# Appendix B

# Discovering Words in Contexts — Emergence of Subpatterns

## B.0 Introductory Notes

This appendix supplements discussions in previous chapters by investigating the nature of words, asking what they are, and where they come from. Discussions that follow will propose a radical reinterpretation of words, thereby revising the traditional conception of them.

The proposed interpretation of words stems from two specifiable sources: one is Rumelhart and McClelland's (1986) idea of **wickelphones**, which is discussed in Section B.1. Another is J. Hill's **classification through word use** hypothesis/model, which is discussed in some detail in Section B.4.4.

## B.1 Words Are Context-sensitive Units

Most theories of language syntax assume that words are units that are combined by certain rules. Rules so required are, by definition, rules of grammar. This view is somewhat artifactual. For example, as I have argued in Appendix A, if selectional properties are incorporated into subcategorization frames, no independent system of rules is necessary. All of these troubles, I argue, stem from the definition of words.

### B.1.1 Remarks on the notion *word*

What are words? This is clearly one of the fundamental questions that have driven a tremendous amount of psycholinguistic research. From the research made so far, I can gain much evidence to show that words are basic units of language, and are even basic units of mind. But there is a bias. As far as I can tell, little research was conducted on their syntax, despite the fact that much research has been focused on their sound and meaning structures. This seems to evidence that most psychologists, let alone linguists, fail to recognize that words may have syntax.

Words are mental units, *par excellence*, but I take a rather phenomenological point of view. Indeed, I propose that, by equating words with subpatterns, words are something "abstracted" from surface formations (or more exactly a single stream of surface forms), and that this is the way that children discover words in speech streams that they are exposed to.

If words are mental, this does not mean that they lack abstract properties that stem from a "perceptual" basis. The perceptual basis of words do not necessarily mean (mainly acoustic) images that their phonetics carry and (mainly visual) images that their meanings carry. Another kind of perceptual property could be a psychological and neurological effect in the memory system to which co-occurrence of words is relevant. Words are not processed one by one. They are processed as parts of a larger "chunk" that a special kind of segmentation operates on. This dynamic relation between parts and whole clearly necessitates the dualism in the representation of words: both words and the whole that they comprise are represented. Thus, a good theory of stream segmentation should play a crucial role in the theory of language acquisition. In such a theory, syntactic contexts are themselves units, of arbitrary lengths.

In the proposed framework, words are more than words in the traditional sense. Words are subpatterns which have internal syntactic structures of their own. Words are **context-sensitive units**, *par excellence*. Take *see* for example. In PMA, the word is equated with a subpattern, *S* ***see*** *O*, in which its lexical syntax is specified by aid of pattern glues, *S* and *O* (roughly for its subject and object), as well as its lexical semantics and phonology are specified by the content of anchor ***see***. The role of glues is not only syntactic but also semantic (and even pragmatic), in that they encode "selectional" restrictions as well as co-occurrence restrictions. This is because glues like *S* and *O* have semantics (and even phonology) of their own. Thus, glues add "relational" informations to "substantial" informations specified in anchors like ***see***.

This conception of words sharply contrasts the traditional conception of them, according to which words are combined by means of "rules" or "templates" extrinsically provided. In this view, *see* is not more than (and not less than) unstructured *see.*

In PMA, words are so structured that they can combine **by themselves**. Words combine to form a phrase, a sentence, or even a larger unit. Words combine like molecules do. This strongly suggests that composition of patterns (e.g., sentences) from sets of subpatterns (e.g., words) is best characterized as **self-organization**. This also manifests a strong need for an **emergent theory of linguistic structure**.

### B.1.2 Some relevant details of wickelphonology

From the PMA perspective, what makes such an emergent theory of linguistic structure possible is nothing but **redundant, distributed representation** of linguistic units. Only under this condition could phrases, sentences, and larger units "emerge" exactly in the modern sense of the term.[1] But it is necessary to substantiate the claim.

Context-sensitivity of words is what PMA makes use of in its description. This is also a property that positively conditions the emergence of syntax, since context-sensitive representations have redundancy. To make this point clearer, turn now to some important details of the proposed context-sensitive encoding.

The context-sensitive encoding of words is inspired by the idea of **wickelphones** that Rumelhart and McClelland (1986) employed to tell their connectionist models to represent phonemes. Wickelphones are based on the insight of Wickelgren (1969) who proposed "trigrams" to solve "the problem of serial order in behavior" raised by Karl Lashley (1951). Because wickelphonology is little known in the linguistic community (even in that of cognitive linguistics), its details deserve a short discussion.

In most familiar versions of phonology, phonological objects comprise units, or segments, called phonemes, putting aside the possibility that phonemes are bundles of features. Phonemes are idealized units of speech in that they stand alone and need to be combined by certain effort. A word is usually represented as a "sequence" or "chain" of such phonemes. Thus, /kæt/ (for *cat*) for example is represented as follows, where + stands for the operator of (con)catenation.

(1) /#kæt#/ = /#/+/k/+/æ/+/t/+/#/

Phonological theory which assumes this kind of representation is sometimes called (con)catenative.

Wickelphonology is not concatenative. In this theory, phonological content of a word is conceived of as a set of context-sensitive allophones which overlap. Thus, the phonology of *cat* is characterized in terms of the following set of three wickelphones.

(2) /#**kæt**#/ = {#**k**æ, k**æ**t, æ**t**#}

Here, Rumelhart and McClelland's original notation $\{{}_{\#}\text{k}_{\text{æ}}, {}_{\text{k}}\text{æ}_{\text{t}}, {}_{\text{æ}}\text{t}_{\#}\}$ is replaced for expository purposes. Since this modification has far from trivial consequences, I will return to the "notational matters" in Section B.1.

A note is necessary. It may seem that wickelphones in (2) are operator-free, which Rumelhart and McClelland suggest, but this is clearly false. It is necessary to make explicit operator $\mathfrak{S}$ of superposition such that:

(3) $W_1 = \mathfrak{S}(U_1)$,
where $W_1$ = #**kæt**# and $U_1$ = {#**k**æt, k**æ**t, æ**t**#}.

Here, $\mathfrak{S}$ defines $W_i$ as **superposition** of units in $U_i$. Clearly, superposition assumes synchronization of wickelphones along time.

It is easy to see that generation of $W_1$, defined as superposition of three units in $U_1$, is encoded as follows:

| | | | | | | |
|---|---|---|---|---|---|---|
| (4) | 0. | # | **k** | æ | **t** | # |
| | 1. | # | **k** | æ | | |
| | 2. | | k | æ | t | |
| | 3. | | | æ | **t** | # |

For more consistency, it is probably necessary to distinguish **actualization** from **potential** of wickelphone sets. The correspondence of a word to a set of wickelphones is like the correspondence of a completed puzzle to a set of puzzle pieces. A set of wickelphones, e.g., $U_1$ = {#**k**æ, k**æ**t, æ**t**#} is not a word any more than a set of puzzle pieces is itself a completed puzzle. A wickelphone set only defines a certain potential for a word, which cannot be realized without additional constructive effort.[2]

### B.1.3 Generalizing wickelphones

As represented in (2), wickelphones generally have the form of $u\boldsymbol{w}v$ where $u$ and $v$ denote immediate left- and right-hand contexts of $\boldsymbol{w}$. Consequently, $u\boldsymbol{w}v$ can only immediately postcede $xu\boldsymbol{w}$ and immediately precede $\boldsymbol{w}vy$ ($x$ and $y$ varying case by case).

Notice that this context-sensitivity is the very property that makes wickelphones free from any template according to which absolute ordering is determined. In this light, wickelphones are rather based on "allophones" than "phonemes". More exactly, every wickelphone is a micro template, and a set of them is a micro system to seek simultaneous satisfaction of local relative positioning specified by $u$ __ $v$.

Also crucially important is that the wickelphonological representation for /#kæt#/ can be identified as "activation" of three wickelphones, at a certain intermediate layer, between a lower-level layer where units represented in a "distributed" fashion in terms of "wickelfeatures", and a higher-level layer, where such units are integrated. See McClelland and Rumelhart (1986) for details.

There is no difficulty in defining subpatterns by relaxing conditions on wickelphones explained above. It goes as follows: wickelphones like #**k**æt, k**æ**t, æ**t**# encode positions by their **overlaps under adjacency**. Note that adjacency is too severe for purposes of syntax and it should be relaxed. For illustration, confirm that given unit $w = ux\boldsymbol{w}yv$, then $w$, $u$ and $\boldsymbol{w}$ and $\boldsymbol{w}$ and $v$ are adjacent, with **precedence** preserved, iff $x = y = Ø$. It is necessary and sufficient to allow $x$ and $y$ to be overt in units of the form $ux\boldsymbol{w}yv$, thereby giving them more flexibility in co-occurrence with units "…$\boldsymbol{u}$…" and "…$\boldsymbol{v}$…" so that adjacency is replaced by (**local enough**) **precedence**. For terminological purposes, I say that in $w = ux\boldsymbol{w}yv$, (i) $\boldsymbol{w}$ is its **anchor**, (ii) *u* and *v* are **glues**, with "...***u***..." and "...***v***..." their **targets**; and (iii) *x* and *y* are **margins** that serve as metavariables.

Thus, the following modifications to wickelphones should suffice to obtaining words as structured units from which complex units such as phrases and sentences can be constructed.

(5) In units of the form *uxwyv*,

A. *u* and *v* may be "schematic" (e.g., *S* and O in *S* ***love*** O), rather than "itemic" units (e.g., *Sam* and *everyone* in *Sam* **love** *everyone*),[3]

B. *x* and *y* in *uxwyv* are allowed to be overt, but:

i. For sufficient **locality**, they may not have too much content, and

ii. For **uniqueness**, they do not contain "fake targets" to yield such structures as ...$u[_x$ ...***u***...]***w***... and ...*u*...***w***$[_y$ ...***v***...]*v*...

For exposition, consider *S* ***invite*** *O*. It this subpattern, glues *S* and O work very much like /k/ and /t/ work in /kæt/. A difference is that, while /kæt/ is contained in **/kæt/**, but not contained in **/bændit/**, *S* ***has*** O is contained in many forms like ***Bob usually has a lot of work***, ***Bob has work***. There are two reason. One is that /kæt/ lacks freedom in specification, or schematicity, to match /bæn.../. This restriction is relaxed by rendering /kæt/ to /𝔒æ𝔆/, where 𝔒 and 𝔆 encode the onset and coda of a syllable. In effect, /𝔒æ𝔆/ matches /bæn.../ if 𝔒 = /b/ and 𝔆 = /n/. Another reason is that *u*, *v* and ***w*** need not be adjacent if precedence among them is concerned.

My interpretation of this fact is that wickelphones are a special case of a more general context-sensitive schema of the form *uxwyv*, where *x* and *y* are parameters for adjacency. Thus, in patterns of the form ***uxwyv*** (e.g., ***Bob*** *usually* ***has*** *a lot of* ***work***, and ***Bob has work***), it is reasonable to say that "***u*** precedes ***w***" (but ***u*** is not adjacent to ***w***). Also, it is also reasonable to say that "***w*** follows ***u***" (but ***u*** is not adjacent to ***w***). For expository purposes, I introduce the distinction between **precedence** and **postcedence.**[4]

### B.1.4 Relevance to the learnability of words

Note that (5)A expresses a condition on glues (*u* and *v* in ***uxwyv***) while (5)B expresses two conditions on *x* and *y*, called margins. Since (5)A is already discussed in Appendix A, my discussion below is concentrated on issues related to (5)B, though I will return to them in Section B.1.

It may seem that conditions in (5)B are extraneous, but this does not mean that they are not essential. The converse is true. In fact, it is very reasonable to assume that a condition on locality, or a ban on too much margin, on the one hand, and a condition uniqueness, or a ban on too much complexity, on the other, both determine the optimal range of dependency between *u* and ***w***, on the one hand, and ***w*** and *v*, on the other.

The locality condition, on which margins *x* and *y* are constrained not to be too long, has to do with inherent limit on human memory, and therefore with the learnability of language syntax. In this regard, it seem reasonable to suppose that crucial task of children, as language learners, is to ignore *x* and *y* by stripping them off from a string ***uxwyv***, where only ***u***, ***v***, and ***w*** are interdependent.

The simplicity condition, on which *x* and *y* are constrained not to be too complex, should have a similar processing root, but it is inadequate to articulate it here because of lack of knowledge.[5]

Clearly, both conditions have important connections to Elman's (1993) emphasis on "starting small" in learning of language (syntax). His results from the connectionist simulation strongly suggest that the basic structure of language learning must be as follows. It begins by finding and absorbing basic dependencies of the form ***u***...***w***...*v* (out of stream *uxwyv*). In his connectionist experiment, Elman succeeds, I construe, in simulating the effect of locality by imposing a restriction short-term memory to process ***uxwyv***. At the first stages of learning, networks are under severe memory restriction, and this restriction is gradually relaxed (appealing to a technique called "gating"). This simulates children's "maturation" in memory and processing ability. Under the initial severe restriction on memory, connectionist networks are unable to find and learn co-occurrence dependencies among ***u***, ***v*** and ***w*** if ***x*** and ***y*** are too long and too complex. Elman argues that this "external" restriction is a necessary condition for successful language learning which is so complex to include multiply center-embedded relative clauses.

### B.1.5 Current status of wickelphones

It is ironical to note that Rumelhart and McClelland's former reliance on wickelphones was a great compromise; they had to encode phonological units on behalf of their connectionist models which were unable to encode them by themselves. Their networks were incapable of "finding structure in time" (Elman 1990). The situation was changed when a new architecture, called "recurrent network", was introduced by Elman (1990), based on Jordan's seminal work (1986*a*, *b*).[6] This was an important innovation. With this architecture, modelers of connectionist networks no longer need to handcraft encoding and decoding networks that interface the kernel network with its "environment". Data given to connectionist networks are no longer wickelphones-like units. Rather, they are "sequences of units", just like humans, supposedly, have access to linguistic data. Recurrent networks learn structure in time by themselves.

Is this the end of wickelphones? I think not. Note first that what this really means is not that interface networks for purposes of encoding and decoding disappear. If something disappears, it is the **distinction of central processing networks** (like pattern associator of Rumelhart and McClelland 1986) **from peripheral processing networks**. This is one of the best characteristics of connectionist models, which show only "weak" modularity. In really complex, nonhybrid connectionist models, interfaces in rigorous sense do not exist; they exist only when viewed from an algorithmic perspective. So, it is better to think that encoding/decoding networks are made implicit by being "incorporated", rather than dispensed with, into the main network.

On this ground, I suspect that even in (simple) recurrent networks, temporally differentiated, context-sensitive units are effective. They are not dispensed with; they are only made implicit. So, it is not unreasonable to think that wickelphone-like context-sensitive units like *vxuyw* are implicitly generated inside the networks, and made use of to represent temporally differentiated abstract structure.

## B.2 How Much Does Notation Matter?

One of the most important contributions that pattern matching analysis can make is its introduction of co-occurrence matrix. This claim requires qualifications, since, in a sense, it is (merely) a notational matter. But, as S.G. Krantz, a distinguished mathematician, points out, "[g]ood notation is extremely important, sometimes as important as a theorem" (1997: 71). I claim that the role of co-occurrence matrix is far from trivial, and PMA could say nothing interesting about language syntax without it. To substantiate my claim, I now turn for a while to the question of how the proposed notational system provides insight into phonology, paying the debt to wickelphonology.

### B.2.1 Notes on the notational convention

Perhaps, the most simple and straightforward way of encoding **(syntactic) context** is to extend the already familiar notation for rewrite rules of the form:

(6) $A \rightarrow Z/X\ \_\ Y$

Note that (6) can be interpreted as an instruction: "replace *A* by *Z* if it occurs between *X* and *Y*".

Reinterpreting this notational system, it is reasonable to encode *Bill*, *hates*, and *gates* in (7) in terms of structured units in (8).

(7) *Bill hates gates.*

(8) i. *Bill/__ hates gates*
ii. *hates/Bill __ gates*
iii. *gates/Bill hates __*

Generalizing this point, it is reasonable to encode the syntactic relationships among subject (= *S*), verb (= *V*), and object (= *O*), as follows:

(9) *S V O*

(10) i. *S/_ V O*
ii. *V/S _ O*
iii. *O/S V _*

Units like *S/ _ V O*, *V/ S _ O*, and *O/S V _* are contextuated units

The notational convention described above is fairly reasonable and well established, but it is far from optimal. Note that symbol " / " is a makeshift device to encode the context $\kappa(u)$ for an identified unit $u$. This means that representation of a unit u is simply a pair $<\zeta(u), \kappa(u)>$, where $\zeta(u)$ and $\kappa(u)$ denote the (lexical) content of $u$ and the (local) context for $u$. Incidentally, $\zeta(u)$ and $\kappa(u)$ correspond to literals "*A*" and "*X* __ *Y*" in *A/X* __ *Y*.

Aware of this fact, it is reasonable to write as follows, where *A/X* _ *Y* is replaced by *XAY*:

(11) ***S V O***

(12) i. ***S*** *V O*
ii. *S* ***V*** *O*
iii. *S V* ***O***

In this notation, context/content differentiation is encoded by gradation rather than separation using symbols " / " and "__".

### B.2.2 Note on typographical variants

For typographical reasons, I will write ***S***–*V*–*O*, *S*–***V***–*O*, and *S*–*V*–***O***, and $\underline{S}$ *V O*, *S* $\underline{V}$ *O*, and *S V* $\underline{O}$, and S 𝔙 𝔒, 𝔖 V 𝔒, and 𝔖 𝔙 O, in place of ***S*** *V O*, *S* ***V*** *O*, and *S V* ***O*** in (12). Of course, they are all typographical variants which all have the same meaning.

### B.2.3 Advantages of the proposed notational system

Based on the notational convention proposed above, the wickelphonology of /#kæt#/, for example, could be simplified and made easy to understand. Consider the following:

(4)

| | | | | | |
|---|---|---|---|---|---|
| 0. | # | **k** | æ | **t** | # |
| 1. | # | **k** | æ | | |
| 2. | | k | æ | t | |
| 3. | | | æ | **t** | # |

As usual, decomposition is carried out in terms of diagonalization. Composition is carried out in terms of column-wise vertical unification. This claims pattern 0 = #**kæt**# is decomposed into subpatterns 1, 2, and 3, on the one hand, and the subpatterns are composed to base pattern 0, on the other.

In (4) and other matrices, it is reasonable to think that boundary symbol # serves as a **mode changer** such that any occurrence of # causes an imaginary parser to shifts mode the mode of morpho(phono)logy, where #*ab*# is interpreted as #*ab*#, instead of #...*a*...*b*...# which is an adequate interpretation on the scale of syntax, thereby imposing the adjacency condition on *a* and *b*.

Note that (4) consists of an "itemic", token-based encoding of #**kæt**#. A schematic, type-based encoding for it is given in (13), where *C* and *V* encode consonant (cluster) and vowel (cluster), respectively:

| | | | | | | |
|---|---|---|---|---|---|---|
| (13) | 0. | # | **k** | **æ** | **t** | # |
| | 1. | # | **k** | *V* | | # |
| | 2. | # | *C* | **æ** | *C* | # |
| | 3. | # | | *V* | **t** | # |

This is a rough conversion from (12), and it does not reflect the syllable structure of #**kæt**#. So, it is reasonable to give the following to implicitly encode the syllable structure missing in the matrix above.

| | | | | | | |
|---|---|---|---|---|---|---|
| (14) | 0. | # | **k** | **æ** | **t** | # |
| | 1. | # | **k** | *V* | *C* | # |
| | 2. | # | *C* | **æ** | *C* | # |
| | 3. | # | *C* | *V* | **t** | # |

This is far from enough. Encoding of syllable is implicit. Thus, more psychological reality can be attained by revising the last matrix as follows, where $\mathfrak{O}$, $\mathfrak{N}$ , and $\mathfrak{C}$ encode onset, nucleus, and coda, respectively, to form a syllable.

| | | | | | | |
|---|---|---|---|---|---|---|
| (15) | 0. | # | **k** | **æ** | **t** | # |
| | 1. | # | **k** | $\mathfrak{N}$ | $\mathfrak{C}$ | # |
| | 2. | # | $\mathfrak{O}$ | **æ** | $\mathfrak{C}$ | # |
| | 3. | # | $\mathfrak{O}$ | $\mathfrak{N}$ | **t** | # |

It is assumed that (i) for any segment *X*, if $X = \mathfrak{N}$ then $X = V^n$ (vowel cluster consisting of *n* vowels); and (ii) for any segment *X*, if $X = \mathfrak{O}$ or $\mathfrak{C}$ then $X = C^n$ (consonant cluster consisting of *n* vowels).

### B.2.4 Reflections on wickelphones

The schematicity in context-sensitive units introduced so far is exactly what Rumelhart and McClelland's wickelphones were criticized for lacking, at least overtly. See Halwes and Jenkins (1971), Pinker and Prince (1988) for more information. It is not clear, however, what happens covertly, for it is very likely that, if interpreted connectionistically, neural units on the hidden layer(s) would create schemas for wickelphones in multi-layer networks.

As discussed briefly above, Rumelhart and McClelland (1986) made use of wickelphonology, understood here as a dual system of wickelphones, as localist units, and wickelfeatures, as distributed units, to make their central processing network, called pattern associator, interface with its environment.

Wickelphones are so-called because, as Rumelhart and McClelland acknowledge, basic insights are due to Wickelgren (1969). He proposed trigrams like /kæt/, which were later dubbed wickelphones, to solve the "problem of serial order in behavior" in the sense of Karl Lashley (1951). The problem is this: associationist theory in general is incapable of handling the "sequential" aspect of behavior, in which certain units of action are "delayed" by others. This is because association in networks embodies no mechanism of temporal differentiation, or control. Basic insight of Wickelgren's is that such problem could be, at least partly, solved by attributing **context-sensitivity** to units of action such as speech. Roughly, if a unit of action (e.g., of speech) consist of (abstract) sequence #123# (e.g., #kæt#) where # stands for either start or end signal, its sequential characteristics is encoded by decomposing it into trigrams #12, 123, 23#, each of which retains the partial, relative sequential order of the whole.

Wickelgren's and Rumelhart and McClelland's treatment of context-sensitivity is not general enough. The most important feature of context-sensitivity is, in my view, rather whether units like them are defined by making use of **overlapping**, with or without adjacency. Thus, there is no a priori reason to limit context-sensitivity to triples. Indeed, there are many other more realistic "modes" of segmentation, one of which is {#12, 23#} (where 1, 2, 3, and 23 are, roughly, unidentifiable as onset, nucleus, coda, and rhyme of a syllable). For purposes of syntactic description, adjacency is not a necessary feature. Thus, it is reasonable to extend the notion of wickelphones by replacing "absolute position" in sequences by "relativized" position in them.

### B.2.5 Comparison with patterns in syntax

For a better understanding, compare the morpho(phono)logical encoding in (15) with (16), which is an encoding on the scale of syntax. This comparison would facilitate seeing that $\mathfrak{R}$ in (15) plays the same role as *V* and *P* in (16), where *X*, *Y*, *Z* are noun phrases of appropriate classes, respectively:

| | | | | | | |
|---|---|---|---|---|---|---|
| (16) | 0. | ***X*** | ***V*** | ***Y*** | ***P*** | ***Z*** |
| | 1. | ***X*** | *V* | *(O)* | | |
| | 2. | *S* | ***V*** | *O* | | |
| | 3. | *S* | *V* | ***Y*** | | |
| | 4. | | | *S* | ***P*** | *O* |
| | 5. | | | *S* | *P* | ***Z*** |

An example of this configuration is given below:

| | | | | | | |
|---|---|---|---|---|---|---|
| (17) | 0. | ***Bill*** | ***invited*** | ***Carol*** | ***to*** | ***his party*** |
| | 1. | ***Bill*** | *V* | *(O)* | | |
| | 2. | *S* | ***invited*** | *O* | | |
| | 3. | S | *V* | ***Carol*** | | |
| | 4. | | | *S* | ***to*** | *O* |
| | 5. | | | *S* | *P* | ***his party*** |

As this matrix claims, *V* = *S* ***invited*** *O* and *P* = *S* ***to*** *O* are patterns *S*–***V***–*O* and *S*–***P***–*O*, respectively, assuming symbols in normal face are syntactic glues.

Even this rough comparison strongly suggests that pattern composition and decomposition play a fundamental role both in syntax and morpho(phono)logy. I note, however, a fundamental difference between syntactic and morpho(phono)logical patterns. As I have pointed out in Appendix A, the basic encoding of (grammatical) subject, verb, and their object, given in (18), is more exactly like in (19), where relativity of positioning is encoded explicitly.

(18) *S* = ***S*** *V (O)*
*V* = *S* ***V*** *O*
*O* = *(S) V* ***O***

(19) *S* = ***S*** … *V* … *(O)*
*V* = *S* … ***V*** … *O*
*O* = *(S)* … *V* … ***O***

In other words, glues in syntax, *S*, *V*, *O* are “stretchable”. But this is not the case with morpho(phono)logical patterns, most of which contain a more restricted sort of glues like $\mathfrak{O}$, $\mathfrak{N}$, $\mathfrak{C}$ (for onset, nucleus, coda), as in (15).[7]

## B.3 Multiple Segmentation Theory, with Special Reference to the Way that Words Are Learned

This section addresses the question of how words are learned.

The problem of how words are represented in the mind (and the brain) has to do with the way they are learned. In this regard, if, as PMA suggests, words are subpatterns that have sufficiently local but schematic syntax of their own in that they "remember" the contexts in which they occur, it follows that:

(20) A. There is a simple, mechanical procedure by which a stream of sounds is converted into subpatterns, thereby enabling encoding of such local syntax; and
B. Such procedure is available for all children who know no language yet.

It should be emphasized that A and B, in conjunction, form a problem of **emergence**. Note that words are acquired by those who know no words yet: in other words, words emerge. This property is not trivial even if the emergence of subpatterns is an automatic consequence of the associative nature of the brain, emphasized by connectionist theories.

The problem of emergence immediately affirms that the nature of words is not exclusively substantive. By substance of words, I mean semantics and phonology of them. It is reasonable to assume that words bear semantics which has a conceptual basis, and phonology which has an articulatory and acoustic basis.

If words have syntax of their own, as PMA claims, then essential properties of words are irreducible to material bases: words are special not because they have semantics paired with phonology (by symbolization), but because they have extensions in time. In fact, there must exist another, more abstract basis, contrasted to substantive bases such as conceptual and phonological ones. I call it a "relational" basis. Only on this abstract basis can words extensions in time.

Where does the relational property comes from? I claim, in conformity with connectionist results such as Elman (1990, *et seq.*), that the relational property is an automatic consequence of the fact that language learners discover words by picking them out of a speech stream. Related to this, it should be stressed that there is a **part/whole relation** between words and (local) speech streams, and in this connection, the acquisition of words possessing syntax of their own is a good example of "embodiment" of experiences at more "abstract" levels than what is learned is substantive, as semantic and phonological structures are. This possibility of highly abstract learning is exactly what Elman's results strongly suggest.

To sum, all words are discovered as parts of certain wholes, and if local segmentations of a speech stream do not serve as such wholes, it would be impossible for words to "remember" their contexts.

### B.3.1 Words as semological units

PMA posits that words are what children find out of a speech stream. This is a claim motivated by criteria A and B in (20). The next question is, What could be a plausible model for the procedure in question that finds out words? The specific model for word discovery that PMA assumes is follows:

(21) i. Speech stream is segmented into smaller, local streams by (mechanistic) memory limitations of the parser (i.e., language learner). This is the segmentation on the largest scale.
ii. Given local streams so defined, which are roughly sentences, a search for smaller units takes place. This search introduces another segmentation on a smaller scale.
ii′. Parsed units are not necessarily continuous.
iii. Search for such units can be nested.

Furthermore, the parsing model sketched here assumes that words can be best characterized as **semological segments**, comparable to phonological segments that are phonological units, in order to capture well-known parallelisms, such as:

(22) i. words (as semological units) ≈ segments (as phonological units)
ii. phrases ≈ syllables
iii. sentences ≈ (prosodic) words
vi. sentences (or clauses) ≈ words (as semological units)

What is strongly suggested is that identification of words results from **semological segmentation** in a way similar to **phonological segmentation**, on the one hand, and that words as semological units as far as sentences, which are parts of a larger speech stream, are semological units, on the other.

Crucially, PMA assumes that phonological and semological segmentations take place in different dimensions. So, as a consequence, the two kinds of parsing processes are autonomous to each other, no matter how strongly they interact with each other.

For illustration, consider a hypothetical segmentation of a (local) speech stream that begins at time *t*. For simplicity, I assume that the local speech stream that begins at *t* is a sentence. Searches for semological and phonological units run in parallel, as illustrated in Figure B.1.

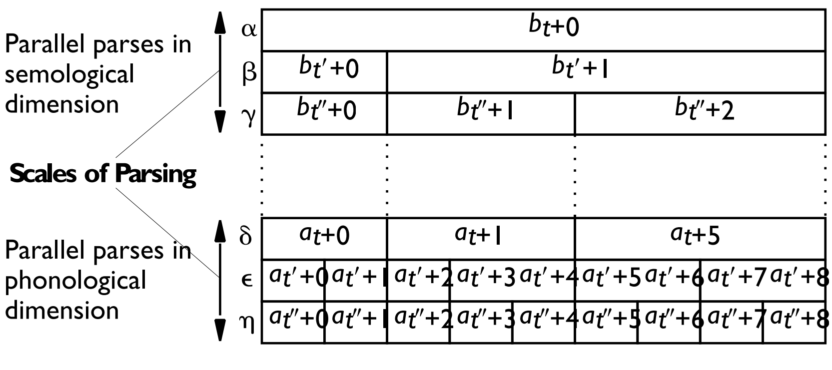


Figure B.1

Here, *a*'s denote phonological segments, and *b*'s denote semological segments.

Effects of nesting are indicated by the number of ′ put on *t*.

The first three scales are semological: α is the scale for sentences; β is the scale for phrases; γ is the scale for words, i.e., semological segments. The last three scales are phonological: δ is the scale for prosodic words. ϵ is the scale for syllables. η is the scale for phonological segments.

The diagram in Figure B.1 illustrates that at least six parallel segmentations begin at time *t*. All of them parse interdependent but autonomously. They are autonomous to each other because upper scale/level process do not "wait for" completion of lower level processes. In fact, higher ones even "guide" lower ones. So, it is inadequate to interpret this diagram as indicating an organization in either exclusively "bottom up" or "top down" fashion.

Higher level processes are "synthetic", and conversely lower level ones are "analytic". Incidentally, one can be aware of a correlation: the higher the level of a process, the more easily one is conscious of it.

Of greatest importance is that the relation between words, $b_{t''+0}$, $b_{t''+1}$, $b_{t''+2}$, on the γ-scale, and phonological chunks $a_{t+0}$, $a_{t+1}$, $a_{t+2}$, on the δ-scale, is one of **correspondence** rather than of analytic/synthetic. Differently put, "words as phonological chunks" and "words as semological units" are different in kind, not in degree.

One should be very aware of this point; otherwise, this admits one to claim, very carelessly, that the nature of language is "symbolic" in that any class of units semantic and phonological poles. This is exactly what Langacker (1987, 1991*a*, *b*) claims for under the rubric of **symbolic view of language**, to which I now turn.

### B.3.2 In what way are units of language symbolic?

It is agreed that language is symbolic in larger, semological scales. But the claim for Langacker's symbolic view of language confuses two different aspects of words: one is that "words as semological units" always have phonological imports. Another is that "words as phonological chunks" happen to bear meaning.

This is because words are essentially "amphibious". Phonologically, they are "chunks" of segments (or of syllables or of certain intermediate units like moras). This means that words are **complex** from a phonological perspective. Semantically, however, words are "units" of meaning (if not ultimate). This means that words are **simplex**.

This contrast can be captured by appealing to the following diagram, where three words, $b_{t''+0}$, $b_{t''+1}$, and $b_{t''+2}$, on the γ-scale in Figure B.1, are represented as sections, [0,1], [1,2], and [2,3], of the line which intersects the phonological and semological planes.

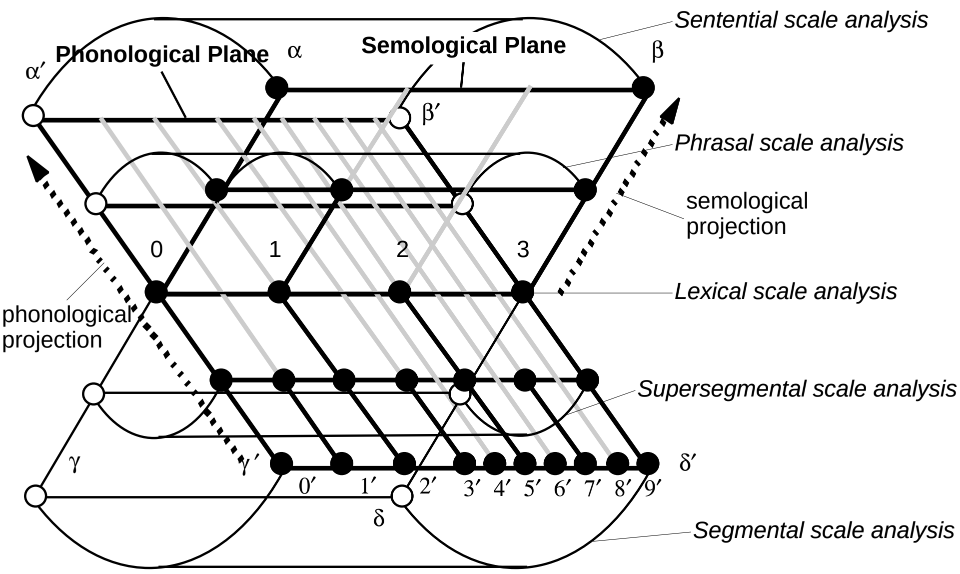


Figure B.2

Here, projections of [0, 1] and [1, 3] match two occurrences of "phrase", on the one hand, and [α, β], as projection of [0, 1, 2, 3], matches a "sentence", on the other.

It should be noted that [α, β] and [α′, β′] have different contents: [α, β] represents a sentence as a string of words in that [α, β] = [0, 1, 2, 3], whereas [α′, β′] represents a sentence as a string of phonological segments in that [α′, β′] = projection of [0′, 1′, ... , 9′] (= [γ′, δ′]).

This diagram in Figure B.2 shows that semological constructs are inside the local region that are circumscribed by sections [0, 3], [α, β], and [α′, β′], whereas phonological constructs in the local region that are determined by line sections [0, 3], [γ, δ], and [γ′, δ′].

My point is this: what Langacker defends under the rubric of "symbolic view of language" (as a "symbolic alternative" to generative linguistics) is adequate only for the upper half region of the structure that the diagram illustrates. Semological constructs "happen" to have phonological contents only there. For this region, it is safe to say that all constructs on the plane closed by [0, 3] and [α′, β′] "symbolize" corresponding constructs on the plane closed by [0, 3] and [α, β] exactly because [0, 1] = $u_1$, [1, 2] = $u_2$ and [2, 3] = $u_2$ are "words as (optimal) symbolic units".

However, asymmetry does exist. The dimension on which there is [0, 1, ...] (= [$t$+0, $t$+1,...] for a certain time $t$) is the "bottom" of symbolic relationship, and such relationship no longer holds in the lower region. So, the lower half of the semological plane must be blank; otherwise, words are no longer "units" of meaning, and therefore there must be certain smaller units of meaning. This is an obvious contraction to the fact. If this were true, the meaning of *dog* could be divided into meanings of /d/, /ɔ/, and /g/ (or of /d/ and /ɔg/ or of /dɔ/ and /g/).

Essential aspects of phonological constructs are determined by the lower region, determined by line sections 0–3, γ–δ, and γ′–δ′. With γ′–δ′ at the bottom line, segments on the bottom line project up to α′–β′.

Note incidentally that the scale of inflectional morphemes (e.g., /z/ in *dogs*) is the upper region, since this kind of forms, though not free-standing words, are lexical units which take words (or phrases) as their "arguments". Such morphemes are "fossils" of phrasal units, and reside in superlexical levels.

### B.3.3 Two ways of contexts affecting parsing

It is natural, then, that context affects words in a dual way. First, words receive a number of phonological (or phonetic) ways of modification. This will be referred to as phonological accommodation. This takes place on the plane closed by 0–3 and α′–β′. Likewise, words receive a number of syntactic and semantic ways of modification. This will be referred to as semological accommodation. This takes place on the plane closed by 0–3 and α′–β′.

### B.3.4 How timing affects parsing

Furthermore, the diagram in Figure B.2 captures the "musical" nature of language, especially rhythmic one. Occurrences of units on different scales are governed by "different timings". Note that timings, $t$, $t'$, $t''$, and $t'''$ are all different timings on different scales. Witness the difference, for example, between $t$ and $t'$ in the following:

(23)

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $t$: | | | | | | | | | | $t'$: | | | |
| $t$+1: | $\boldsymbol{a_1}$ | | | | | | | | | | | | |
| $t$+2: | $a_1$ | $\boldsymbol{a_2}$ | | | | | | | | $t'$+1: | $\boldsymbol{b_1}$ | | |
| $t$+3: | $a_1$ | $a_2$ | $\boldsymbol{a_3}$ | | | | | | | | $b_1$ | | |
| $t$+4: | $a_1$ | $a_2$ | $a_3$ | $\boldsymbol{a_4}$ | | | | | | | $b_1$ | | |
| $t$+5: | $a_1$ | $a_2$ | $a_3$ | $a_4$ | $\boldsymbol{a_5}$ | | | | | $t'$+2: | $b_1$ | $\boldsymbol{b_2}$ | |
| $t$+6: | $a_1$ | $a_2$ | $a_3$ | $a_4$ | $a_5$ | $\boldsymbol{a_6}$ | | | | | $b_1$ | $b_2$ | |
| $t$+7: | $a_1$ | $a_2$ | $a_3$ | $a_4$ | $a_5$ | $a_6$ | $\boldsymbol{a_7}$ | | | | $b_1$ | $b_2$ | |
| $t$+8: | $a_1$ | $a_2$ | $a_3$ | $a_4$ | $a_5$ | $a_6$ | $a_7$ | $\boldsymbol{a_8}$ | | | $b_1$ | $b_2$ | |
| $t$+9: | $a_1$ | $a_2$ | $a_3$ | $a_4$ | $a_5$ | $a_6$ | $a_7$ | $a_8$ | $\boldsymbol{a_9}$ | $t'$+3: | $b_1$ | $b_2$ | $\boldsymbol{b_3}$ |

I guess this rhythmical nature is crucial to the syntactic dimension of language. Clearly, phonology cannot exist without it, and syntax, interpreted here as semology, in analogy with phonology, cannot exist, either.[8]

## B.4 Pattern Matching Account of Language and its Acquisition in Relation to Other Cognitive Accounts

In this section, I will offer comparisons of the PMA account of language and its acquisition with some accounts provided by other cognitive approaches.

### B.4.1 Relevance to Langacker's "usage-based" model

It is not so surprising even if PMA conception of words has a certain resemblance to Langacker's notion of "usage-based, network model" of language structure. Indeed, the PMA conception of words will be compatible with the usage-based view of them if usage of a word means a complex structure that contains its syntactic information. To see this, let me cite a relevant passage from Langacker (1988: 272).

> For a nonlexical example [of usage-based units], consider the analysis of a phoneme as a complex category [...]. Let us suppose that the phoneme /a/ (in a particular language) occurs only in the syllables /a/, /pa/, /ta/, and /ka/. Each preceding consonant induces some phonetic modification of /a/, however minor it might be. This phoneme consequently has at least four allophones, namely [a], [$^{p}$a], [$^{t}$a], and [$^{k}$a] (where [$^{p}$a] is the allophone induced by /p/, and so on.). The allophone [a], which stands alone as a syllable, is plausibly regarded as the basic allophone and equated with the category prototype; the others then function as context-induced extension from this prototype, as diagrammed in Figure 5. Moreover, speakers may well extract a schema to represent the commonality of the various allophones. Shown as [$^{x}$a] in the diagram, the schema is neutral as to whether and how the basic vocalism of [a] is modified by a preceding consonant.

Below is my reproduction of his Figure 5, with slight modification.

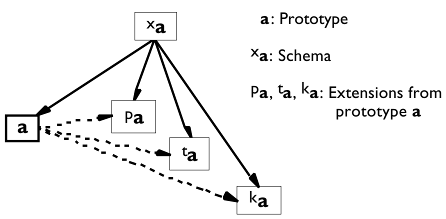


Figure B.3

Though he makes no mention of either Rumelhart and McClelland's (1986) connectionist model or Wickelgren (1969), Langacker's idea of allophonic units [a], [$^{p}$a], [$^{t}$a] and [$^{k}$a] has a striking resemblance to the notion of wickelphones. In this regard, it is expected that PMA shares basic insights with Langacker's usage-based conception of grammar, despite a few disagreements, to which I now turn.

### B.4.2 Disagreement with Langacker's usage-based model

For more information, let me cite some more of the passage. Langacker continues as follows:

... The inventory of conventional units comprising the grammar of a language is structured, in the sense that some units function as components of others. Often, in fact, a unit owes its specific character to a more inclusive structure in which it occurs (at least initially). For instance, the notation in Figure 5 should not be interpreted as implying that [$^{p}$a], [$^{t}$a], and [$^{k}$a] are free-standing units that can occur independently; they occur only in the context of the respective syllabic units [[p][$^{p}$a]], [[t][$^{t}$a]], and [[k][$^{k}$a]], since the preceding consonant induces their distinguishing phonetic properties. The categorizing relationship [a] --> [$^{p}$a] of Figure 5 is thus more adequately represented in Figure 6(a), which shows the extended variant in the environment that determines and supports it.

Figure below is my reproduction of Langacker's Figure 6(a), with slight modification.

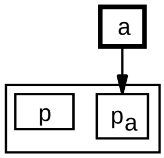

Figure B.4

Relating to the network model diagrammed in Figure B.3, Langacker (1988: 272) remarks:

The network model therefore reconciles two classic views on the nature of a phoneme; that which analyzes a phoneme as set of allophones; and that which treats it as a unitary but necessarily abstract entity (i.e., a schema). The nonreductionist character of the analysis also accords with traditional phonemic descriptions, which provide a list of allophones for each phoneme, and state that environments that condition each derived or nonbasic allophone. The necessity for a nonreductionist account is readily apparent in this domain, since a speaker's phonetic ability does not reside in any single structure. A speaker who fully controls the phonetics of his language is able to pronounce not only the basic allophone, but also the full array of derived allophones, properly distributed. Each implies an articulatory (also an auditory) routine that a speaker masters as part of his internal representation of the linguistic system. These units are properly included in the grammar of a language, for they constitute one facet of a speaker's grasp of linguistic convention.

It is clear that Langacker's conception has far from trivial similarity with the notion of subpatterns. However, some notes are in order.

First, but less importantly, the view of words as subpatterns is directly inspired by wickelphones introduced by Rumelhart and McClelland (1986) rather than Langacker's usage-based, emergent units. So, it is unreasonable to expect that they are the same.

Second, but more importantly, Langacker's theory of "network of extensions from a prototype" is arguably unnecessary, if it is sufficient. PMA does not appeal, unlike Langacker, to a dually based system that alternately relies on the notion of **extension from a prototype** (or **prototypes**), in some cases, and the notion of **schematization**, in other cases. I will show that the entire system is driven by schematization only, thereby dispensing with the notion of extensions from prototype as merely one of its effects.

To make crucial points clearer, note first that it must be accounted for why [a] serves as a "prototype" of [$^{p}$a], [$^{t}$a], [$^{k}$a], thereby making them its "extensions". But is there any reason why [$^{p}$a], for example, is not the prototype instead of [a]? If the reason is that [a] is a prototype, it is clearly circular. This reveals that it is arbitrary to define [a] as a prototype from which extensions to [$^{p}$a], [$^{t}$a], [$^{k}$a] are made.

More specifically, I, like Lakoff (1987), do not take "prototypes" to be something real, while I fully accept "prototype effects". This means that I do not try to take some instances to be primary. Rather, I take so-called prototypes to be emergent properties to be explained rather than something that are given and one must start with.

I ask, "How do prototypes emerge?" rather than "What are such and such instances extensions of?", or "Where do such and such instances extended from?"

A plausible mechanism to replace the notion of successive extensions from a prototype is **clustering effect**. I will discuss relevant details of this notion in Section B.4 relating to J. Hill's **classification through word use** model/hypothesis (Arbib, *et al.* 1987).

In conformity with Hill's model, my account of the prototype effect goes like this. Given segments [ø**a**], [p**a**], [t**a**], [k**a**], which are my reinterpretation of Langacker's [a], [$^{p}$a], [$^{t}$a], [$^{k}$a]. If [**a**] and [p**a**] are generalized through classification through use, we then have schema [{ø, p}**a**], or [$C_1$**a**] where $C_1$ = {ø, p}. Note incidentally, [{ø, p}**a**] behaves as a phonemic unit relative to [ø**a**] and [p**a**]. Prototype effects are observed only when contexts for [**a**] are more typical than contexts for [p**a**], e.g., [**papa**]. If [**a**], [p**a**] and [t**a**] are generalized, we then have schema /{{ø, p}, t}**a**/. If [**a**], [p**a**], [t**a**] and [k**a**] are generalized, we then have schema /{{{ø, p}, t}, k}**a**/. Thus, it is clear that phoneme /**a**/ is best characterized as the most general schema /{{{ø, p}, t}, k}**a**/ in this system.

For better understanding, a diagram is provided to illustrate how schemas emerge in the system.

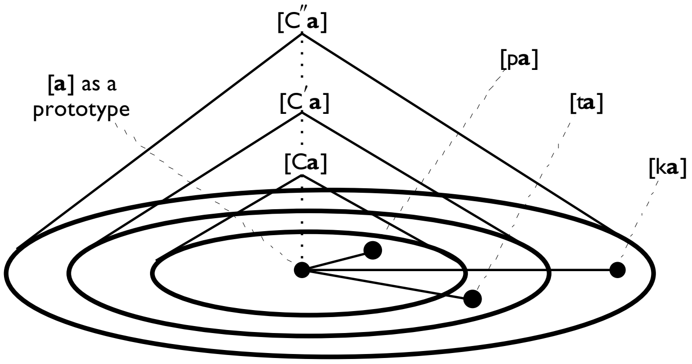


Figure B.5

In this diagram, all circles are on "plane of instances", where extensions are indicated by links from [**a**] at the center. Notably, [**a**] is the prototype because it is the shared center of the projections of all schemas.

Perhaps, structure diagrammed in this figure is what Langacker really intends by his diagram in Figure B.3 above. But even this structure is flawed in some crucial respects.

The basic mechanism in the proposed reinterpretation is **neutralization among instance-level differences** that language learners are always faced with. Simplified greatly, language learners synthesize information of disjunctive nature. The problem is a construction of a simplex representation that holds of both [**a**] and [p**a**]. The most obvious and apparently insightless way of doing this is construct a representation like [{ø, p}**a**]. According to Hill's classification through use model, what children do is exactly this, not more, not less.

Note that it is possible to recognize a number of prototype effects without recognizing such and such "instances" as prototypes. Nothing is necessary than appeal to **relative typicality among instances**. In fact, it is easy to obtain the same prototype effects without assuming any prototypes and extensions from them only if the following relations are true.

(24) i. [t**a**] is more familiar, more "typical" than [k**a**].
ii. [p**a**] is more familiar, more "typical" than [t**a**].
iii. [ø**a**] is more familiar, more "typical" than [p**a**]

But conditions in this set are very weak, since there are many other possible configurations to satisfy these conditions, and for this reason, prototypicality is more complicated than Langacker seems to have in mind. To reveal this, consider a case diagrammed in Figure B.6:

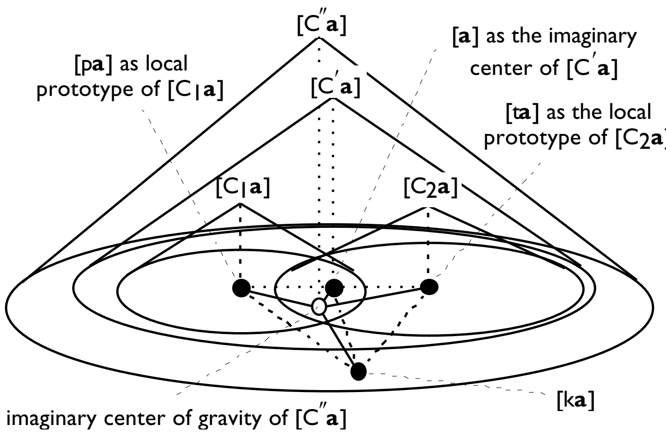


Figure B.6

Note that the structure of the relationship among [**a**], [p**a**], [t**a**], and [k**a**], diagrammed here is compatible with conditions in (24). Nevertheless, it is inadequate to state that [**a**] is the prototype in this structure. This reveals a point of greatest importance, which I make explicit as follows:

(25) Prototypes themselves are mere effects, not causes of prototype effects.[9]

For this reason, prototypes are not real, and extensions from them are an epiphenomenon.

What the structure in Figure B.6 claims is that prototypes shift from one to another as levels of schematicity change, which, in Figure B.6, correlate with the "heights" from the instance plane, where [**a**], [p**a**], [t**a**], and [k**a**] are located. For example, consider the two lowest levels of schematicity. In the domain of [$C_1$**a**], [p**a**] serves as the prototype and [**a**] as an extension of it. Similarly, in the domain of [$C_2$**a**], [t**a**] serves as the prototype and [**a**] as an extension of it. At the level of [C′**a**], however, [**a**] serves as the prototype and [p**a**] and [t**a**] as extensions of it. Furthermore, there is no instance that corresponds to the prototype in the domain [C″**a**].

Why do strange things like this happen? The reason is, again, that prototypes are effects rather than causes, and that prototypes are more adequately characterized as **local centers of gravity**, which may correspond to instances, in certain regions of an *n*-dimensional state space.

Discussions so far suggest the possibility that leads to a crucial point. To be explicit, I claim as follows:

(26) Prototypes need not be real instances.

By holding this, my conception of prototypes is substantially different from Langacker's, and I find this convincing enough to invalidate arguments for a conceptualist position like Langacker's.

### B.4.3 Langacker's constructional schemas reinterpreted

Turn now to another aspect of Langacker's usage-based model. Remember that Langacker appeals to a theoretical construct, called a **constructional schema**, to account for the prototype effect of [a]. In Figure B.4 [= Langacker's Figure 6(a)], repeated here, [a] instantiates a constructional schema [[**p**][$^{p}$a]].

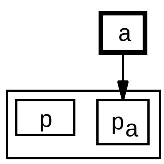

Figure B.4

It is obvious that such constructional schemas are, in some crucial respects, analogous to subpatterns in my sense. Despite this similarity, however, there are still great differences between them, only a few of which I make explicit here.

For reasons discussed in Appendix A, PMA can achieve the same effects as Langacker's constructional schemas by appealing to the following notational scheme:

(27) 0. / **p** **a** /
1. / **p** a /
2. / p **a** /

If this notational system is adequate, then it not only suggests that the descriptive adequacy of Langacker's Figure B.4 is in its being schematic rather than its being image schematic, but also suggests that Langacker's diagram should be replaced by the following:

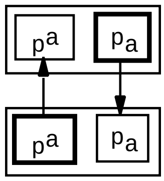


Figure B.7

More generally, the relation of two components in this diagram is a special case of ***C*** = ***AB***, which is a more abstract pattern is decomposed as follows:

(28) 0. ***A*** ***B***
1. ***A*** *B*
2. *A* ***B***

It is possible to diagram this as follows, where ***C*** = ***AB*** is "dissociated" into two subpatterns, $u_1$ = ***A****B* and $u_2$ = *A****B***.

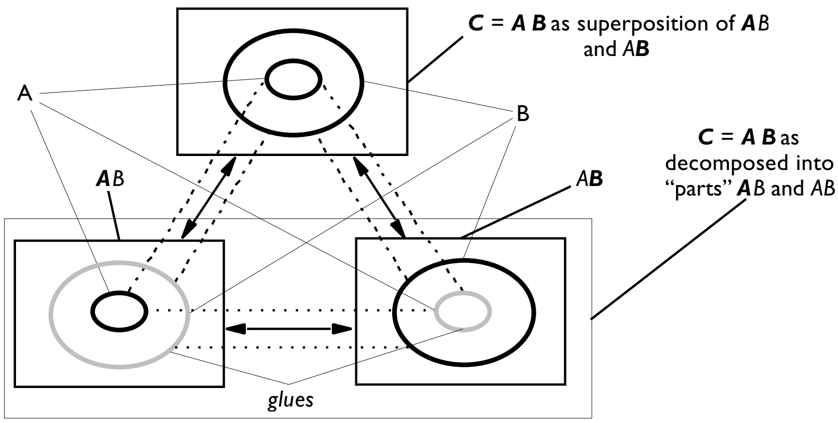


Figure B.8

It is important to note that $u_1$ and $u_2$ are **twins of figure/ground reversal** in that $u_2 = \Re(u_1)$, where $\Re$ is reversal operator.

Relevant here is that the relation between $u_1$ and $u_2$ is analogous to **figure/ground pairs**, provided that boldfaced units correspond to **figure** against specific **ground**. Additionally, the notion of figure/ground pairing has relevance to Langacker's notion of **profile/base relation**. I note that one should be careful in appealing to the profile/base relation. First of all, the profile/base relation is not necessarily conceptual, or even image schematic. Indeed, there is no question that most linguistic units have a conceptual basis, to the extent that image schemas do. But one should be very aware that there is no evidence to claim that word schemas are **exclusively conceptual**. Rather, I suspect that it is not the case as far as such results from Elman (1990, *et seq.*) are concerned.

My interpretation of diagrams in Figure B.8, on the one hand, and of diagrams in Figures B.4 and B.7, on the other, is that they are, first and foremost, **mere graphical representations** of certain abstract properties of activation patterns over neural networks, assuming that neural networks properly characterize the function and structure of the brain/mind. It seems that anything more than this is a forced interpretation.

### B.4.4 Relevance of Hill's *Classification Through Word Use* Hypothesis

Another source of inspiration for pattern composition/decomposition is a special hypothesis that plays an important role in Jane Hill's computational model of language acquisition in the two-year old (see Hill 1982, 1983, 1984; Hill and Arbib 1984; Arbib, Hill, and Conklin 1987). The hypothesis is called **classification through word use hypothesis** (or CWU hypothesis for short). This hypothesis posits that a child learns his language "begin[ing] with as many word classes as he has words, having learned his initial words each as an instantiation of an individual concept" instead of "start[ing] with one large class, *words*, and then proceed to subdivide this class into *noun*, *verb*, *adjective*, and then continue in this way to subdivide again until the correct partitioning is reached for the production and comprehension of adult language" (Arbib, *et al.* 1987: 130).

As I will later discuss, Hill's CWU hypothesis has some important bearings on Elman's series of connectionist results discussed above, on the one hand, and on Langacker's **usage-based model** of language structure (1987, 1988, 1991*a*, *b*), on the other.

The second connection is important in that Hill's model of language acquisition can be seen as a computationally based embodiment of certain basic claims of Langacker's usage-based model, while still contradicting with certain conceptual points. More explicitly, Hill's and Langacker's claims and models of language may go hand in hand as far as the conservative interpretation of the CWU hypothesis is taken, which claims that "it may explain how the child, in interaction with his environment, comes to acquire the class of noun, verb, and adjective during the early stages of language acquisition" (Arbib, *et al.* 1987, p. 131). But the two models should disagree when the CWU hypothesis is radically interpreted so that "the broadest classes—those of *noun*, *verb*, and *adjective*—in their full generality might not be learned at all until encountered in school. This assumes that these classes are not actually necessary in language processing" (p. 131).

Note also that this radical interpretation not only contradicts with the assumption that "the classes *noun*, *verb*, and *adjective* are innate" (p. 131), which is widely held among linguists, whether generative or cognitive. In fact, the radical interpretation is hardly compatible with conceptualist claims such that nouns and verbs are linguistic universals, (merely) because there are **notional/conceptual universals**, which are perhaps formed on the basis that nouns denote "things" and verbs denote their "relations". More specifically, nouns and verbs are (linguistic) universals because they correspond to the classes, which are conceptual universals.

### B.4.5 Cause or correlation: that is a question

Under the strong influence of work of Greenberg (e.g., 1963), linguistics is full of proposals by typologists and functionalists to define notionally word classes such as nouns, verbs, etc, on the one hand, and grammatical classes such as subject and object, on the other. See, for example, Anderson (1989), Dixon (1989), Lyons (1966, 1989). More recently, more cognitively oriented linguists like Langacker make a similar proposal, though from a different perspective. Some of them are really insightful, and I agree that, without deception, they are sure to capture certain interesting properties of human language. Nevertheless, a fundamental question remains unexplained: Are they really explanations in terms of **causal relation**, distinguished from mere **correlation**?

For clarity, let me note a fact. When two independently observable phenomena A and B are given to be explained, there are three possibilities for explanation. First, A is (part of) the cause of B. Second, B is (part of) the cause of A. These two cases are of causal relation. The third possibility is that there is another phenomenon C that is more fundamental and (part of) the cause of both A and B. In this case, the relation of A and B is not causal, but correlational, no matter how intimate it may be. Under this note, let us discuss Langacker's arguments in some detail for a case study.

Langacker argues for notional definition of nouns, by suggesting as follows:

> ... The proposed schematic definition states that a noun **profiles** (i.e., designates) **a region in some domain** [in an "idealized cognitive model" of the world (Lakoff 1987), specifically called "billiard-ball model"], where a **region** is defined abstractly as a **set of interconnected entities**. (Langacker 1991*b*: 15, with emphasis by bold of his own)

Note that a conceptualist position like Langacker's presupposes that classes like nouns, verbs, adjectives, are linguistic universals. But the validity of this presupposition is under attack by the radical interpretation of the CWU hypothesis. It suggests that such so-called universal classes are artifactual. What one is able to discover is only clusters of words, from very small to very large. Allegedly universal lexical classes like nouns, verbs, adjectives are very rough abstractions of them which are ranked as "superordinates" in the taxonomy of language.

So, it is very likely that conceptualists like Langacker's are victims of the **first order isomorphism fallacy** in the sense of Kugler, Turvey, and Shaw (1982), in which one fails to distinguish two different kinds of properties that figure into one's description. One kind are properties inherent to objects to be described by certain means (call this essential properties), and another is properties accidentally inherent to devices (even mental ones like metaphor) used to describe the objects (call this imposed properties). With this distinction, conceptualists like Langacker are very likely to mistake the taxonomy of concepts for the taxonomy of words.

As I argued in detail above, it is dangerous to start description of language by presuming that meaning, understood as something conceivable, is an essential property of language, let alone of its units like words. By this, I do not allude that the characterization is incorrect. Rather, it is dangerous to draw any conclusions under this setting. It is at best successful in capturing only one of a lot of (even abstract) facets of a system called language.

Based on this (still controversial) interpretation, I am inclined to reject strong conceptualist positions like Langacker's. They are too strong. Basically, it is merely a conceptually disguised version of Chomskian innatism, irrespective of whether it is more "cognitive" or not. Note that it does not matter whether it is possible or not to define nouns and verbs on a conceptual basis; it really matters what roles they play on account of language. I will disagree if anyone suggests that the definition is important because it is apparently "cognitive". My point is that the term *cognitive* is not a synonym of *conceptually based*; and there are a number of other ways of being cognitive. To me, such a definition is a generalization for generalization's sake, and perhaps nonsense under the name of generality.

Relevant here is that in a position like Langacker's, higher-level classification of nouns and verbs, for example, is conceptually necessitated by a conceptual route to it. This is what the radical interpretation of the CWU hypothesis rejects. Langacker's position should finally turn out to be circular in the end. Note that even if a word behaves as a noun or a verb because it symbolizes a thing or a (reflexive) relation, there remains a crucial question unsolved: How do children come to know syntactic properties necessary to symbolize such concepts?

### B.4.6 Starting small in language learning

Some connectionist results could help here. Banning overgeneralizations is another form of the problem that children face in language acquisition. Elman (1990, *et seq.*) showed that a class of networks, called simple recurrent networks, are able to learn this same kind of co-occurrence of words (without external, "conceptual" semantics) in this way.[10] His results contain a wealth of evidence to make us convinced that grammar is learnable in this protective, incremental generalization procedure. Roughly put, knowledge of words starts with itemic, token-based encodings; and it seems that such encodings are "unlocked" to generalizations. In this crucial sense, learning grammar is equated with the induction problem of reaching schematic encodings from itemic encodings.

In my interpretation, the most significant aspect of Elman's (1993) results is that language syntax is learnable/discoverable as a hidden, abstract structure among units with no phonological or conceptual contents. Syntax in the abstract sense seems to be carried out by exclusively relying on statistics in the distribution of units. Putting aside the problem of how children "pick out" words, syntax can be characterized as a classification of words according to their relative positioning within a stream of a limited length. Controversially, I posit that Elman's' results confirm a weaker version of the notorious thesis of **autonomous syntax** in that meaning is not necessary for syntax to be learned.

### B.4.7 Semantics internal and external to language

To make clearer certain terminologically and conceptually complicated matters, it is urgent to note that the terms *semantics* and *meaning* are used very differently in different traditions. What some linguists in one school of linguistics call meaning of linguistic expression is not the same in other schools. More specifically, what logically oriented linguists call meaning is not meaning at all in the sense of cognitive linguists or more generally functionally oriented linguists. To avoid terminological complication, I assume the distinction between **semantics internal and external to syntax**, or **internal and external semantics** for short, to be defined below.

What I mean by external semantics is exactly the kind of semantics that linguists like Langacker refer to when they say linguistic units are "symbolic". In this conception, grammar of a language is a collection of "conventional units" which can be equated with "symbolic structure" that are an association of phonetic/phonological formation and conceptual formation, or so-called "conceptualization".

Under the provisional distinction, it is safer to note that what Elman's simulation demonstrated to be unnecessary in language learning is an external kind of semantics. Recurrent connectionist networks need not "know" this kind of semantics at all; what plays a crucial role is internal semantics alone.

What I mean by internal semantics is a "combinatoric" kind of semantics, or roughly a "type" semantics, which could be best captured by the tradition known as "categorial grammars", introduced by Ajdukiewicz (1935), and developed by Lambek (1958, 1959, 1961) and authors contributing to Buszkowski, *et al.*, *eds.* (1988) and Oehrle, *et al.*, *eds.* (1988), or utilized by Montague (1970, 1973) and some of his successors. Along this line of research is Kanazawa's (1994, 1996) proof of an interesting theorem which asserts that the class of "classical" categorial grammars are learnable from positive evidence alone, which I find is one of the most important results in language learning theory since Gold (1967), along with Shinohara's (1990*a*, *b*) result "which says that placing any finite bound on the number of rules used in context-sensitive grammars results in a learnable class" (Kanazawa 1998: 3).[11]

This link to categorial grammars may seem pointless, but is in fact essentially important because it is appropriate to understand the syntactic/semantic property of words in terms of "functors" in the sense of combinatorial categorial grammar. In fact, Elman (1995) discussed his results along this line of interpretation, as follows:

> The status of words in a system of the sort described [as simple recurrent networks] is very different [from the status of data as something to be passively processed by rules in classical architecture]: Words are not the *objects* of processing as much as they are inputs which *drive* the processor in a more direct manner. As Wiles and Bloesch (1992) suggests, it is more useful to understand inputs to networks of this sort as *operators* rather than *operand*. (1995: 207).

The most relevant aspect of words is their status of functors in the sense of categorial grammar fashion.

In $x = z/y$, $z$ is a functor to operate on a word of type $y$ to make it of type $z$. For example, English indefinite article, *a* (and *an*) can be characterized as a functor such that *a* has internal semantics [*i*/*c*], where *c* stands for a "class", and *i* for an "instance" of class *c*, respectively. Thus, it is possible to illustrate, with a few revisions, how internal meaning of *a boy* is derived by relying on such notation as follows:

| | | | | | |
|---|---|---|---|---|---|
| (29) | a. | *a* | *boy* | ⇒ | *a boy* |
| | b. | [*i*/*c*] | [*c*] | | [*i*] |

The basic mechanism assumed is "cancellation", according to which [*c*], combined with [*i*/*c*], "cancels off" [/*c*] to form [*i*]. Thus, *a* of category [*i*/*c*] serves as "instance deriver" in that it combines with a word *boy* of category [*c*] to form a complex unit *a boy* of category [*i*].

Incidentally, it is not surprising to see that this aspect can be well characterized in PMA. It is sufficient to appeal to the following composition/decomposition.

| | | | | |
|---|---|---|---|---|
| (30) | 0. | ***a*** | ***boy*** | [*i*] |
| | 1. | ***a*** | *N* | [*i*/*c*] |
| | 2. | *(D)* | ***boy*** | [*c*] |

In a sense, pattern matching method is more advanced in that it provides a chance to see the following. In ***a*** *N*, ***a*** bears [*i*] and *N* bears [/*c*]. In *(D) boy*, *(D)* bears [*i*/*i*] and ***boy*** bears [*c*]. There are two concurrent but different operations, one by *a* on *boy*, in which [*i*/*c*] is imposed on *boy*; and another by *boy* on *a*, though this results in no overt effect.

### B.4.8 Role of external semantics in learning syntax

Returning to my main point of syntax learning, the next question should take the following form:

(31) Given, hypothetically, children learn syntax in the same way as networks like Elman's simple recurrent networks learn syntax, what is then the exact role of external semantics?

Since networks learn syntax by finding out internal semantics alone, or a set of conditions on co-occurrence called **dependency**, my best guess is that external (conceptual) semantics at best "enhances" the learning of syntax, rather than being a prerequisite, since it is proved that learning syntax may be achieved without making reference to the (external) meanings of words.

This interpretation might be taken to run counter to certain basic claims in cognitive linguistics. Indeed, it contradicts claims by conceptualists like Langacker who are disappointed with this kind of (external) meaning-free definition of linguistic units like words, and then seek for alternative definition:

> ... It is thus a basic doctrine of modern linguistics that nouns and verbs cannot be notionally defined. ... In fact, I would argue that universal categories of such fundamental grammatical significance should be expected to have a conceptual basis. (1991*b*: 15)

I would like to remark that it will be a mistake if one took, without thinking deeply, "having a conceptual basis" to mean that it is "conceptually defined".

But, taken alternatively, this provides an opportunity to drastically rethink what really makes language cognitive. As I have noted above, what makes language cognitive is not its being on a conceptual basis. Music, for instance, is far from conceptually based, but it is a deep deception if one contends that it is not cognitive. In my view, what should be abandoned is the contention that syntax, as well as (external) semantics and pragmatics, is necessarily conceptually based. There is no conceptual and factual necessity for such a contention.

As far as I can see, it seem inevitable to accept the weakest form of the notorious **autonomous syntax thesis**. More specifically, learning syntax is equated with learning words, provided learning words is irreducible to learning their meanings. In any case, this forms a strong counter-argument against Langacker's contention (1997) that syntax of language emerges from the conceptual structuring that underlies it.

Instead, words are learned basically on a "distributional" basis. The result is, I contend, an internalization of subpatterns. Through this process, word classes emerge as "natural classes" as a child generalizes syntactic behavior of words that he or she is exposed to. Thus, one may conclude that it is not always true that so-called "maximalist" accounts of language are better than "minimalist" ones, because learning words need not be based on their external meanings, since it can be attained by abstracting surface distribution of words.[12] This line of interpretation is exactly what Peter Bensch (1991) suggests when relating to Elman's work.

> ... Significantly, [Elman's simple recurrent] networks' output very closely follow the predictions of Harris [(1982)]. (Harris is one of the last remaining practitioners of pre-Chomskyan

> structuralism). The Chomskyan revolution was to some extent precipitated by the lack of sufficient computational tools to meet the goals of linguistic structuralism. Chomsky proposed that the structuralist program of inducing general principles from empirical data would never succeed. As part of his revolution, he advocated a research program based on deduction from general principles to empirical data.
>
> With the emergence of computational tools being developed by Elman, structuralism may again become a viable research program. Further support for this conjecture is provided by the continuing problems encountered by linguists attempting to deduce empirical data from base [*sic*] principles. Thus, a connectionist revolution seems to be emerging. And, this revolution may be fittingly called "neo-structuralism."

On the interpretation of Elman's work suggested here, lexical classes can be inferable by "discovery procedure". Thus, instead of hiding myself into an escape hatch of conceptualism, I would like to assume that there is a correlation such that the more selectively interdependent units are with other units, the less autonomous they are.

## B.5 Concluding Remarks

In this Appendix, two matters are treated. First, I have discussed how the notion of context-sensitivity, skillfully utilized in Rumelhart and McClelland (1986), is incorporated into linguistic theory to change the way that the representation of linguistic units are thought about. If pattern matching analysis provides some insights into language syntax, most of them should come from such a change of the view of language. Second, I examined Langacker's cognitive grammar (1987, 1991*a*, *b*) from the perspective of PMA, to indicate, though indirectly, in what respects the proposed framework is different from other "cognitive" approaches despite some superficial similarity. PMA is "cognitive" not because it is an approach in which substantive properties of language are more emphasized than formal properties, but because it is inspired and perhaps compatible with results in connectionist researches such as Elman (1990, *et seq.*), and aims to be a "realistic" and even "naturalistic" theory of language.

To conclude, it should be emphasized that what PMA tries to describe (and account for, if possible) is so-called "formal" properties of language, which most cognitive approaches dismiss, rather than "substantive" properties, which cognitive approaches cherish. In my view, formal and substantive aspects of language are equally important, and it is deceptive to (try to) reduce one into another. This means that language syntax, one of the most important formal properties, should not be reduced to substantive properties like meaning structure and sound structure. So, there is at least in principle no contradiction in saying that PMA is a new, "realistic" form of generative linguistics in that it seriously aims to describe as many facts of language syntax as possible, though never is it another form of Chomskian linguistics which pretends to "explain" (only part of) the facts in terms of a monster notion Universal Grammar.

## Notes

1. See such works as Gleick (1987), Nicholis and Prigogine (1989), Prigogine and Stengers (1984), Mayr (1998) for more details of the notion of emergence.

2. Incidentally, Wickelgren (1969) relates "spoonerism" to a specific kind of failure in the realization process.

3. For the itemic/schematic distinction, see relevant discussions in Chapter 3.

4. I had to coine a rather awkward term *postcede* and *postcedence* partly because I need a term as useful as a pair *precede* and *precedence*. My trouble is that *follow* lacks proper nominalization except *following* (witness awkwardness of *?*followance*).

5. It is very unlikely that these constraints are part of UG in the sense of Chomsky and his followers. But it is totally unclear why both presumably "neurally based" constraints have nothing to do with competence. If language may be constrained by the architecture of the brain, then Universal Grammar is not anything but an Universal Turing machine.

6. Jordan's and Elman's architectures are different in that in Jordan's sequential network, the activation pattern of output layer is copied and buffered in another layer, and it is fed back to the output layer in next cycle. In Elman's simple recurrent network, the activation pattern of the hidden layer is copied and buffered into another layer (called context units) and it is fed back to the hidden layer in the next cycle.

7. It should be noted that metrical features are mostly stretchable.

8. In this regard, I find it almost absurd that the a majority of cognitive linguists claim that syntax of language is "metaphorically based". I have in mind Lakoff's (1987) **spatialization of form hypothesis**, among others. I am aware that this idea is really stimulating in that it continues to stimulate a series of studies, which include such important works as Deane (1992). No matter how seminal the hypothesis is (so was the reality because a number of works followed), I nevertheless think its validity is highly questionable for a variety of reasons. Some of them are discussed in Chapter 1. At any rate, it will be merely a waste of time if the hypothesis proves to be wrong after all. Be it the case, most of the results have to be, sometimes drastically, reinterpreted. I suspect that no truly seminal idea will ever lead one into such a crash of research program.

9. This point is correctly pointed out by Lakoff (1987), but I cannot accept many of his "solutions" to the problem.

10. Of course, it is up to one's interpretation whether one is or is not willing to say that these networks "learn" English syntax. First, but less significantly, what has "learned" syntax is even not a human after all. Some may claim that networks do not learn language syntax, any more than computers have intelligence even when they passed a Turing test. Second, but more significantly, it is not necessarily clear exactly what conditions make language learning successful. Elman's connectionist networks do not speak. It is fairly likely that on this basis, linguists, mostly generative linguists, claim, either disdainfully or unsympathetically, that the connectionist networks do not learn language syntax at all. I believe these two are minor problems, however. What matters is not that everyone in and out of intellectual traditions arrive at one interpretation. It really matters whether there is one interpretation that the majority of researchers, full of experts, can agree on.

11. Both Shinohara's (1990*a*, *b*) and Kanazawa's (1996) results are rigorously computational, but it must be emphasized that their results are far from trivial and are important from the perspective of learnability theory, no matter how non-cognitive they may seem. I remarked this

because cognitive linguists tend to dismiss, or even disdain without understanding, every formal result of this sort. It is really sad to me to witness anyone around me taking such a disdainful attitude.

12. There is, however, a very subtle point. The data presented to networks are prepared according to the intuition of an experimenter who is a subject of conceptualization. So, the problem is more complicated, for it is necessary to be very aware that conceptual structure, which makes it possible for us to use words, is "implicitly encoded" in the data that networks process. But, in any case, it is proved that conceptual structure *need not precede* syntactic structure, since the most straightforward interpretation of Elman's result is that substantive properties such as conceptual semantics and phonology and/or phonetics are not a prerequisite for language learning.