

DETERMINATION OF WORDS (LEXICAL ITEMS) IN PILIPINO

Curtis D. McFarland
Waseda University
Tokyo

Finding a definition for word is a problem to which linguists have never found a totally satisfactory solution. In fact present-day linguists seem to have lost interest in the problem. But in the course of preparing A Frequency Count of Pilipino (nl) I came face-to-face with it because I had to decide what constituted a word in order to be able to count the frequency of 'words'. Anyone preparing a dictionary also has to decide what 'words' to use as entries.

In general, we use the word word in three different meanings, or senses. For A Frequency Count I worked with a '1,000,000 word' corpus. Students frequency are assigned to write a '500-word' paper, etc. In this sense, as a general rule, a word is any sequence of letters separated from other sequences by one or more spaces. By this definition the preceding sentence (In this sense ...) contains 23 words.

In order to count the frequency of words it is necessary to use a different definition: a word is any given sequence of letters which can be set off by spaces, and which can be understood to occur repeatedly within a text. By this definition the sentence above (In this sense ...) contains only 22 words, since the word a occurs twice. Thus we can speak of the frequency of a word (definition two) within 1,000,000 words (definition one) of text.

We need still one more definition, in order to put words in a dictionary. Different sequences of letters -- or words (definition two) -- are grouped together on the basis that they share the same basic meaning and differ only with regard to tense, number, case, etc. By this definition, the sentence above (In this sense ...) contains only 21 words, because sequence and sequences are understood to be different forms of the same word. This is the definition which gives us the greatest difficulty, since in many cases it is not obvious which forms should be grouped together.

In dealing with Western languages, it seems that we have an intuitive sense of what is or isn't a word. Even here, there are some problems. For example, we have 'two-word verbs,' such as give up. But verbs are a class of words, therefore a two-word verb must be a two-word word. And contractions: are things like I'm, isn't, gonna words or phrases.

In researching third-world languages, especially when this research is conducted by non-native speakers, we don't seem to

have the same intuition about words. We tend to look for rootwords and to group together all forms that share the same rootword. Panganiban's *Tesaurus* (n2) and *Father English's Dictionary* (n3) illustrate this tendency.

The use of rootwords has a number of problems. To name just three: (1) the 'rootword' of an apparently derived form may no longer exist independently in the language, or it may be impossible to extract; (2) The derived form may have a meaning which is not predictable; (3) the derived form may enjoy a higher frequency in language than the rootword.

Another problem for me is that I've always had the feeling there was something 'colonial' about concentrating on rootwords, when we don't pay much attention to them in Western languages (except in etymologies).

Another approach is to disregard rootwords and list all forms separately, with the exception of those which 'obviously' belong together. In Pilipino the 'obvious' combinations include different 'tense/aspect' forms of verbs, plural forms of adjectives, and the like. This is the approach found in *Vicassan's Dictionary* (n4) and the INL's monolingual *Diksyunario*. (n5)

While I favor this latter approach as being more in line with the Western concept of a dictionary and probably easier to use, the dictionaries mentioned have some difficulties. The biggest one is that they don't deal with the problem of focus, which is the problem I will take up in this paper.

Verbs in Pilipino and other Philippine languages are inflected for tense/aspect and focus. While there is some disagreement whether the inflection for time should be considered to be tense or aspect, it is accepted that each focused verb has a basic form or infinitive, and three tense or aspect forms. (The recent past form will be considered below.) For example:

Basic Form:	mag-áral 'to study'
Past (Perfective):	nag-áral 'studied'
Present (Imperfective):	nag-ááral 'studying'
Future (Contemplated):	mag-ááral 'will study'
Basic Form:	bilhin 'to buy'
Past (Perfective):	binili 'bought'
Present (Imperfective):	biníbili 'buying'
Future (Contemplated):	bíbilhin 'will buy'

The construction of these forms is determined by extremely regular rules. All dictionaries I know of list only the basic forms of verbs. No dictionary maker has felt the necessity to list tense/aspect forms separately.

In contrast to the regularity of the tense/aspect inflection, that for 'focus' is highly irregular. In fact, some scholars believe the focus 'system' should be treated as a matter of derivation rather than of inflection. (n6)

The basic idea of focus is illustrated in the following pair of sentences:

Actor Focus:

(1) **Tumúlong si Fidel kay Nita.** 'Fidel helped Nita.'

Direction Focus:

(2) **Tinulungan ni Fidel si Nita.** 'Fidel helped Nita.'

The verb (root: **túlong**) has different forms in the two sentences. In (1) Fidel is preceded by the article **si**, marking it as the Subject. Nita is preceded by the article **kay**, marking it as the Direction, or the recipient of help. In (2) Nita is preceded by **si**, marking it as the Subject. Fidel is preceded by **ni**, marking it as the Actor, or the provider of help. The Subject itself carries no information about the role or case of a particular noun phrase. The 'underlying' case of the Subject is determined by the case of the corresponding noun phrases in other sentences in the same set. **Ni Fidel** in (2) is an Actor; therefore **si Fidel** is an underlying Actor in (1). Similarly, **si Nita** is an underlying Direction in (2). Further, the underlying case of the Subject determines the focus of the verb. Thus, **tumúlong** is an actor-focus verb, and **tinulungan** is a direction-focus verb. (n7)

Pilipino focus differs from 'voice' in English and other Western languages. First, whereas English verbs have two voice forms, active and passive, Pilipino verbs may have as many as six or seven different focus forms.

Second, in Pilipino, in many cases, the passive or non-actor focus form is preferred to the active or actor focus form.

Third, there is no fixed number of focused forms a verb can take. Some intransitive verbs have only one form; others have two, three, or more.

Fourth, the focus of a verb is not immediately apparent from its form, that is from the affixes attached to it. The same focus (depending on the particular verb) may be expressed by different affixes. And the same affix (depending on the particular verb) may also express different focuses.

Fifth, the Pilipino focus system does not have as much freedom as the English voice system. Sentences like (1) and (2) can be interchanged freely, with little or no change in meaning. In many cases this is not true. Consider the following set:

Actor Focus:

(3) **Nagbigay si Leny ng libro kay Manolo.** 'Leny gave a book to Manolo.'

Object Focus:

(4) **Ibinigay ni Leny ang librong ito kay Manolo.** 'Leny gave this book to Manolo.'

Direction Focus:

(5) **Binigyan ni Leny ng libro si Manolo.** 'Leny gave a book to Manolo.'

Sentences (3) and (5) can be exchanged with relative freedom. The subjects in these sentences (and in most Pilipino sentences) have definite reference. (The definite/indefinite contrast is roughly the same as that in English.) Thus the subject of (4)

(ang librong ito) refers to, and must refer to, a particular book. On the other hand, the Object of (3) and (5) (ng libro) does not, and cannot, refer to a particular book.

Thus it might seem that ibinigay has a different meaning from nagbigay and binigyan. However, this is a grammatical difference, rather than a lexical one. This can be shown in a number of ways.

There are a number of constructions or environments in which the choice of focus is not at all free, and in which no difference in 'lexical' meaning can be attributed to the different focus forms. For example we can relativize the predicate in each of the sentences above. But we can only relativize on the subject, that is, we must choose a sentence with the verb in the proper focus. In relativization, the subject is deleted, and the remainder of the sentence can be appended to a noun (followed by the linker na/-ng). For example:

(3a) ang babáing nagbigay ng libro kay Manolo 'the woman who gave a book to Manolo'

(4a) ang librong ibinigay ni Leny kay Manolo 'the book Leny gave to Manolo'

(5a) ang batang binigyan ni Leny ng libro 'the child Leny gave a book to'

[We cannot say *ang librong nagbigay si Leny kay Manolo. A relativized verbal predicate cannot contain a subject.]

In relativized predicates the definite/indefinite restriction does not apply. Thus we can also have:

(3b) ang babáing nagbigay ng librong ito kay Manolo 'the woman who gave this book to Manolo'

(5b) ang batang binigyan ni Leny ng librong ito 'the child Leny gave this book to'

We can also have relativization without a preceding noun, in which case the relativized predicate is preceded by ang (or another article) meaning 'the one that':

(3c) ang nagbigay ng libro(-ng ito) kay Manolo 'the one who gave a (this) book to Manolo'

(4c) ang ibinigay ni Leny kay Manolo 'what Leny gave to Manolo'

(5c) ang binigyan ni Leny ng libro(-ng ito) 'the one Leny gave a (this) book to'

As a general rule, we can form questions with sino 'who' or ano 'what' to request the identity of the various noun phrases or referents in a sentence. The question word is followed by ang and the relativized predicate with a verb in the proper focus. Again the restriction on definiteness does not apply:

(3d) Sino ang nagbigay ng libro(-ng ito) kay Manolo? 'Who gave a (this) book to Manolo?'

(4d) Ano ang ibinigay ni Leny kay Manolo? 'What did Leny give to Manolo?'

(5d) Sino ang binigyan ni Leny ng libro(-ng ito)? 'Who did

Leny give a (this) book to?'

The third relevant construction or environment is the existential construction. As stated earlier, the subject of a Pilipino sentence is generally definite. An 'indefinite subject' is expressed with **may** 'there is' or **walâ** 'there is no'. The existential is followed by a relativized predicate with a verb in the proper focus. (**Walâ** is joined to the predicate by the linker **-ng**.) If there is an Actor in the relativized predicate, it is changed to a Subject in the existential construction. Again the definiteness restriction does not apply:

- (3e) **May nagbigay ng libro(-ng ito) kay Manolo.** 'Someone gave a (this) book to Manolo.'
- (4e) **May ibinigay si Leny kay Manolo.** 'Leny gave something to Manolo.'
- (5e) **May binigyan si Leny ng libro(-ng ito).** 'Leny gave a (this) book to someone.'
- (3f) **Walang nagbigay ng libro(-ng ito) kay Manolo.** 'No one gave a (this) book to Manolo.'
- (4f) **Walang ibinigay si Leny kay Manolo.** 'Leny didn't give anything to Manolo.'
- (5f) **Walang binigyan si Leny ng libro(-ng ito).** 'Leny didn't give a (this) book to anyone.'

The tightness of the system and absence of freedom of choice in these three constructions indicate that we are dealing with a single meaning and a single word. That is, **magbigay**, **ibigay** and **bigyan** (basic forms) are different forms of a single word.

Further evidence is provided by two other constructions: the gerund and the recent past (perfective). The gerund form of **bigay** is **pagbibigay** and expresses 'giving' or 'act of giving'. The gerund construction has no subject; all noun phrases are in their 'underlying' cases:

- (6) (**Nagúlat ako sa pagbibigay ni Leny ng libro(-ng ito) kay Manolo.** '(I was surprised at) Leny's giving a (this) book to Manolo.'

Although the form of the gerund is determined by the form of the actor-focus verb, it has no focus itself and belongs to the set as a whole.

Similarly the recent-past construction has no subject and no focus. Like the gerund, its form tends to be determined by the form of the actor-focus verb, but it belongs to the set as a whole:

- (7) **Kabíbigay lang ni Leny ng libro(-ng ito) kay Manolo.** 'Leny just gave a (this) book to Manolo.'

Furthermore insofar as there is a difference in meaning between **magbigay** and **ibigay**, this difference is rule-governed and applies to thousands of other verbs. It is in no sense idiomatic.

Our conclusion must be that **magbigay**, **ibigay**, and **bigyan** should be treated as forms of a single word, and should have a single dictionary entry. Similar arguments account for hundreds

or thousands of Pilipino verbs. Most of these verbs belong to one of the following large classes:

-um-/ -an (Direction):	tumúlong/tulungan 'help'
mag-/i- (Object):	magbigay/ibigay(/bigyan) 'give'
-um-/ -in (Object):	bumili/bilhin(/bilhan) 'buy'
mag-/ -in (Object):	magsabi/sabihin(/sabihan) 'say'

The **maka-** verb classes are much smaller but contain some very important verbs:

maka-/má- (Object):	makakíta/mákíta 'see'
maka-/má-an (Object):	makaalam/máláman 'know'

There are also smaller classes, such as verbs with **-an** object-focus forms, and **-in** direction-focus forms. And there are also verbs with exceptional focus forms.

As a general rule, it seems to be necessary to memorize what class a verb belongs to; that is, what its primary forms are. There is, however, some semantic motivation for distinguishing one class from another. (n8)

Having established that there is a focus inflection linking different forms into single verbs, it is necessary to point out that there are many cases of derivation and semantic divergence among related verb forms. These are properly to be separated and given separate dictionary entries.

Some derivations follow established patterns and apply to many sets of verb forms. For example there are many pairs of transitive and intransitive verbs such as the following:

-um- (Intransitive):	tumaas 'rise'
mag-/i- (Object):	magtaas/itaas 'raise'

There are other pairs in which the same actor-focus form is both transitive and intransitive:

mag- (Intransitive):	magtágô 'hide (self)'
mag-/i- (Object):	magtágô/itágô 'hide (something)'

The **maka-** verbs mentioned above express events related to the senses or the feelings, such as seeing, hearing, knowing, feeling, etc. They express activities that are 'experienced' rather than 'done'. For most of these verbs there is a corresponding 'intentional' verb. For example:

makakilála/mákilála	'recognize'
kumilála/kilalánin	'give recognition to'
makalímot/malimútan	'forget'
lumímot/limútin	'try to forget'

This derivation is highly irregular both in terms of form and meaning.

In all of these cases, I think we are justified in making separate lexical entries, even where the derivational process seems to be quite regular.

Finally, let us return to the problem of differences in meaning between different focus forms. I showed earlier that the definite/indefinite distinction is grammatical rather than lexical, is rule-governed, and applies to many verbs. There are other differences which are indeed lexical differences.

Humalik, actor focus, in a free sentence expresses 'kiss' in the sense of kissing a relative, a hand, an object:

- (8) **Humalik si Laura sa kamay ni Auring.** 'Laura kissed Auring's hand.'

Halikan, direction focus, expresses romantic kissing in addition to the kind of kissing expressed by **humalik**:

- (9) **Hinalikan ni Ramon si Dolores.** 'Ramon kissed Dolores.'

In relativization, questions, and existential constructions, both meanings are expressed in both focus forms.

Pumunta, actor focus, expresses 'go to', both in the sense of moving to a place, and going to someone for some purpose, such as asking for help, giving consolation, etc.:

- (10) **Pumunta si Belen sa Cebu.** 'Belen went to Cebu.'

- (11) **Pumunta si Belen sa tiya niya.** 'Belen went to (see) her aunt.'

Puntahan, direction focus, in a free sentence expresses only the second meaning, going for some purpose:

- (12) **Pinuntahan ni Belen ang tiya niya.** 'Belen went to (see) her aunt.'

***Pinuntahan ni Belen ang Cebu.**

Again both meanings are expressed in both focuses in relativization, questions, and existential constructions.

These last examples involve very fine semantic distinctions, but there are many similar cases, and a good dictionary should make these distinctions.

There are two big questions I haven't tackled in this paper. First, the other focus forms (benefactive, instrumental, causal, etc.) and other verbal forms such as aptatives and causatives. In general I think these can be considered part of the verbal inflection, although there are many exceptions and much idiomatic divergence.

The second question has to do with where the (combined) entry should be placed in a dictionary. This is an important question in Pilipino, since forms like **magbigay**, **ibigay**, and **bigyan** would be far apart if given separate entries. The candidates are (1) the rootword: this would bring all the derived forms together; (2) the actor-focus form: most verbs have an actor-focus form even if other forms are missing; (3) that form, for each verb, which occurs most freely and best expresses the meaning of the verb. In any case extensive cross referencing is essential.

NOTES

- (n1) Curtis D. McFarland, **A Frequency Count of Pilipino** (1989, Manila: Linguistic Society of the Philippines)
 (n2) Jose Villa Panganiban, **Diksyunaryo-Tesaurus Pilipino-Ingles** (1972, Quezon City: Manlapaz Publishing Co.)
 (n3) Leo James English, **Tagalog-English Dictionary** (1986, Quezon City: Kalayaan Press)
 (n4) Vito C. Santos, **Pilipino-English Dictionary** (1978, Caloocan City: Philippine Graphic Arts, Inc.)

- (n5) **Linangan ng mga Wika sa Pilipinas, Diksyunaryo ng Wikang Filipino** (1989 Mandaluyong: Cacho Hermanos)
- (n6) Personal communication with Lawrence Reid.
- (n7) This procedure is expounded in Paul Schachter and Fe T. Otanes, **Tagalog Reference Grammar** (1972, Berkeley: University of California Press)
- (n8) Cf. Videa P. De Guzman, **Syntactic Derivation of Tagalog Verbs** (1978, Honolulu: University Press of Hawaii)