

MORPHEME-BASED VERSUS WORD-BASED MORPHOLOGY**VIDEA P. DE GUZMAN***The University of Calgary**deguzman@ucalgary.ca*

Several questions have been raised about morphology over the past few decades and various accounts have thereby been proposed. Some linguists favor a morpheme-based morphology whereas others argue for a word-based type. Still others contend with neither one. Two of such types — Lexical Morphology and Seamless Morphology — are explored here to determine which better represents the morphological system of any given language. The three aspects of morphology reviewed in each type are inflection, derivation, and compounding. The cases of affixation and reduplication, being prevalent morphological processes in Philippine languages, are also used to test the plausibility of each theory. Furthermore, this paper also makes reference to some language acquisition and psycholinguistic studies that may shed light on the issue.

1. Introduction

This paper explores two radically different theories, among the variety of theories that have emerged over the past few decades, namely, Lexical Morphology (Lieber, 1990, 1992; Di Sciullo & Williams, 1987; Selkirk, 1982) on the one side and Seamless Morphology (Ford, Singh & Martohardjono, 1997; Singh & Starosta, 2003) on the other. Whereas the former recognizes the MORPHEME as the basic element or unit in morphology, the latter considers the WORD. From this basic difference, we will trace the other conceptual distinctions that follow from these two views particularly in the areas of inflection, derivation, and compounding. As we discuss the strengths and weaknesses in the way each view accounts for the morphological system, we hope to be able to draw some insights that will serve as guide in making an adequate description of the morphological system of a given language.

2. Morpheme-based Morphology

Among those linguists who contend that the basis of morphological analysis is the morpheme, I concentrate on Lieber (1992) because she presents an

innovative treatment of morphology. Her approach differs from the more common statements of ‘word formation rules’ (WFRs as in Aronoff, 1976) or the formulation of ‘word structure rules’ (WSRs as in Selkirk, 1982) even when these earlier works are also both based on generative grammar.¹ For example in English, note the following WFRs (Aronoff, 1976, pp. 49, 63)² (1) which forms N(oun)s from V(erb)s by attaching the suffix *-ee* and (2) which forms negative A(djective)s from A(djective)s by attaching the prefix *un-*:

- | | |
|----------------------|---|
| (1) $[+X]_V + ee]_N$ | (2) $[X]_{Adj} \rightarrow [un\# X]_{Adj, Adj}$ |
| employ + ee | kind → unkind |

From Selkirk (1982, pp. 80, 85)³, compare the following parallel but more general WSRs:

- | | |
|----------------------|------------------|
| (3) $_N[V N^{af}]_N$ | (4) $_A[Af A]_A$ |
| employ-ee | un-kind |

In contrast, Lieber proposes to eliminate entirely a separate morphological component of grammar by generating complex words using only principles from Chomsky’s X-bar theory that are already independently used for syntax. This means that there are no word formation or word structure rules that put together constituents of complex words. In her framework, all morphemes, both stems (free morphemes) and affixes alike, are listed in the lexicon with their syntactic categories, semantic and phonological representations. In addition to these features, affixes bear subcategorization information to indicate the environment in which they can be inserted into unlabeled binary-branching word structure trees. When an appropriate affix attaches to a stem, the structure is projected from these lexical entries following the Projection Principle (Chomsky, 1981). The operation involves two theories which Lieber proposes, namely, the theory of HEADEDNESS and the theory of FEATURE PERCOLATION.

2.1 Headedness

Lieber contends that as in phrasal structures the head of a complex word is followed or preceded by its complement, modifier, or even specifier. Thus, for Tagalog, which Lieber (1992, pp. 40-49) uses as one of the test languages to prove her predictions about headedness in both phrase and word structures, she

¹ This is not to say that there are totally no similarities in the formulation of their respective theories of morphology. It will be observed though that these two types of word formation rules are rather familiar to most linguists.

² Note that with the inclusion of morpheme boundaries, the application of rules in terms of levels in lexical phonology is implied.

³ Note that in rule (4), the constituent AF (for affix) does not bear any category label; in (3) it is surprising that the N constituent is identified as an ‘af’ (for affix) instead of the constituent being an AF and labeled as a N category.

effectively shows it to be left-headed. The heads N of NP, A of AP, or P of PP all appear on the left of their complements, modifiers, and specifiers. Compared with the structure of complex words, she shows the prefix as head of the stem that follows (Lieber, 1992, pp. 44-46):

- (5)a. *ma-* [A ____ [{N,A}] bigat 'weight' mabigat 'heavy'
- b. *mang-dup₁* [N ____ [{N,V}] tanggol 'defend' manananggol 'lawyer'⁴
- c. *pang-* [A ____ [N kamay 'hand' pangkamay 'for the hand'

Being the head, whatever is the category of the prefix becomes the category of the affixed word. However, when a suffix attaches to a base, the suffix carries over its own category to the resulting complex form. This appears then to be right-headed. Lieber identifies the Tagalog suffixes *-an* and *-in* with various functions in the following examples (Lieber, 1992, p. 48)⁵:

- (6)a. *giik*_V 'thresh' giikan]_N 'threshing place'
- b. *mag-away*_V 'fight' awayan]_N 'fighting one another'
- c. *dugo?*_N 'blood' dugu?_{AN} 'covered with blood'
- d. *mag?aral*_V 'to study' aralin]_N 'something to study'
- e. *antok*_N 'sleepiness' antukin]_A 'given to sleepiness'

She offers two explanations for this: first, that the processes may not be particularly productive (no indication of productivity was mentioned by Schachter & Otanes, 1972, her source for these data) and therefore, these items should just be listed separately in the lexicon; and second, that V bases, as above, generally function as predicates, rather than as modifiers, complements, or specifiers in which case a right-headed structure would not be ruled out. We may find these explanations unsatisfactory for at least two reasons: (i) both the locative function of *-an* with either N or V base, such as *la:rú-an* 'playground', *lutu-án* 'place for cooking', *tagu-án* 'place for keeping or hiding something', etc. and the nominal object of the V function of *-in* as in *la:báh-in* 'clothes for laundering', *ta:hí?-in* 'materials ready for sewing', *sulat-in* 'things(to be) written', etc., are deemed productive; (ii) as for the other functions, identified from the examples, such as group action or reciprocity of action, susceptibilities or health conditions, they are naturally constrained by the semantics and pragmatics of the resulting complex forms. For instance, only certain actions can be done to each other; only a minimal number of items can cover or drench a person

⁴ It will be observed that the phonological form of the resulting complex structure is not accounted for here.

⁵ It may be added that the complex forms, with suffixes, usually differ in their stress and/or vowel length patterns from those of the roots used as bases. For example: *gík* vs. *gi:fkan*; *áway* vs. *awayán*; *dugó?* vs. *dugu?_{án}*; *áral* vs. *aralín*; *antók* vs. *antúkin*. This, to my mind, speaks of the need for a morphonological account of these complex forms.

DE GUZMAN

involuntarily. Moreover, besides the functions of these two suffixes identified above with N bases, they are predictably quite productive as verbalizers, not to mention their greatest use with V bases resulting in the objective or locative voice/focus verbs, all exhibiting right-headedness. A few examples, following Lieber's forms above, are:

- | | |
|----------------------------------|--|
| (7)a. bato] _N 'stone' | batuhin]v 'to stone someone or something' |
| b. laro?]]N 'game' | laru?in]v 'to play some game or play with someone' |
| (8)a. sabaw]]N 'soup' | sabawan]v 'to put soup on something' |
| b. damit]]N 'dress' | damitan]v 'to put a dress on someone or something' |
| (9)a. kain]v 'eat' | kainin]v 'to eat something' |
| b. punas]v 'wipe' | punasang]v 'to wipe off something' |

Where the affixes involved are not uniquely prefixes, the left-head principle may be put into question. As it stands, it may be more prudent to state that the tendency of the direction of headedness for Tagalog is towards the left. In this way the existence of forms formed via the opposite direction is not completely ruled out. What is sometimes overlooked in analysis is the extent of alternation or constraints in word order within phrases. For example: N+A ~ A+N as in *mabait na bata/bata-ng mabait* 'good child'; *Maganda-ng umaga po/*Umaga-ng maganda po* 'Good morning *po*', *anak ng mahirap/*mahirap na anak* 'child of the poor'. The alternation given in the first instance is quite pervasive. How will such variation in headedness translate in word structure?

The next set of complex word structure Lieber presents to provide evidence for left-headedness in Tagalog is that of compound words. A combination of two forms each with its given category results in a complex form that bears its own category specification. Again, Tagalog shows itself to be left-headed (Lieber, 1992, pp. 46-47). Not to belabor the point that there are also many instances of right-headed compounds (see De Guzman, 2005), it has to be mentioned that it is an overstatement for Lieber to say that 'Tagalog in fact has no compounds which are either syntactically or semantically right-headed' (1992, p. 47). The first example below is simply to show a right-headed compound⁶ and the succeeding ones from Lieber are meant to highlight the complex forms some constituents of compounds may take:

- (10 a. [A-N]_N
- | | |
|----------------------------|-------------------|
| pang-bansa- <i>ng</i> awit | (pambansang awit) |
| for-nation-linker song | 'national anthem' |

⁶ Other examples of [A N]_N compounds in Tagalog are: *madali-*ng* araw* 'dawn', *hati-*ng* gabi* 'midnight', *huli-*ng* habilin* 'last will/request', etc. With [V N]_N compounds, we have: *basag ulo* 'a fight/brawl', *hanap buhay* 'occupation', *buka-*ng* bibig* 'favorite expression', etc.

b. [V-N]

mag-bigay	galang	
give	respect	'show respect'
mag-bango(n)-ng	puri	
raise	honor	'redeem one's honor'

c. Deverbal (synthetic)⁷

[pang-patid _V] _N	uhaw	(pamatid-uhaw)
cutter	thirst	'thirst quencher'
[pang-tawag _V] _N	pansin	(panawag-pansin)
caller	attention	'attention getter'

Note that in the above examples all the first members of the compound words are complex forms and given that these are not listed or stored in the lexicon, they must be projected individually before each is combined with the other constituent. Undergoing the process every time a complex form is used in compounding, or in any complex word formation for that matter, appears to involve a greater complication following this view.

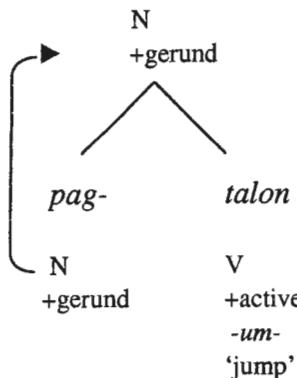
At the end of the exposition on Tagalog, Lieber (1992, p. 49) concedes that the examples given in (6) 'are the only ones which pose any challenge to the predictions' of the conditions on Tagalog headedness. To this we may add the challenge of the existence of productive right-headed suffixes and compounds. On top of these may also be included the consequent changes resulting from strictly morphonological processes applying as the complex words are formed, e.g. secondary stress (or vowel lengthening), stress shift, nasal substitution, etc. There is no provision for morphonology, but these processes are incorporated in a modified version of Lexical Phonology (Lieber, 1992, 178ff., 197ff.).

2.2 Feature percolation

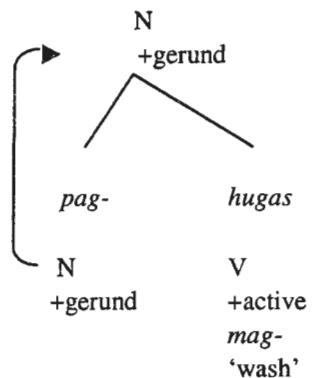
In describing the mechanics for projecting word structures, Lieber employs what is known as feature-percolation. This process accounts for the labeling of word structures projected from lexical entries. She identifies the features which percolate to the dominating node of the head morpheme – category features, morphosyntactic features relevant for a particular category in a given language. If certain relevant features are missing from the head, then those in the co-occurring constituent may serve as back up. For example in Tagalog, from a V morpheme such as *talon* 'jump' there is a corresponding complex N form *pagtalon* 'the act of jumping' with the prefix *pag*. The figure below shows its feature percolation:

⁷I have shown here the complex form of the first word in the compound. The resulting phonological form is in parentheses. The observed nasal assimilation here can be automatically dealt with in the phonological component of the grammar. However, compare (5)c where the same rule does not apply and (5)b where it applies first before reduplication.

(11)a.



b.



In both (11)a and (11)b, the features of the affix *pag-* percolate to the dominating node giving the complex forms their features. While feature percolation ensures that the features of the head be registered in the resulting complex form, there may also be specific features in the nonhead that play a significant role in terms of the resulting complex form. In the above examples, the two resulting complex forms differ. Whereas (11)a is a simple combination of the prefix and the V root, *pagtalon* 'the act of jumping', (11)b is *paghuhugas* 'the act of washing', requiring reduplication of the first CV of the V base. In (11)a, the V base is marked with the voice affix [-*um*-] and does not require any reduplication; (11)b, on the other hand, is marked with the voice affix [mag-]⁸ and manifests reduplication. But this constraint is not shown in the formation or projection. Obviously, there must be a phonological rule that can look into the voice feature of the V base of this type of formation, so that when it is marked [mag-], this triggers the required type of reduplication. This theory may probably resolve the difference between these two forms by creating another entry *pag-dup₁*, except that its subcategorization will still have to specify the voice affix [mag-] that the V base takes, a necessary morphological condition.

The distinction Lieber (1992, p. 112) makes between derivation and inflection is in terms of marking the features. She proposes that 'only stems, bound bases and derivational affixes' will have the full category and morphosyntactic features specified. Inflectional affixes will be marked only with individual features for which they contain specified values. This implies that the distinction between inflectional and derivational affixes is not an issue. Is this true of Philippine languages or of other languages for that matter?

⁸ For ease of presentation, the voice affix [m-] + [pag-] stem will simply be identified as [mag-].

2.3 Remarks

In general, Lieber has shown a fairly plausible position. She has shown quite persuasively that special rules for combining affixes and roots can be dispensed with. With the specification of subcategorizational features along with their word categories and other morphosyntactic features, affixes are already marked for their combinatory potential. Moreover, a lexicon with just morpheme listing is obviously significantly more economical.

Applying this theory to Philippine languages, however, we have already observed certain problems as follows:

(a) It is not certain how stem-forming affixes, say *pag-*, *pang-*, *paki-*, etc., are to be subcategorized for taking multiple-affixed forms, such as causative, abilitative, requestive, etc.

(b) There is a need for some morphonological processes, such as those needed to account for (5)b, nasal substitution before reduplication in [+occupation] forms, and for (11)b, reduplication triggered by the specified [*mag-*] voice feature in gerundive N formation. These particular classes of word formation and other types are morphologically constrained.

(c) Certain phonological rules applying on complex word forms may appear automatic. But some are constrained by semantic features. Compare the nasal substitution in (10)c, *pang-patid* = *pamatid* ‘for cutting or cutter’, *pang-tawag* = *panawag* ‘for calling or caller’ but not in (5)c, *pang-kamay* = *pangkamay* ‘for the hand’ or *pang-simba* ‘for wearing to church’, where the latter forms have the feature [+reservational] in the resulting form rather than [+instrumental]. (See De Guzman, 1978 for a fuller treatment of this constraint.)

In consideration of the independent morphological, morphonological as well as semantic questions that have to be dealt with in accounting for the morphological system of Tagalog, which are most likely true of the majority of Philippine languages, the morpheme-based morphology account has these matters yet to contend with. Besides there are still existing questions on identifying morphemes, particularly in inflecting languages, not to mention the problems relating to conversion and reduplication. On the whole, it seems unnatural and uneconomical to compose each complex word every time one needs to use them or deconstruct them into constituents when she hears them.

3. Word-based Morphology

In *Pace Pāṇini, toward a word-based theory of morphology*, Ford, Singh & Martohardjono ((FSM), 1997, pp. 1-2) provide a sketch of the theory. At the outset, they define morphology as ‘the study of formal relationships amongst words’ and for an expression to be considered a word they identify three properties, which are standard characteristics found also in other morphological theories: (i) a phonological structure, (ii) a syntactic category, and (iii) a semantic

DE GUZMAN

use.⁹ A set of morphologically related words bound by some common features is expressed in what is described as a Morphological Strategy or a Word-Formation Strategy and later, referred to as a word-schema. For example, the English words *cats*, *dogs*, *gems*, *pins*, *bags*, *cards*, and the like, can be shown to be related in the following word-schema that generalizes their similarities:

- (12) [/Xz/]
[N]
[‘plurality of xs’]

What (12) indicates is that the words listed above are all different, expressed as the variable /X/, but that all of them are similar in that they end in /z/, that they belong to the category N, and that they denote plurality of whatever /X/ is. Interestingly enough, the lexicon contains exactly similar words except that the /z/ ending is missing such that the word-schema that captures this set of corresponding words is as follows:

- (13) [/X/]
[N]
[‘x’]

It is significant for morphology that these two sets of words are related in terms of their form, category, and meaning, and such relationship can be captured through the following correspondence:

- (14) [/X_N/] ↔ [/Xz_N/]
[‘x’] [‘plurality of xs’]

This type of morphological relationship is generalized as follows:

- (15) /X_α ↔ /X'_β

In the above schema, X and X' are words, α and β are morphological categories, the double-headed arrow, ↔, indicates an equivalence relation (a bidirectional implication), X' is a semantic function of X, the prime symbol indicates a formal difference between the two poles of the morphological operation. But the prime symbol can be null if α ≠ β. To illustrate a relation where there is no formal distinction between two words belonging to different categories, we have the following rule for pairs of English words such as paint_N/paint_V, spoon_N/spoon_V, brush_N/brush_V, etc.:

- (16) /X_{N, ins.} ↔ /X_{V, use x (as ins.)}¹⁰

The class of Ns on the left are instruments and the related Vs of exactly the same form refer to ‘use x (the instrument expressed by the variable)’. In this particular case, we note readily how much simpler this morphological relation is

⁹ FSM specifically identify ‘semantic use’ in place of semantics or meaning because they don’t believe that the meaning of a word can be simply stated as a bunch of features as proposed in other theories. For ease of exposition here, we will forego the distinction, using semantic use or semantics or meaning alternatively.

¹⁰ The schema given here is an abbreviated one making the category and meaning as subscripts.

stated in contrast to more complicated accounts of conversion in some morpheme-based morphology where a rule of category change may be stated or a zero affix derivation may be posited. The former account may be questioned because in a category change, the formulation of a rewrite rule necessitates directionality implying that the form on the right of the arrow is derived from the form on the left. For example: $[X]_N \rightarrow [X]_V$. However, there are cases in which the direction may be reversed. For instance: $[X]_V \rightarrow [X]_N$ as in English *contact_V*: *contact_N*, and more recently *read_V*: *read_N*, *invite_V*: *invite_N*, etc. The alternative account to this shows that a zero affix with the category V may attach to a N or conversely, following the reverse direction of change of category. Of course, the age-old dissatisfaction with a zero affix, and more so with its bearing a category specification, strikes some linguists that this account is no more than an ad hoc solution. One other recourse is simply to list the two words separately in the lexicon at the cost of losing a generalization in terms of their lexical relation.

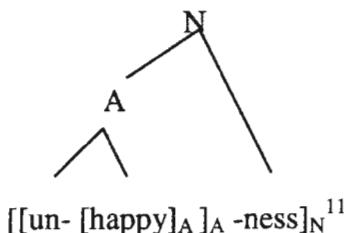
As shown above, a morphological strategy or word-schema captures the morphological relatedness among the words in the lexicon. It should be remembered that in a word-based theory of morphology, only words, not morphemes, are listed in the lexicon. FSM (1997, p. 2) also claim that the WFS 'allows a speaker to morphologically analyze a word she may not have analyzed before or to create a new word that she may not have yet encountered ...' The list of morphological strategies in a given language makes up a part of the grammar of that language. Thus, unlike Lieber's morpheme-based account, FSM set up a morphology with constraints (1997, p. 3) that defy traditional or even current analysis. Some of these constraints are as follows: (a) It does not distinguish between inflectional/derivational, clitic/non-clitic, productive/non-productive, concatenative/non-concatenative, affixation/compounding, etc.; (b) Since the only object of a morphological strategy is the WORD, there are no other smaller units such as ROOT, STEM, LEXEME, etc.; (c) There is no 'morphophonology' that implies a division within morphology; the morphological operation is said to be unified; (d) The bidirectionality of the WFS gives no special status to back-formation or truncation.

With these constraints, the authors have excluded a good number of notions/concepts that we have gotten so used to. Are they really that superfluous so they can be eliminated at no cost? The authors contend that the burden of proof for the concepts that continue to be employed in other morphological theories lies with the proponents of such theories. It doesn't take much then to deduce from these constraints that the word-based theory has made the morphological account much simpler in a number of ways although as will be pointed out later, it does not look totally economical. To determine its feasibility, efficacy, and adequacy, we must explore further other aspects of these Word-Formation Strategies (WFSs).

3.1 Multiple derivational relations

As can be observed in many agglutinating or affixing types of languages, words may contain not only one but two or even more affixes. While a morpheme-based morphology may account for a more complex structure in terms of a hierarchical tree-structure, a word-based morphology that makes use of WFSs will lay out the word-schema of each set of words that are related according to the categories and constant affixes identified in each set without any reference to any kind of structure. (What you see is what you get!) For example, compare the two sketch accounts of the English word *unhappiness*:

(17)a. Morpheme-based account



b. Word-based account

$$\begin{aligned} /unX/A, \text{'not } x' &\leftrightarrow /unXness/N, \text{'state of being not } x' \\ (/X/A \leftrightarrow /unX/A, \text{'not } x') \end{aligned}$$

(17)a shows first that the prefix *un-* combines with the Aroot, *happy* (the head), resulting in a complex word of the same category A, via the projection principle. With the suffix *-ness* marked with the category N and subcategorized for a preceding A category, its specified category N percolates to the topmost node as it combines with the complex form A, *unhappy*. This same hierarchical structure will appear for other similar words such as *unkindness*, *unsoundness*, *uncleanness*, etc. As for (17)b, the word *unhappiness* falls within the first WFS which states the morphological relation between the two sets of words, the */unX/* on the left side bearing the category A and the */unXness/* on the right marked with the category N, differing in form and meaning. The former A form with the beginning portion */un/* represents the negative meaning of what the variable */X/* stands for and, correspondingly, the latter N form has exactly the same form as that of A except for its constant ending, */ness/*, and the complex form meaning 'state of being not x'. Therefore, what this schema covers are all */unX/* forms of the category A being related to */unXness/* forms of the category N. I included a second WFS to show the relation between the A class */X/* and the

¹¹ Just to show the internal structure of the word in terms of tree-diagram and labeled bracketing. For simplicity of presentation, the meaning and subcategorization features have not been included.

negative A class with the form /*unX*/. And showing this implies that a separate WFS exists for the relation between /*X*_A and /*Xness*_N. What is not or cannot be shown in separate rules such as these is that both of these A sets are identically related to the corresponding /*ness*_N forms. I think that to avoid redundancy and at the same time capture the identical relation stated, the WFS should allow for the conflation of the two sets of As as follows:

$$(18) /(\text{un})\text{X}/_A \leftrightarrow /(\text{un})\text{Xness}/_N$$

The only stipulation here would be that the option of referring to /*un*/ complex forms of A requires pairing them also with the same portion in the complex /*ness*/ forms of N.

Another type of generalization involving more than two sets of words may be shown to be morphologically related by identifying each word-schema, separated by a comma and collapsing them all under one pair of curly bracket. For example, given in Haspelmath (2002, p. 51):

attract	:	attractive	:	attraction
instruct	:	instructive	:	instruction
suggest	:	suggestive	:	suggestion
discuss	:		:	discussion
-	:	aggressive	:	aggression
(19)	{	[/X/ _V] , [/Xive/ _A] , [/Xion/ _N] }		
	['do x']	['prone to doing x']		['act of doing x']

It will be observed that there can be gaps in the listing and this is naturally accounted for, not by some rules of exception, but by their non-occurrence in the lexicon. Simply put, they do not exist. Perhaps, later in the future, the words *aggress* or *discussive* may possibly be created on the basis of analogy to the existing similar lexical relations. From a Morpheme-based theory, there need to be two separate projections of the /-ive/_A and the /-ion/_N complex words without stating any relations between them.

3.2 Inflectional relations

The Word-based theory does not make any distinction between derivation and inflection; hence, any relation existing between two sets of what is commonly analyzed as 'inflected' forms is expressed as a WFS. At this point it will be appropriate to consider the account provided for inflected forms in Latin. One significant account is made by Bender (2000, pp. 15-17; 2003, pp. 304-307), who shows quite convincingly that between the WFSs stated for every inflectional relation occurring in each morphological category and his proposed 'paradigmatic strategies' (2000, p. 22), the latter is to be preferred. To illustrate, he compares the two strategies:

- (20)a. /Xô/ 1sg ↔ /Xâtis/ 2pl
 cf. *amô* ~ *amâtis*, *portô* ~ *portâtis*, etc.

b. Conj 1: nonperf pres act ind

	sg.	pl.
1	Xô	Xâmus
2	Xâs	*
3	Xat	Xant

Instead of 15 WFSs to relate the six items in the paradigm, expressing person and number distinctions, his proposed paradigm in (20)b with an asterisk in the center shows in a more succinct way all these relations. More importantly, the paradigm provides a constant value for X, the maximum form shared by the members of the paradigm. On the other hand, this X will vary in the WFSs according to what is common in each pair being related. In sum, keeping the notion of the ‘paradigm’, i.e. ‘the different inflections of one and the same word (or lexeme/lexical item)’ (Bender, 2003, p. 301) and showing at the same time the relations existing between each pair achieves remarkable economy and naturalness. Experience from field work with adult native speakers tells us that they can immediately and unhesitatingly provide the forms of Vs expressing, say, tense or person, traditionally referred to as conjugation. Thus, to ask whether paradigms are part of a language’s design remains a valid question that merits the linguist’s serious attention and consideration.¹²

3.3 Compounding

On this particular class of words, each compound is naturally treated in the Word-based theory as a word, and like all other words, it does not have an internal structure, i.e. it is not made up of two separate words. Rather, there is a word-part that is a constant joined to a variable, with the constant bearing a lexical category. Starosta (1988, pp. 60-64; 2003, pp. 126-130) Singh & Dasgupta (2003, pp. 77-89) expound on this alternative account to the popular rules of compounding a la phrase structure rules, e.g. N → N N. Under this alternative analysis, the morphological constant may look much like an existing word but its category and meaning cannot be associated with the latter (FSM, 1997, pp. 58-59). Starosta illustrates the relation between compounds and one of their ‘constituents’ where the constant part in the compound set is treated as an ‘affix’ (Starosta, 2003, pp. 126-27), restated here according to the WFS format as follows:

(21) N-*house* ‘compounding’

$$/ \underset{\text{N, } \alpha \text{Fi}}{X} / \leftrightarrow / \underset{\text{N, domicile, } \alpha \text{Fi}}{\text{Xhaws}} /$$

¹² For a non-reductionist view, see Aronoff’s *Morphology by itself* (1994). He reinforces the basic idea ‘that morphology has a place of its own in language’. He refers to ‘inflection classes as having a life of their own and existing as independent parts of the grammatical engine’ (1994, p. 62).

The set of words represented on the right side are ‘compound words’ ending in *house* or phonologically /haws/ as in English *doghouse*, *birdhouse*, *cathouse*, *nuthouse*, or even a possible *monkeyhouse*, etc. It states that given any N with the semantic feature [αF_i], shared by this class of Ns, there may be a N ending in the sequence *haws* and having the semantic feature [+domicile, αF_i]. This rule is again based on analogical patterns similar to those stated for both ‘derivationally’ and ‘inflectionally’ related words.¹³

3.4 Remarks

Taking the constraints of Seamless or Word-based Morphology as a great simplification not only in terms of reducing the elements employed in a Morpheme-based account, we may say that the former is more advantageous than the latter. Likewise, having whole words all listed in the lexicon, the speaker has easy access to them in both production and comprehension and even in the creation of new forms on the basis of analogy. Besides, there is always a corresponding check on the actual WFSs that list the various patterns of lexical relations. One other advantage of this theory is that it does not tax the speaker with ‘knowing’ all the details of feature specifications and subcategorization requirements for both stems and affixes that the Morpheme-based account exacts from her. Even without referring to the terms ‘affixes’, they are isolated accordingly in the format of WFSs.

Some concerns about this theory are as follows:

- (a) As shown in (17)b compared with the suggested WFS (18), this theory misses showing the commonality of relation obtaining between positive and negative As in being both related to their corresponding Ns with *-ness*. For all we know, (18) may be judged unacceptable.
- (b) One other source of potential contention is the strategy of identifying one of the members of a compound as a constant, that is, a kind of affix and not as a word. This greatly minimizes its distinction from other derived words, and thus loses a significant morphological generalization. Perhaps, there may even be a corresponding psychological reality to the distinction as will be implied indirectly in the next section.
- (c) I still maintain, as discussed earlier, that morphonology is an essential part of morphology. Certain features in the V base trigger corresponding phonological operations.
- (d) Finally, as the sets of relations that can be covered by each WFS are rather limited, we are looking at an enormous number of WFSs to formulate for

¹³For a fuller exposition of the accounts for ‘compounds’ as well as ‘noun incorporation’, employing Seamless Morphology, see the other papers of Starosta in Singh & Starosta (2003).

all types of lexical relations obtaining in the traditional areas we have been calling ‘derivation, inflection, compounding, etc.’ in the old-time morphological tradition. Ways of avoiding redundancies may still be worked out in this theory.

4. What do Language Acquisition (LA) and Language Processing (LP) say about these two theories?

4.1 Language acquisition

Morphological development studies inform us that children initially produce words that seem to lack any internal structure. The words consist of a single root morpheme. Later, they acquire complex words and treat them on a case-by-case basis as though they are simple forms. Much later, they reflect in their use of certain affixes, e.g. English plural /-z/ or past /-d/, on irregular forms that they have acquired some knowledge of the varying forms of a given word that they have previously encountered. And at a later stage, they learn to distinguish the irregular forms as not being the same in form as the so-called regular ones. Perhaps, we have been so much influenced into thinking that the inception of affix recognition and its application is an absolute sign that the child has acquired the rule of complex word formation, e.g. V root + /-d/ = V, [+past] or N root + /-z/ = N, [+plural]. It may equally be plausible that the English child speaker may have formed *goed or *foots, on the basis of the sets of words, as is represented in a WFS for (regular) past forms and (regular) plural forms, respectively. While the end result in either theory is the same, we cannot determine which approach the child uses. The question is whether the child uses the projection principle, knowing the morphemes and their corresponding features to compose these new complex forms, or on the basis of analogy with other existing forms and word relations she is already familiar with, follows exactly the same correspondence.

What we can be sure of is that children, from whatever language background, are constrained in their initial use of words and even in their coining of words by certain principles, two of which are (i) transparency of meaning and (ii) simplicity of form (Clark, 1993, p. 15). Take for instance, Annie, age 2.8, who named the kind of egg she has been used to eating for breakfast as *circle egg*, because as it is cooked ‘over-easy’ she could dunk her pieces of bread in the soft yolk. Obviously, she knows what a ‘circle’ is and that a fried egg can show a yellow circle, if the yoke is not broken. Both principles are reflected in the creation of this compound. It is perhaps relatively much simpler to process this compound than to learn to say ‘over-easy’ or ‘sunnyside up’. Note also that the constituents of this child’s compound word appear to manifest to a large extent their individual features of meaning.

It is usually reported that during the first stages of LA, children produce words, whether they be full or abbreviated. The focus of attention in the process

is in matching forms with meaning and, generally, both may be short of the adult's corresponding forms and meanings. Unmarked forms are produced preferably over marked ones. Thus, for a Tagalog child in the beginning stage, speaking about 'having eaten' or 'having drunk', she uses *ka?in* or *inom* bereft of any voice affix. Once inflected and/or derived forms are acquired, some forms appear prior to other forms. In a previous study of Tagalog voice/focus affix acquisition, the objective affix *-in* and *i-* are acquired earlier than the active affix *-um-* or *m-(pag)* (Galang, 1982). Here, the function of the affix more than its position in the word is considered more revealing of the learner's preference among the various voice affixes. However, in other studies, it has also been shown that suffixes are learned earlier than prefixes. It may be asked then how this order of acquisition affects the principle of headedness, if it ever does.

In second LA, it is reported that the learner is biased towards natural rules (Karpf & Dringel-Techt, 1995, p. 139), such as conversion or zero derivation in the beginning stages, followed by compounding, which also occurs early in first LA. Following these early morphological processes, affixation without any internal morphonological modifications occurring simultaneously is next. It begins with the most common affixes in the given language, and for nominals in English, for instance, *-er* 'agent; instrument of V', *-ize* 'to make into/put in N', and *-ness* 'state of being A'. The less familiar affixes are learned much later in the development.

In sum, we see from the reports on these LA studies that: (1) the child follows the stages of learning simple forms first before complex forms of words; (2) having been exposed to new forms, she is able to deduce complex forms with similar patterns and use the same pattern(s) to create new forms, i.e. forms she has never heard before; (3) the young speakers are largely influenced in their morphological development by several factors, some stated above, but without informing the psycholinguist analysts with the precise result as to whether the learners are storing morphemes and affixes and none of the complex words, be they inflected or derived.

4.2 Language processing

From Bertinetto's (1995) review of some available psycholinguistic and neurolinguistic research results on morphological processing relevant to the question of compositionality and non-compositionality in the mental lexicon, he says that the evidence gathered is far from allowing conclusive statements. He emphasizes that the cut between these two approaches 'may be relative to any specific language type, rather than fixed for all languages' (1995, p. 11). While there is evidence from the experiments that irregular or even semi-irregular forms of verbs in English and Serbo-Croatian are stored, other sources even state that high-frequency regularly inflected forms are directly accessed (Stemberger & MacWhinney, 1988). In comparison to compositionality, he

suggests that there may also be an alternative strategy of ‘analogical processes’ available to the speakers (as likewise suggested earlier). And even when the same result is reached by either rules or analogy, what matters is the cognitive path along which this result is obtained. Between rule and analogy, he favors the latter because ‘it is more compatible with the sort of fuzziness’ often observed in actual linguistic phenomena (1995, p. 25).

From some recent psycholinguistic research (Libben, 2005, pp. 389-391) on morphological processing, it appears that accessing whole word forms and their constituents (in multimorphemic words) termed POST-LEXICAL DECOMPOSITION has an advantage over the alternative view termed PRE-LEXICAL DECOMPOSITION. In the latter, the constituents of a complex form are computationally scanned and isolated in much the same way as words are isolated individually in hearing or seeing a sentence. In the former, when a complex word form is accessed from the mental lexicon, both the whole word form and its constituent morphemes are activated, resulting in a positive semantic priming.¹⁴

5. Conclusion

It is perhaps not surprising that we do not find direct evidence from either LA or LP as to which of the two theories discussed here is to be recommended not only for its adequacy but also for its plausibility and naturalness. The Morpheme-based theory may be seen as the more elegant and attractive approach for the relationship it establishes between morphology and syntax in terms of being constrained by the same principles. What I find problematic with the theory is the content of the lexicon. If it consists only of morphemes and affixes, and all complex forms are constructed via projection rules, this places a heavy burden on the speaker in both production and comprehension. Granted that the selection of the ‘head’ constituent is already given, barring non-concatenative issues, the assignment of the necessary features for all heads that are in themselves complex may be too cumbersome to follow through individually. As it is, it already requires the speaker to keep close track of the specific categories, subcategorization features, including morphosyntactic and morphonological features, plus any diacritics that may be needed, of all roots, stems, and affixes she has in her mental lexicon. Moreover she has to be adept in following the details of word formation as prescribed for each and every complex form — inflected, derived, or compounded — that she produces in her utterance. LA reminds us of the factors that influence morphological acquisition and this account seems to have disregarded them. The process requires several instantiations of their occurrence in full words beforehand before any generalization or abstraction can take place to make these affixes and stems be listed in the lexicon and be used automatically in word construction. Why can it

¹⁴ The reader is encouraged to refer to other references on these two views.

not be said that once a complex form has been constructed, it becomes part of the speaker's repertoire of lexical items?

In contrast, it appears that retrieving or accessing, rather than constructing, whole complex words that are listed in the speaker's mental lexicon is significantly simpler and involves less of a mental computation for herself. As for new complex words she has not heard before, the WFSs she can associate them with are available for analogical processing. And if this fails, then the new words may be added to her list of vocabulary as such. Likewise, the speaker is also free to create new words on the basis of the WFSs she has already established.

However, one issue I am not convinced of in the Seamless Morphology theory is the lack of generalization that may be stated for two or more sets of forms showing exactly the same relation to another complex set of forms. Likewise, its rejection of the existence of 'paradigms' does not, in my judgment, render Bender's proposal of 'reciprocal paradigm strategies' unsatisfactory. On the contrary, it should be given some serious consideration because the inclusion of this type of word strategies not only makes the morphology component more economical in its statement of WFSs, but more importantly, it captures a natural set of inflectional relations in given languages.

For these reasons, I think that it will be worthwhile to try the Word-based morphology with some modifications on the rigid constraints laid out. The question that remains valid to ask of any morphological theory is whether it is based on the cognitive path that the speakers take in forming complex words.

REFERENCES

- Aronoff, M. (1976). *Word formation in generative grammar*. Cambridge, Mass.: The MIT Press.
- Aronoff, M. (1994). *Morphology by itself, stems and inflectional classes*. Cambridge, Mass.: The MIT Press.
- Bender, B. W. (2000). Paradigms as rules. In V. De Guzman & B.W. Bender (Eds.), *Grammatical analysis (morphology, syntax and semantics): Studies in honor of Stanley Starosta* (pp. 14-29). Oceanic Linguistics Special Publication No. 29. Honolulu: University of Hawai'i Press.
- Bender, B. W. (2003). A perfect strategy for Latin. In R. Singh & S. Starosta (Eds.), *Explorations in seamless morphology* (pp. 301-327). New Delhi: SAGE Publications.
- Bertinetto, P. M. (1995). Compositionality and non-compositionality in morphology. In W. U. Dressler & C. Burani (Eds.), *Crossdisciplinary approaches to morphology* (pp. 9-36). Wien: Österreichischen Akademie der Wissenschaften.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

DE GUZMAN

- Clark, E. V. (1993). *The lexicon in acquisition*. Cambridge: Cambridge University Press.
- De Guzman, V. P. (1978). A case for nonphonological constraints on nasal substitution. *Oceanic Linguistics* 17(2), 87-106.
- De Guzman, V. P. (2005). Aspects of compounding in Tagalog. In H.-C. Liao & C. Rubino (Eds.), *Current issues in Philippine linguistics and anthropology: Parangal kay Lawrence A. Reid* (pp. 282-305). Linguistic Society of the Philippines Special Monograph Issue. Manila: Summer Institute of Linguistics and Linguistic Society of the Philippines.
- Di Sciullo, A. M. & Williams, E. (1987). *On the definition of word*. Cambridge, Mass.: The MIT Press.
- Ford, A., Singh, R. & Martohardjono, G. (1997). *Pace Pānini, towards a word-based theory of morphology*. New York: Peter Lang.
- Galang, R. G. (1982). The acquisition of Tagalog verb morphology: Linguistic and cognitive factors. *Philippine Journal of Linguistics* 13 (2), 1-15.
- Haspelmath, M. (2002). *Understanding morphology*. London: Arnold.
- Karpf, A. & Dringel-Techt, E. (1995). The role of morphological naturalness in second language development. In W.U. Dressler & C. Burani (Eds.), *Crossdisciplinary approaches to morphology* (pp. 131-147). Wien: Österreichischen Akademie der Wissenschaften.
- Libben, G. (2004). Psycholinguistics: the study of language processing. In W.O'Grady & J. Archibald (Eds.), *Contemporary linguistic analysis*, fifth edition (pp. 377-402). Toronto: Pearson Education Canada, Inc.
- Lieber, R. (1990). *On the organization of the lexicon*. New York: Garland Publishing.
- Lieber, R. (1992). *Deconstructing morphology*. Chicago: The University of Chicago Press.
- Schachter, P. & Otanes, F.T. (1972). *Tagalog reference grammar*. Los Angeles: University of California Press.
- Selkirk, E. O. (1982). *The syntax of words*. Cambridge, Mass.: The MIT Press.
- Singh, R. & Starosta, S. (Eds.). (2003). *Explorations in seamless morphology*. New Delhi: SAGE Publications.
- Singh, R. & Dasgupta, P. (2003). On so-called compounds. In R. Singh & S. Starosta (Eds.), *Explorations in seamless morphology* (pp. 77-89). New Delhi: SAGE Publications.
- Starosta, S. (1988). *The case for lexicase*. London: Pinter Publishers.
- Starosta, S. (2003). Do compounds have internal structure? A seamless analysis. In R. Singh & S. Starosta (Eds.), *Explorations in seamless morphology* (pp. 116-147). New Delhi: SAGE Publications.
- Stemberger, J. P. & MacWhinney, B. (1988). Are inflected forms stored in the lexicon? In M. Hammond & M. Noonan (Eds.), *Theoretical morphology, approaches in modern linguistics* (pp. 101-116). San Diego: Academic Press.