

Article-Noun Order

Matthew S. Dryer
University of Alberta

It is almost accepted wisdom (cf. Greenberg 1963, Vennemann 1976, Hawkins 1983) that OV languages tend to place modifiers before nouns while VO languages tend to place modifiers after nouns. But discussions in word order typology rarely consider the order of article and noun. If one adopts the traditional view that articles are modifiers of nouns, then the accepted wisdom regarding word order correlations would predict that articles should tend to follow the noun in VO languages and to precede the noun in OV languages. Based on a sample of 125 languages, I argue in this paper that both of these predictions turn out to be false and that there is in fact a correlation in the opposite direction. In section 1 I discuss my criteria for identifying articles cross-linguistically; in section 2 I present the evidence regarding the correlations with the order of article and noun; in section 3 I discuss the theoretical implications of these correlations; and in section 4, the conclusion, I briefly discuss differences between articles and demonstratives with respect to their word order behaviour.

1. Identifying articles

While there is little difficulty identifying some categories, like nouns, cross-linguistically, it is less obvious what criteria count for identifying a word as an article. I use here two criteria to identify articles: (1) the word indicates definiteness or indefiniteness, or some related discourse notion; (2) the word serves as a noun phrase marker in the sense that noun phrases in the language (other than pronouns) typically occur with one of the words in question. Articles in English, as in many languages, satisfy both of these criteria: except for proper nouns, bare plurals (as in *Dogs bite*), and bare mass nouns (as in *Snow is white*), English noun phrases containing a noun occur with an article (or some other determiner). But I also treat a word as an article if it satisfies either one of these two criteria. Vietnamese has a word *cái*, illustrated in (1), which Binh (1971: 117) describes as an identifier and whose function is something like that of a definite article.

- (1) cái nhà tôi bán hôm qua
 the house I sell day past
 the house I sold yesterday (Binh 1971: 118)

But *cái* does not appear to serve as a noun phrase marker like *the* in English in that it is normal for noun phrases in the language not to contain *cái* or any other word that contrasts with it. It appears to be the case from Binh's description that *cái* is only used when the noun is followed by some other modifier. In other words, while *the* in English belongs to a class of words that is obligatory in certain syntactic environments, *cái* in Vietnamese appears to be simply an optional modifier of the noun.

On the other hand, there are languages which employ words that serve as noun phrase markers but which do not code discourse notions like definiteness or indefiniteness. Many Austronesian languages, for example, have words that occur obligatorily in noun phrases (except for certain syntactic environments) and which vary according to whether the noun is a proper noun or a common noun as well as

various other grammatical features of the noun phrase. Fijian, for example, employs two such words: *a* (with common nouns) and *o* (with proper nouns), illustrated in (2).

- (2) sa tau-ra a drano o Boumaa
 asp hold com lake prop Boumaa
 Boumaa held the lake. (Dixon 1988: 243)

Dixon (1988: 114) says that although these words are called *articles* within the Fijian grammatical tradition, the term *article* is used in a somewhat unusual sense, apparently because they are unlike articles in many other languages in that they do not code notions like definiteness and indefiniteness. But they are like articles in English and many other languages in being noun phrase markers. It is presumably on the basis of this second characteristic that they are called articles within the Fijian grammatical tradition.

For many languages there is some difficulty deciding whether a word is to be considered a definite article or a demonstrative. In languages like English there is little difficulty: the definite article is distinct from the demonstratives and the demonstratives usually carry clearly deictic meaning. But in many languages, the distal demonstrative is used in a way that "comes between" the use of *the* and *that* in English: it is used much more than *that* is used in English, so that many instances of its use seem more like *the* than *that*. The distal demonstratives in Mandarin Chinese are often used in contexts in which English would simply use the definite article. And in Western Tarahumara there is a word that Burgess (1984) glosses as "that" but which in the sample text (pp. 145-149) occurs in most noun phrases referring to the main participants in the text. The existence of such languages should not be surprising, since as both Greenberg (1978: 61) and Givón (1984: 418-419) have noted, distal demonstratives are a common diachronic source for definite articles. What apparently happens is that a word which at one time was used primarily as a true deictic demonstrative is used increasingly in contexts where the spatial meaning is absent, so that eventually its use converges on that of a definite article. If a demonstrative is frequently used in this way, I treat it here as a definite article. In section 2, however, I compare the word order properties of different kinds of definite articles: those which are also demonstratives, those which resemble demonstratives (and which therefore may have developed recently from demonstratives), and those which are quite distinct in form from true demonstratives in the language.

Analogous to the problems in distinguishing definite articles from demonstratives are problems in distinguishing indefinite articles from a numeral meaning "one". In some languages, like French, the indefinite article also functions as a numeral meaning "one". In other languages, the word meaning "one" is used in a subset of the contexts in which the indefinite article is used in English, but still much more widely than the numeral *one* is used in English, its use apparently occurring typically in contexts where the noun phrase is highly pragmatically referential in the sense of Givón (1984: 423-427). Both Mandarin Chinese and Western Tarahumara, discussed above as examples of languages in which the demonstrative is often used as a definite article, are also examples of languages in which the numeral for "one" is often used where English would use an indefinite article. Where this usage appears to be common in a language, I treat the word as an indefinite article. In section 2, however, I provide a breakdown of the data on the basis of whether the indefinite article is identical to the numeral meaning "one",

similar to it, or clearly distinct from it. It should be stressed that even if the indefinite article and the numeral for "one" are identical in form, they may be distinguishable syntactically. For example, Lewis (1967: 54) reports that the word *bir* in Turkish functions as an indefinite article when it occurs between an adjective and the noun, as in *büyük bir tarla* "a large field", but as a numeral meaning "one" when it occurs before an adjective, as in *bir büyük tarla* "one large field".

I should note that I use the terms *definite* and *indefinite article* somewhat loosely in this paper. The actual use of these articles often varies from that of the articles in English, not only in the kinds of cases just discussed where only a subset of definites or indefinites occur with the relevant article, but also in that the relevant distinction in some cases is apparently really specific vs. nonspecific (e.g. West Futuna-Aniwa, Dougherty 1983: 22-23).

Unless otherwise specified, I exclude from consideration inflectional affixes that code the kind of meanings coded by article words, as in the following example from Somali.

- (3) nink-i ad áraktei
 man-def you saw
 the man that you saw (Kirk 1905: 19)

The data I present in this paper does not include such affixes, since including them might obscure the fact that the correlation to be demonstrated is one that holds for article words. On the other hand, I treat bound morphemes as article words rather than as affixes if they are demonstrably clitics in the sense of phrasal affixes. For example in Ngizim the definite article is a clitic that attaches either to the noun or to the last element of the noun phrase, as in (4).

- (4) agwai waara aarawii-gu
 eggs rel white-def
 the eggs which were white (Schuh 1972: 168)

I should note that I do not have data on all the relevant characteristics for all the languages in my sample. To the contrary, for many languages I lack relevant data, either because I am unable to determine the characteristics from my source or because some of the data was collected at an earlier time from sources which I have been unable to obtain access to in order to obtain further data. The data presented here comes from a larger project on word order universals for which data has been collected for over 600 languages (cf. Dryer 1986, 1988a, 1988b). The total number of languages with article words in my data base is 125. It is difficult to determine from this what proportion of the languages of the world employ articles, since some of the languages in my data base for which I have limited data may have articles even though I do not have a record of such. But we can get an idea of how many languages employ articles by comparing the number for which I have data on articles with the number of languages for which I have data on demonstratives, since, except for the relatively few languages in which demonstrative meanings are expressed by affixes, all languages employ demonstratives. Since I have data on demonstratives for 399 languages and on articles for 125 languages, it appears that about a third of the languages of the world employ articles. Table 1 lists the number of languages in my sample with articles of each of the different types I will discuss.

Table 1: Number of languages with different types of articles

	Definite	74		
	Indefinite	51		
	Noun phrase marker	16		
Def=Dem	11	Indef=One	15	
DefLikeDem	21	IndefLikeOne	8	
DefNotDem	20	IndefNotOne	11	

Def=Dem indicates definite articles which are identical to a demonstrative; DefLikeDem indicates definite articles which are similar in form to the demonstrative, but distinct (like *the* and *that* in English); and DefNotDem indicates definite articles which are quite distinct from the demonstrative. The categories for indefinite articles are analogous, except the form is compared to the numeral meaning "one".

It should be noted that there are many languages with either a definite article or an indefinite article but not both. For example, Kobon (Davies 1981: 60) employs an indefinite article but definites are unmarked. In fact, only 31 languages of those listed in Table 1 have both a definite and an indefinite article. Nor if a language has both types of articles need they belong to the same word class. In two languages in my sample (Hausa and Chaha), there is an indefinite article which precedes the noun and a definite article which follows. There are also languages in which an indefinite article can co-occur with a noun phrase marker article, as in the following example from Jacaltec.

- (5) *hune' no' txitam*
 a class pig
 a pig (Craig 1977: 137)

The word *no'* belongs to a class which Craig (1977: 133) calls noun classifiers; these count as noun phrase marker articles in the sense of this paper since it is normal for a noun phrase to contain a word from this class. Hence both *hune'* and *no'* count as articles in this paper.

2. The word order patterns

For the purposes of testing hypotheses, the data is presented here in a manner that is explained and justified in greater detail in Dryer (to appear a). Briefly, the languages in my data base are first grouped into genetic groups I call *genera* which are roughly comparable in time-depth to the subfamilies of Indo-European. These genera are then further grouped into six large geographical areas (Africa, Eurasia (excluding southeast Asia), Southeast Asia & Oceania, Australia & New Guinea, North America, and South America). In order to test whether a given language type is significantly more common than another language type, the number of genera containing languages of each type in each of the six geographical areas is determined. If one of the two types is more common in each of the six areas, then it is concluded that there is a significant linguistic preference for that type over the other. The statistical test is a simple sign test: the chance of all six areas containing more of a given type under the null hypothesis is 1 in 64. Hence if

all six areas do contain more of a given type, the result is statistically significant at a level better than .02.

For example, Table 2 gives the data for prepositions and postpositions in OV languages.

Table 2: Postpositions vs. prepositions in OV languages							
	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
OV&Postp	[15]	[23]	[5]	[16]	[24]	[17]	100
OV&Prep	3	2	0	1	0	0	6

The square brackets indicate the more frequent type within each of the six areas. It can be seen from Table 2 that OV&Postpositional languages are more common than OV&Prepositional languages in each of the six areas. We can therefore conclude that there is a statistically significant preference for OV languages to be postpositional rather than prepositional.

Table 3 give comparable data for article-noun order as opposed to noun-article order in OV languages.

Table 3: Order of article and noun in OV languages							
	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
OV&ArtN	1	2	0	1	4	[3]	11
OV&NArt	[5]	2	[2]	[4]	4	1	18

The data in Table 3 does not provide clear evidence of any pattern for the order of article and noun among OV languages. Although NArt order is slightly more common overall, it is more common in only 3 of the 6 areas. We cannot conclude that there is any tendency for OV languages to be NArt. The examples in (6) illustrate these two orders of article and noun in OV languages.

- (6) a. Nevome (OV&ArtN)
 macco occi an'-t'-igui abamabua
 a/one woman 1sg-perf-irr fondle
 I fondled a woman (Shaul 1982: 66)
- b. Lakota (OV&NArt)
 Mathó ki wa-kté.
 bear the 1sg-kill
 I killed the bear. (Van Valin 1987: 376)

Table 4 gives comparable data for VO languages.

Table 4: Order of article and noun in VO languages							
	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
VO&ArtN	2	[6]	[6]	[1]	[14]	[4]	33
VO&NArt	[9]	0	3	0	2	0	14

A clearer pattern emerges from Table 4 than emerged from Table 3: in 5 of the 6 areas there are more genera containing VO&ArtN languages than there are genera containing VO&NArt languages. One area, Africa, exhibits a preference in the opposite direction. Because VO&ArtN is more common in only 5 of the 6 areas, an overall preference for ArtN order among VO languages falls short of statistical significance. But since 5 of the areas do exhibit such a preference, and since the

total number of VO&NArt genera outside Africa is considerably lower than the total number of VO&ArtN genera (5 vs. 31), we can say that there is a trend toward VO languages being ArtN that falls short of statistical significance. The examples in (1) and (2) above from Vietnamese and Fijian (as well as the articles in English) exemplify the typical ArtN order; (7) from Fulani illustrates the NArt order that is less typical in a VO language.

- (7) Bello hokkii Mamman sheede c'en.
 gave money the
 Bello gave Mamman the money. (Arnott 1970: 28)

Given the fact that there is no statistically significant preference for one order of article and noun among either OV languages or VO languages, we might conclude that there is no evidence of any correlation between the order of article and noun and the order of object and verb. However, if we compare the proportions of genera containing ArtN languages (as opposed to NArt languages) in OV as opposed to VO languages, a statistically significant pattern emerges. The data is given in Table 5.¹

Table 5: Proportions of genera containing ArtN languages

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Average
OV	.17	.50	.00	.20	.50	.75	.35
VO	[.18]	[1.00]	[.67]	[1.00]	[.88]	[1.00]	.79

Although the difference in proportions in Africa is marginal, it is still the case that the proportion of genera containing ArtN languages is higher among OV languages in all six areas. We can therefore conclude that VO languages exhibit a statistically significantly greater tendency to place articles before the noun than OV languages do. Because of that we can say that there is a statistically significant correlation between the order of article and noun and the order of object and verb.

The direction of this correlation, however, is the opposite of what traditional assumptions might lead us to expect. If one adopts the traditional assumption that articles are modifiers of nouns and the common view that modifiers tend to precede the noun in OV languages and to follow in VO languages, then we would expect articles to precede the noun more often in OV languages than they do in VO languages. But the correlation shown in Table 4 is in the opposite direction. I will return to the question of how to reconcile these results with traditional assumptions in section 3.

One question that arises, however, is that of how much the correlation demonstrated in Table 5 is due to one subtype of article and whether the various kinds of words I have treated as articles all exhibit this correlation. These questions are difficult to answer in a rigorous statistical fashion because the small numbers of languages of each sort make it difficult to obtain statistically significant results. For this reason, much of my discussion will be in terms of trends.

Consider first definite articles. Table 6 gives the number of genera containing each of the four possible language types for the different orderings of object and verb and of definite article and noun.

Table 6: Order of definite article and noun

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
OV&DefN	1	0	0	1	3	[2]	7
OV&NDef	[4]	[1]	0	1	[4]	0	10
VO&DefN	1	[5]	[4]	0	[11]	[4]	25
VO&NDef	[8]	0	2	0	2	0	12

The general pattern in Table 6 is the same as the pattern that we saw for articles in general: NDef order is slightly more common among OV languages while DefN order is considerably more common among VO languages, though the trend falls short of statistical significance, primarily because VO&NDef order is much more common in Africa.

Table 7 gives the data on proportions for definite articles, comparable to the data for articles in general given in Table 5.

Table 7: Proportions of genera containing DefN languages

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Average
OV	[.20]	.00	--	.50	.43	1.00	.43
VO	.11	[1.00]	.67	--	[.85]	1.00	.73

The two entries in Table 7 of the form "--" are cases where the proportion cannot be determined since there are no genera containing languages of the given sort in my sample: e.g. there are no OV languages in my sample from Southeast Asia & Oceania with definite articles, so that the proportion would be 0/0. Because of this, we cannot compare proportions for two areas. But in only two of the remaining four areas is the proportion of DefN order higher among VO languages, considerably short of statistical significance. We therefore have no clear evidence that definite articles tend to precede the noun more often in VO languages than in OV languages.

On the other hand, the data in Table 7 includes definite articles which are also demonstratives. If we exclude such words and include only definite articles which are distinct from demonstratives, clearer evidence of a correlation emerges. The relevant data is given in Table 8.

Table 8: Order of definite article and noun for definite articles distinct from demonstratives

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
OV&DefN	1	0	0	0	0	0	1
OV&NDef	1	[1]	0	[1]	[2]	0	5
VO&DefN	1	[4]	[2]	0	[6]	[3]	16
VO&NDef	[5]	0	1	0	2	0	8

While the numbers in Table 8 fall short of statistical significance, the direction of the trend is more striking than that for definite articles in general given in Table 6: the righthand column of Table 8 shows that if we exclude definite articles which are also demonstratives NDef order is overall more common among OV languages by 5 to 1, while DefN is more common among VO languages by 16 to 8. In no area is DefN more common among OV languages and in only one area (Africa) is NDef more common among VO languages.

Table 9 shows further that the more distinct the definite article is from demonstratives, the more likely its position relative to the noun will correlate with the order of verb and object. Because the numbers are rather small, I have not given the breakdown by area but have simply used the total number of genera, though I do give the average proportion of DefN genera within each area.

Table 9: Definite articles by subtype

		Def=Dem	DefLikeDem	DefNotDem
OV	DefN	4	1	0
	NDef	1	3	2
Aver. proportion	DefN	.67	.17	.00
VO	DefN	3	7	10
	NDef	1	6	2
Aver. proportion	DefN	.75	.67	.83
Difference in proportions between VO and OV		.08	.50	.83

We can see from the first column of Table 9 that there is no apparent difference between OV and VO languages in the position of definite articles which are identical to demonstratives. On the other hand, the third column of Table 9 shows that there is the greatest difference between OV and VO languages in the position of definite articles which are quite unlike demonstratives. In fact, out of the 16 languages in my sample of this sort, only two languages, Lelemi and Huave, fail to conform to the correlation. Hence we can conclude that there is a correlation in the case of definite articles and that it increases with the degree to which the definite article is distinct from demonstratives.

One reasonable conclusion one might draw is that it is mistaken to treat demonstratives which often function like definite articles as articles. After all, the evidence here suggests that they do not exhibit the word order properties that "true" definite articles do. Nevertheless, there is evidence that even when such definite articles are identical to demonstratives, they are still subject to different word order "pressures". There are two languages in my sample in which the definite article is identical to the demonstrative in form but in which the two occur on opposite sides of the noun. Or to put it differently, in these two languages, there is a word whose function depends on which side of the noun it occurs. In Swahili, the demonstrative *yule* (or more accurately the demonstrative *-le*, which takes different prefixes depending on the noun class of the noun it is accompanying) either precedes or follows the noun, but when it precedes the noun, "its function corresponds to that of the definite article in English" (Ashton 1947: 59). Hence *mtu yule* means "that man" while *yule mtu* means "the man". The opposite situation holds in Ute (Givón 1984: 419), in which the demonstrative has demonstrative meaning when it precedes the noun, but functions as a definite article when it follows. Significantly the fact that Swahili and Ute are opposite in this respect can be understood in terms of the correlations discussed in this paper. Namely, since articles precede the noun more often in VO languages than they do in OV languages, we might expect that if two languages differed in this respect, the one in which the demonstrative functions as an article when it precedes the noun is more likely than the other language to be VO. And in fact Swahili *is* a VO language while Ute is an OV language. This suggests not only that the position of definite articles

cannot be explained (at least not in all cases) in terms of the position of the demonstratives from which they arose, but also that words with the meaning of definite articles are subject to word order "pressures" that demonstratives are not subject to. As I point out in the conclusion, demonstratives do not exhibit the correlation with the order of verb and object that I have shown articles exhibit.

Consider now indefinite articles. Table 10 gives the data comparable to the data given in Table 6 for definite articles.

Table 10: Order of indefinite article and noun

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
OV&IndefN	1	1	0	0	2	1	5
OV&NIndef	[2]	[2]	0	[3]	2	1	10
VO&IndefN	2	[4]	[3]	0	[4]	[1]	14
VO&NIndef	[3]	0	1	0	0	0	4

The general pattern for indefinite articles in Table 10 is generally similar to that in Table 6 for definite articles: NIndef order is somewhat more common among OV languages, while IndefN order is more common among VO languages, except in Africa, but in both cases we fall short of statistical significance. The data for proportions is given in Table 11.

Table 11: Proportions of genera containing IndefN languages

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Average
OV	.33	.33	--	.00	.50	.50	.33
VO	[.40]	[1.00]	.75	--	[1.00]	[1.00]	.83

Again, because of the absence of certain language types in certain areas in my sample, we cannot compare proportions for all six areas. However, in the four areas for which we can compare proportions, the proportion of genera containing IndefN languages rather than NIndef languages is higher among VO languages. Furthermore, if we treat the instances of "--" in Table 11 as .50 (i.e. treat 0/0 as .50, mathematically inaccurate but with some justification in the present instance), then the proportion would be higher in all six areas. In addition, there is a large difference in the average proportions: .33 for OV languages and .83 for VO languages. For these reasons we can conclude that there is a statistically significant correlation between the order of indefinite article and noun and the order of object and verb.

Again it is worthwhile distinguishing three subtypes of indefinite articles: those which are identical to the numeral meaning "one" (Indef=One), those which are similar to "one" though distinct, like the Dutch indefinite article *een* [ɛn] (cf. *één* [eɪn] "one") (IndefLikeOne), and those which are quite distinct from the numeral meaning "one" (IndefNotOne). Table 12 gives data on such subtypes, analogous to that given in Table 8 above for definite articles.

Table 12: Indefinite articles by subtype

		Indef=One	IndefLikeOne	IndefNotOne
OV	IndefN	5	0	0
	NIndef	2	0	3
Aver. proportion	IndefN	.70	--	.00
VO	IndefN	5	3	5
	NIndef	0	2	1
Aver. proportion	IndefN	1.00	.63	.92
Difference in proportions between VO and OV		.30	--	.92

Although the trend in Table 12 is less clear than it was for definite articles, it appears to be the case that there is the greatest difference between OV and VO languages among indefinite articles which are quite distinct from the numeral "one". This suggests that the correlation between the order of article and noun and that of verb and object may be strongest for articles which are categorially more distinct from other words.

The third class of articles is that of noun phrase markers, illustrated by the Fijian example in (2) above. The number of these is considerably smaller than for definite and indefinite articles: only 16 languages in my sample employ such articles, occurring in a total of 8 genera, and 9 of these languages are Austronesian. But despite the small numbers, two clear patterns emerge: first, all such markers precede the noun; and second, with one exception, all of the noun phrase marker articles occur in VO languages.

3. Explanations

As noted above, the general correlation documented in the previous section between the order of article and noun and the order of object and verb runs counter to what we might expect, at least given two traditional assumptions: that articles are modifiers of nouns, and that modifiers tend to precede the noun in OV languages and to follow in VO languages. The evidence from the previous section is evidence that at least one of these two assumptions must be false. It is not my intention here to argue that one of these assumptions but not the other is false. In fact, I believe that there are good reasons to question both of these assumptions. What I will do in this section is summarize reasons for questioning each of them.

The belief that OV languages tend to place modifiers before nouns while VO languages tend to place them after nouns partly stems from a widespread belief that adjectives tend to precede nouns in OV languages and to follow them in VO languages. I have shown elsewhere (Dryer 1988a), however, that there is no evidence of any correlation between the order of adjective and noun and the order of object and verb, that it is actually more common for OV languages to place the adjective after the noun, and that adjectives precede the noun in VO languages as often as they do in OV languages. I have also shown (Dryer 1986) that there is no evidence that OV languages place demonstratives or numerals before the noun more often than VO languages do. And I have shown (Dryer, to appear b) that plural words, separate words in noun phrases indicating grammatical plural, exhibit a strong tendency to precede the noun in VO languages and to follow the noun in OV languages, the opposite of what one might expect if plural words are modifiers of

nouns. In fact, the only modifiers of nouns that exhibit a correlation are genitives and relative clauses (Dryer 1986). Hence there is no evidence of a general tendency for modifiers to precede nouns in OV languages and to follow in VO languages. Thus, even if one maintains the traditional assumption that articles are modifiers of nouns, there would be no reason to expect them to precede the noun more often in OV languages than in VO languages.

But that leaves unexplained why we in fact find the opposite correlation in the case of articles. I have proposed elsewhere (Dryer 1988a, 1988b) that the word order correlations in general reflect a tendency to consistently order phrasal constituents with respect to nonphrasal ones, that consistent OV languages are consistently left-branching (placing phrasal constituents before nonphrasal ones) while consistent VO languages are consistently right-branching. If one accepts the common assumption in generative grammar that determiners in English combine with an N' constituent, then the correlation documented here for article-noun order is explained: in a noun phrase like *the tall man* the article *the* is a nonphrasal category combining with the phrasal N' *tall man*. In other words, by placing the article at the beginning of the noun phrase in English, the result is a right-branching structure, consistent with the general right-branching nature of the language. On the other hand, if English were to place articles after the noun, then the structure of noun phrases would be left-branching. Such a structure, though it might conform to noun-modifier order, would actually produce inconsistency in the language since English is generally right-branching.

An alternative approach to explaining the correlation discussed in section 2 would be to question the traditional assumption that the article is a modifier of the noun. In fact, a number of linguists (cf. Vennemann & Harlow 1977, Hudson 1984: 90-91, Abney 1987) have questioned this assumption, arguing instead for an analysis under which the article is the head, the N' some sort of dependent of the article. While the arguments are often somewhat theory-dependent and/or English-specific, there are a number of typological facts that lend credence to the idea that articles are the real heads of noun phrases. Under such an approach, the difference between articles and pronouns is rather akin to the difference between transitive and intransitive verbs. In other words, articles and pronouns belong to a single category, which we can arbitrarily call articles, the difference being that articles like *the* are transitive articles, while pronouns are just intransitive articles. There are a number of languages in which, unlike English, a single word functions both as an article and as a personal pronoun. For example, in Jicaltepec Mixtec, it is common for noun phrases to consist of a pronoun followed by a noun in a structure which Bradley (1970: 63) describes as consisting of a pronoun as "center" with the noun as "attribute", illustrated by the word *rá* in (8).

- (8) *číká čaʔa ña saʔma čĩʔi rá ahili*
 thing-that give she clothes to he angel
 That's why she gave the clothes to the angel. (Bradley 1970: 79)

As indicated by Bradley's gloss, *rá* otherwise functions as a pronoun in the language. But except for this fact, *rá* accompanied by a noun is very much like articles in other languages, at least as a kind of noun phrase marker. I therefore treat it as an article in this study. But in doing so, it is not my intention to deny the appropriateness of Bradley's analysis. If one takes the position that articles are simply pronouns modified by nouns, then these words *are* articles, even on Bradley's analysis.

Another phenomenon found in a number of languages that provides further support for treating articles as heads of noun phrases is the following. In some languages, noun phrases that occur as arguments of the verb always occur with an article, but when they occur in predicate position they do not. Thus in the Cebuano examples in (9), the predicate (which is initial) is an N' and does not occur with an article while the subject occurs with an article.

- (9) a. babaye ang duktur.
 woman art doctor
 The doctor is a woman.
 b. duktur ang babaye
 doctor art woman
 The woman is a doctor.

It is in general not normal for nonmaximal projections of a head to occur where they are not a constituent in a maximal projection. If we treat articles as modifiers of nouns, combining with an N', then the occurrence of an N' in predicate position would be an exception to this generalization. On the other hand, if we treat the article as the head of the so-called noun phrase, then the N' would be the maximal projection.

Articles change the category of an N' in a way that modifiers generally do not. While we can say, loosely speaking, that an N' is semantically a predicate, a so-called NP is a referring expression. Hence the semantic category of the article is that of a word that changes a predicate into a referring expression. The Cebuano facts illustrated in (9) are thus not surprising: an NP in argument position is semantically a referring expression while an N' in predicate position is semantically a predicate.

If articles are the heads of NPs, then the correlation demonstrated in section 2 is partly explained: it reflects a tendency for heads to precede dependents in VO languages more often than in OV languages. However, as noted above, there is no general tendency of this sort: the order of adjective and noun does not correlate with the order of object and verb. On the other hand, it turns out that in those cases where we do find a correlation, the dependent is always a phrasal category. If articles are heads, then the N' that the article combines with is a phrasal category. Hence the correlation demonstrated in section 2 may reflect a tendency for *phrasal* dependents to follow their heads more often in VO languages than in OV languages.

4. Conclusion

I have demonstrated in this paper that the order of article and noun correlates with that of verb and object, that the article precedes the noun significantly more often in VO languages than it does in OV languages. But the functional similarity between demonstratives and articles (especially definite articles), the fact that distal demonstratives are a very common diachronic source for definite articles, and the kinds of explanations discussed in the preceding section would all lead us to expect that the order of demonstrative and noun would exhibit the same correlation. But it does not. Table 13 gives the basic data.

Table 13: Order of demonstrative and noun

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
OV&DemN	10	[17]	[4]	8	[21]	[11]	71
OV&NDem	10	1	2	[9]	5	4	31
VO&DemN	4	[8]	7	[5]	[17]	[9]	50
VO&NDem	[24]	1	[11]	0	6	1	43

No pattern emerges from Table 13: there is a trend towards DemN order in OV languages, though the number of OV&DemN genera is larger in only four areas. The totals figures in the righthand column might suggest that the proportion of NDem is higher in VO languages than it is in OV languages, but the majority of VO&NDem genera are in Africa, so the total figure is somewhat skewed by this one area. As before, it is necessary to compare proportions, as in Table 14.

Table 14: Proportions of genera containing DemN languages

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Average
OV	[.50]	[.94]	[.67]	.47	[.81]	.73	.69
VO	.14	.89	.39	[1.00]	.74	[.90]	.68

Table 14 shows no evidence of a correlation: DemN order is more common among OV languages in four areas but more common among VO languages in the other two areas. In addition, the average proportion (shown in the righthand column) is almost the same for OV as it is for VO languages. And even if we were to interpret the evidence as showing a weak trend, the direction of the trend is the opposite from that shown in this paper for articles.

There is clear evidence, then, that demonstratives exhibit very different word order properties from articles. Unfortunately it is beyond the scope of this paper to discuss the possible reasons for this. But this issue is clearly one that warrants attention if we are to understand the causes of word order correlations.

Acknowledgement

I am indebted to the Social Sciences and Humanities Research Council of Canada for Research Grants 410-810949, 410-830354, and 410-850540, which supported the research for this paper.

Notes

¹ To see how the figures in Table 5 are computed, consider the figure .17 for OV languages in Africa given in the upper left hand corner of the table. This figure can be obtained from Table 3, which shows that there is one genus in Africa containing OV&ArtN languages and 5 genera containing OV&NArt languages. The proportion of genera containing OV&ArtN languages is thus 1 out of 6 genera (1+5), or .17.

The characterization given here for the proportion of genera is misleading in one respect. Namely, it ignores the possibility that some genus may contain languages of both sorts. In general, this does not happen, since languages within a genus are typically similar in their word order characteristics. However, it turns out that the example discussed provides an instance of this situation. The sole genus in Africa containing an OV&ArtN language in my sample is Semitic, the language in

question being Tigre. But Semitic is also one of the genera containing an OV&NArt language, namely Chaha. Thus, strictly speaking, there are only 5 genera containing OV&ArtN or OV&NArt languages. The figures in Table 5 can be more accurately characterized as giving the proportion of *subgenera*, where a subgenus is a subset of languages in a genus that are identical with respect to the set of word order characteristics being examined. In general, each genus will contain exactly one subgenus. But in the case of Semitic, there are two subgenera with respect to the order of article and noun among OV languages. Thus Table 5 does give accurate figures for the proportions of *subgenera* containing ArtN languages.

Bibliography

- Abney, Stephen Paul. 1987. *The English Noun Phrase in its Sentential Aspect*. Unpublished MIT dissertation.
- Arnott, D. W. 1970. *The Nominal and Verbal Systems of Fula*. Oxford: Clarendon Press.
- Ashton, E. O. 1947. *Swahili Grammar*. Second Edition. London: Longmans, Green & Co.
- Binh, Duong Thanh. 1971. *A Tagmemic Comparison of the Structure of English and Vietnamese Sentences*. The Hague: Mouton.
- Bradley, C. Henry. 1970. *A Linguistic Sketch of Jicaltepec Mixtec*. Summer Institute of Linguistics.
- Burgess, Don. 1984. Western Tarahumara. In Ronald W. Langacker ed., *Studies in Uto-Aztecan Grammar*, Volume 4. The Summer Institute of Linguistics and the University of Texas at Arlington.
- Craig, Collette. 1977. *The Structure of Jacaltepec*. Austin: University of Texas Press.
- Davies, John. 1981. *Kobon*. *Lingua Descriptive Studies* 3. Amsterdam: North-Holland.
- Dixon, Robert M. W. 1988. *A Grammar of Boumaa Fijian*. Chicago: University of Chicago Press.
- Dougherty, Janet W. D. 1983. *West Futuna-Aniwa: An Introduction to a Polynesian Outlier Language*. University of California Publications in Linguistics, Volume 102. Berkeley: University of California Press.
- Dryer, Matthew S. 1986. Word order consistency and English. In S. Delancey and R. Tomlin, eds., *Proceedings of the Second Annual Pacific Linguistics Conference*. Eugene, Oregon.
- 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*, pp. 93-124. Amsterdam: John Benjamins.
- Dryer, Matthew S. To appear a. Large linguistic areas and language sampling. To appear in *Studies in Language*.
- Dryer, Matthew S. To appear b. Plural words. To appear in *Linguistics*.
- Givón, Talmy. 1984. *Syntax: A Functional-Typological Introduction, Volume 1*. Amsterdam: John Benjamins.
- Greenberg, Joseph H. 1957. Order of affixing: a study in general linguistics. In *Essays in Linguistics*. Chicago: University of Chicago Press.

- Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In J. Greenberg, ed., *Universals of Language*. Cambridge, Mass.: MIT Press.
- Greenberg, Joseph. 1978. How does a language acquire gender markers? In J. Greenberg, ed., *Universals of Human Language, Volume 3: Word Structure*, pp. 47-82. Stanford: Stanford University Press.
- Hawkins, John A. 1983. *Word Order Universals*. New York: Academic Press.
- Hudson, Richard. 1984. *Word Grammar*. Oxford: Basil Blackwell.
- Kirk, J. W. C. 1905. *A Grammar of the Somali Language*. Farnborough, Hants.: Gregg International Publishers Ltd.
- Lewis, G. L. 1967. *Turkish Grammar*. Oxford University Press.
- Schuh, Russell G. 1972. *Aspects of Ngizim Syntax*. University of California, Los Angeles, Ph.D. Dissertation. Ann Arbor: University Microfilms.
- Shaul, David Leedom. 1982. *A Grammar of Nevome*. University of California, Berkeley, Ph.D. Dissertation. Ann Arbor: University Microfilms.
- Van Valin, Robert D., Jr. 1987. The role of government in the grammar of head-marking languages. *International Journal of American Linguistics* 53: 371-397.
- Vennemann, Theo. 1976. Categorical grammar and the order of meaningful elements. In A. Juilland, ed., *Linguistic Studies Offered to Joseph Greenberg on the Occasion of His Sixtieth Birthday*. Saratoga, California: Anna Libri.
- Vennemann, Theo, and Ray Harlow. 1977. Categorical grammar and consistent basic VX serialization. *Theoretical Linguistics* 4: 227-254.