

WPFG

working papers in functional grammar

WPFG no. 14
August 1986

Word order universals revisited: the principle of Head Proximity
Jan Rijkhoff
University of Amsterdam

WORD ORDER UNIVERSALS REVISITED: THE PRINCIPLE OF HEAD PROXIMITY

Jan Rijkhoff

1. Introduction.¹

It has been known for at least a century that certain word order combinations tend to cooccur, but it is only for a few decades that serious attempts have been made to account for these ordering cooccurrences.² This is probably due to the fact that it was not until the nineteen sixties that word order patterns were investigated systematically in a considerable number of languages, for it was then that Joseph Greenberg published "Some universals of grammar with particular reference to the order of meaningful elements" (Greenberg 1966). On the basis of observations in 30 randomly selected languages he proposed 45 universals of language, about half of which related to syntax. For example:

Universal 1. In declarative sentences with nominal subject and object, the dominant order is almost always one in which the subject precedes the object.

Universal 2. In languages with prepositions, the genitive almost always follows the governing noun, while in languages with postpositions it almost always precedes.

Universal 3. Languages with dominant VSO order are always prepositional.

Universal 4. With overwhelmingly greater than chance frequency languages with normal SOV order are postpositional.

Universal 16. In languages with dominant order VSO, an inflected auxiliary always precedes the main verb. In languages with dominant order SOV, an inflected auxiliary always follows the main verb.

Additionally he presented a classification of 142 languages, employing four parameters: 1. the position of the verb relative to nominal subject and object (i.e. VSO-SVO-SOV); 2. the position of the adposition (i.e. preposition/postposition); 3. the position of the adjective (A), and 4. the position of the genitive (G) relative to the noun (N). I.e. VSO/SVO/SOV & pr/po & NA/AN & NG/GN. Thus 24 classes could be set up, 8 for each major language type. Recently Hawkins (1983), using the same parameters, has been able to expand this sample to over 300 languages

and it is this sample that is given here (see Table 1). V-1 stands for V-initial so as to capture some languages that are supposed to have dominant VOS order; the horizontal lines -- indicate the position of the adjective and the genitive/possessor phrase relative to the noun.

1. V-1 & pr & N--	38 lgs.	13. SVO & po & N--	0 lgs.
2. V-1 & pr & ANG	13	14. SVO & po & ANG	0
3. V-1 & pr & --N	1	15. SVO & po & --N	12
4. V-1 & pr & GNA	0	16. SVO & po & GNA	13
5. V-1 & po & N--	0	17. SOV & pr & N--	10
6. V-1 & po & ANG	0	18. SOV & pr & ANG	0
7. V-1 & po & --N	1	19. SOV & pr & --N	2
8. V-1 & po & GNA	0	20. SOV & pr & GNA	0
9. SVO & pr & N--	56	21. SOV & po & N--	11
10. SVO & pr & ANG	17	22. SOV & po & ANG	0
11. SVO & pr & --N	7	23. SOV & po & --N	96
12. SVO & pr & GNA	4	24. SOV & po & GNA	55

Table 1. Classification of languages; adapted from Hawkins (1983).

Observe that some classes have remained empty, which means that these ordering cooccurrences were not attested, whereas other classes are represented quite well numerically (cf. classes 1, 9, 23, and 24). What these data show is that there are, what Greenberg called, harmonic relations (ibid. 1966: 100):

VS & VO & NG & NA & pr.

SV & OV & GN & AN & po.

These ordering patterns are illustrated in the following examples, which are taken from Hawkins (1983: 1 - 3).³

Examples (1) - (4) are from Samoan (VSO & pr & N--).

(1) Ua iloa e Atamu lana ava o Eva (VSO)
 knew Adam his wife Eve
 'Adam knew Eve his wife.'

(2) ole paopao o Tavita (NG)
 the canoe of David

(3) o le teine puta (NA)
the girl fat
'the fat girl'

(4) i le potu (pr.)
in the room

Examples (5) - (8) are from Japanese (SOV & po & --N).

(5) Taroo ga tegami o kaita (SOV)
Taroo letter wrote
'Taroo wrote a letter.'

(6) Taroo no ie (GN)
Taroo's house

(7) kono omoshiroi hon (AN)
this interesting book

(8) Taroo ga zidoosya de Hanako to Tokyo kara ryokoosita (po)
Taroo car by Hanako with Tokyo from travelled
'Taroo travelled from Tokyo with Hanako by car.'

1.1. The order of constituents in Functional Grammar.

Any theory about natural language will have to account for these harmonic relations, and in Functional Grammar (henceforth FG; Dik 1978, 1983) these and other universal ordering tendencies are captured by general Principles of Constituent Ordering. In the functional framework the actual order of constituents in a linguistic expression is determined by Placement Rules, which are applied to fully specified, but unordered underlying predications. These Placement Rules are guided by general Principles of Constituent Ordering (see section 4). The significance of each of these (interacting) principles may vary from language to language, and while some are 'harmonic', others define more or less conflicting ordering preferences. As Dik (in prep.) puts it: "[...] hardly any of these principles has absolute 100 % validity; most of them can be counteracted by one or more of the other principles. They are like forces pulling in different directions, with varying outcomes in different languages". In Dik (1983) twelve such ordering principles are proposed, some of which find their origin in Greenberg's 1966 essay.

Here Greenberg's observations will be interpreted in terms of the Principle of Head Proximity (PHP). I will attempt to show that this proposed universal ordering principle not only accounts for the harmonic relations, but also that it may relate to linguistic phenomena that have so far been treated separately, such as the number of terms in a predication, serialization, 'extraposition', and discontinuous verb complexes in SOV languages. Finally it will be suggested that the PHP could lead to reduction of the number of FG Principles of Constituent Ordering.

2. The Principle of Head Proximity.

First we will have to explain what we mean by the notion Head.⁴ The predication can be regarded as a domain containing one or more subdomains (i.e. Noun Phrases; 'terms' in FG), which, in turn, may also contain subdomains (like the Adjective Phrase). In each of these domains there is one constituent that does not qualify some other element in that domain. This constituent is called the Head of that domain. All other elements in the domain will be called Attributive Material or simply AM, cf. Table 2:

Domain	Head	Attributive Material
Predication	Predicate (V)	Predicate operators (auxiliary verbs and the like), Terms.
Term	Head Noun (N)	Term operators (determiner, quantifier), modifiers (adjective, genitive, etc.).
Adjective Phr.	Adjective (A)	Degree adverbial etc.

Table 2.

Let us now return to Hawkins' Expanded Sample, or rather to a version without the preposition/postposition parameter (Table 3). We can do without this parameter since the position of these elements is determined by FG ordering principles (XI) and (XII), the Relator Principles (see section 4). That leaves us with twelve classes.

Notice also that we have refrained from the use of such syntactic notions as Subject and Object; dominant word order is now indicated by the position of the verb, which yields V-initial, V-second, and V-final languages as major groupings. As for V-final languages that allow certain constituents to appear postverbally, see section 2.2.

I. V-initial & N--	38	V. V-second & N--	56	IX. V-final & N--	21
II. V-initial & ANG	13	VI. V-second & ANG	17	X. V-final & ANG	0
III. V-initial & --N	2	VII. V-second & --N	19	XI. V-final & --N	98
IV. V-initial & GNA	0	VIII. V-second & GNA	17	XII. V-final & GNA	55

Table 3.

If we restrict our attention to expressions that contain one verbal and one nominal Head, it appears that in the best represented V-initial and V-final classes I and XI, no AM (here: adjective and genitive), occurs in between the Head of the term N and the Head of the predication V. Remember that in V-initial and V-final languages the inflected auxiliary does not occur in between these Heads either (cf. Greenberg's Universal 16). This can be represented as follows:

V-initial: Aux V N--

V-final : --N V Aux

On the basis of these observations it could be hypothesized that there is a preference not to have AM in between Heads of domains, or, to put it positively:

THE PRINCIPLE OF HEAD PROXIMITY:

The Head of a domain prefers to be contiguous with the Head of its superordinate domain.

Thus, the adjective prefers to be contiguous with the head noun, which in turn prefers to be contiguous with the main verb, i.e. $A > N > V$ & $V < N < A$. For the sake of the argument we will confine ourselves here to nominal and verbal Heads only and AM will be limited to three categories: adjectives, genitives, and auxiliaries.

When we include some more terms in the structures given above, we get:

V-initial: Aux V N-- N-- N-- N-- (class I)

V-final : --N --N --N --N V Aux (class XI)

Notice that in these linear structures the 'bare' side of every head noun is turned towards the 'bare' side of the Head of the predication, so that if any of the terms had been the only one present, its Head would have been adjacent to the main verb.

We could also frame the constituents in a non-linear structure as in Diagrams 1a-c:

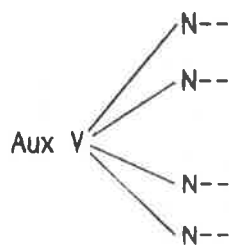


Diagram 1a.

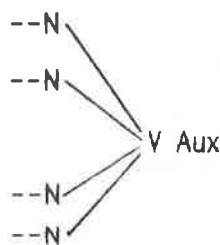


Diagram 1b.

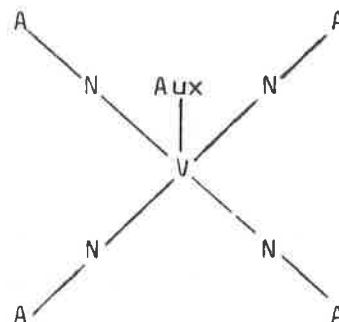


Diagram 1c.

What all these representations have in common is that no AM occurs in between Heads and that the Heads of subdomains are linked directly to the Head of the superordinate domain.

Let us now see how the Principle of Head Proximity relates to the two other major classes V and XII.

2.1. Languages of type V.

Languages of type V can be represented as follows:

V-second: N-- Aux V N-- N-- N--

In these languages one constituent typically appears in preverbal position, so that the adjective and/or genitive, if present, would occur in unpreferred position in terms of the PHP. At this point two FG ordering principles are relevant (see section 4):

- (III) There is a universally relevant clause-initial position P1, which is used for special purposes, including the placement of constituents with Topic or Focus function.
- (IV) Since the Subject is the prime Topic candidate, it will often end up in P1; this may lead to reinterpretation of P1 as the unmarked Subject Position.

Hence we may assume that the clause-initial constituent will very often be the -topical- subject and since topical constituents refer to information that the hearer can easily recover from the linguistic or non-linguistic context (or so the speaker assumes), these constituents are not likely to contain much AM. They provide, in a sense, old information. In fact these constituents might very well consist of a proper name or a pronoun only, which are usually highly restrictive regarding any type of cooccurring AM.⁵ Admittedly these claims need to be substantiated by further research; nevertheless we will now represent languages of type V as follows:

V-second: N Aux V N-- N-- N-- (class V)

The position of the auxiliary remains problematic in terms of head proximity, of course; we will assume that FG principles (III) and (IV) override the PHP in languages of the type that we are concerned with here.

2.2. Languages of type XII.

Other languages that (can) have terms on either side of the main verb are the so called 'leaking' V-final languages. As we saw previously, three major groupings were proposed in Greenberg (1966): VSO, SVO, and SOV languages. The establishment of the latter, however, posed some problems in that in many of the SOV languages the verb need not always be the last constituent in the clause: certain constituents would (typically) appear in postverbal position. In order to preserve his three-way classification Greenberg distinguished between rigid and non-rigid SOV languages (ibid.: 80):

Universal 7. If in a language with dominant SOV order, there is no alternative basic order, or only OSV as the alternative, then all adverbial modifiers of the verb likewise precede the verb. (This is the rigid subtype of III.)

Thus rigid SOV languages are strictly V-final, whereas languages of the non-rigid subtype (can) have constituents positioned after the verb, i.e. they 'leak' (the word was coined by Ross 1973). When we start from the type --N & V, the two subtypes can be represented as:

rigid V-final: --N --N --N --N V Aux

non-rigid V-final: --N --N V Aux --N --N

The non-rigid version clearly shows that the PHP is violated in that both the auxiliary and AM (adjective and genitive) in the postverbal terms would appear in between Heads V and N. However, closer examination will reveal that non-rigid V-final languages rather strongly tend to have their AM in positions that are consistent with the PHP.

2.2.1. The position of the adjective.

Greenberg's original thirty language sample, which provided the data for his universals and which contained eleven V-final languages, enables us to establish the position of the adjective in both V-final subtypes (Table 4; see also Hawkins 1983: 133 ff.).

Rigid V-final		Non-rigid V-final	
Burmese	GNA	Basque	GNA
Burushaski	--N	Chibcha	GNA
Hindi	--N	Loritja	GNA
Kannada	--N	Nubian	GNA
Japanese	--N	Quechua	--N
Turkish	--N		

Table 4.

In sum, the overall tendency is:

rigid V-final: --N --N --N --N V Aux

non-rigid V-final: -N- -N- V Aux -N- -N-

Apparently the -N- (= GNA) pattern is preferred most in the non-rigid subtype and this is in accordance with the PHP, for now in the worst case only one modifier will occur in between Heads V and N (.....NA V GN.....) instead of two (..... V --N.....).

If these ordering tendencies generally hold for the two V-final subtypes, this would be harmonic with the PHP, because modifiers (i.e. adjective and genitive) within the term phrase are arranged in such a way that the potential damage to the PHP is kept to a minimum.⁶ But then there is also the position of the auxiliary to account for. This is discussed in the next section.

2.2.2. The position of the auxiliary.

It was mentioned earlier that in non-rigid V-final languages the auxiliary may find itself 'sandwiched' in between the Head of the predication V and the postverbal term(s), which is not preferred from the point of view of the PHP:

non-rigid V-final:N V Aux N.....

Actually, there are a number of V-final languages that have the auxiliary not following the lexical verb, but in the second position of the clause (see also FG principle (VI) in section 4), as in example (9) from Vata (FUT-A = future auxiliary, FT = future tense particle; from Koopman 1984: 113):

- (9) à nī- k̄ā sāká dŭdŭ kŭ zŭ slé-é mīí
 we FUT-A-FT rice floor on put house-DEF in
 'We will put the rice on the floor in the house.'

This means that the auxiliary would not have appeared in between Heads, had it not been for the constituent in the special P1 position. So here FG-principles (III) and (IV) override PHP again. Another example is Bororo (Crowell 1979: 21 ff.), in which the auxiliary is suffixed to the (sentence-initial) pronominal subject of a transitive agentive clause, which is optionally preceded by a full subject NP. (When the clause lacks an agentive subject, it is suffixed to the verb; the object is cross-referenced on the verb by a pronominal prefix.)

- (10) I- re ipo ø- tawije moto piji. (ibid.: 43)
 1sg-neutral asp. post 3sg-remove ground from
 'I removed the post from the ground.'

- (11) lme e- re areme e- wije. (ibid.: 22)
 men 3pl-neutral asp. women 3pl-advise
 'The men advised the women.'

On the basis of the PHP we expect that an auxiliary will only occur in second position in languages of the non-rigid V-final subtype. In terms of head proximity it is unwarranted for an auxiliary to occur after the P1 constituent in rigid V-final languages since there is no chance of the auxiliary 'obstructing' head proximity there.

Interestingly this hypothesis holds for the Aux-second V-final languages we have come across so far, namely Walbiri (Australia), Papago and possibly other Uto-Aztec languages (North-America), Bororo (South-America), Yata and other Kru languages (West-Africa).⁷ In these languages the adjective follows the head noun, which in this case is also the preferred pattern from the point of view of the PHP. Only in Walbiri, which Hawkins (1983: 285) classified as SOV & po & N--, not only the adjective follows the noun, but the genitive does so too in the most frequent order; in Walbiri word order is extremely free in that A and O may appear on either side of the noun.

It may look as if the languages mentioned above constitute counterexamples to Universal 16, but in fact they do not, since Greenberg's definition only includes auxiliary verbs that are inflected for person and number, thus excluding languages that have no person/number category (like e.g. Japanese) and languages in which such notions as tense, mood and aspect are not expressed by (inflected) auxiliary verbs, but by for instance grammatical particles.⁸

Steele et al. (1981: 21) is an attempt to characterize the category Aux for all languages. Their definition reads as follows:

"Given a set of language internal analyses, those constituents which may contain only a specified (i.e. fixed and small) set of elements, crucially containing elements marking tense and/or modality, will be identified as nondistinct."

This definition includes both verbal and non-verbal auxiliaries and in an earlier study Steele (1978: 35) had also observed that "languages with a second position AUX are predominantly either rigid SVO languages [...] or SOV languages with relatively free word order (Luiseño, Walbiri, Karok)". As far as the evidence goes we are now able to impose a restriction on the distribution of auxiliaries in clause-second position in V-final languages to the effect that an auxiliary will only occur after the P1 constituent in languages of the non-rigid V-final subtype, as was predicted on the basis of the PHP.

3. Supporting evidence.

In the previous sections it has been suggested that the four largest groups, which together contain 247 languages or about 75 % of Hawkins' sample, display ordering patterns that are generally compatible with the idea that head nouns prefer to be contiguous with the verb. The Principle of Head Proximity, however, states that basically this preference is shared by all head nouns, whereas only one or maximally two head nouns can actually occur next to some verb in the linear organization of the clause.⁹ This means that, strictly speaking, every head noun that is not adjacent to a verb is at variance with the PHP. Therefore we may expect languages to display structures that can be interpreted as supporting evidence for the universal ordering principle that is proposed here. From a logical point of view there are three ways languages can deal with non-contiguity between Heads V and N, namely by:

1. restricting the number of terms per clause;
2. employing more verbs in one clause;
3. extraposing elements that would normally occur in between Heads.

These 'strategies' will now be discussed in turn, followed by some psycholinguistic evidence for the PHP.

3.1. The number of terms in a clause.

At first sight it may seem awkward to deal with non-contiguous head nouns by restricting their number per clause. On the other hand, since we argue that all head nouns prefer to appear next to the verb, every head noun that is not presents a problem in terms of head proximity.

Consequently the more terms a clause contains, the more head nouns will be separated from the verb, and the worse this is for the PHP. Conversely, the smaller the number of terms in a clause, the fewer will be the problems for the PHP. Hence we claim that linguistic expressions containing few terms are preferred to those containing many. In practice this often means that speakers will tend to avoid (overtly) expressing information that can be recovered from the context. As a matter of fact in many languages a state of affairs can be communicated by the verb complex alone, which means that all terms are not expressed. In that case the Principle of Head Proximity has become redundant, for when a clause does not contain terms there will not be any problems with respect to head proximity whatsoever. In these languages the entities referred to are usually overtly marked in the morphology of the verb complex. Languages vary considerably in the degree that this is allowed. While in some languages no entity or just one is coded morphologically in the verb complex (most often this will be the principal argument, i.e. Agent, Subject), others allow several entities to be coded there, as in this example from Abkhaz (Hewitt 1979: 51):

- (12) a- xàc'a a- pñ'əs l- y'əza a- š'q'ə
 the-man the-woman her-friend the-book
 (ø-) lə +z- lə - y- te-yt'
 it her +for to-her he give(finite)
 'The man gave the book to the woman for her friend'

All terms in (12) are optionally present. According to a theory developed by De Groot & Limburg (1986) languages like Abkhaz belong to the so-called Appositional Type languages, which -among other things- are characterized by extensive pronominal affixing on the verb. As a rule the entities referred to by affixing are only optionally present in the form of full (pro)nouns. Eventually the pronominal affixes may wear down to agreement markers (and even disappear completely), in which case they also lose their referring potential. At that stage full pronouns will have become obligatory again and the language then belongs to the Free Pronoun Type languages. English, for example, only has third person singular subject marked in the morphology of the verb and full pronouns are necessary. But at some point in time pronouns may cliticize onto the verb (and still keep their referring potential), in which case the language will belong to the Clitic Type languages. Consider the following examples from French:

- (13)a. Jean a donné ton livre à Marie.
 John has given your book to Mary
 b. Il- le-lui- a- donné.
 he-it-her-has-given

In languages of this type intended referents may be added as extra-clausal constituents, i.e. in Theme or Tail position (Dik 1978: Ch. 6 ; see also Harris 1985):

c. Jean, il-le-lui-a-donné, ton livre, à Marie.

Clitic Type languages could eventually turn into languages of the Appositional Type (see above) and the whole process may start again.

Although this summary does not do justice to their theory in that we have only sketched the general idea (which De Groot & Limburg apply to other phenomena as well), it adequately illustrates the statement made above that there might be a universal tendency to restrict the number of overtly expressed terms per clause. Of course this tendency can only go as far as the rules for successful communication allow it to go and it should be noted that as a rule only constituents that are required by the verb (in F0 'arguments'), are signalled by bound morphology on the verb .

There are, however , also ways to reduce the number of constituents that are not required by the verb ('satellites')¹⁰. In a number of Papuan languages this is attained by the employment of extra clauses. In Kobon, for example:

"All of the obligatory and optional arguments can cooccur but if many arguments are involved there is a strong preference for dividing them over two or more clauses which may contain an identical predicate rather than including them all in one clause." (Davies 1981: 45-46).

In Usan different predicates are used (Reesink 1984: 133-134):

"Frequently, because of the restriction on number of terms per predicate Ins [i.e. the semantic function Instrument - JR] is expressed as Go [i.e. the semantic function Goal - JR] of a preceding predication and understood in the predication that follows[...]".

This phenomenon is not restricted to Papuan languages, but can also be attested in Amerindian languages such as Hixkaryana (South-America) and Diegueño (North-America). In Hixkaryana, as in Kobon, the verb is repeated (Derbyshire 1979: 39), whereas in Diegueño, as in Usan, another verb is used: "Diegueño has a 'conspiratorial' tendency to avoid surface configurations of many elements at one level of embedding. The principal device for effecting this structural characteristic is the 'absolute' construction. [...], the construction can be and often is used to indicate what in English or in a simpler Diegueño sentence would be a surface argument of the verb" (Gorbet 1976: 155).

In sum, there seems to be a tendency to avoid the occurrence of many terms in one clause, as was hypothesized on the basis of the PHP.

f. mo fi ggbòn gé igi Yoruba
 I took cleverness cut tree
 'I cleverly cut the tree'

(15) mo mú iwé wá fún o Yoruba
 I took book came gave you (from Stahlke 1970: 63)
 'I brought you a book'

It must be noted, however, that there is no consensus as to the categorial status of these verb-like elements: opinions differ as to whether they are to be regarded as verbs or as relators (adpositions). For instance, Givón (1984: 179-180; examples (14a-f) are his) suggests that verbs in a serial construction (with the exception of the first verb in the series) serve as case role markers of adjacent nouns. Nevertheless, here we will assume that they provide supporting evidence for the Principle of Head Proximity.

3.3. The position of AM of the Head Noun.

Let us first have a look at structures as they could occur in languages of the three major groups.

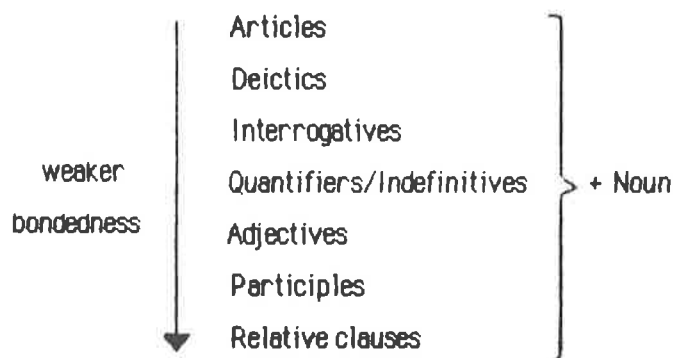
V-initial: Aux V N-- N-- N-- N--
 V-second: N(--) Aux V N-- N-- N-- (cf. section 2.1)
 V-final: --N --N --N --N V Aux

Now suppose that in each of these structures none of the terms can be omitted without a serious loss of information. In that case (at least) two head nouns in the V-second type and three in the other language types are separated from the verb. Their contiguity with the verb is obstructed by one or more other head nouns that may be accompanied by some AM as well.

In this situation there is still a way in which non-contiguous head nouns can appear at a relatively closer distance from the verb, i.e. attain better head proximity, namely by allowing the AM of obstructing head nouns to appear in a position where it will not stand in between some N and V.

Indeed in many languages we find instances of AM occurring outside its proper domain (a phenomenon usually known as extraposition), although not just any type of AM is liable to displacement. This might be related to the Foley's bondedness hierarchy (see below), which is based on the distributional properties of ligatures in Austronesian languages; ligatures are special particles that link AM to the head noun. "The distribution of ligatures is determined by the principle that if a category X in a language employs a ligature, then all categories in the Bondedness Hierarchy below X will use a ligature in that language. The fundamental factor determining the arrangement of these constructions along the hierarchy is the notion of strength

of syntactic bonding. The higher a construction is on the hierarchy, the more tightly bound the adjunct is to the head noun" (ibid.: 2).



Foley's Bondedness Hierarchy.

According to this theory the relative clause is the easiest to separate from the head noun and indeed this construction seems to occur outside its proper domain more often than any of the other categories mentioned above. Here are some examples of extraposed relative clauses in the three major language types.

Jacalteco (V-initial, Craig 1977: 194 ff.):

- (16)a. xitij naj ah hoyom x'apni yet kani boj sc'ahol ixim.
brought cl/the from Todos Santos arrived when last night with his son corn
'The man from Todos Santos who came in last night with his son brought corn'

- b. xitij naj ah hoyom ixim naj x'apni yet kani boj sc'ahol.
brought cl/the from Todos Santos corn cl/the arrived when last night with his son
'The man from Todos Santos brought corn, the man who came last night with his son'

In (16)b. the relative clause appears sentence-finally, albeit with a copy of the antecedent naj, the third person independent pronoun/noun classifier. According to Craig (ibid.), Jacalteco "has a tendency to avoid center embeddings, which arise whenever the relative clause is not embedded under a sentence final NP". To put it differently, relative clauses are not preferred in a position somewhere between N and V.

In English (V-second) the relative clause may appear sentence finally too, provided that this does not result in an ambiguous utterance:

- (17) He bought an old painting at the flea market that is supposed to be an original Rembrandt.

In V-final languages the relative clause is often rather unlike its counterpart in V-initial and V-second languages in that it has no relative marker (such as a relative pronoun) and that the

head noun (the antecedent) is part of the relative construction itself. Commonly the relative clause as a whole is a nominalized structure that may receive case marking. Consider for example this sentence from Mohave (Langdon 1977: 266):

- (18) Mary-c John mal^Yki: hutav- ən^Y hidaw-m
 Mary-subj. John ball hit(nom)-dem catch-tns
 'Mary caught the ball John hit'

In Mojave and other languages of the Yuman family relative clauses may be placed in clause-initial or in clause final-position (examples from Yuma; *ibid.*: 270):

- (19)a. John-ts piilot uukwit- nya Mary-ts adaw-sh
 John-subj. ball hit(nom)-dem Mary-subj. catch-evidential
 'The ball John hit, Mary caught it'
- b. Mary-ts adaw-sh John-ts piilot uukwit- nya
 Mary-subj. catch-evid. John-subj. ball hit(nom)-dem
 'Mary caught it, the ball John hit'

Since the relativized head noun piilot 'ball' is an intrinsic part of the relative construction we will not go into the problem of head proximity here, but merely observe that the relative clause in these examples appears in a position that makes it possible for (other) head nouns to occur at greater proximity to the verb. It must be added, however, that there is also a type of relative construction in the Yuman languages, in which the head noun is repeated after the sentence-initial relative construction and thus appears closer to the verb, as in this Diegueño sentence (Gorbet 1976: 63-64):

- (20) i'pac a'k wi'- m tuc- pu a'k- pu sin^Y- c wyaw
 man bone rock-with hit(nom)-dem bone-dem woman-subj. find
 'The woman found the bone that the man hit the rock with.'

In many respects this construction resembles what has been referred to as a correlative (Keenan 1985: 163 ff.), except that a correlative marker is lacking (at least Gorbet does not mention the presence of such a marker). Consider also the following example from Hindi, which does have this element (*ibid.*):

- (21) Jis a:dmika kutta bema:r hai, us a:dmiko maine dekha
 COREL man GEN dog sick is that man DO I ERG saw
 'I saw the man whose dog is sick' (lit: 'Which man's dog is sick, that man I saw')

It is quite possible that correlatives typically appear in Theme-position, i.e. outside the predication proper (cf. Dik 1978: Ch. 6), so as to yield better head proximity in the clause. It may be interesting to add that the distribution of correlatives is largely restricted to V-final languages. This being the case, we would expect that this must be especially so in rigid V-final languages. In languages of this subtype the clause-final position is not available for the placement of extraposed relatives and since the special clause-initial position P1 is usually reserved for topical or focal constituents, extraposed relatives would thus appear in the position preceding the predication proper, i.e. in Theme position. However, the contrary seems to be true, since in fact they appear to be mostly limited to the non-rigid type (ibid.: 164). At this moment we are unable to account for these puzzling data, but it must be stressed that the distribution of this rare type of relative clause does not provide real counterevidence to the PHP. After all, it does not appear in between Heads in the predication proper. Thus these examples generally indicate that there could be a tendency for at least one category of AM of the head noun not to appear in between Heads V and N.

3.4. Psycholinguistic evidence.

The hypothesis that the Head of a domain prefers to be contiguous with the Head of its superordinate domain is not only substantiated by linguistic data, but there is also some evidence from psycholinguistic research.

It was said that strictly speaking the PHP states that every head noun prefers to be contiguous with the verb, which is of course quite impossible in the linear structure of the clause. It was also suggested that an underlying predication can be represented as a non-linear structure in which all Heads of domains, or rather subdomains, are linked directly to the Head of the superordinate domain, as in the relational network in diagram 1c, which is repeated below as 1c^z.

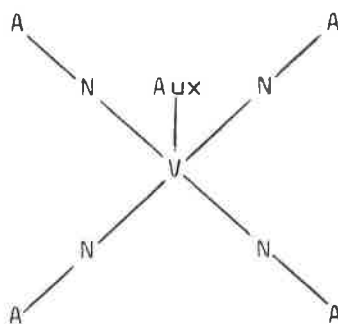


Diagram 1c^z

Given the assumption that this diagram is a more or less adequate mental representation of a predication, then universal word order tendencies would reflect two essential features of this type of organisation: firstly, Heads of domains take up prominent positions in a predication (key-positions one might say); and, secondly, no AM occurs in between these Heads. Interestingly some evidence for this hypothetical view might be provided by language perception research. For example Moore (1972) designed a test in which students had to rate sentences in terms of grammaticality under time pressure. His results suggest that first the relations between the verb and the arguments are processed and then the relations internal to each argument. Thus we might posit that listeners first attempt to identify relations between the head nouns and the verb when they process a linguistic utterance. This would imply that Heads are the relatively most important elements in a predication. Evidence for this view can be found in different places in Clark & Clark (1977: 53), e.g.:

"Listeners have at their command a battery of mental strategies by which they segment sentences into constituents, classify them, and construct semantic representations from them. These strategies rely on the fact that sentences contain elements listeners can use as clues to proper segmentation."

One such strategy, according to Clark & Clark (ibid: 61), goes as follows:

"Strategy 2: After identifying the beginning of a constituent, look for content words appropriate to that type of constituent."

Perhaps this might be paraphrased as: whenever you identify a domain, look for the Head of that domain first, and then try to relate this Head to the Head of the superordinate domain (in Prefield lgs.) or the Head of the subordinate domain (in Postfield lgs.). Proof for such a highly speculative statement does not seem to be available. What appears to be rather transparent, however, is that the PHP, if proven to be a valid ordering principle, will most probably be due to psycholinguistic strategies, especially in view of the nature of the evidence that has been offered in the previous sections.

4. The PHP and ordering principles in FØ.

In the theory of Functional Grammar ordering patterns as they are attested in actual linguistic expressions are determined by Placement Rules that are applied to fully specified, but unordered underlying predications. Placement Rules are language specific in the sense that for any language there are rules that assign a position to every single element that the predication of a particular language contains. These rules are governed by general Principles of Constituent Ordering, which more or less define the extent to which certain ordering patterns are possible in natural language, i.e. they define the space within which Placement Rules can function. The significance of an ordering principle may vary for each individual language and they are not necessarily harmonic, since the ordering that is preferred by principle A may be in conflict with the one that is preferred by principle B. To quote Dik's words again (Dik, in prep.): "[...] hardly any of these principles has absolute 100 % validity: most of them can be counteracted by one or more of the other principles. They are like forces pulling in different directions with varying outcomes in different languages."

The ordering principles that have been proposed so far (see below) are formulated in terms of the Prefield/Postfield typology. In this view languages tend to have Dependents either before or after a given Center, i.e. in the Prefield or in the Postfield respectively. Cf.:

Domain	Center	Dependents
Predication	Predicate	Terms (arguments, satellites).
Term	Head Noun	Other restrictors [†] (adjectival phrase, possessive phrase, relative clause etc.)
Adjective Phrase	Adjective	Arguments (if any), standard (in comparatives), degree adverbial.

Table 5. ([†] The head noun is the first restrictor of a term)

A typical Prefield language will have Dependent elements positioned before the Center (e.g. SOV & AN & GN), and the opposite will be the case in characteristic Postfield languages (e.g. VSO & NA & NG). Apart from the terminology, the similarity with Table 2 (repeated below as 2*) is quite conspicuous. There is, however, one major difference between the two tables: operators are not included in Dik's Dependents, whereas they are in our category of Attributive Material. We will return to this below.

The Principles of Constituent Ordering proposed so far are (Dik 1983):

- (I) A language makes a basic choice between Prefield and Postfield ordering.
- (II) The Subject position precedes the Object position.
- (III) There is a universally relevant clause-initial position P1, which is used for special purposes, including the placement of constituents with Topic or Focus function.
- (IV) Since the Subject is the prime Topic candidate, it will often end up in P1; this may lead to a reinterpretation of P1 as the unmarked Subject position.
- (V) The Prefield is less hospitable to complex material than the Postfield; we may thus expect languages to take measures to relieve the Prefield of excessive complexity.
- (VI) Predicate operators prefer
 - (a) the Counterfield;
 - (b) the second position in the clause.
- (VII) Terms prefer the ordering: Operator - N - Adj - Relative clause
- (VIII) Adjectives may pattern along with term operators.
- (IX) Genitives may pattern along with Adjectives, if these are sensible to Principle (VIII).
- (X) Other things being equal, constituents prefer to be placed in order of increasing complexity: Clitic - Pronoun - Noun Phrase - Adpositional Phrase - Subordinate clause.
- (XI) The preferred position of a Relator is at the periphery of its immediate relatum.
- (XII) The preferred position of a Relator is in between its two relata.

Relators - "elements that serve to link two constituents to each other" (Dik, in prep.)- do not figure in Table 2* (and 5). This implies that they are the only elements that may occur in between Heads. They include coordinators, subordinators, adpositions, and case markers.¹¹ Operators are grammatical or form elements that operate on lexical or content material such as verbs and nouns. A distinction is made between term operators, predicate operators, and predication operators. Term operators (see principle (VII)) typically encode such notions as (in)definiteness and number, which may be expressed formally as determiners and

quantifiers/numerals. Grammatical expressions (auxiliary verbs and the like) indicating tense, mood, aspect, and polarity are referred to as predicate operators (see principle (VI)). Notice that they do not include lexical expressions of temporality, aspectuality, etc. like 'last week' or 'repeatedly'. They do include two notions that are not necessarily required in the category Aux as defined in Steele et al. (1981): aspect and polarity. But as a rule they are part of the verb complex too (see e.g. Givon 1984: 64 ff.). Finally, predication operators are formal expressions of illocutionary values such as declarative, imperative, and interrogative.

We saw above that operators are not sensitive to Dik's basic Prefield/Postfield typology, since they are not included in Table 5. Here a different position is taken, as can be observed in Table 2*, which divides overtly expressed linguistic material in any domain into essentially two categories: Heads and non-Heads, or Attributive Material. The latter category includes Dik's Dependents as well as the operators.

Domain	Head	Attributive Material
Predication	Predicate (V)	Predicate operators (auxiliary verbs and the like), Terms.
Term	Head Noun (N)	Term operators (determiner, quantifier), modifiers (adjective, genitive, etc.).
Adjective Phr.	Adjective (A)	Degree adverbial etc.

Table 2*.

Consequently term operators too must adhere to the PHP (and, by the way, to the Prefield/Postfield typology) and there are in fact languages in which all AM within the term appears either before (in V-final lgs.) or after the head noun (in V-initial lgs.), as is the case in, for example, Japanese (Kuno 1978: 83) and Turkana (Dimmendaal 1982: 206) respectively. We suggest that other, stronger principles or 'forces' will then have to take care of the various ways languages may deviate from the ordering that is preferred in terms of the PHP. With respect to predicate operators we observe that the positions that are claimed for these elements by FG-principle (VI) follow automatically from the preferred ordering in terms of the PHP, provided that we accept a priori that discontinuity of the verb complex, and in other linguistic patterns for that matter, is generally disfavoured. Otherwise the following structures would be possible too:

V-initial: *V N-- N-- N-- N-- Aux

V-final: *Aux --N --N --N --N V

But as we saw in section 2.2 discontinuous verb complexes seem to be found in non-rigid V-final languages only, which is harmonic with the PHP.

The distributional characteristics of predicate operators probably also hold for predication operators, which have only recently been proposed in the FG literature (Dik, in prep.); even to the extent that they can be found in clause-second position in non-rigid V-final languages, as in Nama Hottentot:

- (22) siike ke //naápá kè sii paas /xáa (Hagman 1973: 212)
we decl. there arrived bus with
'We arrived there by bus'

In Nama Hottentot the declarative particle ke is usually present after the first constituent in every declarative sentence. The interrogative sentence does not have this particle, but can have the emphatic interrogative particle kxa in the same position. Imperative sentences also lack the declarative marker; instead the imperative particle re may appear sentence-finally (ibid: 259 ff.). As a rule, however, such values are marked in the verb complex (Givón 1984: 70-71).

In sum, it is suggested here that term operators are to be included in the material that is affected by FG principle (I). But even then two ordering principles -(I) and (VI)- are required to define structures for which one principle may suffice: the Principle of Head Proximity, which was shown to be compatible with the four best represented language classes I-V-XI-XII.

5. Conclusion.

In this paper we have attempted to demonstrate that universal tendencies of constituent ordering can be viewed as a consequence of the Principle of Head Proximity, which could also be related to the occurrence of predicate operators in the second position of the clause in non-rigid V-final languages. It was shown too that there are various other linguistic patterns that make perfectly good sense in the light of head proximity. Hence they were interpreted as providing supporting evidence for the PHP. Although Attributive Material was restricted to only three categories (adjectives, genitives, and predicate operators), the PHP may be compatible with some 75 % of the languages in Hawkins' sample. In the case of AM appearing in between Heads, this must, in our opinion, be ascribed to other, notably stronger ordering principles of the kind proposed in Dik (1983).

Despite the fact that the bulk of this paper is of a highly speculative nature we would like to point out that the PHP compares favourably with earlier proposals that have tried to account for Greenberg's data (see note 2), because of its (potentially) greater descriptive and explanatory power. However, much more research is needed to substantiate the claims that are made here.

FOOTNOTES.

1. This is an abridged and revised version of my MA-thesis (University of Amsterdam 1984). I would like to thank Simon Dik for helpful discussion and comments. The responsibility for the contents, however, remains entirely mine.

2. In chronological order e.g. Bartsch & Vennemann (1972), Lehmann (1973), Keenan (1978), Hawkins (1980), and Maxwell (1984) put forward basic ordering principles to account for cross-categorial correlations of constituent ordering.

3. Japanese examples originally from Kuno (1978), Samoan examples from Marsack (1962).

4. The notion Head can be given different interpretations. For an overview, see Zwicky (1985).

5. So as to provide some evidence for this hypothesis, we checked on the contents of all terms in the first chapter of Hemingway's *The Sun Also Rises* (some 1400 words). It turned out that only about 10 % of all subjects contained any AM, whereas this was the case for as much as 75 % of the other terms.

6. Cf. Dixon (1980: 442), who states that Australian languages are often V-final and that in these languages "peripheral syntactic NPs [...] tend to occur close to the verb, either immediately before or - most often - straight after it. Local NPs will usually follow the verb. [...] Within NPs it is most common for a possessive qualifier to precede and an adjective to follow the head noun."

7. Another common feature of the Kru languages is that finite lexical verbs appear in clause-second position too (cf. Marchese 1979), like in Dutch and German main clauses (i.e.: S Vf .. O .. V ..). However, in the latter two languages the adjective precedes and the possessor phrase follows the head noun, which is why they are not mentioned in the text. This ordering is extremely rare in languages with V-final patterns.

8. Greenberg (1966: 84) defined auxiliaries as "a closed class of verbs [...] inflected for both person and number [...] in construction with an open class of verbs not inflected for both person and number."

9. By definition the preferred position of the Relator, which does not figure in the Head/AM distinction, is between its two relata, e.g. N and V.

10. For the difference between Arguments and Satellites, see Dik (1978: chapter 3).

11. Some specific Relators are discussed in Limburg (1983, 1985) and Kahrel (1985).

REFERENCES

- Bartsch, R., and T. Vennemann. 1972. *Semantic structures*. Frankfurt am Main: Athenäum.
- Bennis, H., and W.U.S. van Lessen Kloeke (eds.). 1983. *Linguistics in The Netherlands 1983*. Dordrecht: Foris.
- Bolkestein, A.M. et al. (eds.). 1985. *Predicates and terms in Functional Grammar*. Functional Grammar Series 2. Dordrecht: Foris.
- Clark, H., and E.V. Clark. 1977. *Psychology and language. An introduction to psycholinguistics*. New York: Harcourt Brace Jovanovich, Inc.
- Corum, C., et al. (eds.). 1973. *You take the high node, and I'll take the low node - Papers from the ninth regional meeting of the Chicago Linguistic Society*. Chicago: CLS.
- Craig, C.G. 1977. *The structure of Jacaltec*. Austin: University of Texas Press.
- Crowell, T. H. 1979. *A grammar of Bororo*. Ph.D. diss. Cornell University. Ann Arbor, Mich.: UMI.
- Davies, J. 1981. *Kobon*. *Lingua Descriptive Studies* 3. Amsterdam: North-Holland.
- Derbyshire, D.C. 1979. *Hixkaryana*. *Lingua Descriptive Studies* 1. Amsterdam: North-Holland.
- Dik, S.C. 1978 (3rd revised ed. 1981). *Functional Grammar*. Dordrecht: Foris.
- Dik, S.C. (ed.). 1983. *Advances in Functional Grammar*. Dordrecht: Foris.
- Dik, S.C. 1983. Two constraints on relators and what do can do for us. In Dik (ed.), 267-298.
- Dik, S.C. (in prep.). *The theory of Functional Grammar. Part 1 - The simple predication*.
- Dimmendaal, G.J. 1982. *The Turkana language*. Ph.D. diss. Universiteit van Leiden. Dordrecht: Foris.
- Dixon, R.M.W. 1980. *The languages of Australia*. Cambridge: Cambridge University Press.
- Foley, W.A. 1976. *Comparative syntax in Austronesian*. Ph.D. diss. University of California, Berkeley.
- George, I. 1975. *A grammar of Kwa-type verb serialization: its nature and significance in current generative theory*. Ph.D. diss. University of California, Los Angeles. Ann Arbor, Mich.: UMI.
- Givón, T. 1984. *Syntax: A functional-typological introduction. Volume 1*. Amsterdam: Benjamins.
- Gorbet, L.P. 1976. *A grammar of Diegueño nominals*. New York: Garland.
- Greenberg, J.H. (ed.). 1966². *Universals of language*. Cambridge, Mass.: MIT Press.
- Greenberg, J.H. 1966. Some universals of grammar with particular reference to the order of meaningful elements. In Greenberg (ed.), 73-113.
- Greenberg, J.H. (ed.). 1978. *Universals of human language - Volume 3: Word structure*. Stanford: Stanford University Press.
- Groot, C. de, and M.J. Limburg. 1986. *Pronominal elements: diachrony, typology, and formalisation in Functional Grammar*. Working Papers in Functional Grammar 12.
- Hagman, R.S. 1973. *Nama Hottentot grammar*. Ph.D. diss. Columbia University. Ann Arbor, Mich.: UMI.
- Harris, Martin C. 1985. *Word order in contemporary French: a functional view*. Working Papers in Functional Grammar 1.
- Hawkins, J.A. 1980. On implicational and distributional universals of word order. *Journal of Linguistics* 16, 193-235.
- Hawkins, J.A. 1983. *Word order universals*. New York: Academic Press.
- Hewitt, B.G. 1979. *Abkhaz*. *Lingua Descriptive Studies* 2. Amsterdam: North-Holland.
- Kahrel, P. 1985. Indirect questions and relators. In Bolkestein et al. (eds.), 165-181.
- Keenan, E.L. 1978. On surface form and logical form. *Studies in the linguistic sciences* 8-2 (special issue), 163-204.
- Keenan, E.L. 1985. Relative clauses. In Shopen (ed.) - Volume II, 141-170.
- Koopman, H. 1984. *The syntax of verbs: from verb movement rules in the Kru languages to Universal Grammar*. Ph.D. diss. Katholieke Hogeschool Tilburg. Dordrecht: Foris.
- Kuno, S. 1978. Japanese: a characteristic OV language. In Lehmann (ed.), 57-138.
- Langdon, M. 1977. Syntactic change and SOV structure: the Yuman case. In Li (ed.), 255-290.
- Lehmann, W.P. 1973. A structural principle of language and its implications. *Language* 49, 47-66.

- Lehmann, W.P. (ed.). 1978. Syntactic typology - studies in the phenomenology of language. Sussex: the Harvester Press.
- Li, C.N. (ed.). 1977. Mechanisms of syntactic change. Austin: University of Texas Press.
- Limburg, M.J. 1983. The parameter preposition/postposition in word order typology. In Bennis and Van Lessen Kloeke (eds.), 149-159.
- Limburg, M.J. 1985. On the notion 'Relator' and the expression of the genitive relation. In Bolkestein et al. (eds.), 147-163.
- Marchese, L. 1979. Tense/aspect and the development of auxiliaries in the Kru language family. PH.D. diss. University of California, Los Angeles. Ann Arbor, Mich.: UMI.
- Marsack, C.C. 1962. Teach yourself Samoan. London: The English Universities Press.
- Maxwell, D.N. 1984. A typologically based principle of linearization. *Language* 60, 251-285.
- Moore, T.E. 1972. Speeded recognition of ungrammaticality. *Journal of verbal learning and verbal behavior* 11, 550-560.
- Noonan, M.P. 1985. Complementation. In Shopen (ed.) - Volume II, 42-140.
- Reesink, G.P. 1984. Structures and their functions in Usan - a Papuan language of Papua New Guinea. PH.D. diss. Universiteit van Amsterdam.
- Ross, J.R. 1973. The Penthouse Principle and the order of constituents. In Corum et al. (eds.), 397-422.
- Shopen, T. 1985. Language typology and syntactic description - 3 volumes. Cambridge: Cambridge University Press.
- Stahlke, S. 1970. Serial verbs. *Studies in African linguistics* 1-1, 60-90.
- Steele, S. 1978. The category AUX as a language universal. In Greenberg 1978 (ed.), 7-45.
- Steele, S. et al. 1981. An encyclopedia of Aux: a study in cross-linguistic equivalence. Cambridge, Mass.: MIT.
- Zepeda, O. 1983. A Papago grammar. Tucson: University of Arizona Press.
- Zwicky, A.M. 1985. Heads. *Journal of linguistics* 21, 1-29.

