

Copyright  
by  
Marie Crevolin Harnisch  
2008

**Investigations of Word Order from a  
Typological Perspective**

**by**

**Marie Crevolin Harnisch, MS**

**Report**

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

**Master of Arts**

**The University of Texas at Austin**

**May 2008**

**Investigations of Word Order from a  
Typological Perspective**

**Approved by  
Supervising Committee:**

---

---

## **Dedication**

To my parents, John and Peggy Harnisch

## **Acknowledgements**

The completion of the present report would not have been possible without the suggestions and insightful comments of I-wen Lai. I am grateful for her help and support, for the hours of intellectually stimulating conversations that allowed me to articulate my ideas, and for her friendship. I am also grateful to the Linguistics Department of the University of Texas at Austin, especially to Pattie Epps and Tony Woodbury.

May 2, 2008

## **Abstract**

# **Investigations of Word Order from a Typological Perspective**

Marie Crevolin Harnisch, MA

The University of Texas at Austin, 2008

Supervisor: Patience L. Epps

This paper, a review of the literature on word order typology, examines in detail a body of work (Comrie 1989; Comrie, Dryer, Gil, Haspelmath 2005; Dryer 1988, 1991, 1992, 2007; Greenberg 1966; Hawkins 1983; Lehmann 1973; Vennemann 1974) that made a major contribution to linguistics by introducing the subfield of typology and the study of word order across the world's languages from a typological perspective. Greenberg's (1966) seminal paper advanced an understanding of cross-linguistic tendencies that had been unknown at the time and which are still being investigated today, especially his three-way typology based on the relative position of V with respect to S and O. Lehmann (1973) and Vennemann (1974) pushed the VO/OV distinction which led to a reanalysis and diminishing of the role of S as an organizing parameter. Two theories, Vennemann's Head-Dependent Theory and Hawkins' Cross-Category Harmony, account for many attested correlation pairs, but neither is as strong as Dryer's Branching Direction Theory in terms of explanatory adequacy, elegance, and adherence to observed cross-linguistic

tendencies. As far as theoretical approaches, we note that generative grammar with its focus on single-language study cannot provide an account of the variations in the world's languages, while the typological approach comes closer to describing universals of language based on empirical data. Finally, I present the idea that investigations of word order from a typological perspective can be successfully undertaken using a functionalist approach within the framework of Optimality Theory.

## Table of Contents

1 Introduction .....	1
1.1 Significance and Background of the Present Study .....	1
2 Questions Central to the Typological Approach to Word Order .....	3
2.1 Determining Basic Word Order .....	4
2.2 Issues of Data Collection .....	6
2.3 Theories of Grammar.....	8
3 Literature on Word Order Typology .....	12
3.1 The Basic Order Typology.....	13
3.2 The VO/OV Typology.....	15
3.2.1 Status of the Subject within VO/OV .....	17
3.2.2 Head-Dependent Theory .....	17
3.3 Argument Against Two-Way Typology.....	21
3.3.1 Cross-Category Harmony .....	22
3.4 Verb and Object Patterners .....	22
3.4.1 Dryer on HDT .....	24
3.4.2 Branching Direction Theory .....	25
3.4.2.1 Adjective-Noun.....	25
3.4.2.2 First Version of Branching Direction Theory.....	26
3.4.2.3 Revisions to BDT: "Revised" and "Alternate" .....	27
3.5 SVO Type .....	29
3.5.1 SVO in the Data .....	30
3.5.2 SVO as VO .....	31
3.6 New Considerations.....	33



4 Further Investigation: Functionalist Constraints within OT.....	34
5 Conclusion.....	36
References .....	38
Vita .....	40

## List of Tables

Table 1:	Greenberg's Basic Order Typology .....	14
Table 2:	Vennemann's Operator/Operand Schema .....	18
Table 3:	Complete List of the Verb- and Object-Patterner Correlation Pairs ... .....	24
Table 4:	Dependent/Head Noncorrelation Pairs .....	24
Table 5:	Correlation Pairs for V-initial, SVO, and V-final languages .....	31
Table 6:	Noncorrelation Pairs for V-initial, SVO, and V-final languages .....	32
Table 7:	SVO as intermediate between V-final and V-initial.....	32

## **List of Figures**

Figure 1: Principle of natural serialization.....	18
---	----

# **1 Introduction**

This paper, a review of the literature on word order typology, examines in detail a set of groundbreaking works (Comrie 1989; Comrie, Dryer, Gil, Haspelmath 2005; Dryer 1988, 1991, 1992, 2007; Greenberg 1966; Hawkins 1983; Lehmann 1973; Vennemann 1974). After addressing several questions central to typological word order studies (determining basic word order, issues of data collection, theories of grammar) we will trace the development of a particular line of inquiry, that of word order patterning within and across languages, beginning with Greenberg's (1966) pioneering study and continuing through Dryer's (2007) contributions. Finally, in light of the findings delineated here, I will put forward for consideration the use of the formal framework of Optimality Theory within a functionalist approach for further investigation in typological word order studies.

## **1.1 SIGNIFICANCE AND BACKGROUND OF THE PRESENT STUDY**

Typology as a subfield brings specific perspectives to bear on linguistic inquiry. While much of linguistics since the Chomskian 1960s overhaul of the field has focused on single-language study (and mostly English for that matter), typology instead widens the net to include the study of many languages in their own right but also with respect to each other. Typology is interested in characteristics of single languages but also those of language families, that is, genetic characteristics, and these studies do not ignore the influence of location and area, language diffusion and borrowing. Thus the questions of typologists may differ from the questions of linguists in the generative tradition.

One tradition of generative linguistics has been to extrapolate universals of grammar by studying a single language in great detail and hypothesizing inviolable

principles that one then supposes to be true for language in general. On the other hand, a typologist might also search out universals of grammar but from the opposite direction, by observing characteristics shared by many languages to gain a perspective on what is possible in human language, and positing certain grammatical tendencies as universals of grammar based on the cross-linguistic empirical evidence.

The study of word order from a typological perspective came into its own when syntax became of burgeoning interest within linguistics in the 1960s. Thus the impetus for the interest in constituent order and clause structure within word order typological studies is in this way tied to the generativist tradition, but because of the typological perspective discussed above these word order studies have contributed in a different way to our understanding of “universals” (i.e., tendencies) of syntax and morphology, for example, that the generativist tradition cannot provide through single-language study. Thus the questions may be similar between the two camps but the two approaches differ greatly and are able to provide very different sorts of evidence. That is, typologists hope to make theories of grammar based on empirical observation while the generativist perspective places less value on empirical data.

Of course the emphasis on empirical data within typology brings its own set of problems, such as how one controls for genetic or areal bias when defining the set of languages for study. Also problematic is the comparison of grammatical categories across languages, for what is the basis of deciding if and how these categories are comparable? Because languages do differ so widely in how grammatical categories are realized (syntactically, lexically, morphologically, phonologically, etc.), how is it possible to know that what we have chosen to compare is in fact viable for comparison? Take the category of subject, for instance: the way the category of subject is realized in a nominative/accusative language will differ from the realization of subject in an

ergative/absolute language, or from that in a language that employs split ergativity. Thus, whether a subject is a semantic or syntactic category (or both, or categorized in another way altogether) is one example of the sorts of issues raised when one studies a set of languages as opposed to a single language (Comrie 1989). But this is evidence of the value of typology, for if we only studied a nominative/accusative Indo-European language like English we might miss out on many other interesting and important ways that grammar is realized around the world and thus miss very important generalizations of human language.

The question of word order within and compared across many languages has captivated typologists since Greenberg (1966) brought it to the table. Greenberg (1966) notes that certain orders of words that are observed in a particular language and group of languages (for example the order of a preposition relative to its object noun) correlate with other orders of words (such as the relative order of verb and object). Word order patterns within a language have been observed for centuries but correlating them with each other across many languages and presenting these correlations as typological tendencies brought new light and vigor to this type of study. To make his study possible, Greenberg had to posit a basic word order for each language in his database, and when many orders might be possible and even common in just one language, this raises the question of how basic order might be systematically established for any and all languages so that comparisons can be made across languages.

## **2 Questions central to the typological approach to word order<sup>1</sup>**

In this section we will address questions that are central to the general typological approach to word order. Sections 2.1 through 2.3 present a discussion of the problems and working solutions for issues of determining basic word order and defining grammatical categories for the purposes of typological studies; methods of data collection; and finally, the ways in which typology both depends on and produces theories of grammar.

### **2.1 DETERMINING BASIC WORD ORDER**

Although we retain the term “word order typology” as introduced by Greenberg, typologists recognize that “word” is not a technically accurate term since what is of interest is the order of constituents, which may not necessarily be realized as words but rather as phrases (Comrie 1989:80). For example, an object noun could consist of a word, a phrase with several words, a clitic or an affix. Also typologists are interested not only in orders of constituents but also in the order of morphemes smaller than a word, such as affixes or stems. Essentially what Greenberg termed “word order” is the order of significant grammatical categories such as verb, object and subject, when these categories are in fact realized as words (or words within constituent phrases). In addition, Comrie 1989 (and many others but especially Du Bois 1987 and Mithun 1987) notes that another problem for deciding basic word order in any one language is not at the word or even phrase level but at the clause or sentence level, since it is precisely the verb, object, and

---

<sup>1</sup> I am grateful to Pattie Epps who provided to me her prepared class notes from her graduate course on Linguistic Typology at the University of Texas at Austin, Spring 2007, which helped me tremendously in developing this paper.

subject categories that move in response to pragmatic constraints such as focus and topic, and which are more free to move than a constituent like a morpheme that is smaller than a word.

That said, it is generally agreed that in determining basic word order we must define our grammatical categories as full noun phrases and not consider for example an object marked on a verb or a pronominal object as indicative of the object category, since one or both may be treated differently in the grammar than an object expressed as a full noun phrase. A language like Spanish gives the perfect example where SVO order is employed in the case of an object noun phrase (“la tortilla” in (1)) and SOV order in the case of a pronominal object (“la” in (2)):

1) La mujer hace la tortilla. [SVO]

2) La mujer la hace. [SOV]

Or since subject is morphologically marked on the verb in Spanish, we have SVO order as in (1) above for a full noun phrase subject, but OV order when the object is pronominal and the non-lexical subject is dropped (but marked on the verb stem: *hac-e* make-3s/pres) as in (3):

3) La hace. [OV]

Generally, when subjects and objects are expressed as full noun phrases, then other words in the clause, such as the verb, have less morphological marking. Thus minimal morphological marking is one consideration in defining a “basic” word order. Also presumably it is less complicated to determine relative order of categories when they are expressed lexically rather than morphologically and so this makes the task of determining a basic order more straightforward.

When determining basic word order, the consensus is to take the main clause as basic, since some languages may have a different order in subordinate clauses. German is



the classic example where there are two common orders, SVO in the main clause as in (4) and SOV in the subordinate (or dependent) clause as in (5):

4) Die Frau sah den Mann. [SVO - main]

5) Ich denke, daß die Frau den Mann sah. [SOV - dependent]

In addition, we restrict our attention to affirmative main clauses, not negated or interrogative clauses as these may exhibit different patterns from an affirmative clause. In terms of noun phrases, they should be similar in definiteness and animacy since more definite or more animate noun phrases tend to precede other elements. Also when there are competing orders within a language such as both Adjective-Noun and Noun-Adjective orders, the order that is less syntactically restricted is determined to be more basic than the one that is more restricted in its distribution. For example in French a small number of adjectives precede the modified noun (AN) but the majority appear after the noun (NA) so this order is considered more basic. Finally, we consider most basic the order that is seen most frequently across texts, but only those clauses that are pragmatically least marked (e.g. no topic/focus fronting) and that fit all the considerations above: affirmative main clauses that contain full noun phrases and are the least morphologically marked.

Of course considering only sentences with the above characteristics is problematic. A main clause that has both O and S is necessarily transitive, and combined with the fact that we consider only those sentences in which categories are expressed as full noun phrases, it means we're considering as basic a clause that is cross-linguistically much less common than a main clause that is intransitive and or a transitive one where one or both noun phrases are expressed pronominally or marked on the verb, which is akin to ignoring the most common and basic clauses out there.

In terms of defining grammatical categories, that of subject is potentially problematic in cross-linguistic comparison since the subject in a nominative/accusative language is realized syntactically whether it is semantically an Agent or a Patient, while the realization of subject in an ergative/absolutive language differs based on agency. That is, a subject that is semantically an Agent in an ergative grammar is comparable to a subject in a nominative grammar, but a subject that is semantically a Patient in an ergative language is grammatically treated as an object and is not comparable to a subject in a nominative system. To complicate the issue more, languages that have split ergativity employ either the nominative system of subject-marking or the ergative system, depending on the environment or type of clause, so it is not only problematic to define S but also whether an ergative or a nominative clause should be considered in determining basic word order (Comrie 1989).

There are languages in which case marking allows for flexible word order such that a basic word order cannot be determined. While this might seem to be problematic, Comrie (1989:82) notes that languages which lack a clear basic order of S, O and V pose no problems for typological word order studies in that they are irrelevant to the topic.

## **2.2 ISSUES OF DATA COLLECTION**

Because typological studies are concerned with large numbers of languages, sampling bias is a general issue in typology as a subfield, not only in word order studies. If one is to claim tendencies are cross-linguistic then attention must be paid to such issues as locations of languages, shared history, borrowing, size of the set, and resources used to obtain data from individual languages.

Comrie (1989) recognizes that Greenberg's work caused typologists to be interested in using a large database of languages. Greenberg's (1966) main findings are

based on a sample of 30 languages (within a larger set upon which more general observations were made). At that time, examining the grammars of 30 languages in detail was a significant undertaking as well as a novel approach within linguistics as a whole. Among other contributors to the study of word order, Lehmann (1973) and Vennemann (1974) used Greenberg's original set of languages; Hawkins (1983) used the findings of Greenberg's set as a starting point but expanded the set to over 300 languages; unlike the others, Dryer (1988, 1991, 1992, 2007) does not use Greenberg's set and instead uses a database of over 600 languages, and explicitly controls for genetic and areal bias.

The languages Greenberg selected are: Basque, Serbian, Welsh, Norwegian, Modern Greek, Italian, Finnish (European); Yoruba, Nubian, Swahili, Fulani, Masai, Songhai, Berber (African); Turkish, Hebrew, Burushaski, Hindi, Kannada, Japanese, Thai, Burmese, Malay (Asian); Maori, Loritja (Oceanian); Maya Zapotec, Quechua, Chibcha, Guarani (American Indian). Greenberg states that he selected languages with which he was familiar or for which there was an adequate grammar available, and he tried to have wide genetic and areal coverage and thus more representative findings. However, Dryer (1992) notes that to be statistically significant, items in the sample should be independent of each other, a requirement which Greenberg's set does not satisfy, since his sample contained "languages within the same language family ...[that] share a given characteristic due to mutual inheritance" (1992:83). So although the 30-language set was large and the first of its kind, there are questions of areal and genetic bias due to geographical proximity and mutual inheritance.

Another aspect of the methodology is whether to count single languages or language families and how many languages of any given family to include in a sample. Perkins (1980) and Bybee (1985) advocate including one language per family in a sample for a study to be nonbiased. Greenberg's sample was unrepresentative in that he counted

languages and not families and the majority come from Europe and Africa (Dryer 1991), while Dryer (1988) includes many languages per family but only counts each family.

In his 603-language database, Dryer (1991) labels as “genera” the language families in his set which are based on published findings of genetic groups (from Campbell & Mithun 1979). Dryer’s term genus (pl. genera) refers to a genetic group “roughly comparable in time depth to the subfamilies of Indo-European” (1992:84). Dryer controls for the most severe genetic bias by counting genera and not single languages. Dryer’s (1992) database of 625 languages (of 252 genera) provides “statistical significance” in that, as much as is possible, the methodology controls for areal and genetic bias. Thus Dryer shows that it is possible to control for various biases.

### **2.3 THEORIES OF GRAMMAR**

Although typology emerged from under the auspices of the generative grammar movement, each is distinct from the other in focus. Generativists extract the notion of “possible human language” as defined by Universal Grammar from the in-depth study of a single language, usually English, as advocated by Chomsky (1965) (Hawkins 1983:6). In other words, “the generative approach is founded on the proposition that by uncovering the basic formal and substantive properties of the grammar of any one language, we discover the universals of language in general” (Hawkins 1983:6). Generativists seek to discover exceptionless (“absolute”) universals of language while typologists create classifications based on systematic variations between languages (Hawkins 1983:7) and find tendencies just as interesting as absolute universals.

Hawkins (1983) noted a shift among generativist work, from hypothesizing universals based on ideal structures, towards defining parameters on variation. For example X-bar theory (Jackendoff 1977, Lightfoot 1979) makes predictions on co-

occurring and non-co-occurring phrase structure rules that a grammar can have, thus defining possible and impossible grammars. However, a critique to that approach is that it is still based on single-language study (of English) and thus cannot be justified as a component of universal grammar without sufficient cross-linguistic work.

The generativist theory of competence maintains that grammar is innate in humans and denies the role of performance and processing. Typologists, on the other hand, posit that the realization and use of grammar is inextricably tied to performance (including such factors as memory, complexity and ease of processing). From the typologist's perspective, the generativist theory of competence may be elegant, but is ultimately too costly if performance and processing don't matter. As Hawkins puts it, "the role of processing difficulty (a performance notion) within linguistic theory appears

in a new light. The kinds of processing considerations that Chomsky used to motivate the irrelevance of performance for the competence grammar of English can now be used to explain the form and functioning of numerous grammatical rules across the world's languages" (1983:11).

Thus the typological perspective of the importance of performance in driving grammar may be less elegant than the generativist theory of competence, but it is more robust in that it is empirically supported by large language databases. Comrie (1989) discusses the issue of elegance versus accounting for the data, noting the struggle within typology to extrapolate typological universals while staying true to the data. Comrie also expresses the need for a more intuitive organizing principle behind the word order patterns that typologists have observed. But at least typologists recognize the connection between performance and human cognition (cf. Lehmann (1973), Dryer (1991, 1992)), a recognition that is absent from generativist theory.

Typically typology as a subfield is more interested in empirical findings than in theories of grammar in the sense that ideally these empirical typological findings would

not be based on theories of grammar but would be the basis for them, that is, typologies would drive theories and not vice versa. However, in order to identify structures and define grammatical categories cross-linguistically we must necessarily rely on grammatical theory and it becomes very difficult to tease apart typologies from theories. Some cross-linguistically problematic categories include the subject, as discussed above, as well as adjectives, genitives, and others. For these categories, whether we justify a category on syntactic or semantic grounds, and how, may very well depend on one grammatical theory over another.

Hawkins argues for typologies generating theories in that “cross-language frequencies provide a set of relevant data which inform the construction of a theory of

rule complexity (and markedness) in the area of cross-categorical generalizations. A single language provides sufficient data for the beginnings of such a theory, but cross-language data provide an empirical means of refining it, and of making theoretical decisions which otherwise cannot be made” (1983:11).

Hawkins also argues that typologies contribute to the theory of universal grammar in two ways: single-language analysis can attest to possible combinations of grammatical properties but only cross-language data can attest to impossible combinations; with data from large databases of languages we can create hierarchies of frequencies of combinations of grammatical properties in order to define “distributional universals of language” (1983:10) which grammatical theories based on single-language study cannot do. Hawkins mentions Comrie’s (1981) critique of the generative paradigm, “...that one cannot define the notion ‘possible human language’ on the basis of a study of single languages” (1983:7) since there is such wide variation among the languages of the world. In other words, “single languages underdetermine the range of possible and impossible language types” (Hawkins 1983:8).

There are other ways in which generative theory may sometimes be at odds with a typological perspective. In the case of German having two common word orders, SVO in main clauses and SOV in subordinate clauses, Comrie (1989) discusses the debate as to which of these is “basic”: the generative camp would opt for the subordinate clause as representing a deep structure and thus be more basic, while typologists would say that the surface structure of the main clause is basic. Which we choose becomes a theoretical question which then influences our interpretation of the data.

### **3 Literature on Word Order Typology**

In this section we will trace the development of the inquiry into word order patterning within and across languages. Greenberg (1966) initiated the study of word order in typology and thus is our starting point for discussion. Then we will consider the contributions of Lehmann (1973) and Vennemann (1974), namely the move to consider only V and O as the categories that organize other word order patterns. Then we will consider Hawkins’ (1983) Cross-Category Harmony as well as Dryer’s (1988, 1991, 1992, 2007) many contributions to typological word order studies.

One agreement among scholars in this area is that the order of constituents of different categories in a language tends to follow a particular pattern, that of the modified element generally following the modifying element, or vice versa (cf. Greenberg’s (1966) modifier-modified observation, Vennemann’s (1974) operands and operators; Hawkins’ (1983) Cross-Category Harmony which specifies that verb-initial languages are head-initial and verb-final languages are head-final; Dryer’s (1988) Branching Direction Theory). However, scholars differ substantially on what exactly defines the modifier and the modified, for example whether the motivation for the category is syntactic or

semantic. Scholars also debate the importance of the verb as a grammatical category after which other categories pattern themselves; the cross-linguistic status of subject as an important category; and the status of SVO as a type.

Greenberg (1966) established word order typology as a major concern in linguistic inquiry by presenting his empirical findings on a large set of languages. Greenberg presents observations, his 45 universals, while a smaller section discusses possible explanations for the universals. However, these universals are not “absolute,” rather they are universals in the sense of typological tendencies. (Many of these statements are unidirectional implicational universals that take the form “given  $x$  in a particular language, we always find  $y$ ” (1966:73)). Vennemann (1974) notes that although observations of word order of individual languages are not new, Greenberg’s main contribution was typological-comparative, in that his implicational universals cut across languages and language families and was the first to do so.

One of the fundamental notions of Greenberg’s work is that languages tend to place modifying elements before modified, or they do the opposite (modified elements before modifying), and he points out that linguists in general have been aware of the tendency to favor one order over the other. Hawkins also acknowledges this point, but notes that even though this principle has been refined over the years, “the word order predictions remain essentially as they were” (1983:3). However, with respect to the issues of data collection discussed previously, we must reevaluate this claim since these findings are based on Greenberg’s original data set and thus subject to the same genetic and areal biases. Obviously the influence of Greenberg’s work is not to be downplayed as he was the creator of this genre of linguistic study but since his innovation of typological word order studies, many of his original statements have been called into question in the



light of new data gathered with less biased sampling methods that were unknown at the outset.

### 3.1 THE BASIC ORDER TYPOLOGY

As for Greenberg’s (1966) findings, he proposes a basic order typology that employs three sets of criteria: 1) the use of prepositions (Pr) or postpositions (Po) in a language; 2) the relative order of subject (S), verb (V) and object (O) in a declarative sentence when subject and object are nominal; and 3) the position of a qualifying adjective (A) in relation to the noun (N) it modifies. Of these criteria, Greenberg forms his typology based on the second, the relative order of S, V and O. Of the six possible orders of S, V, and O, three (SVO, SOV, and VSO) occur as the most common types of dominant orders, while the other three (VOS, OSV, and OVS) were unknown at the time. Since Greenberg, VOS, OSV and OVS languages have been “discovered” but they remain quite rare relative to SVO, SOV and OVS (Comrie 1989:81).

From his set of 30 languages, Greenberg formulates his Universal 1, that “...the dominant order is almost always one in which the subject precedes the object” (1966:77), or SO order, and then the position of the verb V with respect to SO builds the typology: Type I is VSO, Type II is SVO, and Type III is SOV. Greenberg correlates these three types with criteria 1 (Pr/Po) and 3 (AN/NA) above, giving us twelve logical possibilities which are presented as representative of the worlds’ languages in Table 1:

	<b>I</b>	<b>II</b>	<b>III</b>
<b>Po-A<sup>2</sup></b>	0	1	6
<b>Po-N</b>	0	2	5

<sup>2</sup> A signifies AN order and N is NA order.

<b>Pr-A</b>	0	4	0
<b>Pr-N</b>	6	6	0

Table 1: Greenberg's Basic Order Typology

Greenberg finds that SVO (Type II) is the most common order cross-linguistically with SOV (Type III) next common, while VSO (Type I) is the least common of the three types; Greenberg also finds that the Pr/Po distinction more closely correlates with the three types and AN/NA order correlates less.

Greenberg (1966) was the first to bring attention to the idea that the verb V, subject S, and object O are all crucial to the ordering of other categories and that of these V is most crucial; he did so not explicitly but by naming of the types I, II, and III based on the three possible positions of V with respect to S and O.

### 3.2 THE VO/OV TYPOLOGY

By establishing a typology on the basis of V, Greenberg (1966) implicitly recognizes the verb as consequential to word order patterns. Vennemann (1974) credits Lehmann (1972) with explicitly positing the verb as the organizing parameter, and in first establishing a split between verb-final languages (OV languages) and all others (VO languages).

Lehmann (1973) distinguishes consistent OV and VO languages, and for Lehmann a consistent language satisfies two criteria: a basic word order is easily established for the language, and the patterning of other categories following from OV or VO is very regular. This is true for both Japanese (OV) and Portuguese (VO), for example: Japanese is a consistent OV language because it correlates with the following word orders: RelN, AN, and GN; conversely Portuguese is a consistent VO language and

has the orders NRel, NA, and NG. Lehmann (1973) says that a VO language with OV-type characteristics (such as AN order) would not be a consistent VO language. He posits diachronic change as a reason for this sort of inconsistency, where languages change from one type into another to become more consistent in their word order correlations. For example, English is a VO language that has AN order. This example is problematic for Lehmann's claim, however, since English has both AN and NA order, but the older NA order is giving way to the much more common AN order, moving away from type consistency, not toward it.

Lehmann (1973) predicts that OV languages are more likely to be agglutinative and VO languages are more likely to be isolating. He defines an agglutinative language as one in which categories such as tense/aspect (past, progressive), modality (causative, passive), and negation are affixes on the verb root where the affix and root remain distinct from each other phonologically; Japanese is his example of an OV agglutinating language. For Lehmann an inflectional language marks morphological categories as affixes which may modify the root, or be infixes or create phonological changes such as consonant gemination (1973: 47); Hebrew is an example of a VO inflectional language. However here we see a potential confusion of order in word-level versus morpheme-level categories since in his theory Lehmann does not treat morphemes and words differently.

Lehmann establishes the "Principle of Placement for Modifiers: modifiers are placed on the opposite side of a basic syntactic element from its primary concomitant" (1973:48). This principle deals solely with verbal modifiers (such as negation, causation, reflexives) which the principle predicts are placed after verb roots in consistent OV languages, and before verb roots in consistent VO languages, meaning O is the primary concomitant and the verbal modifier is on the other side of V (the verb root) from O.

Comrie (1989) critiques Lehmann along these lines: since he's generalizing from Greenberg's data, his findings work best for verbs but not so well for noun phrases that aren't object noun phrases; and the fact that he doesn't distinguish between words and affixes is not problematic for verbs, since post-verbal auxiliaries correlate with having suffixes, and pre-verbal auxiliaries with affixes. But his theory is problematic for nouns since noun modifiers either don't have these correlations or show the opposite tendency, in languages in which possessors precede their head noun but possessive affixes are suffixed.

### **3.2.1 Status of the Subject within VO/OV**

As for the subject category S, Lehmann notes that although many linguists swayed by Western logic may assume that the subject is mandatory, this assumption may be misleading: "As we may note from consistent OV languages like Japanese and from

consistent VO languages like Hebrew, subjects are by no means primary elements in sentences. Including them among the primary elements, as in the attempt to classify SVO and SOV languages as major types in the same way as VO and OV languages, has been a source of trouble for typologists as well as for linguistic theorists in general...[and] typological study accordingly supports this point of view by illustrating that the S in SVO formulae is far less significant than are the categories represented by V and O" (1973:51).

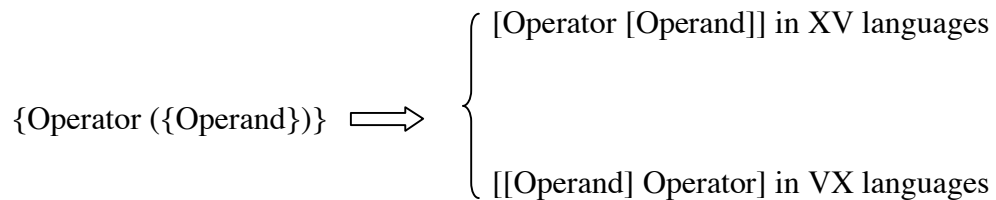
Lehmann's arguments against S as primary element are both convincing and consequential, but they are problematic for several reasons: he makes no distinction in his principle or in the argument about S between a full noun phrase, a lexical item, and a morpheme marked on the verb. Another problem is one touched on earlier, that the treatment of subject will differ greatly depending on if a language is nominative/accusative, ergative/absolutive, or has split ergativity. For nominative patterning, on the one hand, the subject category can signify agency and as attention to

agency is central in the human understanding of event structure, it seems unlikely that subjects in nominative patterning wouldn't be "primary elements". On the other hand, agency is signified in other ways besides lexically or by order which would support his argument that subject is not a primary element where agency is not grammatically encoded in word order.

### 3.2.2 Head-Dependent Theory

Vennemann (1974) follows Lehmann in his OV/VO distinction but prefers XV/VX with X standing for the verb complement, presumably because it is not important to his purposes whether that complement is the object or the subject. From Greenberg's implicational statements, Vennemann formulates the Principle of Natural Serialization (more often called his Head-Dependent Theory) that drives the word order patterns, as follows in Figure 1:

Figure 1: Principle of natural serialization (1974: 10)



Semantically, the operator is the specifier of the operand and in that way corresponds to Trubetzkoy's (1939) *déterminé* and *déterminant* for French; the category of the operator-operand constituent corresponds to the syntactic category of the operand.

A smaller version of Vennemann's operator/operand table (Comrie 1989:92) follows in Table 2:

<b>OPERATOR</b>	<b>OPERAND</b>
Object	Verb
Adjective	Noun
Genitive	Noun
Relative clause	Noun
Noun phrase	Adposition
Standard of comparison	Comparative adjective

Table 2: Vennemann’s Operator/Operand Schema

Vennemann (1974) notes Lehmann’s (1973) principle stating that verbal modifiers appear on the opposite side of V from O. This fits with the principle of natural serialization in that verb roots are operators on their modifiers (e.g. auxiliary, modal, intensional verb) and noun phrases are operators on their modifiers (e.g. adposition). He further argues that SXV order has two advantages: that the primary case for the expression of topics (the subject) is at the beginning of the clause, and the verb is at the end (sentence-marginal). Admittedly verb-initial (such as VSX) languages also have the verb in a sentence-marginal position but without the added advantage of the subject being first. However, we might wonder how this fits with SVO having been recognized as the most common word order worldwide, since it wouldn’t by Vennemann be the most advantageous order.

Lauding the elegance of Vennemann’s operator-operand schema, Comrie (1989) takes a similar position but admits that the schema does not perfectly predict the facts of Greenberg’s data, and while Greenberg’s approach fits more closely with the data, it doesn’t present a clear schema or “coherent conceptual whole” like Vennemann’s elegant principle (1989:93). Comrie (1989) praises Greenberg for having been cautious in his

claims of universals but he notes that with regard to Greenberg's "less intuitively plausible universals ... one senses a certain tension between, on the one hand, empirical validity without a coherent conceptual system, and, on the other, plausible coherent conceptual systems which ... lack empirical validity" (1989:88). This critique is applicable in the first instance to typology and generativist theory in the second.

Comrie (1989) notes that out of Greenberg's six logically possible types (SOV, SVO, VSO, VOS, OVS, OSV) the distribution in the world of languages with these patterns is skewed towards the first three and most heavily toward the first two, SOV and SVO. He gives a schema of the four types out of Greenberg's possible 24 that are by far the most represented in the world's languages (Comrie 1989:89), as follows in (a-d):

- a. VSO/Pr/NG/NA
- b. SVO/Pr/NG/NA
- c. SOV/Po/GN/AN
- d. SOV/Po/GN/NA

Comrie notes that (a) and (b) are identical except for the position of the subject, and removing S as an important category would collapse VSO and SVO into the type VO. In the same vein (c) and (d) are identical except for the position of the adjective and "the only embarrassment to this generalization [is] the widespread occurrence of NA basic order in OV languages" (Comrie 1989:89). (See Section 3.4.2.1 below for Dryer's (1988) argument against adjective-noun as a correlation pair.)

From the four types (a-d) we can create 2 major types (e-f) which are the inverse of each other:

- e. VO, Pr, NG, NA
- f. OV, Po, GN, AN

These results are associated with the work of Lehmann and Vennemann, the former who first posited V and O as the primary organizers of other word orders, and the latter whose Head-Dependent Theory stated that operands (heads) pattern like verbs and operators (dependents) pattern like objects. Comrie labels type (e) as operator-operand (or head-adjunct) and type (f) as operator-operand (or adjunct-head) following Vennemann.

Vennemann's "operator" corresponds to the traditional structuralist syntactic term adjunct (or dependent) and "operand" to the traditional term head and this could pose a problem since there is cross-theory lack of agreement as to what is a head and what is a dependent. Comrie notes, "the assignment of operator (adjunct) and operand (head) status

is in most instances uncontroversial, though some linguists have been less comfortable with declaring the head of an adpositional phrase to be the adposition, rather than the noun (phrase)...but this assignment can be justified, for many languages, by the usual structuralist syntactic test of substitution" (1989:92).

### **3.3 ARGUMENT AGAINST TWO-WAY TYPOLOGY**

Rather than V as the principle organizing category, Hawkins (1983) notes that whether a language has prepositions or postpositions more accurately predicts the other word order patterns and correlation pairs in that language. However, he also maintains that languages simply do not fall under neat divisions of type in which they share a set of word order properties, such that "the whole notion of a 'word order type' becomes more abstract" (1983:16). Hawkins argues against languages being typecast when the empirical data points to gradual rankings of preferences for certain word order pairs to correlate with others. More specifically, he states that "SVO is no longer a type indicator; [and] VSO and SOV are type indicators, but limited ones" (1983:16)<sup>3</sup>.

---

<sup>3</sup> But Dryer (1988) finds that Hawkins' claim (that adposition is a better predictor of other word order patterns than object and verb) is a Eurasian areal phenomenon.



Hawkins (1983) argues that Vennemann's principle of serialization is both too strong and too weak precisely because it is bidirectional. Greenberg's unidirectional implicational statements make predictions in one direction, for example, that OV order implies Postpositions. Vennemann's principle predicts this but also in the opposite direction, that Postpositions imply OV, which is not borne out by the data, and which Greenberg did not posit. Also Greenberg's unidirectional statements restrict possible correlations, but Vennemann's bidirectional statements are not restrictive enough, such that his principle overpredicts correlation pairs that are not attested. In other words, "the modifier - modified generalization, in all its forms, does not succeed in distinguishing between those word order co-occurrences that are never found and those that are" (Hawkins 1983:3).

### **3.3.1 Cross-Category Harmony**

Hawkins (1983) posits languages that are V-initial, V-medial, and V-final, predicting that V-final languages most commonly place all modifiers before nouns, V-initial languages place all modifiers after nouns, and SVO places some modifiers before and some after nouns. This is captured by his Principle of Cross-Category Harmony (hereafter CCH). A refinement of Vennemann's Head-Dependent Theory (HDT), Hawkins' CCH posits that head-dependent orders will harmonize across verb and noun categories, that is, if in a language a noun dependent precedes the head noun then a similar proportion of verb dependents will precede the verb in that language. To explain apparent exceptions such as NRel order in both OV and VO languages (where the HDT predicts NRel for just VO), Hawkins appeals to ease of processing as it is easier to process a long modifier phrase (the relative clause) when the noun comes first and is established as the concept to be modified.

### 3.4 VERB AND OBJECT PATTERNERS

Dryer (1991) follows the two-way distinction VO versus OV generalized by Lehmann (1973, 1978) and Vennemann (1974, 1976) from Greenberg's three types. Using the VO/OV typology as a starting point in relation to his database, Dryer (1992) establishes that certain grammatical categories pattern after the verb position with respect to the object, while others pattern after the object. The definition of verb patterner and object patterner is as follows: "If a pair of elements X and Y is such that X tends to precede Y significantly more often in VO languages than in OV languages, then <X,Y> is a correlation pair, and X is a verb patterner and Y an object patterner with respect to this pair" (Dryer 1992:87).

Following in Table 3 is a complete list of the correlation pairs that the verb patterner/object patterner distinction accounts for (Dryer 1992:108):

VERB PATTERNER	OBJECT PATTERNER	EXAMPLE
verb	object	<i>ate + the sandwich</i>
verb	adposition	<i>(there) entered + a tall man</i>
adposition	NP	<i>on + the table</i>
copula verb	predicate	<i>is + a teacher</i>
'want'	VP	<i>wants + to see Mary</i>
tense/aspect	auxiliary verb	<i>has + eaten dinner</i>
negative auxiliary	VP	<i>7 in 4.2<sup>4</sup></i>
complementizer	S	<i>that + John is sick</i>
question particle	S	<i>8 in 4.4<sup>5</sup></i>

---

<sup>4</sup> Diegueño:      ?-u·ya·w-x      ?-əma·w-x  
                          1-know-FUT      1-not-FUT  
                          'I won't know.'

adverbial subordinator	S	<i>because + Bob has left</i>
article	N'	<i>the + tall man</i>
plural word	N'	<i>9 in 4.7<sup>6</sup></i>
noun	genitive	<i>father + of John</i>
noun	relative clause	<i>movies + that we saw</i>
adjective	standard of comparison	<i>taller + than Bob</i>
verb	PP	<i>slept + on the floor</i>
verb	manner adverb	<i>ran + slowly</i>

Table 3: Complete List of the Verb- and Object-Patternner Correlation Pairs

Dryer (1992) establishes that the clearest correlation pair in his database of 252 genera is Adposition with Noun where noun is an object patternner and adposition (preposition or postposition) is a verb patternner. That is, as Hawkins noted and Dryer corroborates, OV languages tend to be postpositional and VO languages tend to be prepositional, and more than any other word order correlation, cross-linguistically this is the most widely attested.

### 3.4.1 Dryer on HDT

Dryer calls HDT the most accepted explanatory view (at that time) of correlation pairs; in light of his distinction between verb- and object-patternners he redefines HDT as the following: “Verb patternners are heads and object patternners are dependents. That is, a

---

<sup>5</sup> Mokilese (Austronesian): *a koah sihkei?*  
Q you well  
'Are you well?'

<sup>6</sup> Gbeya: *ó tú wí-ré*  
PLURAL black person  
'black people' (Samarin 1966:81)

pair of elements X and Y will employ the order XY significantly more often among VO languages than among OV languages if and only if X is a head and Y is a dependent (Dryer 1992:87). Under this definition HDT accounts for the correlation pairs listed above. However, as a theory borne of Greenberg’s data set, it suffers from the same shortcomings of sampling bias as previously discussed. Following in Table 4 are pairs that HDT predicts to correlate as dependent and head, but which are not attested cross-linguistically and therefore are noncorrelation pairs (1992:108):

<b>DEPENDENT</b>	<b>HEAD</b>	<b>EXAMPLE</b>
adjective	noun	<i>tall + man</i>
demonstrative	noun	<i>that + man</i>
intensifier	adjective	<i>very + tall</i>
negative particle	verb	<i>not + go</i>
tense/aspect particle	verb	<i>examples in 6 in 3.5<sup>7</sup></i>

Table 4: Dependent/Head Noncorrelation Pairs

### 3.4.2 Branching Direction Theory

Dryer has contributed significantly to typological word order studies, especially for certain questions that scholars have debated, such as claims about Adjective-Noun order, the importance of verb and subject as organizing categories and the status of SVO as a type. He proposes the Branching Direction Theory (a refinement of both the Head-Dependent Theory and Cross Category Harmony) to account for the patterns in the data.

---

<sup>7</sup> Yapese: *gamow*            *raa*            *guy-eem.*  
 1PL.EXCL        FUT        see-2SG  
 ‘We will see you.’ (Jensen 1977:194)

### **3.4.2.1          *Adjective-Noun***

As discussed above, Lehmann (1973) and Vennemann (1974), generalizing from Greenberg (1966), claimed that OV languages correlate with Adjective-Noun (AN) order and VO languages correlate with NA order. But [as noted above] NA basic order in OV languages is widespread (Comrie 1989:89) and not AN order as Greenberg's data indicates. Dryer asserts that the apparent correlation, again, is due to areal and genetic sampling bias: in looking at Greenberg's data set, Dryer finds that the OV languages of Eurasia do tend to have AN order while non-Eurasian OV languages tend to have NA order so this seems to be a Eurasian areal phenomenon (1988:208). Dryer's data (1988), which is controlled for genetic and areal sampling bias, includes 11 families where OV-AN is dominant, 13 families where OV-NA is dominant, and 6 where neither is dominant (1988:185). Thus he shows that contrary to Greenberg's (and others') data and findings, there is no clear worldwide cross-linguistic correlation between the order of O and V and the order of A and N.

### **3.4.2.2          *First Version of Branching Direction Theory***

The fact that there is no correlation between object and verb and adjective and noun is problematic for Greenberg's (1966) theory that "word order correlations reflect a tendency towards consistent ordering of head and dependent". To that end Dryer proposes (his first version of) the Branching Direction Theory: "Languages tend toward consistent left-branching or consistent right-branching" (1988:191). In other words, in a left-branching language, branching categories *precede* nonbranching categories, while in a right-branching language, branching categories *follow* nonbranching categories. In fact this is a significant refinement of HDT in that BDT looks internally to the language and

does not arbitrarily depend on external theory like HDT does as to what defines head and dependent.

He motivates this tendency toward consistency in branching by appealing to ease of processing, saying that language structures that consistently branch in one direction are easier to process than those that branch to both the right and the left. To exemplify this concept, Dryer (1992:129) gives a sentence with multiple center-embedding of relative clauses as in (6):

6) The cheese [that the rat [that the cat chased] ate] was rotten.

Just as these nesting relative clauses are difficult for English speakers to process because they interrupt constituency, constituency would be disrupted by recursive categories that do not consistently branch to one direction since they would create multiple nesting by generating to the inside of the phrase, as in the following Postpositional phrase in (7) (Dryer 1992:129):

7) color [flowers [vase [table on] in] of]

Dryer is defining branching versus non-branching categories prototypically, in that cross-linguistically certain categories are more likely to be branching than others, although possibly from one language to the next and certainly from one construction to the next we can find exceptions. In the case of adjectives, presumably languages in which modifying concepts are verbs or nouns would not be considered to have adjectives at all and so would be exempt from any generalizations about NA word order (1988:198).

### **3.4.2.3        *Revisions to BDT: “Revised” and “Alternate”***

In an effort to better account for the data, Dryer (1992) revisits BDT and posits what he refers to as the “revised version”, then later an “alternate version”. The revised version follows: “Verb patterners are nonphrasal categories or phrasal categories that are

not fully recursive, and object patterners are fully recursive phrasal categories in the major constituent tree. That is, a pair of elements X and Y will employ the order XY significantly more often among VO languages than among OV languages if and only if X is not a fully recursive phrasal category in the major constituent tree and Y is a fully recursive phrasal category in the major constituent tree” (1992:114)<sup>8</sup>.

In other words, a recursive and phrasal category is a verb patterner, while nonphrasal nonrecursive category is an object patterner. For example, in English a PP is an object patterner because it is a fully recursive phrasal category that dominates another, namely NP. A major constituent tree depicts a fully recursive phrasal category while a minor constituent tree depicts a phrasal constituent that is of the same category as its dominating node and serves as the head of that node (Dryer 1992:113).

The alternate version of BDT does not refer to major or minor constituents but to head and dependent: “Verb patterners are heads and object patterners are fully recursive phrasal dependents. i.e., a pair of elements X and Y will employ the order XY significantly more often among VO languages than among OV languages if and only if X is a head and Y is a phrasal dependent of X” (1992:116).

This BDT is an alternate to the revised BDT above and is advantageous because it is more elegant and does not require any reference to major or minor constituents. But as it relies on the head/dependent distinction it is problematic because it is theory-dependent. Dryer makes no claim as to which is better, simply stating that more research needs to be done for that decision to be made.

However BDT is able to deal with adjectives better than HDT or CCH: in languages that have the adjective category, adjectives are object patterners. For example Spanish is a VO language and has basic NA order: *el carro rojo* ‘the car red’. English is problematic according to Lehmann and Vennemann in that it is a VO language but has

---

<sup>8</sup> Dryer states that a fully recursive phrasal category is one in which the phrases are formed by rules that feed into the rest of the grammar and that can dominate other major phrasal categories (1992:111).

basic AN order, but Dryer's BDT can deal with this if we say that although adjectives are generally not phrasal, in English they combine with nouns which is fully recursive phrasal category which would render the adjective a verb patterner.

Manner adverbs could be problematic for BDT in that they could be classified as object patterners, but since they are nonphrasal and not fully recursive, they are verb patterners under Dryer's definition and thus do fit the predictions of BDT. Also there are certain categories, such as demonstratives and numerals, that are semantically similar cross-linguistically, but how they are realized in individual grammars can differ quite widely in terms of their grammatical categories. This fact makes the predictions of BDT harder to test or perhaps less relevant for these particular categories (Dryer 1992:122).

### **3.5 SVO TYPE**

Whether SVO is a type has been a question in word order studies since Greenberg first established it as one of his three types. In Greenberg's 30-language sample SVO languages seem to occupy a category that is intermediate between VSO and SOV languages in terms of which word order patterns they exhibit (such as Pr/Po, NAdj/AdjN, NDem/DemN, and AdjIntens/Intens/Adj). This has led scholars down a path of positing that SVO, unlike VSO and SOV that have their own characteristics, is not a true type but is an amalgam of the two more polar types. The Lehmann/Vennemann move towards simplifying the three-way typology to the two-way VO/OV distinction then implied that the position of the subject is less important suggesting that SVO languages are of type VO (Dryer 1991).

For Hawkins (1983) since V and O are not organizing parameters, SVO ceased to be a type indicator at all. Comrie on the other hand agrees with the collapse of VSO and VOS into VO type, but not SVO, because he claims that it is sufficiently different from



the other two as to inhabit its own type. Comrie goes so far as to say “...knowing that a language is SVO, we can predict virtually nothing else” (1989) with respect to its other word order parameter values.

Although many scholars (Comrie 1981, 1989; Mallinson & Blake 1981; Siewierska 1988; Payne 1990; Hawkins 1980, 1983) have argued that SVO languages are intermediate between VO and OV languages, Dryer (1991) categorizes SVO as a VO language since he has found significant differences in word order patterns between VO and OV but that SVO exhibits very similar patterns to VO. Of course, as previously noted, Greenberg’s (1966) sample provided the data for the initial findings on SVO which is the result of areal bias and accident due to a small sample (Dryer 1991).

### **3.5.1 SVO in the data**

Surprisingly, despite all the work done on this issue, Dryer was the first to question the data from which the findings come. Greenberg’s (1966) database contains 19 SVO languages with Postpositions and 33 SVO languages with Prepositions, which leads one to believe that the SVO correlation with adposition is intermediate between VSO and SOV. Meanwhile, Dryer’s (1991) database contains only 6 genera containing SVO-Po languages and a full 42 genera containing SVO-Pr languages, an overwhelming majority. Questioning the discrepancy between his data and Greenberg’s, Dryer finds that more than half (10 out of 19) of Greenberg’s are languages from Africa, evidence that SVO-Po is an areal phenomenon.

Upon closer inspection of the individual 19 languages, Dryer finds that perhaps due to inaccurate data at least 8 of them are either not SVO or not Postpositional, including Chinese, the Algonquian family, Zoque, Tonkawa, Songhai, the Mande family, the Ijoid family, Nupe, and the Daghestan family. Another 10 languages are clearly SVO-

Po, or cannot be shown to be otherwise, including Finnish, Estonian (both Finno-Ugric), Kru (Kru), Twi, Gã, Guang, Ewe (all Kwa), Guarani (Tupian), Rutul (Daghestan), and Kirma (Voltaic/Gur). In addition, these 10 languages belong to only 6 genera, four of which are accounted for in Dryer's database, thus showing that the set of languages brought forth findings that do not reflect cross-linguistic tendencies. Dryer (1991) notes that there are two other genera containing SVO-Po languages in his database, Xu (Northern Khoisan) and Iquito (Zaparoan). This is further evidence that SVO patterns with V-initial in the correlation with Prepositions.

### **3.5.2 SVO as VO**

Dryer (1991) goes through characteristics that were supposed to support the claim that SVO is a category separate from VO and refutes that claim with his findings, and shows that SVO and V-initial<sup>9</sup> languages share a strong tendency to have the following properties: Prepositions; noun + relative clause (NRel) order; adjective + standard of comparison (AdjSt) order; copula + predicate (CopPred) order; adverbial subordinator + main clause (SubS) order; plural word + noun (PlurN) order; verb + adpositional phrase (VAdp) order; verb + manner adverb (V-MAdv) order; tense/aspect auxiliary verb + verb (TAauxV-V) order; negative auxiliary + verb (NegAux-V) order.

Dryer's overall proportions for each of those properties are summarized (1991:451-460) in the Table 5:

---

<sup>9</sup> Dryer (1991) defines V-initial languages as those in which the verb comes before the lexical subject and object but in which other elements (such as negative words, tense-aspect particles, and question particles) may even precede the verb itself; V-final languages are those in which verb generally follows the lexical subject and object but in which those other elements may follow the verb.

	<b>V-final</b>	<b>SVO</b>	<b>V-initial</b>
<b>Postp</b>	0.96	0.14	0.09
<b>Rel-N</b>	0.43	0.01	0.00
<b>St-Adj</b>	0.82	0.02	0.00
<b>Pred-Cop</b>	0.85	0.26	0.39
<b>S-Sub</b>	0.70	0.06	0.06
<b>N-Plur</b>	1.00	0.24	0.13
<b>Adp-V</b>	0.90	0.01	0.00
<b>MAdv-V</b>	0.91	0.25	0.17
<b>V-TAuxV</b>	0.94	0.21	0.13
<b>V-NegAux</b>	0.88	0.13	0.00

Table 5: Correlation Pairs for V-initial, SVO, and V-final languages

From the table above it is clear that SVO patterns much like VO.

For other orders, such as AdjN, DemN, and IntensAdj, Dryer (1991) finds that there is no significant difference even from OV to VO languages because these are not correlation pairs. His findings are summarized in Table 6 below:

	<b>V-final</b>	<b>SVO</b>	<b>V-initial</b>
<b>AdjN<sup>10</sup></b>	0.40	0.41	0.34
<b>DemN</b>	0.68	0.74	0.58
<b>IntensAdj</b>	0.61	0.57	0.51

Table 6: Noncorrelation Pairs for V-initial, SVO, and V-final languages

---

<sup>10</sup> Testing the predictions of Hawkins' CCH against his own data, Dryer (1991) finds that for SVO languages CCH does not make the correct predictions (specifically with respect to Noun pairings with Gen, Adj, Dem, and Num). Again we note that although Hawkins had a large database, it was based on Greenberg's original data, and thus does not empirically match up with Dryer's data.

### 3.5.3 Three exceptions

However, there are 3 pairs of elements which do position SVO as intermediate between VO and OV: that of genitive and noun, yes/no question particles (Q) and clause (S), and interrogative (*wh-*) words and clause. His findings are summarized in Table 7:

	<b>V-final</b>	<b>SVO</b>	<b>V-initial</b>
<b>GenN</b>	0.89	0.59	0.28
<b>SQ</b>	0.73	0.30	0.13
<b><i>Wh-in situ</i></b>	0.71	0.42	0.16

Table 7: SVO as intermediate between V-final and V-initial

Thus Dryer (1991) shows that the Lehmann-Vennemann distinction between VO and OV holds and that SVO patterns with VO languages for the majority of the correlation pairs, and that only for a very small defined number of correlation pairs does SVO appear to be somewhat intermediate between OV and VO.

Specifically addressing Comrie's (1989) claim that although word order patterns can be predicted for VO and OV languages nothing can be predicted if a language is SVO, Dryer (1991) has shown that to the extent that word order patterns can be predicted for VO and OV languages, word order patterns can be predicted for SVO as well, and that the patterns that emerge lead us to classify SVO as a VO type.

## 3.6 NEW CONSIDERATIONS

Dryer (2007) notes that it is often difficult to establish a language as one of Greenberg's six types: SOV, SVO, VSO, VOS, OVS, and OSV, due to difficulty

establishing either subject or object, or when a language displays several orders and none is more basic than another. He acknowledges however that it is much easier to determine a language's basic order of VS/SV in intransitive clauses and VO/OV in transitive clauses when the subject (or agent) is marked on the verb. Thus he argues that languages be classified according to the relative positions of S and V, and O and V. Of course this means that instead of the canonical noun subject and noun object we are looking at clauses in which S or O is not a word but an affix on the verb, and this is not the traditional approach to typological word order studies.

However, considering that transitive clauses with both a noun S and a noun O do not occur very often in most languages, frequency criteria justify classifying a language based on clauses with S or O affixed to the verb. In fact, we are missing an important generalization of the language if we do not consider those very clauses. For example, Spanish undeniably has SOV order in nominal transitive clauses, and SV order in nominal intransitive clauses, but VS order in intransitive clauses is almost as common as SV, and to ignore this is to miss an important fact of the grammar.

#### **4 Further Investigation: Functionalist Constraints within OT**

More recently there has been a movement within typology as a subfield to use Optimality Theory as a framework to capture typological tendencies. Many linguists have found that by using language-specific rankings of the violable principles of OT (which theoretically constitute universal properties of language) we are able to more elegantly and systematically account for cross-linguistic preferences in a way that is not possible in a structuralist or generativist framework that posits inviolable rules (Haspelmath 1999).

I think it would be fruitful to approach the questions of word order typology from a functionalist perspective using the formal framework OT. That is, we could compose constraints from a functionalist perspective and use the preferential ranking system of OT to explain the empirical data, which in this line of study would include observations of surface word order in many different languages.

For example, we might propose a violable constraint that says if you have a rich morphology and the arguments are obligatorily marked on the verb, put the verb first in the clause, precisely because it has all the crucial information<sup>11</sup>. If you have an impoverished morphology and it's necessary to express subject and object lexically, put the verb last in the clause. This uses a functionalist approach in that the constraint itself expresses the explanation for a particular grammatical strategy by way of the function of the structure. Then we would rank particular languages to the extent that they violate the constraint. Precisely because its rules are violable, OT provides the perfect mechanism to capture cross-linguistic tendencies in word order correlations. This would allow us depict frequency distributions of universal constraints on word order in the world's languages. OT provides an elegant way to account for the instances in which languages exhibit a particular characteristic or pattern. It is less optimal, for example, to have basic OSV order because an object is lower on the animacy/agency hierarchy, which is why this order is not widely attested cross-linguistically. Or for correlation pairs, we can take a functionalist approach and show within an OT that it is more optimal for an SOV language to have Prepositions than Postpositions.

Thus although Prince & Smolensky (1993) originally formulated OT for phonology, OT is well suited for capturing typological tendencies through rankings of highly general universal constraints to describe grammatical optimality. Haspelmath

---

<sup>11</sup> Suggested by I-wen Lai, personal communication.

argues that grammatical optimality is “largely parallel” to user optimality (1999:186), an appeal to ease of processing and production that other theories of grammar cannot incorporate.

Using the OT framework within typology, Aissen notes that “differences between languages can be described as differences in the ranking of universal constraints” (1999:673). Aissen develops an OT account of the interplay between syntactic markedness and the agency/animacy hierarchy of Silverstein (1976), exploring how Silverstein’s generalization might be expressed in a formal theory of grammar and how it can play a role in individual grammars; in short, to “reconceive hierarchies as sets of ranked constraints” (1999:679).

Interestingly she finds that morphological marking coincides with syntactic markedness (which could be evidence that word order studies were on the right track in not considering as basic those clauses in which subject or object is not expressed as a noun but morphologically marked on the verb). Aissen also finds that structural positions in the clause (such as subject and object) interact with the dimensions of person, animacy and agency, such that a first- or second-person subject is less marked than a third-person subject, while conversely a third-person nonsubject is less marked than a first- or second-person nonsubject (1999:681).

While Aissen (1999) does not explicitly remark on typological word order studies, her findings and methodology both are important for typology as a subfield, as well as the use of an OT framework to interpret the data and to form a theory of grammar that, in Comrie’s (1989) words, forms a clear conceptual whole while at the same time staying true to the empirical data. As Aissen eloquently states, “the important point is that while

the ranking of constraints within a subhierarchy is universally fixed..., cross-linguistic variation arises because the interpolation of constraints from distinct

subhierarchies can yield distinct overall constraint rankings, i.e., distinct grammars” (1999:684).

She rightly points out that “to construct grammars which succeed not only in characterizing the facts of individual languages, but do so within a universal theory of markedness...is striking”, and that such insights “have remained largely unexpressed in generative approaches to grammar” (1999:708).

## **5 Conclusion**

The body of work investigated in this paper made a major contribution to linguistics by introducing the subfield of typology and the study of word order across the world’s languages from a typological perspective. Greenberg’s (1966) seminal paper advanced an understanding of cross-linguistic tendencies that had been unknown at the time and which are still being investigated today, especially his three-way typology based on the relative position of V with respect to S and O. Lehmann (1973) and Vennemann (1974) pushed the VO/OV distinction which led to a reanalysis and diminishing of the role of S as an organizing parameter. Two theories, Vennemann’s Head-Dependent Theory (compatible with the two-way typology of VO/OV), and Hawkins’ Cross-Category Harmony account for many attested correlation pairs, but neither is as strong as Dryer’s Branching Direction Theory in terms of explanatory adequacy, elegance, and adherence to observed cross-linguistic tendencies.

As far as theoretical approaches, we noted that generative grammar with its focus on single-language study cannot provide an account of the variations in the world’s languages, while the typological approach comes closer to describing universals of language based on empirical data.



Finally, I present the idea that investigations of word order from a typological perspective can be successfully undertaken using a functionalist approach within the framework of Optimality Theory.

## References

- Aissen, Judith. 1999. Markedness and subject choice in Optimality Theory. *Natural Language and Linguistic Theory* 17: 673-711.
- Comrie, Bernard. 1989. *Language universals and linguistic typology*. 2nd edn. Chicago: University of Chicago Press.
- Comrie, Dryer, Gil, Haspelmath. 2005. *Word Atlas of Language Structures*. Oxford University Press. Online at <http://wals.info/>
- Dryer, Matthew S. 1988. Object-verb order and adjective-noun order: dispelling a myth. *Lingua* 74: 185-217.
- Dryer, Matthew S. 1991. SVO languages and the OV:VO typology. *Journal of Linguistics* 27: 443-482.
- Dryer, Matthew S. 1992. The Greenbergian word-order correlations. *Language* 68: 81-138.
- Dryer, Matthew S. 2007. *Word Order. Clause Structure, Language Typology and Syntactic Description*, Vol. 1, ed. by Timothy Shopen. Cambridge University Press.
- Du Bois, John W. 1987. The Discourse Basis of Ergativity. *Language* 63: 805-855.
- Greenberg, Joseph. 1966. Some universals of grammar with particular reference to the order of meaningful elements. *Universals of Language*, ed. by Joseph Greenberg, 73-113, Cambridge, MA: MIT Press.
- Haspelmath, Martin. 1999. *Zeitschrift für Sprachwissenschaft* 18.2: 180-205.
- Hawkins, J.A. 1983. *Word Order Universals*. NY: Academic Press.
- Lehmann, W.P. 1973. A Structural principle of language and its implications. *Language* 49: 47-66.
- Mithun, Marianne. 1987. Is basic word order universal?. In *Coherence and Grounding in Discourse*, ed. by Tomlin, Russell S., Amsterdam: Benjamins.

Vennemann, Theo. 1974. Theoretical word order studies: results and problems. *Papiere zur Linguistik* 7: 5-25.

This document does not include the vita page from the original.