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** (n1) 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 **Tesauro** (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) **Tinulúngan 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 **tinulúngan** 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 bátang 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 bátang 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 **síno** '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) **Síno 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) **Síno ang binigyan ni Leny ng libro(-ng ito)?** 'Who did