Handbook on lexicography for Philippine and other languages. *Linguistic Society of the Philippines, 36* (Special Monograph Issue). Manila: Linguistic Society of the Philippines.

Svensen, B. (2009). *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making.* Cambridge: Cambridge University Press.

Yap, F.A. (2010). *A Look at Global Filipino Orthography: Towards Modernization and Standardization.* Paper presented at the 2nd International Conference Filipino as a Global Language Transformation through Empowerment in Filipino Language and Culture, Mission Valley Hilton Hotel, San Diego, California, U.S.A., January 15-18, 2010). Retrieved from https://larc.sdsu.edu/images/larc/yap-a-look-@-global-filipino-orthography.pdf