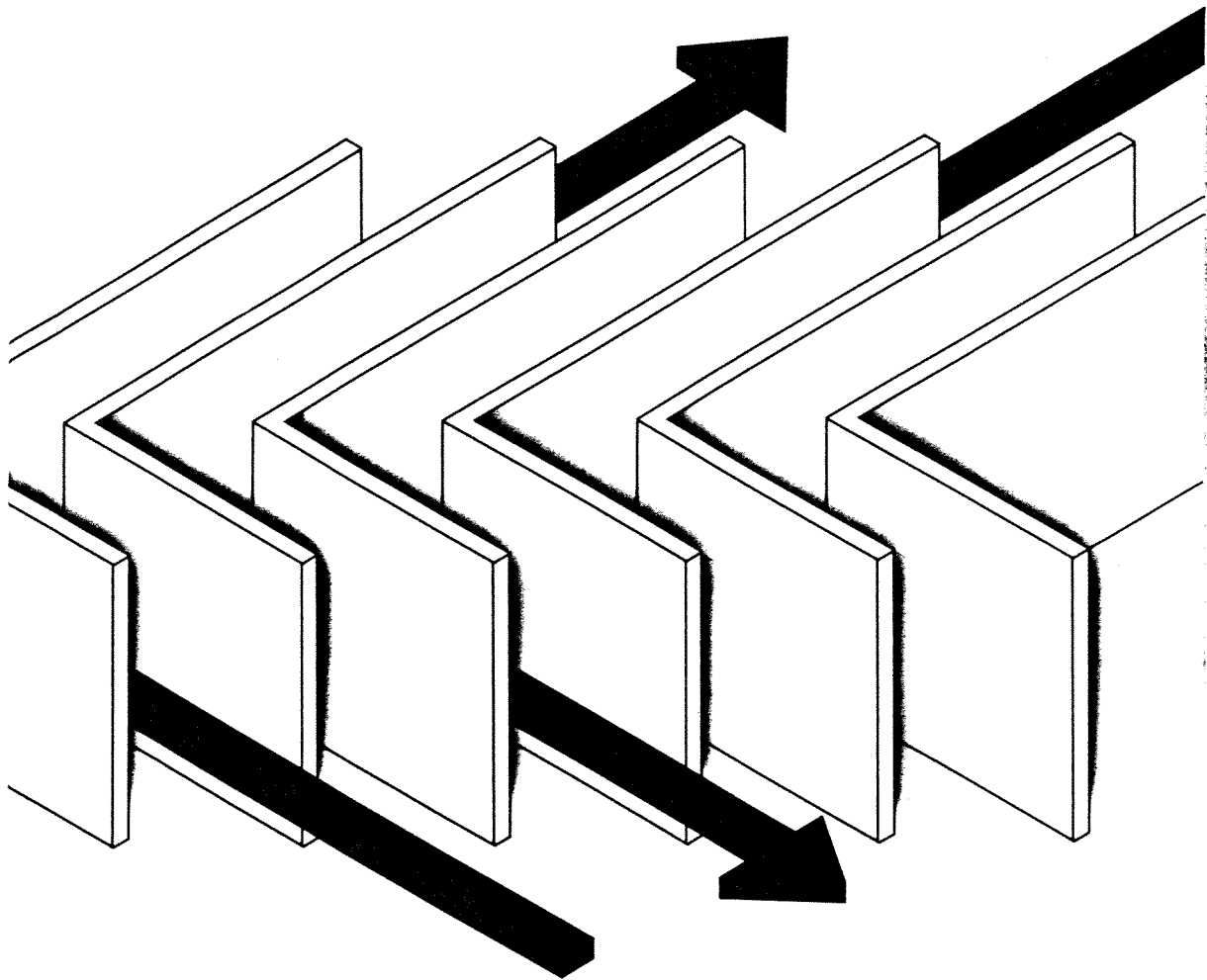# UNIVERSALS

OF LANGUAGE

Michel Kefer - Johan van der Auwera

Werner Abraham - Georg Bossong - Bernard Comrie - Matthew Dryer
Ronald Geluykens - Joseph Greenberg - Roland Noske
Jan-Ola Östman - Hansjakob Seiler - Willy Van Langendonck

Tiré à part

DISCOURSE-GOVERNED WORD ORDER
AND WORD ORDER TYPOLOGY

Matthew S. DRYER

State University of New York at Buffalo

The tradition of word order typology associated with Greenberg (1963) assumes, or is often taken to assume, that every language can be assigned a basic word order from the orders SOV, SVO, VSO, VOS, OSV and OVS.[1] There is also a tradition (e.g., Thompson 1978) that argues that we must distinguish two kinds of languages, those in which word order is syntactically determined and those in which word order is determined by discourse factors. In the former kind of language the notion of basic word order is relatively straightforward, since in such languages one order is clearly the dominant order. In languages of the latter sort, it is often less clear whether the notion of basic word order is a useful one. In such languages, all orders of subject, object, and verb are typically grammatical, and native speakers may not have clear judgments that one order is neutral or predominant. While text counts may reveal that one order in texts occurs more frequently than others, legitimate questions arise as to whether it is useful to designate this order as the basic order.

Mithun (1987: 309) offers a useful further distinction between two subtypes of languages with discourse-determined word order. On the one hand, there are languages like Czech in which word order is governed by discourse factors but in which one order is clearly the neutral or unmarked word order. In such languages, that one unmarked order can be considered the basic word order, since other orders will be considerably less frequent and will be associated with special discourse situations. On the other hand, there are languages (Mithun cites Cayuga as an example) in which it is much more difficult to isolate one order as neutral or unmarked, and in which the different orders occur with sufficient frequency that it is difficult if not impossible to isolate one order as the dominant one.

One conclusion that one might draw from this is that in classifying languages for the purposes of word order typology we should first distin-

guish those languages in which there is **no** basic word order from those in which there is a basic word order. The general enterprise of word order typology, as reflected in the work of Greenberg, Lehmann, Vennemann, and Hawkins, would be applicable only to languages of the latter sort. Questions about implicational universals, be they statistical or exceptionless, and questions about the general correlations associated with VO and OV languages, would not be considered applicable to languages in which one cannot isolate a single order as basic.

I will argue in this paper, however, that although the view outlined in the previous paragraph seems to be a common-sense response to the situation, there is a body of empirical evidence that suggests that traditional word order typology is applicable to at least some languages with highly flexible discourse-governed word order which seem to be instances of languages of the sort that lack a basic word order in Mithun's sense.

I will assume in this paper that relative frequency provides an adequate basis for determining basic word order. While there exist arguments in the literature for distinguishing basic word order from most frequent order (cf. Brody 1984, Siewierska 1988: 8), these issues go beyond the scope of this paper. For the purposes of this paper, I will essentially assume a notion of basic word order that is defined in terms of frequency: the basic word order in a language is that which is used most frequently. Whether there are other notions of basic word order which are not necessarily linked to frequency is an open question which I will not discuss here. But most of the empirical evidence in word order typology derives from studies like Greenberg (1963) which are based on grammatical descriptions that contain statements which describe some word order as the normal or most common order in the language. Hence a notion of basic word order that is linked to frequency is at least one for which there is empirical evidence of its usefulness.

One possible problem with treating the most frequent word order as the basic word order is provided by Mithun (1987:312). She plausibly questions the utility of treating one word order in a language as basic if that one order is only slightly more frequent than other orders. After all, in a language with highly flexible discourse-governed word order, one order may indeed occur slightly more often in texts simply because the discourse environments conditioning that order occur slightly more often. I suspect that where frequency differences are sufficiently small, it is probably better to say that the language lacks a basic word order. But the question then arises how great a difference in frequency is necessary in order to justify saying that the more frequent order is basic. Though any decision is necessarily an arbitrary division in what is really a continuum, I will assume in this paper that where one order is twice as frequent as any other order that is sufficient basis for treating the order as basic. Later in the paper I will present empirical

evidence that provides justification for this assumption: I will cite evidence from a number of languages in which OV order is more frequent than VO order by a ratio of only 2 to 1, but which otherwise exhibit typical OV characteristics. If we assume that the presence of OV characteristics in these languages is related to the fact that OV order is the more frequent order, then we have some justification for treating OV as the basic order in those languages. But I will assume that languages in which the frequency differences are much less than 2 to 1 are best treated as lacking a basic order.

My primary argument in this paper is that languages with discourse-governed word order are relevant to word order typology even when frequency differences are as low as 2 to 1. But it is worth noting first that even languages which lack a basic word order of subject, object, and verb in this sense may still be relevant to word order typology. At first sight, this might seem paradoxical: if a language lacks a basic word order, how can such a language bear on universal claims? One possibility arises from the fact that many universals that have been posited do not refer to the position of the subject or object. Consider, for example, Greenberg's Universal 2: "In languages with prepositions, the genitive almost always follows the governing noun, while in languages with postpositions it almost always precedes". (Greenberg 1963: 78). There are languages which exhibit highly flexible word order at the clause level, so that it is difficult to assign a basic word order at that level, but in which the order is less flexible at the phrase level, so that it is possible to identify a basic or normal order for other pairs of elements. For example, my counts of texts in Hupa (an Athapaskan language of California) in Goddard (1903-4) reveal the frequencies given in Table 1.

SV 14 (54 %)		OV 32 (53 %)
VS 12 (46 %)		VO 29 (47 %)

Table 1: Hupa

On the basis of the frequencies in Table 1, it is clear that word order in Hupa is highly flexible, and that the language is apparently one that lacks a basic word order. Despite this high flexibility at the clause level, there is considerably less flexibility at the phrasal level. The language is genitive-noun (GenN) (Goddard 1911:103, Golla 1970:296) and postpositional (Goddard 1911:103, 150) and text counts of Goddard (1903-4) show that the language is demonstrative-noun, numeral-noun, and noun-adjective. Hupa is thus relevant to universals like Greenberg's Universal 2. Since Hupa is postpositional and GenN, it conforms to this universal, even though it cannot be assigned a basic order at the clause-level. Similarly, Scancarelli (1986, 1988: 173-198) shows that the order of constituents at the clause level in Cherokee is highly flexible and governed by discourse factors. But she also

reports (1988: 173-174) that the language is rigidly GenN and that it has postpositions rather than prepositions; hence Cherokee too conforms to Greenberg's Universal 2.

Another way in which languages without a basic order of subject, object, and verb might still be relevant to word order typology is that one might be able to determine a basic order for the object and verb, even if one cannot determine a basic order for clauses containing a subject, object, and verb. The text counts in Table 2 for Korana, a Khoisan language spoken in South Africa, based on texts in Meinhof (1930), suggest that while it may be difficult to identify a basic order for subject and verb, the order OV is sufficiently predominant to justify saying that the basic order of object and verb is OV, and hence that the language is an OV language.

Table 2: Korana

SV 28 (61 %)	OV 24 (89 %)
VS 18 (39 %)	VO 3 (11 %)

This is significant since the claims of Lehmann and Vennemann are largely formulated in terms of the order of object and verb, ignoring the order of subject and verb. And in fact, when we examine the other word order characteristics of Korana, we find that it exhibits characteristics that are typically associated with OV languages: it uses postpositions (Meinhof 1930: 55-56); it is normally genitive-noun (p. 40); it uses a clause-final question particle (or clitic) (p. 61); it places subordinate conjunctions at the end of adverbial clauses (p. 61); it places the negative after the verb (p. 54-59) (as shown in Dryer 1988b, postverbal negatives are far more common in OV languages than in VO languages); and it can place relative clauses before the noun (p. 62) (while it is not clear this order is basic in Korana, it is shown in Dryer (1986) that RelN order as a basic order is generally found only in OV languages). Hence, despite the problems in identifying a basic order for subject and verb, classifying Korana as OV allows one to account for these other word order characteristics.

There are many other languages in which it is possible to identify one order of object and verb as most frequent, but which are difficult to classify according to the traditional typology of SOV, SVO, etc. The text counts in Table 3 from Papago, reported by Payne (1987), show that its word order is sufficiently flexible that no one order of subject, object, and verb emerges as the dominant one:

Table 3: Papago

SVO 8 (36 %)
OVS 5 (23 %)
VOS 5 (23 %)
VSO 3 (14 %)
SOV 1 (5 %)

Although SVO order is the most common order in this body of texts, it constitutes a clear minority of these clauses. In fact, from such small numbers we cannot tell whether the differences are simply due to chance. But even if the same proportions were maintained over a much larger sample, it would seem misguided to classify the language as SVO when the other orders occur with relatively high frequency. On the other hand, if we simply examine the order of subject and verb and the order of object and verb, as in the counts in Table 4 from Payne, a clearer pattern emerges.

Table 4: Papago

SV 48 (23 %)	OV 44 (29 %)
VS 158 (77 %)	VO 108 (71 %)

These counts show that both subjects and objects more commonly follow the verb in Papago, providing a possible basis for classifying the language as VS and VO. There are two reasons why we can be more comfortable with such a classification compared to classifying the language as SVO on the basis of the figures in the Table 3. For one thing, the total numbers in Table 4 are much larger: there is less likelihood that the greater numbers of VS and VO are simply due to chance. And more importantly, the orders VS and VO each occur in more than two-thirds of the clauses in question, while SVO occurred in only 36 % of the clauses in question.

Despite these arguments, there are other considerations, discussed by Payne, that seem on the surface to provide a good argument against describing Papago as VS or VO. Payne's data in Table 5 shows that the order of subjects and objects with respect to the verb correlates highly with whether they are definite or indefinite:

Table 5: Papago

	preverbal	postverbal
indefinite	125	26
definite	6	278
pragmatically marked	38	2
other	1	4

It is clear from this data that the position of subjects and objects is not determined by their syntactic status. There is thus no reason to say that the syntax of Papago states that the language is VO and VS. Rather, the natural conclusion is that the grammar of the language will contain rules linking the definiteness of nominals (or some related discourse properties) with their order with respect to the verb. The fact that VS and VO order are more common, as shown in Table 4, can presumably be explained as being simply due to a greater frequency of definite subjects and objects (at least in the texts examined). In other words, Table 5 provides data that is a direct reflection of the grammar of Papago, while Table 4 provides data that does not directly reflect anything about the grammar of Papago. Rather the greater frequency of VS and VO order in Papago is simply a side effect of two things: the discourse rules reflected in Table 5 and the relative frequency in actual discourse of definite as opposed to indefinite subjects and objects. These considerations would seem to provide a good reason not to describe Papago as VS and VO.

But it is ultimately an empirical question whether languages like Papago exhibit word order properties that are associated with VO or OV languages. A priori, we would certainly expect not: if the word order correlations are to be explained in terms of the form of grammars, then we would not expect Papago to exhibit VO characteristics if there is nothing in the grammar of Papago that says that it is VO. But I will argue in the remainder of this paper that languages with discourse-governed word order like Papago do tend to exhibit characteristics that correlate with the order of object and verb that is most common in the language.

What word order characteristics are typical of languages which are VS and VO? According to Greenberg's Universal 17, languages with VSO order are overwhelmingly noun-adjective (NAdj). I have shown elsewhere (Dryer 1988a), however, that this universal does not hold up against a larger sample than the 30-large sample on which Greenberg based this generalization. It turns out that adjectives follow the noun no more often in VSO languages (or languages which are VS & VO) than they do in SOV languages: although there is a weak tendency for adjectives to follow the

noun in VS & VO languages, the same in true for SOV languages. On the other hand, two characteristics which are typical of VS & VO languages are that they are prepositional and noun-genitive (NGen). Evidence for this is given in Tables 6 and 7, based on a total sample of 604 languages. The method used for presenting the data is explained in detail and justified in Dryer (to appear). The languages in the sample are grouped into language groups comparable in time depth to the subfamilies of Indo-European. These groups are further grouped into six large geographical areas: Africa, Eurasia (excluding southeast Asia), Southeast Asia & Oceania, Australia-New Guinea, North America, and South America. The numbers represent the number of groups containing languages of the given sort in each of the six areas. Square brackets indicate the predominant type within each area.

Table 6: Prepositions vs. Postpositions in VS & VO Languages

	Africa	Eurasia	SEAsia & Oc	Aus-NewGui	NAmer	SAmer	Total
Prep	[5]	[2]	[6]	[1]	[16]	[3]	33
Postp	0	0	0	0	2	2	4

Table 7: Order of Genitive and Noun in VS & VO Languages

	Africa	Eurasia	SEAsia & Oc	Aus-NewGui	NAmer	SAmer	Total
NGen	[6]	1	[6]	0	[19]	[3]	35
GenN	0	1	1	[2]	2	2	8

The data in Table 6 shows that in all six areas, there are more groups containing VS & VO & Prepositional languages than there are groups containing VS & VO & Postpositional languages. There are only four groups in my sample containing languages of the latter sort. Thus there is clear evidence that VS & VO languages tend to be prepositional rather than postpositional. The data in Table 7 shows that there is a similar tendency for VS & VO languages to be NGen, although the tendency is not as strong. What this means is that if it is correct to treat Papago as a VS & VO language then we would expect it to be prepositional and NGen.

At first sight, the word order characteristics of Papago would seem not to support this prediction. Two recent descriptions of the language, Saxton (1982) and Zepeda (1983), differ somewhat in their descriptions of these word order characteristics. Both orders of genitive and noun are permitted, as in (1), from Saxton (1982:185):

(1) (a) NGen: g   kii-ǰ   g   huan
art house-poss art Juan
'Juan's house'

(b) GenN: g   huan   kii
art   Juan   house
'Juan's house'

Although the GenN order illustrated in (1b) is grammatically unmarked (since (1a) includes two morphemes that are absent in (1b)), I have not seen data on the relative frequency of the two orders, and thus have no reason for designating one of the two orders as basic on the ground of frequency. Saxton (1982: 185) refers to the NGen order as basic, but the discussion in Zepeda (1983: 131-132) suggests that it might be best to designate neither order as basic, and in fact Zepeda describes only the GenN order in the earlier chapter in which she introduces the possessive construction, implying, perhaps, that GenN is basic. Furthermore, Pamelo Munro (personal communication) reports that the GenN order is the usual one in the Pima dialect. For these reasons, it seems best to treat neither order as basic, and thus to treat the language as GenN/NGen.

A very similar situation obtains with adpositions. There are two constructions for adpositions, one in which the adposition precedes the noun phrase, and is therefore a preposition, as in (2a), and another construction in which the same morpheme occurs after the noun phrase, and is therefore a postposition, as in (2b) (see Saxton 1982: 189):

(2) (a) Prep: wiima-ǰ   g   huan
with-poss   art   Juan
'with Juan'

(b) Postp: huan   wiim
Juan   with
'with Juan'

The two structures differ in the same way as the genitive constructions in (1), presumably indicative of the fact that the adpositional structure is at least historically a possessive structure. Whether one or the other of these two structures is basic is unclear and again it seems best to classify the language as indeterminately Pr/Po.

As discussed above, VO languages exhibit a strong tendency to be prepositional and NGen rather than postpositional and GenN. Papago fails to conform to these tendencies, since it exhibits both orders, both for adpositions and for genitives. Although it is not really an exception, since these generalizations are intended to apply to languages that are clearly prepositional or clearly postpositional, the fact that Papago exhibits these characteristics would seem to suggest that the word order correlations do not apply to languages like Papago which are VS & VO only in the weak sense that VS and VO are more common as a side effect of discourse factors. When we examine the language in a historical context, however, there is reason to believe that the language has recently undergone a word order change, that the higher frequency of VS and VO order is a relatively recent innovation (perhaps even within the past few hundred years), and that the language has also been undergoing a more recent change in the order of genitive and noun and adposition type.

The large sample of languages that provides the basis for the data in Table 6 and 7 above contains two other languages in the Pimic branch of Uto-Aztecan to which Papago belongs. One of these, Northern Tepehuan, is described by Bascon (1982: 273) as VSO, and the examples cited are generally postpositional and GenN. The third Pimic language in my sample, Nevome, is described by Shaul (1982: 10-11) as SOV, in fact as being rigidly verb-final. It otherwise exhibits typical OV characteristics, being GenN and postpositional. Various considerations point to OV as the original Pimic order with the VS and VO order of Northern Tepehuan and Papago as an innovation. First, Nevome is an extinct dialect of Pima Bajo; Shaul's description is based on documents dating from the 17th century. Second, non-Pimic Uto-Aztecan languages are more commonly OV and exhibit the GenN and postpositional properties typical of OV languages. Table 8 lists the Uto-Aztecan languages from my sample, with basic clause order and basic order of genitive and noun and adposition type:

Table 8: Word Order in Uto-Aztecan Languages

Group	Language				
Pimic	Papago	VS	VO	GN/NG	Po/Pr
	N. Tepehuan	VS	VO	GN	Po
	Nevome	SV	OV	GN	Po
Numic	N. Paiute	SV	OV	GN	Po
	Shoshoni	SV	OV	?	Po
	Comanche	SV	OV	GN	?
	Ute	SV/VS	OV	GN	Po
	Chemehuevi	SV	OV	GN	Po
Takic	Luiseño	SV	OV?	GN	Po
Hopi	Hopi	SV	OV	GN	Po
Taracahitic	Yaqui	SV	OV	GN	Po
	Tarahumara	SV	OV	GN	Pr/Po
Aztecan	Pipil	VS	VO	NG	?
	Huasteca Nahuatl	VS	VO	NG	Pr
	Michoacan Nahuatl	SV/VS	VO	GN/NG	Pr
	North Pueblo Nahuatl	SV	VO	NG?	?
	Tetelcingo Nahuatl	SV	VO	NG	Pr
Coric	Cora	VS	VO	GN/NG	Po
	Huichol	SV	OV	GN	Po

Most of these word order properties are based on grammatical descriptions rather than text data; it is possible that examination of texts would alter some of these characterizations. The general picture here is that Numic, Takic, and Taracahitic languages, as well as Hopi, are OV, GenN, and postpositional (though Tarahumara also has prepositions), while the Aztecan languages are VO, NGen, and prepositional. The Aztecan languages are spoken in the Meso-American linguistic area (cf. Campbell, Kaufman and Smith-Stark 1986) in which only VO order is found (and in which VS order is wide-spread), so the word order in Aztecan languages is plausibly due to diffusion from other languages in the area. Cora (Casad 1984) is the only language exhibiting characteristics reminiscent of those we find in Papago, being VSO, GenN/NGen, and postpositional; Huichol, the other Coric language in my sample is OV, SV, GenN, and postpositional. While Papago and Cora lie outside the Meso-American area, it is not impossible that the rise in VS and VO order in these languages has been influenced by languages to the south. At any rate, except for the Aztecan languages and Cora, the pattern outside Pimic is not only OV but also GenN and postpositional. The natural interpretation is that Proto-Uto-Aztecan was OV, GenN, and postpositional

(cf. Langacker 1977: 24 for a similar conclusion), that Proto-Pimic was the same, and that Papago has undergone recent changes in word order, not only at the clause level, but in innovating NGen order and prepositions along with the older GenN order and postpositions. Note in addition that the NGen order in Papago, illustrated in (1a) above, is grammatically marked by comparison with GenN order, suggesting that it is the more recent order. The change in the grammar of Papago governing the apparent increase in VS and VO order has apparently been one by which the rules associating discourse functions with word order have changed so that while the effect of these rules at one time was that SV and OV were more common, VS and VO are now more common. But the fact that NGen order and prepositions have been increasing in frequency of occurrence is most likely connected to the apparently recent increase in VS and VO order. What this means is that although the higher frequency of VS and VO order in Papago does not appear to be a fact about the grammar of Papago, but only a side effect of a discourse rule and the higher frequency in discourse of definite NP's, this higher frequency has been sufficiently strong to trigger other word order changes in the language, towards NGen order and prepositions. While there is no evidence that these orders will become dominant, the apparent fact that they have become frequent as a result of the high frequency of VS and VO order suggests that there is a sense in which we can usefully describe Papago as VS and VO, not in the sense that these are properties of the language that are reflected in its grammar, but in the sense that Papago apparently "counts" as a VS and VO language for the purposes of word order typology, as reflected in the Greenbergian word order correlations. While it is clear that there are certain word order characteristics associated with languages that are VO and VS, it is not obvious what is the relevant class of VS and VO languages which tend to exhibit these characteristics, i.e. what counts as a VS and VO language for the purposes of the correlations. The evidence from Papago argues that it does count as a VS and VO language for the purposes of word order typology, in the sense that the class of languages that tend to exhibit the characteristics typical of VS and VO languages includes languages like Papago which are VS and VO only in the sense that these orders are more common as a side effect of dicourse rules.

At this point we are faced with an apparent paradox. A widely held assumption is that generalizations about languages are to be formulated as generalizations about grammars. But the account I have given for Papago word order is at odds with this assumption, since the Greenbergian word order correlations are generalizations about languages and, as I have argued, Papago counts as a VS and VO language for the purposes of the Greenbergian word order correlations, despite the fact that there is no reason to believe that the grammar of Papago says that the language is VS and VO. Insofar as the grammar of Papago has changed, what has apparently changed is the rules that associate particular discourse functions with particular word order.

orders. In other words, it appears that modern Papago has descended from a language in which word order was determined by discourse factors whose effect was such that OV order occurred more often, and that these rules have changed in such a way that VO order is now more common. Thus, Papago has changed from OV to VO only in an extragrammatical sense. But if the Greenbergian word order correlations are interpreted as generalizations over grammars, then we would not expect this change to trigger a change in the order of genitive and noun or in adposition type.

How do we circumvent this paradox? One possibility is that, contrary to Payne's claims, the grammar of Papago does specify that the language is now VO. Although there is no way to disprove such a proposal, there is no independent reason to believe that such is the case, no reason to doubt that Papago word order is determined entirely by discourse factors of the sort discussed by Payne.

A second possibility is that the order of genitive and noun and that of adposition and NP are determined by the same discourse factors as those determining the order of nominals relative to the verb. Under this hypothesis, the change in the order of genitive and noun and of adpositions and NP is not a separate change from that of nominal and verb but simply part of a change in a single rule of discourse grammar; i.e., it might be that the same discourse factors that govern the order of nominals with respect to the verb also govern the order of nominals with respect to nouns and adpositions. If this is correct, the change would not really be a change in the order of genitive and noun – and that of adposition and NP – which was triggered by a change in the order of subject/object and verb, and thus not really an example of an instantiation of the Greenbergian word order correlations. Payne does not discuss the matter, so the question of what discourse factors may condition the order of genitive and noun and that of adposition and NP remains uninvestigated. But there are a variety of considerations that would lead us to expect that these other alternations are not governed by discourse factors in exactly the same way as the order of nominals with respect to the verb. In many other languages in which the order of nominals with respect to the verb is determined by discourse factors, the order of genitive and noun is determined by other factors. In many languages, in fact, the order at the clause-level is flexible but the order at the phrase-level, at least with respect to the order of genitive and noun, is not. Cherokee, as noted above, and Turkish (Lewis 1967: 239-244) are apparent examples of languages of this sort. Flexibility in the adpositional phrase is cross-linguistically uncommon and Papago is relatively unusual in allowing both orders of adposition and NP. And while both orders of genitive and noun are found in many languages, the alternation is often governed by factors other than discourse factors: English exhibits an alternation (*John's book, picture of John*) which is governed by a complex interplay of semantic and syntactic factors. Dutch

exhibits a similar alternation, but only human genitives can occur in the GenN order (Shetter 1958: 50). And even in languages in which the order of nominals with respect to the verb is governed by discourse factors, and in which the order of genitive and noun is flexible, the two alternations are often governed by different factors. Payne (1985: 201-235) discusses the factors determining the position of nominals with respect to the verb in Yagua, but it is clear from her evidence that the order of genitive and noun is not governed by the same factors. For one thing, GenN order is much more common than NGen order in Yagua (by over 10 to 1 according to Payne's text count), while nominals more commonly follow the verb. Hanis Coos (Dryer 1983) exhibits an alternation between GenN and NGen order, but again the alternation is not conditioned by the same factors that condition the position of nominals with respect to the verb: the latter correlates with definiteness is a way very similar to Papago, but while it is not common for definite subjects and objects to precede the verb, definite genitives often precede the noun. This is not to deny that there may be languages in which the order of genitives with respect to the noun is governed by the same factors as those governing the order of nominals with respect to the verb. But it seems more common cross-linguistically for alternations in the order of genitive and noun to be governed by factors distinct from those governing the order of nominals with respect to the verb. Nor should this be surprising: the clause is an informational unit in a sense that phrases are not. Thus discourse-determined word order within the clause involves arranging the parts of that information in a way that seems less useful as a way of organizing parts of a phrase.

There are a few further considerations specific to Papago that suggest that the order of genitive and noun and the order of adposition and NP are not governed by the same discourse rule that governs the position of nominals with respect to the verb. The hypothesis that they are implies that the increase in NGen order and the use of prepositions should have been simultaneous with the change towards VO and VS order. But various facts suggest that the increased use of prepositions and NGen order has taken place later than the increase in VO and VS order. One of these is the fact that the closely related language Northern Tepehuan, as described by Bascom (1982), is VSO while still being postpositional and GenN. This suggests that the change in clause order has occurred in Northern Tepehuan without the change in phrase order occurring. The word order in Northern Tepehuan may be the same as that of an earlier stage of Papago. In addition, Pamela Munro (personal communication) reports that in the Pima dialect, which is mutually intelligible with Papago (and which Hale (1958) separates from Papago at a time depth of about 200 years), GenN order is clearly preferred over NGen. Finally, if the position of genitives in Papago were determined in the same way as nominals at the clause level, we might expect NGen order to be more common. For one thing, genitives are probably overwhelmingly

definite: Du Bois (1988:8) reports that a text count of genitives in Sacapultec revealed only 4.2 % of genitives to be new information. If genitives are overwhelmingly definite, and if the order of genitive and noun in Papago were determined by the same factors determining the order of nominals with respect to the verb, then we would expect NGen order to predominate. But there is no evidence that it does, and examples of GenN order in Zepeda (1983) and Saxton (1982) all apparently involve definite genitives. All of these considerations make it seem unlikely that the order of genitive and noun and that of adposition and NP in Papago are governed by the same factors as those governing order at the clause level; therefore, it is unlikely that the change in the order of genitive and noun and in adposition type is simply the reflection of a change in a single rule of the language associating certain discourse factors with the order of nominals with respect to verb, noun or adposition.

A third solution to the paradox, and the one that I will propose here, is that although the greater frequency of VO and VS order may not be something in the grammar of Papago, or even something that follows directly from the grammar of Papago, it may be something that speakers of the language are (at least unconsciously) aware of, and this extragrammatical awareness may be enough to trigger other changes in word order in the language. In other words, the solution to the paradox is to question the assumption that generalizations over languages are necessarily to be understood in terms of generalizations over grammars.

There are independent reasons to believe that speakers may be aware to varying extents of the frequencies of different constructions. Certainly, speakers often have conscious awareness of major frequency differences between words or constructions, such as the difference between actives and passives in English. The psychological literature on word recognition provides considerable evidence that awareness of the relative frequency with which different words occur influences word recognition (Morton 1970) and that the relative frequency of different meanings of a given word plays a role in sentence comprehension (Onifer and Swinney 1981). Oldfield and Wingfield (1965) provide evidence that relative frequency of different words plays a role in production. In addition, the literature on sentence processing makes various kinds of appeals to speaker awareness of different frequencies, at least with respect to competing subcategorization frames for specific verbs (cf. Ford, Bresnan and Kaplan 1982:745, Schubert 1984). In addition, Fodor, Bever and Garrett (1974:344-348) argue that perceptual strategies motivated by highly frequent structures play a role in sentence processing: the basic idea is that the human parser attempts first to parse incoming sentences in terms of the kinds of structures that are encountered most often. The knowledge – conscious or unconscious – of frequency differences between different constructions is extragrammatical in the way

that strategies in playing a game like chess are distinct from knowledge of the rules.

My suggestion then about the apparent changes in Papago word order is that the discourse rules governing the order of nominals with respect to the verb changed, and that although the increase in postverbal order for such nominals was simply a side effect of that change, speakers were sufficiently aware of the higher frequency of postverbal nominals that it gradually triggered a change in word order at the phrase level, with the innovation and increased use of NGen order and prepositions. If this line of explanation is correct, it would illustrate that one cannot characterize the word order correlations in purely grammatical terms, since the class of languages which are VO languages for the purposes of the correlations would not constitute a natural class in terms of the form of their grammars, but only in terms of characteristics that are shared by the languages and are not shared by their grammars. In addition, if correct, this line of explanation further demonstrates the impossibility of explaining the form of grammars in purely grammatical terms: the motivation for the innovation of prepositions and NGen order in Papago cannot be understood in terms of the form of the grammar at the time this innovation took place.

I have discussed the case of Papago at some length since it presents a relatively clear illustration of the theme of this paper: the fact that word order in a language is governed by discourse factors does not constitute a reason not to classify it as OV or VO on the basis of frequency of order, even if the frequency difference is demonstrably attributable to a side effect of the discourse rules. I will turn now to a briefer examination of a number of other languages which seem to provide further support for this conclusion.

Ute is another Uto-Aztecan language on which we have considerable evidence from Givón (1983b), both on the relative frequency of the different orders of subject and object with respect to the verb and on the discourse factors governing the variable order. The frequency of the different orders of subject and object with respect to the verb is given in Table 9.

Table 9: Ute

SV 39 (61 %)	OV 34 (72 %)
VS 25 (39 %)	VO 13 (28 %)

Although the difference in frequency between the two orders of subject and verb is not great, there is a fairly clear preference for OV over VO, by a ratio of over 2 to 1. Despite the differences from the frequencies cited for Papago, the general direction of the discourse factors determining the alternation is

similar. Namely, VO is associated with significantly lower referential distance (cf. Givón 1983a); the mean distance in number of clauses from a given NP back to a previous NP referring to the same individual or object is significantly lower for postverbal objects. As with Papago, indefinite objects usually precede the verb. The reason that OV is more common in Ute than it is in Papago is that definite objects are common both before and after the verb. Givón's data does not provide much evidence bearing on what determines the position of definite objects, except that those with a high referential distance are more likely to precede the verb. At any rate, as with Papago, there seems to be little reason to believe that the grammar of Ute states that the language is OV, or even that it states that OV is the unmarked order. Rather, the word order is determined by discourse rules, but the rules are different from those in Papago in a way that results in OV being more common. And significantly, Ute otherwise exhibits typical OV characteristics, particularly in being GenN and postpositional.

In Tlingit (Dryer 1985:8), at least for the dialects other than the Auk dialect, OV order occurred in 88 out of 131, or 67% of the clauses containing objects. This is barely a difference of 2 to 1, yet Tlingit otherwise exhibits distinctive OV characteristics: it is GenN, postpositional, and places the relative clause before the noun (cf. Naish 1979:102). This last characteristic is not an OV characteristic in the sense that most OV languages are RelN – in fact, both orders of relative clause and noun are common among OV languages (cf. Dryer 1986:102). But it is an OV characteristic in the sense that prenominal relative clauses are generally only found in OV languages. Since RelN order is presumably linked to the fact that OV order is more common in the language. But VO order is sufficiently common that it seems likely that the choice between OV an VO order is determined by discourse factors.

Text counts of Huallaga Quechua by Weber (1983:18-19) show OV outnumbering VO by just over 2 to 1: out of 276 clauses, 190 (or 69%) were OV while 86 (or 31%) were VO. Again the relatively high frequency of VO order implies that the alternation between OV and VO is discourse-governed, since no syntactic basis for the alternation was found. But Huallaga Quechua otherwise exhibits typical VO characteristics, being postpositional, GenN, and RelN.

I will now give data for a number of further examples of languages in which word order appears to be discourse-governed but in which one order of object and verb is at least twice as common as the other. Unless otherwise noted, the text counts were performed by myself, so they should be considered tentative. The very small totals in some cases is further reason to consider the conclusions tentative. In addition, I do not know in these cases

	Frequency of OV	OV characteristics
Ute	72 %	GenN, Po
Tlingit	67 %	GenN, Po, RelN
Huallaga Quechua	69 %	GenN, Po, RelN
Trumai	65 %	GenN, Po
Koryak	66 %	GenN, Po
Tacana	86 %	GenN, Po
Takelma	85 %	GenN, Po, Clause-final subordinator

Table 10: Summary of frequency of OV order and OV characteristics

what discourse factors govern the alternations. Rather, I assume, since both orders occur relatively frequently, that the alternation is discourse-governed. It is always possible that the alternation in some of these cases there may be syntactic factors, but the text counts are restricted to nominal objects, excluding pronouns and clausal objects. In none of these cases is there any reason to believe that the order is syntactically determined.

In Trumai (Monod-Becquelin 1975), a language of Brazil classified by Greenberg (1987) as an isolate within Equatorial, my small text count revealed 11 OV clauses and 6 VO clauses. But the language is GenN and postpositional, consistent with the fact that OV is more common. My counts for texts in Koryak (a Chukchee-Kamchatkan language of eastern Siberia) in Bogoras (1917) revealed 21 OV clauses and 11 VO clauses, again a difference of about 2 to 1. Again Koryak otherwise exhibits the OV characteristics of being GenN and postpositional. In Tacana, a language of Bolivia classified by Greenberg as Pano-Tacanan within Ge-Pano-Carib, my counts of texts in Ottiviano (1980) revealed 12 OV and 2 VO. The numbers are such that it is possible that the language is syntactically OV, but if the alternation is discourse-governed, OV is apparently more common. Again the language otherwise exhibits OV characteristics, in being GenN and postpositional, and in using clause-final subordinators in adverbial clauses. And in Takelma (Sapir 1909), a putative Penutian language of Oregon, my counts found 11 OV clauses and 2 VO. Again, Takelma is GenN and postpositional.

Table 10 summarizes the evidence presented other than that from Papago:

It is important to point out that in five of the seven languages in Table 10 the ratio of OV order to VO order in the texts examined was only about 2 to 1.

This provides justification for the assumption that frequency differences of 2 to 1 provide enough basis for determining basic word order.

It must be noted that there are other languages in which the word order characteristics are not what one might expect, given the more frequent order of object and verb. From a synchronic point of view Papago does not conform since it is VS and VO but not basically NGen and prepositional. But the diachronic evidence that these latter characteristics have become more common with the greater frequency of VS and VO makes Papago a supporting example. Yagua (Payne 1985) is a language with flexible order in which VS and VO are more common due to discourse factors. But it is GenN and postpositional. VS and VO are also more common in Hanis Coos (Dryer 1983), but GenN order is somewhat more common than NGen. Cree, an Algonquian language spoken in Canada, appears to be more commonly VO from my text counts, but is clearly a language with discourse-governed order; but in so far as it has adpositions, these are postpositions (Rich Rhodes, personal communication). Thus at best we can say that languages with discourse-governed word order **tend** to exhibit the word order characteristics associated with the order that is more common as a result of the discourse factors governing order in the language.

The general theme of this paper is that languages with discourse-governed word order often exhibit the word order characteristics associated with the more frequent word order in the language. This is the case apparently despite the fact that there is no reason to believe that there is any rule in the grammar of these languages that states that the more frequent order is in any sense the basic order. Rather, this order is apparently more frequent due to the interaction of the rules associating discourse functions with particular orders and the relative frequency with which the relevant discourse contexts occur in typical discourse. The strongest evidence I have given for this conclusion is based on apparent word order changes in Papago. But the evidence I have given is weak in a variety of respects. First, although I have cited a number of languages that conform, I have also noted a number that do not. Unless the number of languages that conform significantly exceeds the number that do not, the evidence is inconclusive. But the number of languages I have examined here is insufficient to draw any firm conclusion at this time. There is a clear need to examine a number of additional languages from the perspective of the issues in this paper. Second, the evidence cited here for a number of languages is rather weak, both because the text counts are small and were performed by someone who was not an expert in the language and because we do not know for sure that the alternations are governed by discourse factors. Third, except for the diachronic evidence from Papago, all of the supporting languages are languages in which OV order is more common and which exhibit OV characteristics. This is particularly disturbing since all of the nonconforming languages are

languages in which VO order is more common but which exhibit OV characteristics. Furthermore there are other languages I have examined, such as Hupa (discussed above), in which OV and VO occur with roughly equal frequency, but in which other characteristics are ones associated with OV languages. Thus the evidence cited here is actually more consistent with an hypothesis according to which languages with discourse-governed word order tend to exhibit OV characteristics, regardless of which order of object and verb is more common. Only the diachronic evidence from Papago and a priori assumptions cast doubt on this hypothesis. A final weakness in the evidence presented here is that it is based entirely on languages of the New World, except for Koryak, which is spoken in Siberia just across the Bering Strait from the New World. I have shown elsewhere (Dryer to appear) that there are dangers in trying to draw typological inferences from languages from just one area of the world. The reasons for the areal bias in the set of languages examined in this paper is twofold. First, there is generally more accessible text data for languages of the New World. And second, languages with discourse-governed word order are far more common in the New World; except for Australia, languages in the Old World less commonly have discourse-governed word order, and if they do, one order is often clearly the unmarked order. No languages of Australia are discussed here for a variety of reasons. For one thing, many languages of Australia exhibit so much flexibility that it is difficult to identify even a more frequent order, or at least an order of object and verb that occurs as much as twice as often as the other order. In addition, many Australian languages lack adpositions altogether, or other constructions relevant to word order typology, so it is more difficult to test whether they have OV or VO characteristics. Nevertheless, until the claims of this paper are tested against a wide variety of languages, they will remain speculative.

A final point that should be made is that the evidence from this paper shows that it is useful to determine the most frequent order in a language without examining the particular discourse factors that govern word order. But this does not mean that the latter kind of question is unimportant. In fact, from the perspective of describing a language, the relative frequency of different orders is far less important than the discourse factors governing the alternation. What this paper has shown is that it is also useful to know the relative frequency of the different orders.

NOTE

REFERENCES

BASCOM, Burton. 1982. Northern Tepehuan. In R. Langacker (ed.), Studies in Uto-Aztecan Grammar, Volume 3: Uto-Aztecan Grammatical Sketches, 267-293. Summer Institute of Linguistics.

BOGORAS, Waldemar. 1917. Koryak Texts. Publications of the American Ethnological Society. Leyden: Brill.

BRODY, Jill. 1984. Some problems with the concept of basic word order. Linguistics 22, 711-736.

CAMPBELL, Lyle, Terrence KAUFMAN, and Thomas C. SMITH-STARK. 1986. Meso-America as a linguistic area. Language 62, 530-570.

CASAD, E.H. 1984. Cora. In R. Langacker (ed.), Studies in Uto-Aztecan Grammar, Volume 4: Southern Uto-Aztecan Grammatical Sketches, 151-459. Summer Institute of Linguistics.

DRYER, Matthew S. 1983. Coos word order. Paper presented to the annual meeting of the Western Conference on Language (WECOL), Eugene, Oregon.

——. 1985. Tlingit: an object-initial language? Canadian Journal of Linguistics 30, 1-13.

——. 1986. Word order consistency and English. In S. Delancey and R. Tomlin (eds.), Proceedings of the Second Annual Pacific Linguistics Conference, 97-106. Eugene: Dept. of Linguistics, University of Oregon.

——. 1988a. Object-verb order and adjective-noun order: dispelling a myth. Lingua 74, 77-109.

——. 1988b. Universals of negative position. In M. Hammond, E. Moravcsik, and J. Wirth (eds.). Studies in Syntactic Typology. Amsterdam: John Benjamins.

——. To appear. Large linguistic areas and language sampling. Studies in Language.

Du Bois, John W. 1988. Discourse as pattern model for grammar: the possessor = ergator affiliation. University of California, Santa Barbara.

FODOR, J.A., T. G. BEVER, and M.F. GARRETT. 1974. The Psychology of Language: An Introduction to Psycholinguistics and Generative Grammar. New York: McGraw-Hill.

FORD, Marilyn, Joan BRESNAN, and Ronald N. KAPLAN. 1982. A competence-based theory of syntactic closure. In J. Bresnan (ed.), The Mental Representation of Grammatical Relations, 727-796. Cambridge, Massachusetts: MIT Press.

GIVÓN, Talmy. 1983a. Introduction. In Givón (ed.), Topic Continuity in Discourse: A Quantitative Cross-Language Study, 1-41. Amsterdam: John Benjamins.

——. 1983b. Topic continuity and word-order pragmatics in Ute. In Givón (ed.), Topic Continuity in Discourse: A Quantitative Cross-Language Study, 141-214. Amsterdam: John Benjamins.

GODDARD, Pliny Earle. 1903-4. Hupa Texts. American Archaeology and Ethnology. Volume 1, No. 2. University of California Publications, 89-378.

——. 1911. Athapascan (Hupa). In F. Boas (ed.), Handbook of American Indian Languages, Bureau of American Ethnology Bulletin 40, Part 1, 85-158. Washington. Bureau of American Ethnology.

GOLLA, Victor Karl. 1970. Hupa Grammar. University of California at Berkeley dissertation.

GREENBERG, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In J. Greenberg (ed.). Universals of Language, 73-113. Cambridge, Massachusetts: MIT Press.

——. 1987. Language in the Americas. Stanford University Press.

HALE, Kenneth. 1958. Internal diversity in Uto-Aztecan: I. International Journal of American Linguistics 24, 101-107.

HAWKINS, John A. 1983. Word Order Universals. New York: Academic Press.

LANGACKER, Ronald. 1977. Studies in Uto-Aztecan Grammar, Volume 1: An Overview of Uto-Aztecan Grammar. Dallas: Summer Institute of Linguistics.

LEHMANN, Winfred P. 1973. A structural principle of language and its implications. Language 49, 42-66.

LEWIS, G.L. 1967. Turkish Grammar. Oxford University Press.

MEINHOF, Carl. 1930. Der Koranadialekt des Hottentotishen. Berlin: Verlag von Dietrich Reimer (Ernst Vohsen) A.-G. [Reprinted 1969, Nendeln, Liechtenstein: Kraus].

MITHUN, Marianne. 1987. Is basic word order universal? In R. Tomlin (ed.), Coherence and Grounding in Discourse, 281-328. Amsterdam: John Benjamins.

MONOD-BECQUELIN, Aurore. 1975. La pratique linguistique des Indiens Trumai, Tome 1. Paris: Centre National de la Recherche Scientifique.

MORTON, J. 1970. The interaction of information in word recognition. Psychological Review 76, 176-178.

NAISH, Constance. 1979. A Syntactic Study of Tlingit. Summer Institute of Linguistics.

OLDFIELD, R.C., and A. WINGFIELD. 1965. Response latencies in naming objects. Quarterly Journal of Experimental Psychology 17, 273-281.

ONIFER, W. and D.A. SWINNEY. 1981. Accessing ambiguities during sentence comprehension: effects of frequency of meaning and contextual bias. Memory and Cognition 9, 225-236.

OTTAVIANO, Ida de. 1980. Textos Tacana. Riberalta, Bolivia: Instituto Lingüístico de Verano.

PAYNE, Doris. 1985. Aspects of the Grammar of Yagua: A Typological Approach. University of California at Los Angeles dissertation.

——. 1987. Information structuring in Papago narrative discourse. Language 63, 783-804.

SAPIR, Edward. 1909. Takelma Texts. University of Pennsylvania Anthropological Publications of the University Museum, Volume 2, No. 1.

SAXTON, Dean. 1982. Papago. In R. Langacker (ed.), Studies in Uto-Aztecan Grammar, Volume 3: Uto-Aztecan Grammatical Sketches, 93-266. Summer Institute of Linguistics.

SCANCARELLI, Janine. 1986. Pragmatic roles in Cherokee Grammar. Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistics Society, 224-234.

——. 1988. Grammatical Relations and Verb Agreement in Cherokee. University of California at Los Angeles dissertation.

SCHUBERT, Lenhart K. 1984. On parsing preferences. Proceedings of the 10th International Conference on Computational Linguistics (COLING 84), 247-250.

SHAUL, David Leedom. 1982. A Grammar of Nevome. University of California at Berkeley dissertation.

SHETTER, William Z. 1958. Introduction to Dutch. The Hague: Martinus Nijhoff.

SIEWIERSKA, Anna. 1988. Word Order Rules. London: Croom Helm.

THOMPSON, Sandra. 1978. Modern English from a typological point of view: some implications of the function of word order. Linguistische Berichte 54, 19-35.

VENNEMANN, Theo. 1974. Analogy in generative grammar: the origin of word order. Proceedings of the Eleventh International Congress of Linguists, 79-83.

WEBER, David John. 1983. A Grammar of Huallaga (Huanaco) Quechua. University of California at Los Angeles dissertation.

ZEPEDA, Ofelia. 1983. A Papago Grammar. Tucson: University of Arizona Press.