DOCUMENT RESUME

ED 428 566		FL 025 765

AUTHOR	Sypniewski, Bernard Paul
TITLE	Linguistic Units.
PUB DATE	1998-00-00
NOTE	10p.
PUB TYPE	Opinion Papers (120)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Grammar; *Language Patterns; Language Research; *Language Universals; *Linguistic Theory; Research Methodology
IDENTIFIERS	*Applicative Universal Grammar; *Linguistic Units

ABSTRACT

Applicative Universal Grammar (AUG) is a linguistic theory based on combinatorial logic. This paper expands on the notion of "linguistic unit." Linguistic unit, as a notion, is generalized and the linguistic hierarchy supporting all natural languages is shown to be a hierarchy of linguistic units. It is argued that, on the genotype level, natural languages manipulate linguistic units rather than words or sentences. In this paper, the term "linguistic unit" refers to units on all levels of the linguistic hierarchy. AUG says that natural languages are bi-stratal; each natural language has phenotype and genotype strata, the genotype strata being universal to all languages. Units inhabit the genotype level and thus are universal to natural languages. It is proposed that the purpose of the phenotype level is to prepare input for the genotype level, i.e., to prepare "raw" linguistic input to be assembled into units. (Author/MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 428 566

LINGUISTIC UNITS

O.B.O., Inc.

Bernard Paul Sypniewski

syp@jerseycape.com

ABSTRACT

Applicative Universal Grammar (AUG) is a linguistic theory based on combinatorial logic. Shaumyan 1987 This paper will expand the notion of *linguistic unit*, Shaumyan and Sypniewski 1995 (S & S). *Linguistic unit*, as a notion, is generalized and the linguistic hierarchy supporting all natural languages is shown to be a hierarchy of linguistic units. In this paper, I argue that, on the genotype level, natural languages manipulate linguistic units rather than words or sentences.

In S & S, linguistic units were discussed in the context of words. In this paper, I generalize the notion of linguistic unit so that the term refers to units on all levels of the linguistic hierarchy. AUG says that natural languages are bistratal; each natural language has a *phenotype* and *genotype* strata, the genotype strata being universal to all natural languages. Units inhabit the genotype level and, hence, are universal to natural languages. I propose that the purpose of the phenotype level is to prepare input for the genotype level, i.e., to prepare "raw" linguistic input to be assembled into units.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Bernard Sypniewski
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

BEST COPY AVAILABLE

Linguistic Units

A *linguistic unit* (or, simply, a *unit*) is a linguistic entity which is treated as a whole rather than as an accumulation of parts regardless of its internal composition. For example, a *word* is a unit, even though it consists of one or more syllables or phonemes; a *sentence* is a unit, even though it consists of one or more words. A unit has *elements*, i.e., components, which make up the unit. A word, seen as a unit, has syllables or phonemes as elements; the words which make up a sentence are the elements of the sentence unit. Even for a unit which has only one element (a *minimal unit*), the element and the unit are not the same. The unit is always treated as indivisible even if it may be decomposed into smaller units or elements. Units actually have a hierarchical structure. A sentence unit may have other units as elements as well as words. Here is a familiar sample sentence:

(1) *The cat sat on the mat*

In (1), the sentence unit has three elements, which I will parenthesize for convenience:

(2) (*The cat*),

(3) *sat*,

and

(4) (*on* (*the mat*)).

(4) is a unit consisting of two elements: *on* and (5) (*the mat*). (5) is a unit which, in turn, consists of two elements: *the* and *mat*. If we wish, we can carry this expansion further to determine the elements of the individual word units.

In the sense mentioned above, every unit is recursively hierarchical. To coin a phrase, big units have little units. The hierarchical structure is the same on all levels of the linguistic hierarchy: a *nucleus*, the indispensable part of a unit, and zero or more satellites of the nucleus. We can see the following as rules.

Nucleus rules

- the nucleus of a term unit is a term;
- the nucleus of a predicate unit is a predicate;
- the nucleus of a sentence unit is a predicate;
- the nucleus of a word unit is a root.

Satellite rules

- the satellite of a term is a term modifier;
- the satellite of a predicate is a predicate modifier;
- the satellite of a sentence is a term;
- the satellite of a word is an affix.

In the rule lists above, word units are considered graphically rather than phonetically. Similar rules may be constructed for the phonetic representation of a word. I will leave aside the question whether the hierarchy can be extended "upward" (super-sentence units) or "downward" (sub-word units).

Units are the consequence of combination. When an operator is applied to its operand, a unit results which is more complex than the operand[1]. For example, in (2), when *the* is applied to *cat*, the term *the cat* results. *Yellow car* is a more complex unit than either *the* or *cat* because (2) consists of both words combined. The combination of elements resulting from the application of, in this case, **Ott** to **t** results in a single linguistic entity which has a single genotype. *The cat* is type **t**. All units have a single type for the entire unit regardless of the internal complexity of the unit. During genotype parsing, the application of an operator can be seen as an attempt to assign single genotypes to the largest unit possible.

Units relate to each other when they share a *context*. By *context*, I mean the linguistic environment in which a unit exists. A unit may exist in a *minimal context* such as a one word sentence[2]. More often, the context of a unit can be described as the relation between units, for example, the relation between a term and a predicate. In the case of a minimal context, the unit relates to itself. A minimal context is, in a sense, a self-referential context. A single word existing in a minimal context is superposed to a sentence by the minimal context which is, therefore, a superposer.

Units provide natural languages with a significant benefit. The number of unit elements in a context is usually larger than the number of units in the context. When a language manipulates units, a language has less to manipulate which means that the language can manipulate units faster than it can manipulate the elements of units. The genotype of a unit is the result of the operations which produced the unit. On the genotype level, there are three fundamental types: **O** (an *operator*), **t** (a *term*), and **s** (a *sentence* or *s-type*). All operators are one place operators. AUG's general operator notation is in the form *Oxy*. *X* is the *operand* of the operator, i.e., *x* is what *O* operates on. *Y* is the *resultant* of the application of the operator to *x*, i.e., *y* is what is produced. The resultant of a operation is either **t**, **s**, or another function which can be applied further. For example, a predicate function which takes a primary and secondary term and reduces them to a sentence has the type **OtOts**. This notation means that the operator operates on an operand which has type **t** and produces a resultant with type **Ots**, which is another operator. Here is another sentence:

(8) *John bought a watch*

John is the primary term with type **t**, *a watch* (not just *watch*) is the secondary term with type **t**, and *bought* is a predicate with type **OtOts**. Applying *bought* (**OtOts**) to *a watch* (**t**), results in the creation of a new predicate function *bought (a watch)* which has type **Ots**. *Bought (a watch)* is a unit; it has a single type: **Ots**. When it is applied to *John* (**t**), the resultant is (*bought (a watch) (John)*) which is a unit with type **s**. I have used parentheses, even though they are not strictly necessary, to highlight the units in the resultant. Every word or group of words in common parentheses are a unit. The word or group of words in the parentheses are the elements of that unit. So, we have this reduction:

(9) bought	unit type:	**OtOts**
(10) a watch	unit type:	**t**
(11) John	unit type:	**t**
(12) bought (a watch)	unit type:	**Ots**
(13) (bought (a watch) (John))	unit type:	**s**

The process of reduction is, essentially, the process of assembling elements into a unit. The process of reduction is carried out through the repeated application of the predicate to the remaining elements. When there are no remaining uncombined elements, the application ceases and a unit, in this case, a **s-type**, results.

The Relations between Units

Context is the relation between units. A sentence results from the application of the nuclear predicate to its satellite terms. A great benefit supplied by units is the simplification of syntax. A fifty word sentence may only be composed of three or four units. Research shows that the human parser uses units in the parsing of sentences. Ferreira and Henderson (1990) studied the role played by verbs in the parsing of ambiguous sentences. The results of their study indicate that, if possible, a sentence is treated as a whole; if this is not possible then phrase-structure-sized units are used (minimal attachment parsing). Only if an "error" occurs does the human parser break down these units into smaller parts to determine how these parts interact with each other (garden path parsing). In the following quote from their paper, Ferreira and Henderson use the phrase "interactive models" to refer to traditional linguistic theories:

The lack of initial use of verb information by the parser is inconsistent with the general class of interactive models of parsing, which presume that all *sources of information, whether lexical, syntactic, semantic, or discourse-contextual, communicate in an unconstrained fashion to produce the most plausible reading of a sentence at the earliest stages of sentence comprehension ... Our results add to the growing literature ... suggesting that the parser does not operate in this unconstrained fashion. Initially, the parser uses only phrase-structural information to construct a syntactic representation. If an error occurs in this initial analysis (whether the error is signaled by syntactic or semantic/pragmatic anomaly), the parser then uses whatever information is available, including verb information, to come up with a more acceptable analysis. This difference in timing suggests a different architecture for the sentence processing system. In interactive models, the architecture does not clearly distinguish among components responsible for assigning structure at different levels of representation (lexical, syntactic, semantic). In models such as the garden path model, there are distinct modules ... within the language system, each using its own representational vocabulary ... These modules communicate but in a constrained fashion.*

Ferreira and Henderson (1990:565-566, citations omitted)

If we see language as a hierarchy of units, then the parsing described by Ferreira and Henderson uses several levels of the linguistic hierarchy during a parse. Parsing starts at the highest possible level, L_h, the level of complete sentences. Let us assume that a parse on this level begins by assuming that all words in the sentence to be parsed, *S*, have their primary types, i.e., that none of the elements of *S* are superposed. The parser constructs a pattern of genotypes P_g on this assumption. Let us further assume that the parser has available to it previously encountered sentence patterns, $P_1 \ldots P_n$, which have proved from experience to be the patterns of well-formed sentences according to the grammar which the parser uses (French, Mandarin, etc.) The parser then compares P_g to some P_i based on some internal rules, e.g., that P_g and P_i have the same length, i.e., the same number of elements. If $P_g = P_i$, i.e., if each of their elements, taken in order, have the same genotypes, the parser applies the

ERIC

nuclear predicate (main verb) to the term units. When the operation has been completed, an s-type unit has been formed.

If P_g and P_i do not match, the parser has several options. If the difference between P_g and P_i is sufficiently great, the parser may wish to compare P_g with P_{i+1}. In this case, parsing recommences at L_h. If the difference between P_g and P_i is small, the parser drops to L_{h-1}, a level on which the parser can examine the elements of *S* rather than *S* as a whole. The parser tries to resolve the difference between patterns by assigning a secondary type to the differing word *w* in P_g. By doing so, the parser superposes *w*. If $P_g = P_i$ after the superposition, the parser proceeds with the reduction.

P_m is a list of the genotypes of the elements of a unit. The first step in a parse is to determine whether *S* is properly formed by comparing P_g to some P_i as was mentioned above. The next step in the parse is to form the elements of *S* into units. Once accomplished, the parser applies the predicate[3] to the units within *S*. The elements of *S* concern us only when it is necessary to be concerned with them, e.g., when there is an error in the match between P_g and P_i. The elements of *S* have information which we may need. If $P_g = P_i$, we do not need the information which they have even though it is available. If we know that the phrase *the yellow car* has the pattern **Ott Ott t** and we know that that pattern is acceptable and reduces to **t** once the operators are applied, we do not need to know more about the syntactic make up of the phrase. We can treat it as a simple term, even though it has a complex internal structure without investigating its internal structure any further. I take it as a principle that a parser such as I have described uses the minimum amount of syntactic information possible to resolve syntactic anomalies.

Units interact if they are on the same level of the linguistic hierarchy. In a compound sentence, sentences may act as terms if they are superposed to terms <**s qua t**>. Operators have two broad functions regarding units. Operators may produce a resultant which is on the same level as its corresponding operand, e.g., **Ott** or **OOtOtsOtOts**, or an operator may produce a resultant which is on a different level than its operand, e.g. **Ots** or **Ost**. Every unit has a nucleus and zero or more satellites. If term and sentence are treated as zero-place operators Shaumyan (1987:196), the nucleus of every unit is an operator. If the nucleus of a unit U is a zero-place operator, U cannot be reduced to another unit with a type different than its nucleus except by an operator which is not an element of U. If the nucleus of U has a valence greater than zero, U can be reduced to a type with a zero-place operator (either term or sentence) by the application of the nucleus of U to its other elements unless U consists only of a nucleus and modifiers of the nucleus. Thus there are two broad types of units which we may call *active* and *passive*. An *active unit* U_a is one which can be reduced to a zero-place operator by the application of its nucleus to its other elements. A *passive unit* U_p consists of a nucleus and nucleus modifiers only and cannot be reduced to a zero-place operator by the application of its nucleus to its other elements. A passive unit may become active if placed in context with other passive units to which U_p may be applied. I call this *latent operation* if the nucleus of U_p has a valence of zero. Superposition activates latent operation. The term *fish* is the name of a class of animals and is, therefore, a term. Placed in the appropriate context with another term, such as *fish tank*, *fish* becomes superposed to an operator such as a term modifier <**t qua Ott**> or, in other circumstances, such as *I fish for flounder*, a predicate <**t qua Ots**>. Without the latency of operation, it would be difficult to explain why superposition is possible. We can thus see that superposition activates a latent syntactic function while also providing the semantic information available via the primary type of the element. Superposition modifies both the grammatical and lexical meanings of the element. In the case of *fish tank*, the superposing context has changed the grammatical meaning of *fish* from term to term modifier while also causing *fish* to narrow the semantic scope of *tank* (which is what a modifier does).

A unit has the combined grammatical and lexical meanings of its elements. The elements of a unit interact to produce new lexical meanings by modifying the lexical scope of the nucleus of the unit. The grammatical meaning of the unit is either the type of its nucleus, if U_p, or the type produced by applying the nucleus to its satellites, if U_a. A further discussion of the lexical meanings of units is beyond this paper.

---

[1] The most obvious exception to this observation is an application of the identity operator.

[2] It is theoretically possible to have a context without words at all. For example, a semiotic device, such as a shrug, may very well be such a context. I do not consider such contexts in this paper.

[3] The difference between a verb and a predicate is that a predicate is the "complete" verb considered as a unit.

ERIC

FL 025765

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Linguistic Units

Author(s): Bernard Paul Sypniewski

LACUS presentation? ___ yes ✓ no If not, was this paper presented at another conference? ___ yes ✓ no Specify: working paper

Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY ___ Sample ___ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY ___ Sample ___ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY ___ Sample ___ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1 [✓]	Level 2A [ ]	Level 2B [ ]
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: [signature]	Printed Name/Position/Title:	
Organization/Address: DBO Inc, PO Box 513, Woodbine, NJ 08270	Telephone: 609-861-5181	FAX:
	E-Mail Address: bps@jerseycape.com	Date: 7/27/98

ERIC Full Text Provided by ERIC

(over)