

This is a nonfinal draft of a paper that appeared  
in 1991 in *Journal of Linguistics* 27: 443-482

## SVO Languages and the OV:VO Typology

Matthew S. Dryer  
University at Buffalo

Within word order typology, two positions have been taken on the status of SVO languages. One position is that SVO languages pattern like V-initial languages and that there is thus a fundamental split between OV and VO languages. The other position is that SVO languages are an inconsistent type, intermediate between V-final and V-initial languages, and that the OV:VO typology is thus inaccurate. Evidence is presented here showing that, although there are a few respects in which SVO languages exhibit properties intermediate between those of V-final languages and V-initial languages, they in general pattern very much like V-initial languages, thus supporting the OV:VO typology.

## 1. INTRODUCTION

Greenberg (1963) formulated a number of implicational universals that refer to the order of various syntactic elements. He classified languages on the basis of their order of subject, object, and verb into three types, which he labeled I, II, and III, and which correspond to what are commonly known as VSO, SVO, and SOV languages, respectively. Since that time, evidence for the existence of the three other logically possible orders, VOS, OVS, and OSV, has been presented (cf. Keenan 1978, Derbyshire and Pullum 1981, 1986), thus leaving us with a typology of six types. Lehmann (1973, 1978) and Vennemann (1974, 1976) collapsed these six types into two types OV and VO. Implicit in this move is the idea that the fundamental parameter is the order of verb and object and that the position of the subject is less important. In treating VSO, VOS, and SVO as subtypes of the general type VO, the claim is that these three types are similar to each other in their other word order characteristics and different from OV languages.

More recently, a number of linguists, including Comrie (1981: 90, 94-95; 1989: 96, 100-101), Mallinson and Blake (1981: 379), Siewierska (1988: 18-19), and Payne (1990: 19) but most particularly Hawkins (1980: 199; 1983: 30), have criticized Lehmann and Vennemann for collapsing VSO, VOS, and SVO languages into a single category VO. They all argue that the available evidence does not support the claim that SVO languages pattern like VSO and VOS languages. The purpose of this paper is to argue that, although some of these criticisms are not without merit, Lehmann and Vennemann were largely right: with certain well-defined exceptions, the word order properties of SVO languages differ little from those of VSO and VOS languages. In short, it will be shown that with respect to a large number of word order characteristics, we do find a basic split between VO and OV languages.

I will use the term *v-INITIAL* to refer to languages in which both subjects and objects generally follow the verb. This includes, not only languages which are clearly VSO or VOS, but also languages like Fijian and Shuswap in which both VSO and VOS are common and for which there appears to be little basis for choosing one or the other order as basic, as well as languages which are clearly VS and VO, but for which I have been unable to determine from my sources what order, if any, is basic for clauses containing both a lexical subject and a lexical object. It should be noted that many of these languages are not literally *v-initial*, in that various elements other than the subject and object, such as negative words, tense-aspect particles, and question particles, normally precede the verb if they occur. I use the term *v-FINAL* in an analogous way, to refer to languages in which both subjects and objects generally precede the verb. This includes both SOV languages and languages which are clearly SV and OV, but for which I have been unable to ascertain a basic order for clauses containing both a lexical subject and a lexical object. Most of the languages of the latter sort are probably SOV, but my sources do not provide unequivocal evidence for this. Again, many of these languages are not literally *v-final*, since again various elements may normally follow the verb. This includes a few SOVX languages, in which adpositional phrases follow the verb.

I will distinguish three kinds of arguments in the literature against collapsing V-initial and SVO languages into a single type VO. The first argument is that while exceptionless generalizations can be made about V-initial languages (and perhaps V-final languages), this is not the case with SVO languages. The second argument makes a similar claim with reference to STATISTICAL generalizations. The third argument is specifically due to Hawkins (1982, 1983), who argues, as part of his principle of Cross-Category Harmony, that while V-initial languages most commonly place all dependents after their heads and while V-final languages most commonly place all dependents before their heads, SVO languages most commonly place some dependents before their heads and some after. I will address each of these three arguments in sections 2, 3 and 4 respectively.

All of the arguments in this paper are based on evidence from a large crosslinguistic database on word order that contains word order data for 603 languages (cf. Dryer 1988a, 1988b, 1989a, 1989c).<sup>1</sup> It will be shown that previous conclusions about SVO languages were based, to some extent, on unrepresentative samples of languages, and that while some of the arguments had some merit given the available evidence, they do not survive the evidence that is now available.

## 2. THE LACK OF EXCEPTIONLESS GENERALIZATIONS ABOUT SVO LANGUAGES

The first argument offered against collapsing SVO languages with V-initial languages into a single type VO is that while there are exceptionless universals about V-initial languages and V-final languages, there are no such exceptionless universals about SVO languages, and hence no exceptionless universals about VO languages. Hawkins (1979: 199) notes that the thirteen implicational universals in Greenberg (1963) that refer to the order of subject and object relative to the verb apply either to VSO languages or to SOV languages, never to SVO languages. Comrie (1981: 90; 1989: 96) pursues the same point:

Knowing that a language is VSO or VOS, we can predict its values for other word order parameters; knowing that a language is SOV, we can with considerable reliability predict its other word order parameter values; knowing that a language is SVO, we can predict virtually nothing else.

Comrie's wording implies a strong claim about V-initial languages, that one can predict their other word order characteristics, apparently with greater than 'considerable reliability'. What Comrie appears to be claiming here is that there are a large number of exceptionless generalizations to be made about V-initial languages.

### 2.1. *Exceptionless generalizations in Greenberg's 30-Language sample*

If one examines the six V-initial languages in Greenberg's 30-language sample (Berber, Hebrew, Maori, Maasai, Welsh, and Zapotec), one finds that they are indeed

consistent with respect to a wide range of characteristics. The properties in (1) are exceptionless among these six V-initial languages.<sup>2</sup>

- (1) a. prepositional  
 b. adjective follows noun  
 c. genitive follows noun  
 d. verbal auxiliary precedes verb  
 e. intensifier (words like *very* and *more* in English) follows adjective  
 f. question particle marking yes/no question occurs in initial position rather than final position  
 g. *wh*-word occurs in sentence-initial position rather than *in situ* (or other non-initial position)  
 h. relative clause follows noun  
 i. adjective-standard order in comparative construction<sup>3</sup>

However, when we examine a larger sample of V-initial languages, we find many V-initial languages which are inconsistent with respect to the characteristics in question. Although my database contains no exceptions to the last two characteristics in (1), it does contain exceptions to the first seven characteristics, listed in Table 1.<sup>4</sup>

Exceptional Type	Number of exceptions in database	Language	Word Order	Family	Location
Postpositional	4	N. Tepehuan	VSO	Uto-Aztecan	N Mexico
		Cora	VSO	Uto-Aztecan	C Mexico
		Guajajara	VSO	Tupian	E Brazil
		Yagua	VSO	isolate (Macro-Carib?)	Peru
AdjN	21	Gude	VSO	Chadic	Nigeria, Cameroon
		Car Nicobarese	VOS	Mon Khmer	Bay of Bengal
		Rukai	VS, VO	Austronesian	Taiwan
		Chamorro	VS, VO	Austronesian	W Pacific
		Central Agta	VSO	Austronesian	Philippines
		Manobo	VSO	Austronesian	Philippines
		Wembawemba	VOS	Pama-Nyungan	SE Australia
		Quileute	VSO	Chimakuan	NW USA
		Kwakiutl	VSO	Wakashan	W Canada
		Squamish	VSO	Salishan	W Canada
		Coast Tsimshian	VSO	Penutian	W Canada
		Yokuts	VSO	Penutian	California
		Lower Chinook	VOS	Penutian	NW USA
		Northern Sahaptin	VSO	Penutian	NW USA
		Aleas	VS, VO	Penutian (?)	NW USA
		Huamelultec Chontal	VS, VO	Tequistlatecan	S Mexico
Jacaltec	VSO	Mayan	Guatemala		
Papago	VS, VO	Uto-Aztecan	SW USA		

		Pipil	VOS	Uto-Aztecan	El Salvador
		Otomi	VOS	Oto-Manguean	S Mexico
		Cayuvava	VOS	isolate (Equatorial?)	Bolivia
GenN	7	Kilivila	VS, VO	Austronesian	E Papua
		Garawa	VOS	Australian	NE Australia
		Wembawemba	VOS	Pama-Nyungan	SE Australia
		Alsea	VS, VO	Penutian (?)	NW USA
		Northern Tepehuan	VSO	Uto-Aztecan	N Mexico
		Guajajara	VSO	Tupian	E Brazil
		Yagua	VSO	isolate (Macro-Carib?)	Peru
VAux	1	Island Carib	VSO	Maipuran	Belize
IntensAdj	15	Scots Gaelic	VSO	Indo-European	Scotland
		Chamorro	VS, VO	Austronesian	W Pacific
		Mamanwa	VSO	Austronesian	Philippines
		Toba Batak	VS, VO	Austronesian	W Indonesia
		Yapese	VSO	Austronesian	W Pacific
		Wembawemba	VOS	Pama-Nyungan	SE Australia
		Coast Tsimshian	VSO	Penutian	W Canada
		Jacaltec	VSO	Mayan	Guatemala
		Papago	VS, VO	Uto-Aztecan	SW USA
		N. Tepehuan	VSO	Uto-Aztecan	N Mexico
		Huasteca Nahuatl	VSO	Uto-Aztecan	S Mexico
		Otomi	VOS	Oto-Manguean	S Mexico
		Tlapaneca	VSO	Oto-Manguean	S Mexico
		Cayuvava	VOS	isolate (Equatorial?)	Bolivia
		Yagua	VSO	isolate (Macro-Carib?)	Peru
Final Q-particle	6	Karimojong	VSO	Nilotic	Uganda
		Lamang	VSO	Chadic	Nigeria, Cameroon
		Gude	VSO	Chadic	Nigeria, Cameroon
		Jicaltepec Mixtec	VSO	Mixtecan	S Mexico
		Peñoles Mixtec	VSO	Mixtecan	S Mexico
		Chatino	VSO	Oto-Manguean	S Mexico

<u>wh-in situ</u>	13	Karimojong	VSO	Nilotic	Uganda
		Pokot	VSO	Nilotic	Kenya, Uganda
		Rukai	VS, VO	Austronesian	Taiwan
		Bikol	VS, VO	Austronesian	Philippines
		Toba Batak	VS, VO	Austronesian	W Indonesia
		Kiribatese	VOS	Austronesian	C Pacific
		Fijian	VS, VO	Austronesian	S Pacific
		Niuean	VSO	Austronesian	S Pacific
		Tongan	VS, VO	Austronesian	S Pacific
		Samoan	VS, VO	Austronesian	S Pacific
		Tahitian	VSO	Austronesian	S Pacific
		Hawaiian	VSO	Austronesian	N Pacific
		Papago	VS, VO	Uto-Aztecan	SW USA

Table 1

Exceptions to generalizations about V-initial languages that are exceptionless in Greenberg's 30-Language sample

Thus, despite the lack of exceptions in Greenberg's 30-language sample, we see that we can in no way predict with certainty these word order characteristics for V-initial languages. There are three characteristics that are not attested of V-initial languages in my database, namely RelN (relative clause before noun), PP-V (adpositional phrase before verb), and Standard-Adjective (in comparative structures).<sup>5</sup> Since we cannot predict the other properties of V-initial languages with certainty, V-initial languages do not differ substantially from SVO languages in this respect. At most we can say that SVO languages differ from V-initial languages in that there exist a few generalizations about V-initial languages for which no exceptions are known, while this is not the case for SVO languages. It is worth noting, however, that these three characteristics not attested for V-initial languages are also very rare among SVO languages: the Chinese languages are the only ones in my database which have these three properties. The fact that the only languages with these characteristics among VO languages happen to be SVO rather than V-initial may just be a coincidence or may simply reflect that fact that there are more SVO languages in the world so we have a greater chance of finding an exception. In short, the available evidence provides no basis for claiming that we can predict with certainty the properties of V-initial languages in a way that we cannot with SVO languages.

It should be noted that for 6 of the 9 properties in (1) that are exceptionless among the V-initial languages in Greenberg's 30-language sample, Greenberg (1963) himself made no claim that they were exceptionless. He did claim that VSO languages are always prepositional (Universal 3), that they always place inflected auxiliaries before the main verb (Universal 16), and that they always place interrogative words or phrases first in interrogative word questions (Universal 12). The examples in Table 1 under these categories constitute counterexamples to the claim that these generalizations are exceptionless. His Universal 17 makes a statistical claim about the order of noun and adjective in VSO languages, claiming that they are generally NAdj. I have argued in

Dryer (1988a) that V-initial languages are no more likely than V-final languages to be NAdj; in other words, Greenberg's Universal 17 is apparently not correct even as a statistical generalization. Greenberg made no explicit claim that VSO languages tend to have the other five properties in (1), though his Universal 2 claims that languages with prepositions are almost always NGen, which, combined with his claim that VSO languages are always prepositional, predicts that VSO languages will usually be NGen.

## 2.2. *Hawkins' implicational universals*

While the data in Table 1 show that there exist many exceptions to generalizations about V-initial languages that are exceptionless in Greenberg's 30-language sample, this does not fully address the argument. Hawkins (1983) observes that there are many other generalizations which are exceptionless even in his expansion of Greenberg's appendix. These generalizations are formulated as implicational universals that refer to three or more characteristics. Thus it might be the case that it is easier to formulate such complex implicational universals with reference to V-initial languages than with reference to SVO languages. But again there is little evidence that this is the case.

Hawkins proposes five implicational universals which are exceptionless in his data and which apply to V-initial languages. One of these, his (XIII), refers to ' $\sim$ SOV', in other words SVO and V-initial languages.

(XIII)  $\text{Prep} \supset (\sim\text{SOV} \supset \text{NRel})$

But since this universal applies to both SVO and V-initial languages, it provides no argument that it is more difficult to formulate exceptionless implicational universals with respect to SVO languages.<sup>6</sup> Furthermore, ignoring object-initial languages (of which far too few have been attested to usefully test universals about them), this universal could just as well be formulated with reference to VO languages as follows:

(XIII')  $\text{Prep} \supset (\text{VO} \supset \text{NRel})$

Since this universal applies to VO languages, it provides no argument against the utility of the class of VO languages.

Of the remaining four implicational universals in Hawkins (1983) that apply to V-initial languages, my database yields exceptions to three, listed in Table 2.<sup>7</sup>

	Universal	Exception	Type
(II')	$V-1 \supset (NA \supset NG)$	Kilivila Garawa Yagua Guajajara	VOS&NA&GN VOS&NA&GN VSO&NA&GN VSO&NA&GN
(III')	$Prep \ \& \ \sim SVO \supset (NA \supset NG)$	Kilivila Garawa	Prep&VOS&NA&GN Prep&VOS&NA&GN
(VII')	$Prep \ \& \ \sim SVO \supset (NDem \supset NG)$	Kilivila	Prep&VOS&NDem&GN

Table 2  
Exceptions to Hawkins' universals referring to V-initial languages

Only one of Hawkins' implicational universals that applies to V-initial languages, namely (VIII'), remains exceptionless in my database:

(VIII')  $Prep \ \& \ \sim SVO \supset (NNum \supset NG)$

There are other generalizations that apply to SVO languages for which I am aware of no exceptions. These include the use of clause-initial complementizers and the absence of head-internal relative clauses. There is thus little reason to say that SVO languages are different from V-initial and V-final languages in not allowing exceptionless generalizations. Rather, SVO languages are like V-initial and V-final languages in that few exceptionless generalizations can be made about them.

### 3. ARE SVO LANGUAGES INTERMEDIATE BETWEEN V-INITIAL AND V-FINAL LANGUAGES?

While there may be few exceptionless generalizations to be made about V-initial languages and V-final languages, one might still retreat to a position like that expressed with respect to SOV languages in the quotation from Comrie (1981, 1989) cited above: perhaps we can predict the properties of V-initial languages and V-final languages with 'considerable reliability' in a way that we cannot with SVO languages. This claim is somewhat more difficult to test: under exactly what conditions would we conclude that a property can be predicted with 'considerable reliability' for a given language type? Because of this vagueness, I will ask a somewhat different question: to what extent do SVO languages differ from V-initial languages with respect to their other word order characteristics? Are they intermediate between V-final and V-initial languages with respect to other word order characteristics, as Comrie (1981, 1989) and Mallinson and

Blake (1981) imply, or do they pattern like V-initial languages, as Lehmann (1973, 1978) and Vennemann (1974, 1976) assume? I will show that for a large number of characteristics, SVO languages in fact differ very little from V-initial languages, that insofar as there is a tendency for V-initial languages to exhibit a particular property, SVO languages exhibit the same property. And since these properties are typical of both V-initial languages and SVO languages, they are typical of VO languages in general, thereby providing justification for the basic distinction between OV and VO languages.

### 3.1. *Properties shared by SVO and V-initial languages*

#### 3.1.1. *Prepositions*

The evidence presented here is based on the database of 603 languages described above. Table 3 presents data from this database on the relationship between clause order and adposition type (prepositions vs. postpositions) in a manner that is explained and justified in detail in Dryer (1989c). First, the languages are assigned to genetic groups called GENERA, which correspond roughly to the subfamilies of Indo-European. The counts below represent numbers of genera that contain languages of the given sort rather than numbers of languages. Second, the genera are divided into six large continental areas.<sup>8</sup> Table 3 and similar tables below show the number of genera in each area that contain languages in my database of the given type. For example, the 15 in the upper lefthand cell of Table 3 indicates that there are 15 genera in Africa that contain languages in my database which are V-final&Postpositional, while the 3 below it indicates that there are 3 genera in Africa that contain languages which are V-final&Prepositional.

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
V-final&Postp	<b>15</b>	<b>22</b>	<b>5</b>	<b>16</b>	<b>23</b>	<b>16</b>	97
V-final&Prep	3	2	0	0	0	0	5
Proportion Postp	.83	.92	1.00	1.00	1.00	1.00	Avg.=.96
SVO&Postp	3	1	0	0	0	2	6
SVO&Prep	<b>15</b>	<b>5</b>	<b>12</b>	<b>4</b>	<b>4</b>	2	42
Proportion Postp	.17	.17	.00	.00	.00	.50	Avg.=.14
V-initial&Postp	0	0	0	0	2	2	4
V-initial&Prep	<b>5</b>	<b>1</b>	<b>6</b>	<b>1</b>	<b>15</b>	<b>3</b>	31
Proportion Postp	.00	.00	.00	.00	.12	.40	Avg.=.09

Table 3  
Adposition type

Key: The numbers indicate the number of genera containing languages of the given type in the given area. The larger of the two numbers for each area and for each order of verb

and object is enclosed in a box. Africa includes Semitic languages of southwest Asia; Eurasia = Europe and Asia, except for southeast Asia, as defined immediately; SEAsia&Oc = Southeast Asia (Sino-Tibetan, Thai, and Mon-Khmer) and Oceania (Austronesian); Aus-NewGui = Australia and New Guinea, excluding Austronesian languages of New Guinea; NAmer = North America, including languages of Mexico, as well as Mayan and Aztec languages in Central America; SAmer = South America, including languages in Central America that are neither Mayan nor Aztec.

In general, languages within a genus are identical with respect to the typological features being examined. Occasionally, however, a genus will contain more than one of the types being examined. In such cases, a genus will be represented in more than one cell in the table. For example, there are 2 genera in Africa that contain languages that are V-final&Postp as well as languages that are V-final&Prep. These 2 genera are thus included both among the 15 genera containing languages that are V-final&Postp and among the 3 genera containing languages that are V-final&Prep. This should be borne in mind in understanding the third line of Table 3, which gives the proportion for V-final&Postp. This figure is computed by taking the figure on the first line as a proportion of the sum of the figure on the first line and the figure on the second line. In the case of the third line of Table 3, this proportion is computed by calculating 15 as a proportion of 18 (15+3), namely .83. I will refer to this figure as the PROPORTION OF GENERA. It is important to stress that since both figures may include the same genus, this proportion is not, strictly speaking, the proportion of genera that contain languages which are V-final&Postp among genera containing languages which are V-final. Since there are 2 genera in Africa containing languages which are V-final&Postp as well as languages which are V-final&Prep, the total number of genera in Africa containing languages which are V-final (and for which I have data on adposition type), is 16, not 18. Nevertheless, for ease of computation, I use the sum of the number of genera, in this case 18, in computing the proportion. For this reason, given two types A&B and A&~B, the proportion of genera that are B should be interpreted as the number of genera containing languages that are A&B as a proportion of the sum of the number of genera containing languages that are A&B plus the number of genera containing languages that are A&~B.

The motivation behind the use of genera is discussed in detail in Dryer (1989c). Counting genera, rather than languages, controls for the most severe kind of genetic bias; differences in numbers of languages can reflect a type that happens to be dominant in a small number of genetic groups that contain a large number of languages. In determining the genera, I have attempted to identify groups whose time depth is not more than 4000 years or so, but whose immediate subgroups do not have a time depth greater than 3500 years. Clearly, our current knowledge about the time depth of most genetic groups is sufficiently unknown that there is considerable guesswork on my part in identifying genera. My decisions regarding genera are based on published estimates of time depths, informal estimates from experts in particular groups, and my own impressions about the rough genetic distance between groups, as reflected both in the descriptions of the languages themselves and in the discussions in the literature regarding the genetic classification of various groups. In general, genera represent groups whose time depth is

sufficiently shallow that few people if any would doubt the genetic relatedness of the languages within the group. Thus, if I find evidence in the literature of linguists questioning whether the languages within a putative group are genetically related, I take this as evidence that the group must be of a time depth greater than that of a genus and thus must contain more than one genus. For example, within North America, Campbell and Mithun (1979) provide a list of the the minimal genetic groups whose validity nobody questions. I therefore conclude that no group subsuming more than one of these groups is a genus. In fact, most of these groups I treat as a genus. In a few cases, however, published estimates of the time depth of the group and/or my own observations of the typological variation within the group have led me to conclude that the groups contain more than one genus: Salish and Uto-Aztecan are examples of groups like this, and I treat the generally accepted subgroups of these families as genera. It is worth noting that Nichols (1990) employs the term FAMILY in a way that is apparently equivalent to my notion of genus, and her guesses as to which groups are families (arrived at independently of my work) are remarkably similar to my guesses. Clearly, however, my decisions regarding which groups are genera are subject to dispute, and some of them are undoubtedly wrong.

The effect of counting genera rather than languages is that where closely-related languages exhibit the same characteristics, they are not counted twice. This controls for the most severe type of genetic bias. Breaking the genera down into six areas of the world allows one to control for the risk of one area having an inordinate effect on overall numbers: a difference in overall numbers can often be shown to reflect the properties of languages in a single area of the world rather than a general property of language. Consider a case where there are more genera in the entire world containing languages with property A than there are genera containing languages with property B. If the number of genera containing languages with property A is greater than the number of genera containing languages with property B in each of the six areas of the world, then we have reason to believe that the difference in numbers between the two types reflects a general property of language. If, on the other hand, the difference between the number of genera of the two types can be attributed entirely to genera in one area of the world, then we should be concerned that the difference in numbers may just reflect something about that one area, rather than a general property of language. The fact that the number of genera containing V-final&Postp languages is greater than the number of genera containing V-final&Prep languages in each of the six areas of the world, as shown by the first two lines of Table 3, provides us with reason to believe that that there is a preference for postpositions among V-final languages. The average of these proportions (shown in the righthand column), in this case .96, is often the most revealing single statistic. What this means is that if we took a random V-final language that is not in the database and not related to languages in the database then the chance of its being postpositional can be estimated as 96%.

The three clause order types that are compared in Table 3 are V-final, SVO, and V-initial. As noted already, the first three lines show a preference for postpositions among V-final languages. The final three lines give comparable data for V-initial languages. They show that V-initial languages exhibit a strong tendency to be

prepositional, the opposite of what is found in V-final languages: in all six areas of the world, there are more genera containing languages which are V-initial&Prepositional than there are genera containing languages which are V-initial&Postpositional. The average proportion for V-initial&Postpositional over areas is .09, showing that there is only a 9% chance of a random V-initial language employing postpositions rather than prepositions. The middle lines of Table 3 give the data for SVO languages. It can be seen that SVO languages tend to be prepositional almost as overwhelmingly as V-initial languages: the average proportion over areas that are postpositional is .14, indicating that the chance of a random SVO language being postpositional is only 14%, almost as low as the figure for V-initial languages. While this figure is literally intermediate between the 96% for V-final languages and the 9% for V-initial languages, it is clearly much closer to the figure for V-initial languages. In fact, this difference between SVO and V-initial languages is not statistically significant and may just be due to random variation.<sup>9</sup> We can conclude, therefore, that at least with respect to adposition type, SVO languages pattern very similarly to V-initial languages. The appropriate generalization about adpositions SHOULD make reference to the distinction between OV and VO languages: OV languages tend to be postpositional, while VO languages tend to be prepositional.

### 3.1.2. *Noun + relative clause*

Table 4 shows analogous data for the order of noun and relative clause among the three language types.<sup>10</sup>

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
V-final&RelN	5	11	2	2	2	3	25
V-final&NRel	8	4	2	4	11	3	32
Proportion RelN	.38	.73	.50	.33	.15	.50	Avg.=.43
SVO&RelN	0	0	1	0	0	0	1
SVO&NRel	19	5	11	3	2	2	42
Proportion RelN	.00	.00	.08	.00	.00	.00	Avg.=.01
V-initial&RelN	0	0	0	0	0	0	0
V-initial&NRel	5	1	3	0	9	3	21
Proportion RelN	.00	.00	.00	--	.00	.00	Avg.=.00

Table 4  
Order of noun and relative clause

The first three lines of Table 4 show that contrary to a common view (cf. Lehmann 1973: 48), V-final languages do not tend to place the relative clause before the noun. As noted by Hawkins (1990: 241), both orders are common in V-final languages. Table 4, shows that a random V-final language has a likelihood of only 43% of being RelN, and only in

Eurasia are there more genera containing languages that are V-final&RelN than V-final&NRel. On the other hand, the last three lines of Table 4 show that V-initial languages are overwhelmingly NRel: there are in fact no instances of V-initial RelN languages in my database. The middle section of Table 4 shows that SVO languages again pattern much more like V-initial languages: there is only one instance in my database of a genus containing languages that are SVO&RelN, namely Chinese. While the likelihood of a random language being RelN is 43% if the language is V-final, it is only 1% if it is SVO. Again SVO languages pattern very similarly to V-initial languages and very differently from V-final languages. And while, if a language is V-final, we cannot predict the order of noun and relative clause, if the language is VO, we can predict with a very high level of probability that it will be NRel.

### 3.1.3. *Adjective + standard of comparison*

The order of standard of comparison with respect to the adjective also provides clear evidence of how SVO languages pattern like V-initial languages. The relevant data is given in Table 5.

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
V-final&StAdj	3	14	3	0	4	3	27
V-final&AdjSt	2	0	0	0	0	3	5
Proportion StAdj	.60	1.00	1.00	---	1.00	.50	Avg.=.82
SVO&StAdj	0	0	1	0	0	0	1
SVO&AdjSt	5	5	1	0	2	1	21
Proportion StAdj	.00	.00	.11	.00	.00	.00	Avg.=.02
V-initial&StAdj	0	0	0	0	0	0	0
V-initial&AdjSt	2	1	3	0	5	3	14
Proportion StAdj	.00	.00	.00	--	.00	.00	Avg.=.00

Table 5  
Order of standard of comparison and adjective

Table 5 shows a very clear difference between V-final languages on the one hand and the SVO and V-initial languages on the other. While the average proportion for StAdj among V-final languages is .82, the figure for SVO languages is only .02, close to the zero figure for V-initial languages.

### 3.1.4. *Copula + predicate*

A further pair of elements for which SVO languages are not intermediate between V-initial and V-final languages is that of copula and predicate. While many languages lack a copula, many other languages use a copula when the predicate is adjectival,

nominal, or locative. Table 6 gives the data for the relationship between clause order type and the order of copula and predicate.

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
V-final&PredCop	8	16	2	7	9	13	55
V-final&CopPred	3	0	1	1	1	1	7
Proportion PredCop	.73	1.00	.67	.88	.90	.93	Avg.=.85
SVO&PredCop	2	1	0	0	1	1	5
SVO&CopPred	15	5	10	0	1	1	32
Proportion PredCop	.12	.17	.00	--	.50	.50	Avg.=.26
V-initial&PredCop	0	0	0	1	4	1	6
V-initial&CopPred	3	1	1	0	7	0	12
Proportion PredCop	.00	.00	.00	1.00	.36	1.00	Avg.=.39

Table 6  
Order of copula and predicate

While the correlation is not as strong for this pair of elements as it is for a number of other pairs of elements, there is a clear difference between V-final and SVO languages. Furthermore, in this case it is V-initial languages rather than SVO languages which exhibit a pattern that is intermediate between the other two types: the average proportions that are PredCop are .85 for V-final, .26 for SVO, and .39 for V-initial.<sup>11</sup> However, the figure for V-initial is much closer to SVO than it is to V-final.

### 3.1.5. *Adverbial subordinators*

By ADVERBIAL SUBORDINATOR, I mean words, like *when* and *because* in English, which introduce or follow adverbial subordinate clauses, indicating their semantic relationship to the main clause. Table 7 shows that while such words more often follow the clause (SSub) in V-final languages, they almost always precede (SubS) in both SVO and V-initial languages. Again, we have clear evidence of SVO and V-initial languages patterning the same, but very differently from V-final languages.

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
V-final&SSub	4	7	2	5	9	10	37
V-final&SubS	5	6	2	1	1	0	15
Proportion SSub	.44	.54	.50	.83	.90	1.00	Avg.=.70
SVO&SSub	0	0	0	0	0	1	1
SVO&SubS	15	5	10	2	3	2	37
Proportion SSub	.00	.00	.00	.00	.00	.33	Avg.=.06
V-initial&SSub	0	0	0	0	0	1	1
V-initial&SubS	4	1	4	1	10	2	22
Proportion SSub	.00	.00	.00	.00	.00	.33	Avg.=.06

Table 7  
Order of clause and adverbial subordinator

### 3.1.6. *Plural words*

The final word order characteristic which I will discuss in detail that supports the claim that SVO languages pattern like V-initial is that of the order of plural words with respect to the noun. Plural words, discussed in detail in Dryer (1989d), are words found in a minority of the languages of the world whose function is similar to that of plural affixes on nouns, but which are separate words, as in (2).

- (2) Gbeya:      *ó*    *túÇ*    *wí-ré*  
                   *plur* black person  
                   black people (Samarin 1966: 81)

The numbers for plural words in Table 8 are small, reflecting the fact that they are not common. Table 8 shows, however, that these words more often precede the noun in SVO languages, as in V-initial languages, but quite unlike V-final languages, in which these words invariably *follow* the noun in my data.

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
V-final&NPlur	2	0	3	5	0	3	13
V-final&PlurN	0	0	0	0	0	0	0
Proportion NPlur	1.00	---	1.00	1.00	---	1.00	Avg.=1.00
SVO&NPlur	5	0	0	0	0	0	5
SVO&PlurN	2	0	5	1	0	0	8
Proportion NPlur	.71	---	.11	.00	---	---	Avg.=.24
V-initial&NPlur	0	0	2	0	0	0	2
V-initial&PlurN	0	0	3	0	2	1	6
Proportion NPlur	---	---	.40	---	.00	.00	Avg.=.13

Table 8  
Order of noun and plural word

### 3.1.7. *Other pairs of elements for which SVO patterns with V-initial*

In the last six sections I have provided detailed evidence for six pairs of elements, showing that their order patterns close to that found in V-initial languages and different from that found in V-final languages. In this section I briefly summarize similar evidence for five other pairs of elements. Table 9 gives the average of proportions of genera for four of these pairs for the three word order types.<sup>12</sup>

	V-final	SVO	V-initial
Adpositional Phrase-V	.90	.01	.00
Manner Adverb-V	.91	.25	.17
V-Tense/Aspect Auxiliary Verb	.94	.21	.13
V-Negative Auxiliary	.88	.13	.00

Table 9  
Average of proportions of genera over areas for other pairs of elements  
showing SVO patterning like V-initial

Key: A tense/aspect auxiliary verb is an auxiliary verb whose primary function is to indicate tense or aspect, like the auxiliary *have* in English. A negative auxiliary is a negative word which is itself a type of auxiliary verb (rather than being a nonverbal particle like *not* as in English) (cf. Dahl 1979).

For all four pairs of elements in Table 9, the average of proportions for SVO is intermediate between V-initial and V-final, but the figure for SVO is much closer to V-initial in all four cases.

A final pair of elements that is apparently relevant here is that of complementizer and clause. I do not have detailed data of the same sort to report at this time for this pair of elements, but all of the languages that I am aware of with clause-final complementizers are V-final (cf. Dryer 1980). SVO languages apparently share with V-initial languages the property of never having clause-final complementizers.

There are therefore a total of 11 pairs of elements whose order in SVO languages closely resembles that found in V-initial languages and contrasts with that found in V-final languages. These pairs of elements thus provide support for the notion of a VO type. At most, it must be admitted that for 8 of these 11 pairs the pattern for SVO languages is, strictly speaking, intermediate between V-initial and V-final. Nevertheless, the pattern for SVO is much closer to V-initial than it is to V-final. The mean difference between the averages of proportions for V-initial and SVO for these 8 pairs is .06, while the mean difference between the averages of proportions for SVO and V-final is .71. In other words, the difference between SVO and V-final is almost 12 times as large as the difference between SVO and V-initial. In fact, the difference between SVO and V-initial is often so small as to be well within the range of chance; we do not want to attribute great significance to differences that may easily be due to random variation.

### 3.2. *Some unrepresentative properties of Greenberg's 30-language sample*

#### 3.2.1. *Postpositional SVO languages*

Given the evidence presented in section 2 demonstrating that SVO languages are much more like V-initial languages than they are like V-final languages, we might ask where the idea arose that SVO is a mixed type, intermediate between V-initial and V-final languages? Hawkins (1983: 30) provides a table that summarizes the relationship between various pairs of elements and clause order in Greenberg's 30-language sample, and he refers to this table in supporting his claim that SVO is a mixed type. And for a number of word order parameters, SVO languages do indeed exhibit a pattern in Greenberg's 30-language sample that is intermediate between V-initial languages and V-final languages. Among these is adposition type. The distribution for this in Greenberg's sample is given in Table 10.



It turns out, however, that, despite possible appearances to the contrary in Greenberg's sample, the three pairs of elements in Table 11 are all ones which do not exhibit any correlation with the order of object and verb, for which there is no significant difference between V-initial and V-final languages.

### 3.2.2.1. *Order of adjective and noun*

It is shown in Dryer (1988a) that the order of adjective and noun does not correlate with the order of object and verb despite a widely-held belief that it does. Table 12 gives the data for adjective and noun in the same format as Tables 3 to 8.

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
V-final&AdjN	7	19	2	4	9	7	48
V-final&NAdj	16	4	5	13	16	11	65
Proportion AdjN	.30	.83	.29	.24	.36	.39	Avg.=.40
SVO&AdjN	2	4	2	4	2	1	15
SVO&NAdj	22	2	11	1	3	2	41
Proportion AdjN	.08	.67	.15	.80	.40	.33	Avg.=.41
V-initial&AdjN	1	0	3	1	12	1	18
V-initial&NAdj	6	1	3	1	5	4	20
Proportion AdjN	.14	1.00	.50	.50	.71	.20	Avg.=.34

Table 12  
Order of adjective and noun

The overall averages of proportions that are AdjN for the three clause types are close and well within the realm of accident: .40 for V-final, .41 for SVO, and .34 for V-initial. There is thus no evidence of any correlation between the order of adjective and noun and the order of object and verb. While Greenberg's sample also reflects the overall preference for NAdj order, as shown in Table 11, his sample is unrepresentative in suggesting that V-initial languages are less likely to be AdjN than V-final languages.

### 3.2.2.2. *Order of demonstrative and noun*

Table 13 gives the data for the order of demonstrative and noun.

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
V-final&DemN	10	17	4	7	19	12	69
V-final&NDem	10	1	2	9	5	4	31
Proportion DemN	.50	.94	.67	.44	.79	.75	Avg.=.68
SVO&DemN	3	6	5	4	5	4	27
SVO&NDem	22	0	10	0	0	0	32
Proportion DemN	.12	1.00	.33	1.00	1.00	1.00	Avg.=.74
V-initial&DemN	2	0	4	2	14	5	27
V-initial&NDem	4	1	3	0	5	1	14
Proportion DemN	.33	.00	.57	1.00	.74	.83	Avg.=.58

Table 13  
Order of demonstrative and noun

As with the order of adjective and noun, we see no significant difference among the three clause types as far as order of demonstrative and noun is concerned: the average proportions that are DemN are .68 for V-final, .74 for SVO, and .58 for V-initial. Insofar as there is a difference, SVO is not intermediate between the other two orders.

### 3.2.2.3. *Order of intensifier and adjective*

We come now to the order of intensifier (like *very* in English) and adjective. Greenberg's sample, as shown in Table 11, suggests a strong correlation, but my data, as shown in Table 14, shows a very different pattern.

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
V-final&IntensAdj	4	13	2	4	7	7	37
V-final&AdjIntens	2	0	2	8	4	6	22
Proportion IntensAdj	.67	1.00	.50	.33	.64	.54	Avg.=.61
SVO&IntensAdj	0	5	4	0	4	1	14
SVO&AdjIntens	7	0	6	1	0	0	14
Proportion IntensAdj	.00	1.00	.40	.00	1.00	1.00	Avg.=.57
V-initial&IntensAdj	0	1	3	1	6	2	13
V-initial&AdjIntens	2	1	3	0	3	3	12
Proportion IntensAdj	.00	.50	.50	1.00	.67	.40	Avg.=.51

Table 14  
Order of intensifier and adjective

While there is a slight trend in Table 14 in the same direction as Greenberg's data, and while SVO is intermediate between the other two types, the differences among the three clause types are small and well within the realm of accident. This is strikingly different from Greenberg's sample: both V-initial languages for which Greenberg found a basic order are AdjIntens, while all 10 V-final languages are IntensAdj.<sup>13</sup> In addition, SVO languages are intermediate between V-initial and V-final in his sample, with 6 IntensAdj languages and 6 AdjIntens languages. To a certain extent the different results reflect the fact that 20 of the 30 languages in Greenberg's sample are spoken in Africa and Eurasia. The languages in my sample from these two areas do exhibit a pattern somewhat more like the languages in Greenberg's sample. Hence, to some extent, the difference reflects the areal bias in Greenberg's sample. But to a large extent, the differences simply reflect the accidental properties that can arise in a small sample like Greenberg's.

We can conclude then that the pairs of elements discussed in the last section do not provide any reason to believe that SVO languages are intermediate between V-initial and V-final languages, since these are pairs of elements for which there is no significant difference between V-initial and V-final languages. They therefore provide no evidence against the claim that there is a basic dichotomy between VO and OV languages.

### 3.3. *Characteristics for which SVO is a mixed type*

In the evidence presented so far, we have seen that SVO languages pattern very much like V-initial languages. The question then arises whether there is any reason to treat SVO as a separate category. It turns out that while SVO languages pattern like V-

initial languages on the whole, there are some ways in which they do pattern differently. In this section, I will discuss three word order characteristics for which SVO languages do exhibit properties that are intermediate between those of V-final languages and those of V-initial languages. This evidence shows that the criticisms of the OV:VO typology are partly right, since there do exist some ways in which SVO languages are indeed a mixed type. However, the other evidence presented in this paper shows that the OV:VO typology is justified since apart from the three characteristics discussed in this section, SVO languages do pattern very much like V-initial languages.

### 3.3.1. *Order of genitive and noun*

One characteristic for which SVO languages behave as a mixed type is the order of noun and genitive, illustrated in Table 15.

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
V-final&GenN	<u>17</u>	<u>18</u>	<u>5</u>	<u>15</u>	<u>28</u>	<u>19</u>	102
V-final&NGen	5	3	1	2	0	0	11
Proportion GenN	.77	.86	.83	.88	1.00	1.00	Avg.=.89
SVO&GenN	5	3	4	<u>5</u>	2	<u>3</u>	22
SVO&NGen	<u>20</u>	3	<u>9</u>	0	2	0	34
Proportion GenN	.20	.50	.31	1.00	.50	1.00	Avg.=.59
V-initial&GenN	0	0	1	<u>2</u>	2	2	7
V-initial&NGen	<u>6</u>	<u>1</u>	<u>6</u>	0	<u>17</u>	<u>3</u>	33
Proportion GenN	.00	.00	.14	1.00	.11	.40	Avg.=.28

Table 15  
Order of noun and genitive

The first three lines of Table 15 show that V-final languages exhibit a strong preference for GenN order: the average proportion is .89. The last three lines of Table 15 show that V-initial languages exhibit a preference for NGen order, with an average proportion for GenN of .28.<sup>14</sup> Among SVO languages, the two orders of noun and genitive are about equally common. While the total number of genera is higher for SVO&NGen (34 vs. 22), this figure is higher only because of the large number of genera containing such languages in Africa: 20 out of the 34 genera containing such languages are in Africa. The average proportion for NGen is actually less than that for GenN (.41 vs. .59), suggesting that there might even be a preference for GenN order among SVO languages, but this difference is sufficiently small that it could well be due to chance.

In short, the order of noun and genitive is one word order characteristic for which SVO languages are intermediate between V-final and V-initial languages. A natural interpretation of this is that the order of noun and genitive exhibits a correlation with the order of verb and subject as well as with the order of verb and object. A plausible explanation for this is that it reflects the parallels between the relationship of a genitive to a noun and the relationship of a subject to a verb, as demonstrated by other morphosyntactic similarities between genitives and subjects in many languages (cf. Allen 1964, Greenberg 1963: 99). The crucial point for our purposes, however, is that the order of genitive and noun is one characteristic for which SVO languages do not pattern the same as V-initial languages, for which SVO languages ARE intermediate between V-initial and V-final languages.

### 3.3.2. *Position of question particles*

A second characteristic for which SVO languages are intermediate between V-initial and V-final languages is the position of question particles in yes-no questions. I restrict attention here to question particles which appear in clause-peripheral position, ignoring particles which occur in other positions in the clause. The data for such particles is given in Table 16. 'SQ' denotes a sentence-final question particle while 'QS' denotes a sentence-initial question particle.

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
V-final&SQ	3	5	3	7	5	5	28
V-final&QS	1	2	0	3	4	3	13
Proportion SQ	.75	.71	1.00	.70	.56	.63	Avg.=.73
SVO&SQ	9	0	7	1	0	0	17
SVO&QS	4	2	2	2	1	1	12
Proportion SQ	.69	.00	.78	.33	.00	.00	Avg.=.30
V-initial&SQ	2	0	0	0	2	0	4
V-initial&QS	3	1	2	0	6	1	15
Proportion SQ	.40	.00	.00	--	.25	.00	Avg.=.13

Table 16  
Position of question particles

Table 16 shows that while V-final languages predominantly use final question particles (SQ) and V-initial languages initial question particles (QS), both positions of question particles are common among SVO languages. In terms of overall numbers, final particles (SQ) are more common, though because the SVO languages with final particles are primarily concentrated in two areas, Africa and Southeast Asia & Oceania, the average of

proportions over areas for final particles is only .30. In other words, averaging over areas, we find initial particles about twice as common among SVO languages as final particles. Despite this, final question particles are clearly more common among SVO languages than they are among V-initial languages, and hence we can say that SVO languages are intermediate between V-initial and V-final languages for this characteristic.

### 3.3.3. *Position of wh-words*

The third characteristic for which SVO languages are intermediate between V-initial and V-final languages is the position of interrogative words (*wh*-words) or phrases. As Greenberg (1963) observed, this characteristic correlates with clause order type. According to Greenberg's Universal 12,

If a language has dominant order VSO in declarative sentences, it always puts interrogative words or phrases first in interrogative word questions; if it has dominant order SOV in declarative sentences, there is never such an invariant rule. (p. 83)

While the two parts of this universal were exceptionless in Greenberg's 30-language sample, my database contains many exceptions to both parts. The exceptions to the first part are discussed in section 1 above. Table 17 gives the data on this parameter for the three clause order types.

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
V-final& <i>Wh</i> -in situ	9	14	4	11	11	7	56
V-final&Initial- <i>wh</i>	4	1	1	3	8	9	26
Proportion <i>Wh</i> -in situ	.69	.93	.80	.79	.58	.44	Avg.=.71
SVO& <i>Wh</i> -in situ	15	1	12	2	0	1	31
SVO&Initial- <i>wh</i>	3	5	1	4	5	3	21
Proportion <i>Wh</i> -in situ	.83	.17	.92	.33	.00	.25	Avg.=.42
V-initial& <i>Wh</i> -in situ	1	0	4	0	1	0	6
V-initial&Initial- <i>wh</i>	4	1	2	2	11	3	23
Proportion <i>Wh</i> -in situ	.20	.00	.67	.00	.08	.00	Avg.=.16

Table 17  
Position of interrogative words and phrases

Table 17 shows that SVO languages are intermediate between V-initial and V-final languages with respect to the position of interrogative words or phrases: while such

phrases more commonly occur in situ in V-final languages and in sentence-initial position in V-initial languages, the two types are both common among SVO languages.

### 3.4. *Summary*

We see then that while SVO languages pattern like V-initial languages for some characteristics, there are other characteristics for which SVO languages pattern between V-initial and V-final languages. It is useful therefore to separate out those typological characteristics that treat SVO languages like V-initial languages from those characteristics that treat SVO languages differently from V-initial languages. Once we do so, we see that the Lehmann-Vennemann distinction between OV and VO languages and the claim that SVO languages are intermediate between V-initial and V-final languages are both – to a point – valid. The Lehmann-Vennemann distinction between OV and VO languages is justified by the large number of word order characteristics for which SVO languages do pattern like V-initial languages. On the other hand, the critics of that typology are right to the extent that there do exist characteristics for which SVO languages are indeed much more a mixed type. But their criticisms of the Lehmann-Vennemann typology miss the point that the greater inconsistency among SVO languages is localized in certain characteristics. The existence of a large number of other characteristics for which SVO languages do pattern like V-initial languages provides justification for making a distinction between OV and VO languages.

There may, of course, be various other ways in which SVO languages differ from V-initial languages. Payne (1990) discusses a number of characteristics that Edward Keenan suggested in unpublished work in the 1970's were typical of V-initial languages. While many of these are characteristics I have discussed here, many of them, specifically a number of characteristics that do not refer to word order, have yet to be systematically investigated, such as his suggestion that V-initial languages always have a passive voice. Even if it turns out that V-initial languages and SVO languages do differ in a number of ways not involving word order (and they probably do), this will have no bearing on the question at hand, since the content of Vennemann and Lehmann's OV:VO typology is that there are a number of ways in which both SVO languages and V-initial languages tend to exhibit the opposite WORD ORDER characteristics from those found in V-final languages.

We can now specifically address Comrie's (1981: 90; 1989: 96) claim that one can predict the word order characteristics of V-initial and V-final languages with considerable reliability but that with SVO languages we can predict 'virtually nothing else.' The evidence presented in this section shows that there are a few characteristics (such as order of genitive and noun) which can be predicted of a language in a probabilistic sense if one knows that it is V-initial or V-final but not if it is SVO, but that in other respects we can predict the other characteristics of SVO languages about as well as we can predict the other characteristics of V-initial or V-final languages. Some word order characteristics (like order of adjective and noun) do not correlate with clause order; these characteristics cannot be predicted for languages of any clause order type. Other characteristics (such as adposition type) can be predicted for SVO languages with about

as much confidence as they can for V-initial or V-final languages. And at least one characteristic (order of relative clause and noun) can be predicted with considerable reliability for SVO and V-initial languages but not for V-final languages.

#### 4. HAWKINS' PRINCIPLE OF CROSS-CATEGORY HARMONY

A third argument that SVO languages are intermediate between V-initial and V-final languages is implicit in Hawkins' (1982, 1983) Principle of Cross-Category Harmony. This principle predicts that while it should be most common for V-final languages to place all modifiers BEFORE nouns and for V-initial languages to place all modifiers AFTER nouns, it should be most common for SVO languages to place some modifiers before and other modifiers after nouns.<sup>15</sup> If this is correct, it provides a clear sense in which SVO languages ARE intermediate between V-initial and V-final languages. I will argue here, however, that insofar as this prediction is correct, it is largely if not entirely attributable to the single fact observed in the previous section that SVO languages are intermediate between V-initial and V-final languages with respect to the order of genitive and noun.

Hawkins defines and tests his principle of CCH on the basis of raw numbers of languages, using either Greenberg's 30-language sample or the set of languages in Greenberg's appendix. While the 30-language sample suffers from a certain amount of genetic bias (e.g. 6 of the languages are Indo-European) and areal bias (20 are spoken in Africa or Eurasia, applying the latter term as it is used in this paper, to exclude southeast Asia), the set of languages in Greenberg's appendix suffers from much more severe biases. As argued in Dryer (1989c), employing samples which include large numbers of closely related languages may yield results which reflect properties of those particular genetic groups rather than general properties of language.<sup>16</sup> Because of the difficulty interpreting different numbers of language types in a set that exhibits biases like those in Greenberg's appendix, I will test the predictions of Hawkins' CCH using the set of languages in my database, using averages of proportions over language areas as the primary statistic rather than raw numbers of languages. If CCH reflects general properties of language, rather than accidental properties of the genetic groups and areas that are overrepresented in Hawkins' sample, it should be supported by evidence calculated in this way, since this method is specifically designed to control for genetic and areal bias.

##### 4.1. *Test #1 of CCH: Adjectives and genitives*

Hawkins' CCH is a general principle that applies across different phrasal categories. The claim that SVO languages should be most commonly inconsistent within other phrasal categories is only one of many predictions this principle makes. It is not the purpose of this paper to evaluate the principle itself, but only the specific predictions it makes about SVO languages. Furthermore, since it makes a large number of different predictions about SVO languages, it is beyond the scope of this paper to examine all of

these predictions. I will restrict attention here to two predictions it makes regarding the order of modifiers with respect to the noun. The first of these predictions involves two modifiers of the noun, the adjective and genitive, with respect to the noun. Hawkins (1983: 140, 149) himself uses this set of types as a basic set to illustrate his CCH. And I will focus on one set of predictions that is relevant here, namely that SVO languages should differ from V-final and V-initial languages in most often having one modifier before the noun and one after rather than both preceding or both following.

Consider first V-final languages. CCH predicts that the most frequent language type among V-final languages should be the one in which both modifiers precede the noun, and that the least frequent type should be the one in which both modifiers follow the noun. Table 18 gives the relevant data. (Following Hawkins, N<sub>3</sub> means that both adjective and genitive precede the noun, N<sub>2</sub> means that one precedes the noun while the other follows, and N<sub>1</sub> means that both modifiers follow the noun.<sup>17</sup>)

	No. of languages in Hawkins' data	No. of genera in my database	Average of Proportions over areas
V-final&N <sub>3</sub>	29	45	.40
V-final&N <sub>2</sub>	24	53	.51
V-final&N <sub>1</sub>	11	9	.09

Table 18  
Cross-Category Harmony and V-final languages

The first column in Table 18 shows the data that Hawkins (1983: 149) presents in support of his claim that the most common type of V-final language is N<sub>3</sub>, in which both modifiers precede the noun: this is the most common type, with 29 languages. In my database, whether one examines total number of genera (shown in the middle column of Table 18) or average of proportions over areas (shown in the righthand column), this type is less common than the type V-final&N<sub>2</sub>, in which one modifier precedes the noun and one follows: the average proportion for the type N<sub>3</sub> (which CCH predicts should be most frequent) is only .41, while the average proportion for N<sub>2</sub> (in which one modifier follows the noun) is .51. The difference between Hawkins' result and mine is attributable to the fact that the set of languages that Hawkins used includes a disproportionate number of languages from Eurasia, in which V-final languages are most often GN&AN: 16 of the 29 V-final&N<sub>3</sub> languages in his sample are spoken in Eurasia (excluding southeast Asia). Hence the prediction of CCH that V-final languages should most often place both modifiers before the noun is not supported. Only the prediction that V-final&N<sub>1</sub> should be least common is borne out. But we can explain the lower frequency of this type by appeal to the fact that it is the one type in which the genitive must follow the noun, contrary to the general pattern found for V-final languages.

Consider next V-initial languages. CCH predicts that the most frequent type should be the type  $N_1$ , in which both modifiers follow the noun, and that the least frequent type should be  $N_3$ , in which both modifiers precede the noun. The relevant data is given in Table 19.

	No. of languages in Hawkins' data	No. of genera in my database	Average of Proportions over areas
V-initial& $N_3$	2	3	.18
V-initial& $N_2$	5	19	.41
V-initial& $N_1$	19	16	.41

Table 19  
Cross-Category Harmony and V-initial languages

Again we find that while the number of languages in Hawkins' data supported CCH, the data from my database do not: the type in which both modifiers follow the noun ( $N_1$ ) is no more common than the type  $N_2$  in which one modifier precedes and the other follows. CCH correctly predicts that the type  $N_3$  should be least common, but again we can explain this by appealing to the fact that  $N_3$  is the one type in which the genitive necessarily precedes the noun, contrary to the general preference for NGen order among V-initial languages.

Consider finally SVO languages. Here CCH predicts that the most common type should be  $N_2$ , in which one modifier precedes the noun and one follows. The data is given in Table 20.

	No. of languages in Hawkins' data	No. of genera in my database	Average of Proportions over areas
SVO& $N_3$	9	8	.26
SVO& $N_2$	22	16	.42
SVO& $N_1$	21	29	.33

Table 20  
Cross-Category Harmony and SVO languages

Again, the lefthand column of Table 20 gives the evidence that Hawkins presented in support of the predictions of CCH: SVO& $N_2$  is most common. And again the figures in the middle column from my database do not make the same prediction: SVO& $N_1$  is more common. But the crucial figures, I claim, are the averages of proportions, given in the righthand column, and at first sight, the data in that column seem to support the claim of CCH: the most frequent type is  $N_2$ , in which one modifier precedes the noun and one

follows. Does this therefore support the claim of CCH that SVO languages are intermediate between V-initial and V-final languages in most often placing one modifier before the noun and one after? There are a couple of reasons to hesitate before drawing such a conclusion. First, as noted above, this type is also very common in V-initial and V-final languages. Second, it is important to realize that there are two types of SVO&N<sub>2</sub> languages, SVO&NG&AN and SVO&GN&NA. But there is only SVO&N<sub>1</sub> type (namely SVO&NG&NA) and only one SVO&N<sub>3</sub> type (namely SVO&GN&AN). What this means is that if there were no association at all among the clause type SVO, the order of genitive and noun, and the order of adjective and noun, then we would expect there to be twice as many SVO&N<sub>2</sub> languages as SVO&N<sub>1</sub> or SVO&N<sub>3</sub>: if there were no association we would expect approximately equal numbers of SVO&GN&AN, SVO&GN&NA, SVO&NG&AN, and SVO&NG&NA, and two of these types fall into the general type SVO&N<sub>2</sub>. So the observed average of proportions over areas, .42, though highest of the three figures, is actually LESS than what we would expect due to chance. We would expect the average proportion over areas for SVO&N<sub>1</sub> to be .25, but the observed average proportion over areas is actually higher than this, namely .33. In other words, we can conclude that SVO&N<sub>2</sub> is less frequent than we would expect due to chance, while SVO&N<sub>1</sub> is more common. Hence, while the data in Table 20 may literally support CCH, it provides no evidence of any tendency for SVO languages to place one modifier before the noun and one after the noun MORE OFTEN THAN CHANCE.<sup>18</sup>

It is useful, then, to consider the distribution of types in which the two subtypes of N<sub>2</sub>, GN&NA and NG&AN, are kept separate. Table 21 presents the data in this fashion.

	No. of languages in Hawkins' data	No. of genera in my database	Average of Proportions over areas
SVO&GN&AN	9	8	.26
SVO&GN&NA	12	11	.32
SVO&NG&AN	10	5	.10
SVO&NG&NA	21	29	.33

Table 21

Cross-Category Harmony and SVO languages, with N<sub>2</sub> broken down into GN&NA and NG&AN

The numbers in the first two columns of Table 21 suggest that the most common type is SVO&NG&NA. However, when we examine the average of proportions in the righthand column, we see that while SVO&NG&NA is highest, it is only marginally so (.33 vs. .32 for SVO&GN&NA and .26 for SVO&GN&AN).<sup>19</sup> Nevertheless, neither of the two subtypes of N<sub>2</sub> are most common, and we cannot conclude that SVO languages place one modifier before the noun and one after to an extent that is any better than chance.

It is also insightful to compare SVO languages to V-final and V-initial languages with respect to the frequency of languages in which one modifier precedes the noun and one follows. Table 22 gives the averages of proportion for the three clause types from the righthand columns of Table 18 to 20 (using a single type  $N_2$ ).

	V-final	SVO	V-initial
$N_3$	.40	.26	.18
$N_2$	.51	.42	.41
$N_1$	.09	.33	.41

Table 22  
Averages of proportions over areas for clause type by noun-position type

It is clear from Table 22 that the type  $N_2$  is no more common among SVO languages than it is among V-final or V-initial languages. The average of proportions for  $N_2$  is less for SVO languages (.42) than it is for V-final languages (.51) and about the same as it is for V-initial languages (.41). Hence,  $N_2$  languages are no more common among SVO languages than they are among V-final or V-initial languages.

The progression of values in Table 22 for  $N_3$  and  $N_1$  might seem to support the CCH: the figures get progressively smaller as one moves along the  $N_3$  line from V-final to V-initial (.40 > .26 > .18) and the figures get progressively larger as one moves along the  $N_1$  line from V-final to V-initial (.09 < .33 < .41). If we compute the success of the predictions of CCH in the manner of Hawkins (1983: 149), we find that it makes the correct prediction in 22 out of 26 cases, a success rate of 85%. It is possible, however, to explain this pattern entirely in terms of the tendencies governing the order of genitive and noun. The preference for GN order among V-final languages, the dispreference for GN order among V-initial languages, and the absence of any preference one way or the other among SVO languages predicts that  $N_3$  should be most common among V-final languages and least common among V-initial languages. An analogous account can be given for the opposite progression for  $N_1$ . Hence while these two lines of Table 22 do show SVO languages intermediate between V-final languages and V-initial languages, this can be explained entirely by the fact observed in section 2 that SVO languages are intermediate between V-final and V-initial languages with respect to the order of genitive and noun.

#### 4.2. Test #2 of CCH: Adjective, genitive, demonstrative, and numeral

It has not been the purpose of this section to argue against CCH in general, but only to question the specific claim of CCH that SVO languages are intermediate between V-initial and V-final languages in their placement of modifiers with respect to the noun. I have addressed this claim so far only with respect to the order of genitive and adjective with respect to the noun. It would be beyond the scope of this paper to completely

evaluate the predictions of CCH for various other sets of elements. I will, however, present data from my database for one larger set of noun modifiers, namely adjectives, genitives, demonstratives, and numerals. Table 23 gives the data analogous to that in Table 22 but for all four of these modifiers.  $N_5$  means that all four of these modifiers precede the noun,  $N_4$  means that three of these modifiers precede the noun while one follows, and so on down to  $N_1$ , which means that all four of the modifiers follow the noun.

	V-final	SVO	V-initial
$N_5$	.36	.34	.12
$N_4$	.12	.13	.26
$N_3$	.29	.25	.17
$N_2$	.19	.13	.37
$N_1$	.04	.15	.08

Table 23

Averages of proportions over areas for clause type by noun-position type for four noun modifiers (Gen, Adj, Dem, Num)

Hawkins does not formulate CCH in such a way that it makes exact predictions with four modifiers, but it is clear from the spirit of CCH that we should expect  $N_5$  to be most common for V-final languages,  $N_1$  to be most common for V-initial languages, and  $N_3$  to be most common for SVO languages. The data in Table 23 support only one of these predictions:  $N_5$ , in which all four modifiers precede the noun, is most common among SOV languages. On the other hand,  $N_1$ , in which all modifiers follow the noun, is actually more common among SVO languages than among V-initial languages, contrary to the CCH; in fact it is the LEAST common type among V-initial languages. And  $N_3$ , in which two modifiers precede the noun and two follow, is more common among V-final languages than it is among SVO languages. The overall predictions of the CCH are best evaluated by characterizing each type (as in Hawkins 1983) in terms of the number of CCH deviations of that type; CCH predicts that the types with fewer deviations should be more common than the types with more deviations. The data from Table 23 is reorganized along these lines in Table 24.

	CCH Deviations	Avg. Proportion over Areas
V-final & N <sub>5</sub>	0	.36
SVO & N <sub>3</sub>	0	.25
V-initial & N <sub>1</sub>	0	.08
V-final & N <sub>4</sub>	1	.12
SVO & N <sub>4</sub>	1	.13
SVO & N <sub>2</sub>	1	.13
V-initial & N <sub>2</sub>	1	.37
V-final & N <sub>3</sub>	2	.29
SVO & N <sub>5</sub>	2	.34
SVO & N <sub>1</sub>	2	.15
V-initial & N <sub>3</sub>	2	.17
V-final & N <sub>2</sub>	3	.19
V-initial & N <sub>4</sub>	3	.26
V-final & N <sub>1</sub>	4	.04
V-initial & N <sub>5</sub>	4	.12

Table 24  
Predictions of Cross-Category Harmony in terms of number of deviations for  
data in Table 23

If we calculate (in the manner of Hawkins 1983) the number of instances in which a type with fewer CCH deviations is more frequent than a type with more CCH deviations and compare this to the number of instances of the opposite situation, we find that there are 50 instances in which CCH makes the correct prediction and 38 instances in which the CCH makes the wrong prediction.<sup>20</sup> In other words, it has a success rate of 57%. On the other hand, if we isolate those predictions for which one of the two types is SVO, there are 24 correct predictions and 24 incorrect predictions. In other words, with respect to its predictions regarding SVO languages, CCH does no better than chance. It does better than chance only with respect to V-final and V-initial languages. Again, we find no evidence to support the claims of CCH with respect to SVO languages.

## 5. CONCLUSION

I have argued in this paper that SVO languages exhibit properties that are consistent with the typology of Lehmann and Vennemann in which the basic dichotomy is between OV and VO languages. I have argued that there is no reason to believe that

there are fewer exceptionless generalizations to be made about SVO languages than there are about V-initial and V-final languages. I have shown that while there are a few characteristics (like the order of genitive and noun) for which SVO languages pattern intermediate between V-initial and V-final languages, there are a large number of other word order characteristics for which SVO languages pattern very much like V-initial languages. And I have argued that the claim of Hawkins' principle of Cross-Category Harmony that SVO languages place modifiers on opposite sides of the noun more often than V-initial and V-final languages is not supported by the available evidence.

#### APPENDIX: POSTPOSITIONAL SVO LANGUAGES

As noted in section 2, Greenberg's Appendix II lists a much higher frequency of postpositional SVO languages than in my database: he lists 19 SVO&Po languages or groups, compared to 33 SVO&Pr languages or groups, while my database exhibits an average proportion over areas of prepositions among SVO languages as .86, with 42 genera containing SVO&Pr languages but only 6 genera containing SVO&Po languages. The purpose of this appendix is to examine the source of this discrepancy. There appear to be two sources to the discrepancy. One is that Greenberg's Appendix contains a higher concentration of languages from Africa than from many other areas of the world: 10 of the 19 languages or groups he lists as SVO&Po are in Africa. The other source is a number of apparent inaccuracies in Greenberg's data. There is reason to believe that at least 8 of these 19 cases are either not SVO or not postpositional. It is worth examining the particular languages more closely, with citations of evidence from my sources of contrary classifications:

(1) Chinese. (Mandarin) Chinese word order has been analysed in a number of different ways. Greenberg treats it as SVO&Po. Hawkins (1983) treats it as indeterminately SVO/SOV and Pr/Po. I treat it as SVO&Pr. Sun & Givón (1985) present convincing evidence that it is SVO. While it does have some postpositions, the primary adpositional function is apparently served by what Li and Thompson (1981) call 'coverbs', many of which can also function as verbs but which are nevertheless distinct from verbs, others of which are more clearly prepositional. Since they are not identical to verbs, it is accurate to treat them as prepositions, in which case Chinese is prepositional. See Li and Thompson (1974, 1981: 356-369) for discussion.

(2) Algonquian. Greenberg affixes '(probably)' to the listing of Algonquian as SVO/Po/GN/AN; it is not clear exactly which characteristics he intended the hedge to be attached to. Algonquian is a family, not a language, and there are some differences among the Algonquian languages as far as word order is concerned. But in general the Algonquian languages exhibit fairly flexible word order and it is not clear that any of them can be classified into one of the six orders of subject, verb, and object. While VO order is apparently somewhat more common than OV in Blackfoot, Cree, and Ojibwa, the order of subject and verb is more flexible. D. Frantz (p.c.) reports that SVO is the most common order in Blackfoot; but Blackfoot does not appear to employ adpositions. Cree does have some postpositions (R. Rhodes, p.c.; Hive 1948: 32), though they apparently do not play a major role in the language. My own informal counts of a Cree text in Hive

(1948) found 2 instances of SVO, 2 instances of VOS, and 1 instance of SOV, illustrating the flexibility of word order in the language. Although it is possible that further evidence might provide a basis for treating Cree as SVO, the language is probably best left unclassified for basic clause order. Hence it is unclear that there is an SVO&Po Algonquian language.

(3) Zoque. There are a number of distinct Zoque languages, but none of them appears to be SVO. Campbell, Bubenik, and Saxon (1988: 212) report that Copainalá Zoque is VOS (cf. Harrison, Harrison, and García 1981: 402). Engel and Longacre (1963: 336) say that transitive subjects in Ostuacan Zoque precede and follow the verb with equal frequency, implying that this particular Zoque language is not specifically SVO. Again there is no evidence of an SVO&Po Zoque language.

(4) Tonkawa. Campbell, Bubenik, and Saxon also dispute the classification of Tonkawa as SVO, claiming instead that it is SOV. My own informal text counts of a set of texts in Hoijer (1972) uncovered 11 instances of SOV clauses and 7 of SVO, certainly casting doubt on the claim that the language is SVO, though leaving it unclear that it is SOV. On the other hand, my counts also showed 86% OV, suggesting that if Tonkawa can be assigned a basic order, it is probably SOV, not SVO.

(5) Songhai. According to Prost (1956: 79, 124), the order of words in a transitive clause in Songhai is Subject + Conjugation Particle + Object + Verb. The conjugation particles code tense/aspect, mood, and negation. Unless one interprets the conjugation particle as the verb, which is not Prost's analysis, the language is SOV, not SVO. Postpositional phrases do follow the verb in Songhai; if one were to treat these as objects, then the language would be SVO. But DIRECT objects, whose position is generally assumed to be the defining characteristic in determining whether a language is SVO, precede the verb, so on standard usage the language is SOV, though we might more specifically describe the language as SOVX.

(6) Most Mandingo (i.e. Mande) languages. All five of the Mande languages in my sample are SOV, not SVO. Dwyer (1989: 57) reports that Mande languages in general are SOV. A number of Mande languages (e.g. Bambara, Vai) share both features of Songhai described in the preceding paragraph: a tense-aspect word occurs between the subject and object, and adpositional phrases follow the verb. As with Songhai, either of these properties might lead one to describe a language as SVO, though not by the conventional use of the term. Furthermore, while the tense-aspect word that occurs between the subject and the object sometimes seems nonverbal, like the 'conjunction particle' in Songhai, one of these words in Vai is a copula verb functioning as a verbal auxiliary (Welmers 1976: 88-89), exhibiting an S-Aux-O-V structure. If we were to treat the Aux as the verb, ignoring the content verb, then we might say Vai is SVO. However, the most likely analysis is presumably one whereby the Aux is in construction with the combination of object+verb, rather than just the object, so it seems unnatural to describe such a language as SVO.

(7) Most Voltaic (i.e. Gur) languages. My database contains four Gur languages, with varying characteristics. Prost (1964: 274) implies that Toussian is OV. My own informal counts of Tenyer texts in Prost (1964) show only OV examples. Bimoba (Jacobs 1970) is SVO, but it has both prepositions and postpositions (pp. 143-149), and it is not clear that either is basic. The one Gur language in my database that may be SVO&Po is Kirma. While the evidence I have examined in the description in Prost (1964) is insufficient to establish that it is SVO, the examples suggest that it is, in which case it would be SVO&Po. But it does not seem true that most Gur languages are such.

(8) I<sup>a</sup>jo<sup>a</sup>. Williamson (1965: 33) implies that the Kolokuma dialect of I<sup>a</sup>jo<sup>a</sup> is SOV, not SVO. Jenewari (1989: 117) reports that the Ijoid languages in general are SOV.

(9) Nupe. This language is not in my database, but Hawkins lists it as Pr/Po, citing Larry Hyman (personal communication) as his source. I.e. while it is SVO, it is not clear that it is postpositional.

(10) & (11) Rutul(ian) and other Daghestan languages. Rutul is not in my database, but five other Daghestan languages are and all five are OV according to my sources, not SVO: Avar, Lak, Dargwa, Lezgi, and Archi. I have not examined Rutul, but Hawkins (1983) does list it and three other Daghestan languages I have not examined as SVO, citing original sources.

This leaves eight remaining cases of SVO&Po, none of which I dispute: Finnish, Estonian, Kru, Twi, Gã, Guang, Ewe, and Guarani. If we include Rutul, some other Daghestan languages, and Kirma, none of which I have seen convincing evidence for, but which are probably valid cases, that makes a total of ten SVO&Po languages. But Finnish and Estonian both fall within the same genus (Finno-Ugric), as do Twi, Gã, Guang, and Ewe (all Kwa). Similarly, the four Daghestan languages Hawkins cites (Rutul and three others), are all in the same genus. In other words, we have only six genera that clearly contain SVO&Po languages (four of which are represented in my database): Finno-Ugric, Kru, Voltaic (= Gur), Kwa, Tupian (Guarani), and Daghestan. My database also includes two other genera with SVO&Po languages, namely Northern Khoisan and Zaparoan, the languages being Xu and Iquito respectively. In short, even when we examine the large set of languages that Greenberg lists as SVO&Po, we find at most two additional instances of genera containing languages with this order. But my database contains a much larger set of SVO&Prepositional languages than occur in Greenberg's Appendix II, 42 genera containing 97 languages. The following is a list of these languages, by area and by genus. (The name of the genus is given first with member languages in parentheses; genera consisting of single languages are listed without repeating the name of the language.)

*SVO&Prepositional Languages:*

## Africa:

Northern Atlantic (Fulani)  
 Adamawa (Mbum)  
 Ubangi (Gbeya Bossangoa, Sango)  
 Defoid (Yoruba)  
 Edoid (Engenni)  
 Igboid (Izi)  
 Platoid (Jukun)  
 Kainji (Duka)  
 Bantoid (Ewondo, Bobangi, Swahili, Nkore-Kiga, Luvale, Zulu)  
 Nilotic (Dholuo, Acooli, Bari)  
 Kresh  
 Bongo-Bagirmi (Sara-Ngambay, Bagirmi, Yulu)  
 Biu-Mandera (Tera)  
 West Chadic (Hausa, Kanakuru, Angas, Ngizim)  
 Semitic (Colloquial Egyptian Arabic)

## Eurasia:

Albanian  
 Italic (Rumanian, French, Spanish)  
 Germanic (Danish, Swedish, English)  
 Baltic (Lithuanian)  
 Slavic (Russian, Polish)

## Southeast Asia &amp; Oceania:

Chinese (Mandarin, Hakka, Cantonese)  
 Karen  
 Miao-Yao (Miao)  
 Khasi  
 Palaung-Khmuic (Palaung)  
 Viet-Muong (Vietnamese)  
 Bahnaric (Stieng, Chrau, Sre, Brao)  
 Khmer (Cambodian)  
 Kam-Tai (Nung, Thai)  
 Philippine Austronesian (Palauan)  
 Sundic (Sundanese, Indonesian)  
 Central-Eastern Malayo-Polynesian (Mor, Kaliai-Kove, Patep, Nissan, Tolai, Tigak, Halia, Lenakel, Iai, Dehu, Kusaiean, Mokilese, Ponapean, Puluwat, Sonsoral-Tobi, Sakao, Nguna, Rotuman, Futuna-Aniwa)

## Australia-New Guinea:

Torricelli (Arapesh)  
 Tiwi  
 Yiwaidjan (Maung)  
 Gunwinyguan (Gunbalang)

## North America:

Yurok

Totonacan (Totonac)  
 Palaihnihan (Achumawi)  
 Aztecan (Tetelcingo Nahuatl)  
 South America:  
 Guaicuruan (Abipon)  
 Mataco

The large number of genera containing SVO&Pr languages in this list, compared to the small number of genera containing SVO&Po languages described above, illustrates that SVO languages exhibit a stronger preference for prepositions than Greenberg's data would suggest.

#### ACKNOWLEDGEMENT

The research for this paper was supported by Social Sciences and Humanities Research Council of Canada Research Grants 410-810949, 410-830354, and 410-850540 and by National Science Foundation Research Grant BNS-9011190. I am indebted to two anonymous *Journal of Linguistics* referees, Nigel Vincent, and Lindsay Whaley for comments on an earlier draft of this paper and to Lyle Campbell for comments on portions of an earlier draft of the paper.

#### Footnotes

<sup>1</sup> My database does not contain data on all 603 languages for all the word order characteristics discussed in this paper for a variety of reasons. First, in many cases my source (usually a published grammatical description of the language) does not provide clear evidence regarding the normal order of certain pairs of elements. Second, in many cases, the order of a given pair of elements will be sufficiently flexible in a given language that there seems to be little basis for treating one or the other order as basic. And third, certain word order parameters, such as the order of article and noun, do not apply to languages lacking one of the categories in question.

<sup>2</sup> Not all of the six languages have all of these properties in Greenberg's data, either because he was unable to determine a basic order for the language or because the characteristic does not apply to the language because it lacks the relevant construction. For example, only three of the six languages are listed as having verbal auxiliaries in Greenberg's data. Greenberg lists only two of the languages as placing the intensifier after the adjective; he does not give an order for two of the languages (apparently because he did not have the relevant data) and two of them allow both orders. One of the languages does not have a question particle. The point is that for each of these properties there is a generalization that is exceptionless in Greenberg's sample according to which the V-initial languages do not have the opposite property from that listed in (1).

<sup>3</sup> Greenberg and others often include position of the marker of comparison in discussion of comparative structures. Thus the V-initial languages in his sample are not only

Standard-Adjective, but more specifically Standard-Marker-Adjective. My database contains one instance of a V-initial language which is Adjective-Standard-Marker, namely Yagua. This language is Adjective-Standard, as one would expect of a V-initial language, but unexpectedly Standard-Marker. I view the order of standard, marker, and adjective as two distinct parameters, the order of standard and marker and the order of standard and adjective, and restrict attention to the latter here.

<sup>4</sup> The classification of some of the languages in Table 1 as V-initial may be controversial. The classification of a language as VS, VO indicates that both subject and object normally follow the verb, but that the language is not clearly VSO or VOS. Some of these languages are indeterminately VSO/VOS. A few of them are languages in which the order SVO is frequent in clauses containing a nominal subject and a nominal object, but in which VS order is more frequent overall. The classification of Papago as V-initial is questioned by Payne (1987) and Campbell, Bubenik, and Saxon (1988: 211-212) but defended by Dryer (1989b). Campbell et al also question (p. 213) Hawkins' (1983) characterization of Chontal as VSO, claiming that it is really SVO. However, Campbell et al's comments apply at best to Tequistlatec, the Chontal language spoken in the mountains; Hawkins' characterization apparently refers to Huamelultec Chontal, the lowland language, since he cites Waterhouse (1962) as source. The description by Waterhouse (1962) implies it is V-initial: she says that the usual order is 'predicate followed by concomitant' (p. 18) and the examples, both with intransitive verbs and with transitive verbs, are almost entirely VS. The only example that I have found with both a nominal subject and a nominal object is VOS.

Campbell, Bubenik, and Saxon discuss a few additional exceptions not in my database: they list Copainalá Zoque as VOS&Postp and they list Chontal of Tabasco (a Mayan language, not to be confused with Chontal of Oaxaca, which refers to the non-Mayan Tequistlatecan languages Huamelultec Chontal and Tequistlatec), Xinca (spoken in El Salvador), and Copainalá Zoque all as VOS&AN.

<sup>5</sup> In Dryer (1988b), I report a further exceptionless generalization about V-initial languages: they always place negative words before the verb in unmarked word order, never after. Since writing that paper, I have found an exception to this generalization: Lamang, a Chadic language spoken in Nigeria and Cameroon (Wolfe 1983: 172). While there may be a negative word preceding the verb in Lamang, there is always a postverbal negative. I am not aware, however, of any V-initial languages with postverbal negative auxiliary verbs, though this may be due to the small number of V-initial languages in my database which employ negative auxiliaries.

<sup>6</sup> On the other hand, I consider Chinese to be an exception to (XIII). See the discussion of Chinese in the appendix.

<sup>7</sup> More detailed information on these exceptions can be found in the following sources. Campbell, Bubenik, and Saxon (1988: 216, 224) observe that Tigre (a Semitic language of Eritrea in Ethiopia) is an exception to Hawkins' Universal I (which is equivalent to Greenberg's Universal 5). See also Raz (1983). Payne (1986) discusses Yagua (an isolate spoken in Peru) as an exception to II, which is like II' but refers specifically to

VSO languages ( $V-1 \supset (NA \supset NG)$ ). Harrison (1986) observes that Guajajara is an exception to II; see also Bendor-Samuel (1972: 78, 110). Senft (1986: 108-112) discusses main clause order in Kilivila (an Austronesian language of the Trobriand Islands off Papua New Guinea): while SVO is a common alternative, he describes VOS as the 'prime order' and says that the initial position of the subject in SVO clauses 'implies emphasis' (p. 110). Kilivila is GN (pp. 103-104), NA (pp. 104-105) and prepositional (p. 93). Garawa is a non-Pama Nyungan Australian language in the Garawan group. My classification of Garawa's word order characteristics is more tenuous, since it is based primarily on examples in Furby and Furby (1977), though Blake (1988) also lists it as V-initial. Among eight examples in various parts of Furby and Furby (1977) containing a lexical subject and a lexical object, 7 are VOS and one is VSO; all six examples involving a lexical genitive are GenN; 12 of 16 examples with an adjective are NAdj; and while the language has only two adpositions, they are both prepositions (p. 33-34). Only Yagua and Guajajara are exceptions to Hawkins' Universal II (the stronger form of II'), which applies specifically to VSO languages.

<sup>8</sup> The method in this paper is slightly different from that of Dryer (1989c), where only five areas were used instead of six. Southeast Asia & Oceania was treated as part of Eurasia in that paper. Using six areas rather than five makes it possible to test the statistical significance of generalizations in a more conservative way. I do not employ statistical tests in this paper, except in footnote 9, since the primary thesis is a negative one, that SVO languages do not differ substantially from V-initial languages.

<sup>9</sup> Using the test for statistical significance discussed in greater detail in Dryer (1989c), we can compute the level of statistical significance for the difference between SVO and V-initial languages in Table 3 to be  $p > .31$ , which is not remotely statistically significant: it says that if there is no difference between SVO and V-initial languages with respect to the use of prepositions as opposed to postpositions, then we would have more than 31 chances in 100 of finding a difference as great as that observed in Table 3. This level of statistical significance can be computed by a simple binomial test: in 3 of the 4 areas where we find a difference between SVO and V-initial languages, the proportion of genera containing languages with postpositions is higher for SVO languages. The problem is logically equivalent to the likelihood of obtaining 3 or more heads on 4 tosses of an unbiased coin. As discussed in detail in Dryer (1989c), we cannot apply statistical tests (like Chi-square) to the total number of genera of each of the types since the genera are not genetically or areally independent and the relevant tests require that the tokens be independent.

<sup>10</sup> The figures in Table 4 are lower than those in Table 3 because I have been able to obtain data on relative clauses for a much smaller number of languages.

<sup>11</sup> While it is unclear why the V-initial languages exhibit an intermediate pattern, the high average proportion for PredCop order among V-initial languages may be due, at least in part, to the small number of V-initial languages for which I have data on the order of predicate and copula. In four of the six areas, I have data for only one genus; if we

were to base our average only on the two areas for which I have data for more than one genus, the average would have been .18, below the figure for SVO languages.

<sup>12</sup> It must be borne in mind that ‘V-final’ and ‘V-initial’ are being used here in a technical sense, to refer to the position of the verb relative to the subject and object. An SOVX language (where X=PP) is V-final by the definition given in section 1, since both subject and object precede the verb.

An anonymous referee objects to the inclusion of the order of adpositional phrase and verb here, since the adpositional phrase is itself a dependent of the verb. It is true that intra-categorical correlations like this one, i.e. a correlation between the order of verb and object and the order of the verb with respect to some other dependent of the verb, are less dramatic than cross-categorical ones, which involve dependents of some category other than the verb. Nevertheless, there is no logical reason why SOVX languages should be less common than SOV languages in which the PP precedes the verb. Nor is there any logical reason why there are no attested instances of VSO or VOS languages in which the PP precedes the verb. And there is certainly no logical reason why SXVO languages should be any less common than SVOX languages. The fact that the PP almost always occurs on the same side of the verb as the object rather than the subject in SVO languages means that the order of subject and verb is irrelevant for predicting the order of PP and verb, and that only the order of object and verb is relevant. But that is precisely what the OV:VO typology claims.

<sup>13</sup> Greenberg lists two of the languages as allowing both order of intensifier and adjective and lists another two as languages for which he had no data on the order of intensifier and adjective.

<sup>14</sup> The figure .28 for GenN in V-initial languages is somewhat higher than the overall figures might suggest. This is due to the fact that the only V-initial languages in my sample from Australia-New Guinea (Garawa and Wembawemba) are both GenN. The average over the other five areas is much less, .13. This figure is probably more representative of the likelihood of a random V-initial language being GenN.

<sup>15</sup> Hawkins formulates the principle to refer to different categories in general, not just to noun phrases. I will restrict attention here to its predictions with respect to the order of modifiers with respect to nouns.

<sup>16</sup> Greenberg himself did not use counts of languages in his appendix; his counts are based entirely on his more reliable 30-language sample. Although his 30-language sample includes 6 Indo-European languages, all are from different branches of the family. But the set of languages in his appendix includes Norwegian, Swedish, Danish, Icelandic, German, and Dutch, as well as various other instances of multiple sets of closely related languages, each of which is counted as a separate language in Hawkins' counts based on this set. In addition, within this set of 142 languages, 84 are spoken in Africa or Eurasia (again using the latter term in the narrow sense of this paper). On the other hand, this set contains only 7 languages spoken in South America, and one of those, Papiamento, is an

Indo-European-based creole. It should be stressed that while Eurasia covers a fairly large area (though not as large as conventional map projections suggest, since they typically exaggerate areas in the Northern Hemisphere), there are probably more languages spoken in South America than in Eurasia, exhibiting greater typological variation. Voegelin and Voegelin (1977) list 581 languages in South America, including known languages which are now extinct. This does not include languages which have become extinct since the arrival of the Europeans but which we have no record of. Voegelin and Voegelin list 613 languages within Eurasia as a whole, but only 256 in the narrower sense of Eurasia that excludes southeast Asia.

<sup>17</sup> The idea behind the notation is that  $N_1$  means that the noun comes first in the set of noun, adjective, genitive; in general  $N_i$  means that N occurs in the  $i^{\text{th}}$  position in the set of noun and modifiers.

<sup>18</sup> It should be noted that in a number of the cases in which CCH is applied in Hawkins (1983), only one of the two types of  $N_2$  was permitted by his implicational universals. For example, while in theory there are two types of  $\text{Pr}\&N_2$ , the type  $\text{Pr}\&N\&A$  is ruled out by his (statistical) universal (III) ( $\text{Pr}\&N \supset (\text{NA} \supset \text{NG})$ ). In these cases, the problem of there being two types of  $N_2$  languages does not arise. It is only a problem in cases like  $\text{SVO}\&N_2$ , where neither type of  $N_2$  is ruled out by an implicational universal.

<sup>19</sup> The high frequency of  $\text{SVO}\&N\&A$ , both in Hawkins' data and in the total number of genera in my database, reflects the large number of such languages in Africa: 9 of the 21 languages of this type in Hawkins' data and 18 of the 29 genera in my data are from Africa.

<sup>20</sup> Following the practice of Hawkins (1983), I treat the one instance of a tie as a correct prediction.

## REFERENCES

- Allen, W. Sidney (1964). Transitivity and possession. *Language* 40.337-343.
- Bendor-Samuel, David (1972). *Hierarchical Structures in Guajajara*. Norman, Oklahoma: Summer Institute of Linguistics.
- Blake, Barry (1988). Redefining Pama-Nyungan: Towards the prehistory of Australian languages. *Aboriginal Linguistics* 1.1-90.
- Campbell, Lyle, Vit Bubenik, and Leslie Saxon (1988). Word order universals: refinements and clarifications. *Canadian Journal of Linguistics* 33.209-230.
- Campbell, Lyle, and Marianne Mithun (1979). Introduction: North American Indian historical linguistics in current perspective. In Lyle Campbell and Marianne Mithun (eds.), *The Languages of Native America: Historical and Comparative Assessment*. Austin: Univ. of Texas Press. 3-69.
- Comrie, Bernard (1981). *Language Universals and Linguistic Typology*. First Edition. Chicago: University of Chicago Press.
- Comrie, Bernard (1989). *Language Universals and Linguistic Typology*. Second Edition. Chicago: University of Chicago Press.
- Dahl, Osten (1979). Typology of sentence negation. *Linguistics* 17.79-106.
- Derbyshire, Desmond, and Geoffrey Pullum (1981). Object-initial languages. *International Journal of American Linguistics* 47.192-214.
- Derbyshire, Desmond, and Geoffrey Pullum (1986). Introduction. In Desmond Derbyshire and Geoffrey Pullum (eds.), *Handbook of Amazonian Languages 1*. Berlin: Mouton de Gruyter. 1-28.
- Dryer, Matthew S. (1980). The positional tendencies of sentential noun phrases in universal grammar. *Canadian Journal of Linguistics* 25.123-195.
- Dryer, Matthew S. (1988a). Object-verb order and adjective-noun order: dispelling a myth. *Lingua* 74.77-109.
- Dryer, Matthew S. (1988b). Universals of negative position. In M. Hammond, E. Moravcsik, and J. Wirth (eds.), *Studies in Syntactic Typology*. Amsterdam: John Benjamins. 93-124.

- Dryer, Matthew S. (1989a). Article-noun order. *Papers from the 25th Regional Meeting of the Chicago Linguistics Society*, 83-97.
- Dryer, Matthew S. (1989b). Discourse-governed word order and word order typology. *Belgian Journal of Linguistics* 4.69-90.
- Dryer, Matthew S. (1989c). Large linguistic areas and language sampling. *Studies in Language* 13.257-292.
- Dryer, Matthew S. (1989d). Plural words. *Linguistics* 27.865-895.
- Dwyer, David J. (1989). Mande. In John Bendor-Samuel (ed.), *The Niger-Congo Languages*. Lanham, Maryland: University Press of America. 47-65.
- Engel, Ralph, and Robert E. Longacre (1963). Syntactic matrices in Ostuacan Zoque. *International Journal of American Linguistics* 29.331-344.
- Furby, E. S., and C. E. Furby (1977). *A Preliminary Analysis of Garawa Phrases and Clauses*. (Pacific Linguistics, Series B, No. 42.) Canberra: Australian National University.
- Greenberg, Joseph H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In Joseph Greenberg (ed.), *Universals of Language*. Cambridge, Mass: MIT Press. 73-113.
- Harrison, Carl H. (1986). Verb prominence, verb initialness, ergativity, and typological disharmony in Guajajara. In Desmond C. Derbyshire and Geoffrey K. Pullum (eds.), *Handbook of Amazonian Languages 1*. Berlin: Mouton de Gruyter. 407-439.
- Harrison, Roy, Margaret Harrison, and Cástulo García H. (1981). *Diccionario Zoque de Copainalá*. Mexico City: Summer Institute of Linguistics.
- Hawkins, John A. (1979). Implicational universals as predictors of word order change. *Language* 55.618-648.
- Hawkins, John A. (1980). On implicational and distributional universals of word order. *Journal of Linguistics* 16.193-235.
- Hawkins, John A. (1982). Cross-category harmony, X-bar and the predictions of markedness. *Journal of Linguistics* 18.1-35.
- Hawkins, John A. (1983). *Word Order Universals*. New York: Academic Press.
- Hive, Rev. H. E. (1948). *A Cree Grammar*. Toronto: The Missionary Society of the Church of England in Canada.

- Hoijer, Harry (1972). *Tonkawa Texts*. (University of California Publications in Linguistics 73) Berkeley: University of California Press.
- Jacobs, Gillian (1970). *Bimoba Syntax: A Syntagmatic Analysis*. (Language Monographs No. 1) Ghana: University of Ghana.
- Jenewari, Charles E. W. (1989). Ijoid. In John Bendor-Samuel (ed.), *The Niger-Congo Languages*. Lanham, Maryland: University Press of America. 105-118.
- Keenan, Edward (1978). The syntax of subject-final languages. In Winfred P. Lehmann (ed.), *Syntactic Typology*. Austin: University of Texas Press. 267-327.
- Lehmann, Winfred P. (1973). A structural principle of language and its implications. *Language* 49.42-66.
- Lehmann, Winfred P. (1978). The great underlying ground-plans. In Winfred P. Lehmann (ed.), *Syntactic Typology*. Austin: University of Texas Press. 3-55.
- Li, Charles, and Sandra Thompson (1974). Co-verbs in Mandarin Chinese: verbs or prepositions? *Journal of Chinese Linguistics* 2.3.257-278.
- Li, Charles N., and Sandra A. Thompson (1981). *Mandarin Chinese: a Functional Reference Grammar*. Berkeley: University of California Press.
- Mallinson, Graham, and Barry J. Blake (1981). *Language Typology*. Amsterdam: North-Holland.
- Nichols, Johanna (1990). Linguistic diversity and the settlement of the New World. *Language* 66.475-521.
- Payne, Doris (1987). Information structuring in Papago narrative discourse. *Language* 63.783-804.
- Payne, Doris L. (1986). Basic constituent order in Yagua clauses: implications for word order universals. In Desmond C. Derbyshire and Geoffrey K. Pullum (eds.), *Handbook of Amazonian Languages 1*. Berlin: Mouton de Gruyter. 440-465.
- Payne, Doris L. (1990). *The pragmatics of word order: typological dimensions of verb initial languages*. Berlin: Mouton de Gruyter.
- Prost, R. P. A. (1956). *La Langue Sonay et ses Dialectes*. (Memoires de l'Institut Français d'Afrique Noire 47) Dakar.
- Prost, R. P. A. (1964). *Contribution à l'Étude des Langues Voltaiques*. (Memoires de l'Institut Français d'Afrique Noire 70.) Dakar.

- Raz, S. (1983). *Tigre Grammar and Texts*. Malibu: Undena.
- Senft, Gunter (1986). *Kilivila, The Language of the Trobriand Islanders*. Berlin: Mouton de Gruyter.
- Siewierska, Anna (1988). *Word Order Rules*. London: Croom Helm.
- Sun, Chao-Fen, and T. Givón (1985). On the so-called SOV word order in Mandarin Chinese. *Language* 61.329-351.
- Vennemann, Theo (1974). Analogy in generative grammar: the origin of word order. *Proceedings of the Eleventh International Congress of Linguists (1972)*. 79-83. Bologna.
- Vennemann, Theo (1976). Categorical grammar and the order of meaningful elements. In A. Juillard (ed.), *Linguistic Studies Offered to Joseph Greenberg on the Occasion of His Sixtieth Birthday*. Saratoga, California: Anna Libri. 615-634.
- Voegelin, C. F., and F. M. Voegelin (1977). *Classification and Index of the World's Languages*. New York: Elsevier.
- Waterhouse, Viola (1962). *The Grammatical Structure of Oaxaca Chontal*. (International Journal of American Linguistics, Vol. 28, No. 2, Part 2)
- Welmers, Wm. E. (1976). *A Grammar of Vai*. University of California Publications in Linguistics 84. Berkeley: University of California Press.
- Williamson, Kay (1965). *A Grammar of the Kolokuma Dialect of Ijo*. (West African Language Monographs 2.) Cambridge University Press in association with The West African Languages Survey and The Institute of African Studies, Ibadan, Nigeria.
- Wolff, Ekkehard (1983). *A Grammar of the Lamang Language*. Glückstadt: Verlag J.J. Augustin.