

Pre-final draft - J. Rijkhoff - May 2014 (PDF)

Bibliographic details of the final version:

Rijkhoff, Jan. 2015. Word Order. In James D. Wright (editor-in-chief), *International Encyclopedia of the Social & Behavioral Sciences*, 2nd edition, Vol. 25, 644–656. Oxford: Elsevier.
Available in both print and online editions.

Available online at:

<http://www.sciencedirect.com/science/article/pii/B9780080970868530311>

Abstract

The way constituents are ordered in a linguistic expression is determined by general principles and language specific rules. This article is mostly concerned with general ordering principles and the three main linguistic categories that are relevant for constituent order research: formal, functional and semantic categories. The general principles appear to be motivated by cognitive considerations, which are deemed to facilitate language processing and which can all be regarded as manifestations of iconicity: non-arbitrary relations between the form and the content of a linguistic expression. Three major iconic ordering principles are the principles of Domain Integrity, Scope, and Head Proximity.

Word order

Jan Rijkhoff

1. Introduction¹

The way words or groups of words are ordered in any particular language is always meaningful, so when a language allows for alternative ordering patterns, each variant should be regarded as communicatively significant. For instance, the English sentence ‘The woman called the doctor’ does not mean the same as ‘The doctor called the woman’. To give another example, in Spanish one can find both *mi amigo viejo* (lit. ‘my friend old’) and *mi viejo amigo* (lit. ‘my old friend’), but *mi amigo viejo* (with the noun preceding the adjective) means ‘my elderly friend’ (‘my friend who is old’), whereas *mi viejo amigo* (with the adjective preceding the noun) expresses emotional content: ‘my old/dear/longtime friend’. Even in so-called ‘free word order’ languages, it is not the case that each possible order is fully grammatical or acceptable under all circumstances (Dik 1997: 18, 394; Heine and König 2010: 87).

Speech is inherently linear as we can only pronounce one word at a time, but from a cross-linguistic perspective it appears that some ordering patterns are selected on a language-specific basis to structure discourse, whereas in other cases the linear organization of speech appears to be motivated or constrained by general cognitive considerations (e.g. for communicative purposes or to facilitate language processing), which would explain why we find certain ordering tendencies in languages across the globe. Thus, whether the object argument precedes the verb (as in Turkish) or follows it (as in English) in the unmarked declarative sentence of a language is probably not due to any general ordering preference, as in this particular case there is no obvious reason why one order should be preferred over the other in human communication. Furthermore, Object-Verb order and Verb-Object order are more or less evenly distributed in languages across the globe (which might also be due to chance, of course).

On the other hand, there are several reasons why the subject argument commonly precedes the object argument in the languages of the world: ‘Subject before Object’ mirrors the fact that language users prefer to process GIVEN before NEW information, i.e. WHAT IS BEING TALKED ABOUT (Topic or Theme) before WHAT IS BEING SAID ABOUT IT (Focus or Rheme). The insight that constituents which help the hearer to interpret the rest of the clause normally appear at the beginning is commonly attributed to the Prague School of Linguistics in the first half of the 20th century, which divided the communicative structure of the sentence into a thematic and a rhematic part, but statements to the same effect can already be found in Weil (1844: 19).

As may be clear from these introductory paragraphs, the current article discusses word order from a functional-typological perspective, which means that, in addition to formal aspects, due attention is also paid to semantic and functional (pragmatic) factors.

Before we continue, we should mention two issues. First of all, since word order studies are concerned with the order of INDIVIDUAL WORDS as well as GROUPS OF WORDS that function as a single unit in a larger linguistic expression, it would be more appropriate to use the term ‘constituent’. For this reason ‘constituent order’ will be the preferred term in the remaining part of this article.

Secondly, the current contribution is restricted to constituents that are properly contained in a linguistic unit or domain (section 5.1) like a clause or a noun phrase (NP). Not all studies on constituent order indicate whether the constituents they discuss are actually an integral part of a clause or an NP (rather than e.g. an afterthought or an apposition), but failing to make

¹ I am grateful to Eva van Lier for helpful comments on an earlier version.

this distinction gives an unreliable picture of the true ordering possibilities within a phrase or clause. For example, a complete clause often contains one or more extra-clausal constituents, as in (with extra-clausal constituents in italics):

- (1) *Well*, what do you think, *Michelle*?
 (2) *By the way*, I shouldn't have said that.

In the case of phrases, a distinction is often made between 'tight' and 'loose constituency'. An extreme case of 'loose constituency' is attested in Kalkatungu (Australia). In this language "there are in fact no NPs, but [...] where an argument is represented by more than one word we have nominals in parallel or in apposition. [...] Each word is a constituent of the clause [...]" (Blake 1983: 145).

In other languages only postposed elements are regarded as loosely connected or 'appositional' elements. For example, in the Caucasian language Georgian a postposed modifier behaves more like a separate NP, or apposition, in that it inflects for the full set of cases and numbers (Testelec 1997: 652). This is illustrated in (3b).

Georgian

- (3) a. *am or-Ø lamaz-Ø kal-s* b. *kal-eb-s lamaz-eb-s*
 that:OBL two-DT nice- DT woman-DAT woman-PL-DAT nice-PL-DAT
 'to those two nice women' 'to the nice women'

Thus, constituents such as 'well' and 'Michelle' in (1), 'by the way' in (2), and Georgian *lamaz-eb-s* (nice-PL-DAT) in (3b) will not be taken into account.

In addition to offering a brief outline of main developments in the history of constituent order research (section 3), this article discusses some problems and recent developments in this area, in particular the way in which constituents are identified and compared across languages. Since certain semantic categories (like basic color terms) or form-based categories (word classes such as adjectives or prepositions) are not attested in all languages, and since the same constituent may have different discourse functions (and conversely, since the same function can be fulfilled by constituents belonging to different form classes), we need semantic, formal as well as functional categories to account for constituent order phenomena in all the world's languages (section 4). Constituent order studies have been mainly concerned with the relative order of (i) subject, object, and verb in the clause and (ii) certain adnominal modifiers (typically adjectives, numerals and demonstratives). The present survey is somewhat biased towards the NP, as constituent order in this phrase has been investigated in considerable detail in some large cross-linguistic studies (Hawkins 1983, Dryer 1992, Rijkhoff 2004, Cinque 2005).

2. Investigating constituent order across languages: some general issues

Constituent order is not equally relevant for all languages. The study of constituent order presupposes that there are constituents that appear in certain linear patterns, but what is expressed as a free constituent in one language can be expressed as a morphologically bound form in another languages. This is particularly true in the case of incorporating or polysynthetic languages, as illustrated by this example from the Australian language Tiwi (Blake 2006: 591):

Tiwi

- (4) *Pitiwuliyondjirrurlimpirrani*
Pi-ti-wuliyondji-rrurlimpirr-ani.
3PL-3SG.FEM-dead.wallaby-carry.on.shoulders-PT.HABIT
'They would carry the dead wallaby on their shoulders.'

It is also possible to find polysynthesis in phrases. Since in Nivkh (Siberia, Russia) "dependent-head synthesis is obligatory with nominal heads" (Mattissen 2002: 152), demonstratives, numerals and adjectives all appear as bound forms, as in (Mattissen 2002: 156):

Nivkh

- (5) *t'-vila-Gan*
2SG-big-dog
'your big dog'

A cross-linguistic comparison of constituent order can also be affected by the fact that languages do not necessarily share the same set of categories. For example, the Amazonian language Pirahã is deemed to have no color terms and adnominal demonstratives are absent in Hixkaryana (another Amazonian language from Brazil). Furthermore it appears that true numerals are not attested in all languages and that in some languages the expression for a cardinal number must be accompanied by a gesture; in fact in some languages (e.g. the Papuan language Kobon) the gesture alone can suffice.

Then there is the problem that the translational equivalents of some form or construction can be too varied for cross-linguistic comparison (section 3.2.2). For example, a distinct class of true adjectives is lacking in many of the world's languages, in particular outside Europe. Such languages typically employ phrasal or clausal structures headed by a stative verb ('be.pretty_V', 'be.silent_V') or an abstract noun ('prettiness_N', 'silence_N') to express adjectival notions in the NP, as shown in these examples from Kiribati (Kiribati, Micronesia) and Makwe (Palme, Mozambique).

Kiribati (Ross, 1998: 90)

- (6) a. *E tikiraoi.* b. *te uee ae e tikiraoi*
3SG.S be.pretty ART flower REL 3SG.S be.pretty
'She/He is pretty.' 'a pretty flower'

Makwe (Devos, 2008: 136; see also section 3.3.2)

- (7) *muú-nu w-á=ki-búúli*
NC1-person PP1-GEN=NC7-silence
'a silent person' (lit. 'person of silence')

Similar things can be said about the various ways numeral modifiers are expressed across the globe, as in quite a few languages cardinal numerals are regarded as morphological verbs or nouns. Notice also that many African languages (e.g. Igbo, Hausa) only have a minor, closed class of non-derived adjectives.

Another reason why certain constituent orders may not be found in all the world's languages has to do with the fact that speakers of a language fail to realize certain ordering sequences in normal discourse, even though these patterns are deemed grammatically possible. According to Lyle Campbell (personal communication), there would be nothing ungrammatical about a noun phrase containing a demonstrative, a numeral and an adjective in

Pipil (El Salvador), in which case the order would be [dem num A N], but apparently this never happens in actual discourse. Some languages have structural constraints that prevent certain constituents from appearing simultaneously in the same phrase or clause. Thus, a lexical NP can accommodate only one modifier in Yimas (New Guinea). Any other modifying material must take an agreement suffix and occur as an appositional phrase (Foley, 1991: 184-8).

When such factors (polysynthesis, apposition, co-occurrence restrictions etc.) are taken into consideration, quite a few languages are excluded from a cross-linguistic investigation of constituent order. For example, because of the issues mentioned above (but also because of e.g. ‘dependency reversal’; cf. Malchukov, 2000), the relative order of demonstrative, numeral, adjective and head noun could only be established for less than half of the languages in a representative sample of the world’s languages (Rijkhoff, 2004: 327-329).

3. Traditional constituent order typology

Especially since Greenberg (1963), constituent order typology has mainly been concerned with two issues:

- (i) the classification of languages on the basis of the relative order of the verb (V) and its arguments subject S and object O in a transitive clause (which gives six logically possible patterns: SOV, SVO, VSO, OSV, OVS, VOS - see section 3.1);
- (ii) establishing correlations between ordering patterns (e.g. between the order of verb and object on the one hand and between noun and relative clause on the other; see section 3.2).

Section 3.3 is concerned with some problems in traditional constituent order typology.

3.1. Classification of languages in terms of ‘basic constituent order’

In traditional constituent order typology, it was initially assumed that V, S and O are labels for categories which are relevant for all languages and that each language uses one of the six logically possible ordering patterns of V, S and O as the ‘basic word order’. Based on a large-scale investigation, Tomlin (1986) proposed that some 45% of the world’s languages have SOV as the basic order and that in 42% SVO and in 9% VSO is the basic order (for discussion, see Dryer 2005 and Song 2012). Only a very small percentage of the world’s languages have Object before Subject as the basic order: 3% of the languages are deemed to employ VOS order and 1% OVS order. There were no OSV languages in Tomlin’s sample, but the Amazonian language Nadëb, for instance, has since been claimed to have OSV as the unmarked order (Weir, 1994: 292).

In spite of these results, the classification of languages in terms of the basic order of V, S and O can be problematic, especially from a cross-linguistic perspective (Song, 2012: 14-15). For example, the notion ‘subject’ is deemed irrelevant for ‘topic prominent’ languages, languages without an active-passive opposition and languages with an ergative-absolutive alignment system (Dik, 1997: 247-89). It has also been claimed that there are languages without a distinct class of verbs (Rijkhoff and van Lier eds., 2013).

Furthermore, ‘basic order’ is commonly defined as “the order that occurs in stylistically neutral, independent, indicative clauses with full noun phrase (NP) participants, where the subject is definite, agentive and human, the object is a definite semantic patient, and the verb represents an action, not a state or an event” (Siewierska 1988: 8). However, clauses in which both arguments of a transitive verb are expressed as definite lexical phrases (as in ‘*The dog chased the cat*’) are very infrequent in natural language, especially in the spoken variety (Siewierska, 1988: 8-14).

3.2. Constituent order correlations

The first serious investigation of correlations between constituent order patterns is undoubtedly Greenberg's (1963) seminal cross-linguistic survey, which led to the formulation of implicational universals. Greenberg classified languages on the basis of the unmarked order of nominal subject (S), verb (V) and nominal object (O), which resulted in three groups: languages with basic order VSO, SVO or SOV. This is shown in Table 1, which is based on Hawkins's (1983) more recent study. Notice that here 'V-1' ('Verb-initial') is used, rather than 'VSO' as in Greenberg's original publication (A = adjective, G = genitive/possessor phrase).

1. V-1 & N—	38	5. SVO & N—	56	9. SOV & N—	21
2. V-1 & ANG	13	6. SVO & ANG	17	10. SOV & ANG	0
3. V-1 & —N	2	7. SVO & —N	19	11. SOV & —N	98
4. V-1 & GNA	<u>0</u>	8. SVO & GNA	<u>17</u>	12. SOV & GNA	<u>55</u>
	53 lgs.		109 lgs.		174 lgs.

Table 1. Classification of languages (based on Hawkins 1983; '—' signals that both A and G precede or follow the head noun N).

Table 1 shows, among other things, that most SOV languages have both the adjective and the genitive preceding the noun (type 11), whereas most V-1 and SVO languages have these modifiers following the noun (types 1 and 5).

Greenberg (1963) also formulated 45 'universals of grammar with particular reference to the order of meaningful elements'; for example:

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.

Universal 17. With overwhelmingly more than chance frequency, languages with dominant order VSO have the adjective after the noun.

Universal 18. When the descriptive adjective precedes the noun, the demonstrative and the numeral, with overwhelmingly more than chance frequency, do likewise.

In the same article Greenberg (1963: 96-104) offered tentative explanations for some of his findings, suggesting that several ordering sequences could be accounted for in terms of two competing motivations: Dominance and Harmony. For example, a language that has the head-initial pairs VS – VO – NA – NG (where V and N are the head of the clause and the NP, respectively) is 'harmonic' and the same goes for languages with the head-final pairs SV – OV – AN – GN. The notion 'dominance' can be exemplified with Hawkins's Universal VI (see (9) below), as summarized in Table 2 (Rel = relative clause).

	NG	GN
NRel	+	+
RelN	-	+

Table 2: Schematic representation of Universal VI (Hawkins 1983: 83)

Since GN order occurs with both NRel and RelN order, GN is regarded as the dominant order (Tigre, mainly spoken in Eritrea, seems to be an exception with RelN and NG order).

Greenberg added, however, that his theory was far from complete and emphasized that disharmonic and non-dominant (or: ‘recessive’) patterns should not be ignored.

Subsequently Lehmann (1973) and Vennemann (1975) offered more radical accounts for Greenberg’s findings, essentially reducing his three-way typology (VSO, SVO, SOV) to two ‘word order types’ (see Song 2012: 17-21, 29-54) for detailed discussion of two-way ordering typologies). Lehmann’s *Fundamental Principle of Placement* suggested it is possible to predict certain ordering pairs (such as the order of adjective and noun), if one knows that the language has OV or VO as the basic order. According to Vennemann’s *Principle of Natural Serialization* constituents are either Operators or Operands, which tend to be serialized either with Operator before Operand or with Operand before Operator. Examples of Operators would be ‘object’ and ‘adverbial’, both of which have ‘verb’ as their Operand. Even though there are many languages with ordering patterns that deviate from Lehmann’s or Vennemann’s ideal types, such languages were largely ignored as they were deemed to be in the process of changing from one type to another due to internal development or contact.

Greenberg’s original tripartite typology (Table 1) was restored in Hawkins’s (1983) monograph on constituent order universals. Using a sample containing over 300 languages, Hawkins formulated some new, often exceptionless universals (but see Dryer (1992) and Rijkhoff (2004: 227) for counterexamples). For instance (Hawkins 1983: 64, 83):

- (8) If a language has OV order, then if the adjective precedes the noun, the genitive precedes the noun; i.e. $OV \supset (AN \supset GN)$.
- (9) If a language has noun before genitive, then it has noun before relative clause; i.e., $NG \supset NRel$ (equivalently: $RelN \supset GN$).

He then demonstrated that various implicational universals could be collapsed into statistical implications with a preposition or postposition as ultimate antecedent (Hawkins 1983: 75, 83):

(10) Prepositional Noun Modifier Hierarchy:

$Prep \supset ((NDem \cup NNum \supset NA) \& (NA \supset NG) \& (NG \supset NRel))$

(11) Postpositional Noun Modifier Hierarchy:

$Postp \supset ((AN \cup RelN \supset DemN \& NumN) \& (DemN \cup NumN \supset GN))$

He also attempted to account in a more principled way for the many languages with ordering patterns that do not quite fit the ‘ideal’ two-way classification, which had been the focus of Lehmann’s and Vennemann’s proposals. For example, the *Heaviness Serialization Principle* (Hawkins, 1983: 90-91) deals with the fact that many languages have adnominal modifiers on both sides of the head noun (section 3.3.3).²

Almost a decade later Dryer (1992) used an even bigger sample (625 languages) in an attempt to investigate which constituent order pairs actually correlate with the order of object and verb (OV/VO), because assertions about such correlations had never been substantiated by systematic research (but cf. Perkins 2001: 432). Dryer concluded there are indeed several constituent order pairs that more or less correlate with the OV/VO distinction (Table 3), among them the pair HEAD NOUN - POSSESSOR NP (NG/GN) and the pair HEAD NOUN - RELATIVE CLAUSE (NRel/RelN). However, he did not find evidence for a correlation between

² Cf. also the *Mobility Principle* (Hawkins, 1983: 93-97).

OV/VO order and, for instance, the order of noun and adjective (AN/NA) or the order of noun and demonstrative (demN/Ndem).

<u>VERB PATTERNS</u>	<u>OBJECT PATTERNS</u>	<u>EXAMPLE</u>
verb	manner adverb	<i>ran + slowly</i>
adposition	Noun Phrase [NP]	<i>on + the table</i>
copula verb	predicate	<i>is + a teacher</i>
‘want’	Verb Phrase [VP]	<i>wants + to see Mary</i>
noun [N]	genitive [G]	<i>father + of John</i>
noun [N]	relative clause [Rel]	<i>movies + that we saw</i>
adjective [A]	standard of comparison	<i>taller + than Bob</i>

Table 3. Some of Dryer’s correlation pairs (Dryer, 1992: 108; see also Dryer 2009)

Thus, we tend to find NG and NRel order in VO languages and GN and RelN order in OV languages. To account for the correlation pairs, Dryer proposed the *Branching Direction Theory*, according to which there is tendency for phrasal categories (such as G and Rel) to either precede or follow a member of a non-phrasal category (such as N, A and other ‘verb patterners’). The theory makes no ordering predictions about non-branching categories like demonstratives or numerals. However, the theory failed to account for some correlation pairs, such as the order of verb and manner adverb (both of which are non-phrasal). Another problem of the *Branching Direction Theory* concerns the categories that are used to make the cross-linguistic comparison (see Rijkhoff (2004: 278-291; Rijkhoff 2009) and Song (2012: 29-54) for discussion). This is discussed in sections 3.3.2 and 3.3.3.

3.3. Some problems in traditional constituent order typology

From a cross-linguistic or typological perspective, the notion ‘basic constituent order’ is problematic for various reasons, casting some doubt on the assumption that all languages have a basic order. Here we will focus on two problems: (i) the establishment of a basic constituent order (3.3.1), and (ii) the employment of semantic categories to identify constituents for comparative (i.e. cross-linguistic) syntactic purposes (3.3.2). Section 3.3.3 deals with the problem of categorial confusion and section 3.3.4 shows that both meaning-based and form-based categories are relevant in the area of constituent order research, provided they are kept separate from each other.

3.3.1. The notion ‘basic constituent order’

Even if it would always be possible to find for each language a representative number of stylistically neutral (active, positive), independent, indicative clauses that contain a transitive verb and lexical subject and object arguments (section 2), there would still be quite a few languages that statistically speaking display no clear preference for just one of the six possible orderings of V, S and O. To give some examples: in Dutch, where the finite (main or auxiliary/Aux) verb normally appears after the subject in the main clause (SV_{finite} ...), the basic order would be SVO in the absence of an auxiliary verb (*De kat ving the vogel* ‘The cat caught the bird’), but SAuxOV if the clause contains a finite auxiliary verb, as in: *De kat heeft_{Aux} de vogel gevangen_V* (lit. ‘The cat has the bird caught’) ‘The cat has caught the bird’. Notice also that in subordinate clauses the whole verbal cluster appears after S and O (SOAuxV or SOVAux), as in *Ik hoorde dat de kat de vogel heeft_{Aux} gevangen_V / gevangen_V heeft_{Aux}* (lit. ‘I heard that the cat the bird has caught / caught has’).

To give another example, Ket (a language isolate spoken in Siberia) is typically classified as having two ‘basic’ orders (SVO/SOV). Furthermore, since this language makes

considerable use of incorporation, many transitive verbs appear without lexical subject or object arguments. Constituent order is also claimed to be remarkably free in Yimas, “even in comparison to the relative freedom exhibited by other Papuan languages. Within the clause, there is no set word order pattern at all (in fact, the majority of clauses consist of just a verb, so word order is not much of a question in these cases anyway). [...] Word order in Yimas is so free that NPs need not even form a constituent” (Foley, 1991: 4).

These remarks also highlight another problem touched upon earlier: that it is rather unusual to find a transitive verb with two full (i.e. lexical) subject and object arguments where the subject is definite, agentive and human, the object is a definite semantic patient (Siewierska, 1988: 8), particularly in normal spoken discourse. As was already mentioned in section 2, in polysynthetic languages (but also in other languages) the arguments of a transitive verb are often not expressed as free constituents, but by clitics or affixes, if at all. These problems could indicate that in quite a few languages basic constituent order has been established on the basis of an elicitation procedure, i.e. at the explicit request of the linguistic field worker who provides the native speaker informant with a constructed example, which is, of course, far from ideal.

3.3.2. *The employment of semantic criteria in constituent order research*

In a cross-linguistic investigation of constituent order patterns, it is important that the constituents involved are sufficiently similar to be identified and compared across languages (Song, 2012: 10-15). Due to the enormous degree of morphological and syntactic diversity that is attested in the world’s languages, form-based definitions of constituents have generally been avoided in constituent order studies. For example, to say that a noun is a word that is inflected for number is quite irrelevant for the many languages across the globe in which number marking is absent.

Because formal categories are on the whole too language specific, typologists have usually followed Greenberg (1963: 74), who applied semantic criteria to identify the various kinds of constituents in his cross-linguistic investigation of ordering patterns. Thus, his semantic category Adjective essentially covers all the forms and constructions that translate as an adjective in English, even if a language does not have a dedicated class of adjectives, but instead uses abstract nouns such as ‘bigness_N’ (as in certain Chadic and Bantu languages), or stative verbs like ‘be.big_V’ (as in e.g. many SE Asian languages) to express adjectival notions (section 2.3.4).

Two decades after Greenberg’s publication, Hawkins (1983: 12) also claimed that semantic criteria ‘suffice to make the cross-linguistic equation’ and when Dryer (1992: 120) investigated ‘Greenbergian word order correlations’, he also identified ‘different categories largely on the basis of semantic criteria’ (more on this in section 3.3.3). Thus, their studies suggested that, by employing semantically defined categories, the same constituent ordering patterns (e.g. the order of adjective and noun) can be investigated in all the world’s languages, even when it is clear from the grammatical description of a language that the constituents involved belong to different form classes, whose members can have their own, form-based ordering preferences (see sections 3.3.4 and 4.3). In other words, whereas formally defined categories are usually too narrow for cross-linguistic research into constituent order, semantically defined categories can be too wide: they may include forms or constructions that are too varied to allow for cross-linguistic comparison of constituent ordering patterns. In fact, this was already acknowledged in Greenberg’s (1963) investigation of language universals. In his discussion of comparative constructions (as in English ‘*Mary is TALLER THAN BILL*’), he found that the semantic category Comparison is expressed rather differently in languages across the globe. As later shown by Stassen (1984), there are at least five major construction types to express the semantic category of Comparison. For example, whereas

English uses an adjective followed by the standard introduced by the marker *than* (as in the example above: Adjective/*taller* – Marker/*than* – Standard/*Bill*), Duala (a Bantu language from Cameroon) employs two predicates: one expressing the adjectival notion (here *kolo* ‘big’), the other is a verb meaning ‘to exceed’ or ‘to surpass’ (Stassen 1984: 151).

Duala

(12) *Nin ndabo e kolo buka nine.*
this house it big exceed that
‘This house is bigger than that.’

Even though the ‘exceed construction’ constitutes the second largest type in Stassen’s 110-language sample, Greenberg writes that that he could only formulate *Universal 22* if he left this construction out of consideration.³

“A number of languages are not entered in [Table 8 on the same page - JR] because they utilize a verb with a general meaning ‘to surpass’. This is particularly common in Africa (e.g., Yoruba): ‘X is large, surpasses Y’. Loritja, an Australian language which has ‘X is large, Y is small’, is likewise not entered.” (Greenberg 1963: 88)

Two further problems concerning the employment of so-called semantic (or ‘conceptual’) criteria in traditional constituent order research are, firstly, that the semantic criteria are not explicitly listed and, secondly, that in some cases constituents are essentially identified and compared on the basis of functional rather than semantic criteria.⁴ Thus, the reason why adjectives and relative clauses can sometimes be grouped together is not because they belong to the same semantic category (see sections 3.3.4 and Table 4 below on genuine semantic features like ‘±animate’ and categories like ‘color’), but rather because members of both formal categories can serve as a qualifying modifier in the NP, specifying a more or less inherent property of an entity, such as its age, weight or size (section 4).

3.3.3. *Categorial confusion: mixing ‘semantic’ and formal categories*

In spite of the problems concerning the application of ‘semantic’ criteria in traditional constituent order research (section 3.3.2), these criteria continue to be used in syntactic typology, including large-scale projects such as *The World Atlas of Language Structures* (Rijkhoff 2009a). Thus, Dryer essentially treats the translational equivalents of English adjectives in all languages as members of the ‘semantic’ category Adjective, but admits that ‘there are many languages in which what I call adjectives are really verbs, and ‘adjectives’ modifying nouns are really just a kind of relative clause’ (Dryer, 1992: 196 fn. 12). However, Dryer’s correlation pairs also include Relative Clause (Rel) as an independent category (Table 3). If relative clauses can belong to two ‘semantic’ categories (Adjective and Relative Clause), one cannot properly interpret the *Branching Direction Theory* (section 3.2), where adjectives and relative clauses are treated as members of distinct, non-overlapping formal categories. Relative clauses are branching or phrasal categories and ‘object patterners’; adjectives are non-branching categories and the order of

³ ‘*Universal 22*. If in comparisons of superiority the only order or one of the alternative orders is standard-marker-adjective, then the language is postpositional. With overwhelmingly more than chance frequency, if the only order is adjective-marker-standard, the language is prepositional’ (Greenberg. 1966: 111).

⁴ For this reason, the notion ‘semantic’ as employed in traditional constituent order research will henceforth appear in single quote marks. Notice also that semantic categories are not always distinguished from conceptual categories (especially in cognitive linguistics), but there are good reasons why the two categories should not be conflated.

Adjective and Noun does not correlate with the order of Object and Verb. Whereas the correlations pairs in Table 3 are presumably all ‘semantic’ categories (Dryer, 1992: 120), the two categories referred to in the *Branching Direction Theory* (viz. phrasal/branching and non-phrasal/non-branching categories) are unambiguously syntactic (i.e. formal) categories. In sum, it remains unclear what happened with the phrasal or branching members of the ‘semantic’ category Adjective in the *Branching Direction Theory*.

The same confusion between ‘semantic’ and formal categories is found in Hawkins (1983), where Greenbergian ‘semantic’ categories (which may contain members of different form classes) are used to identify constituents and compare ordering patterns in many different languages (Hawkins, 1983: 9-12). However, some of the same category labels (Adjective, Genitive, Relative Clause) are also used when constituents are discussed as part of the Heaviness Hierarchy, which is based on formal criteria (Hawkins, 1983: 90-91):

Rel \geq_R Gen \geq_R A \geq_R Dem/Num

“where \geq_R means: ‘exhibits more or equal rightward positioning relative to the head noun across languages’. That is heavier noun modifiers occur to the right.”

One constituent is heavier than another, if one or more of the following factors imposes an ordering between them: (i) length and quantity of morphemes, (ii) quantity of words, (iii) syntactic depth of branching nodes, (iv) inclusion of dominated constituents. The different degrees of ‘heaviness’ would explain why relative clauses, which are comparatively ‘heavy’, are more likely to appear after the head noun in prepositional languages than adjectives, which are relatively ‘light’. In other words, names for ‘semantic’ categories (A, Rel etc.) are now used as labels for categories defined on the basis of formal criteria. Here, too, it remains unclear how verbal and nominal expressions of adjectival notions (see examples (6) and (7)) are categorized in terms of heaviness. Presumably the original members of the single semantic category Adjective are now distributed over several formal categories (including at least A and Rel).

We will see in section 4 that the confusion between ‘semantic’ and formal categories in constituent order studies is at least partly due to two problems: (i) the ‘semantic’ criteria used in traditional constituent order research are in fact sometimes functional criteria in that constituents are occasionally identified and compared on the basis of the function they serve in a linguistic expression (e.g. QUALIFYING MODIFIER), (ii) the same function (e.g. QUALIFYING MODIFIER) can be fulfilled by constituents that belong to different form classes (e.g. adjective, prepositional phrase, relative clause), and vice versa, that the same constituent (e.g. relative clause) can have different functions in actual discourse.

3.3.4. *The relevance of meaning and form-based categories for constituent order research*

Both semantic and formal criteria are needed to account for the way constituents are linearized in a linguistic expression, but it is very important that they are kept separate. For example, the semantic features ‘±definite’ or ‘±animate’ (often in tandem with other meaning-based or pragmatic notions like ‘prominence’, ‘topicality’ or ‘salience’) can have various effects on the grammar of a language (Dik, 1997: 34-41), including the way constituents are linearized. Thus, in some languages the animacy hierarchy determines the order of the two object phrases in clauses with a di-transitive verb (Heine and König, 2012). The following examples are from Kinnauri (Nepal). In (13a) both objects (child, maternal aunt) refer to animate entities, in which case the direct object precedes the indirect object. In (13b), however, the order of the objects is determined by animacy: the animate entity (uncle) precedes the inanimate entity (letter) in the unmarked order (Kittilä, 2006: 2).

Kinnauri

- (13) a. *amas anu chan-u ane-pən rano*
mother:ERG she:REFL child-ACC maternal aunt:ACC send
'The mother sent her son to the maternal aunt'
- b. *gəs än bapu-pən cithi cemigduk*
1SG:ERG 1SG:GEN uncle-ACC letter write:S
'I have to write a letter to my uncle'

Alamblak (Papua New Guinea) provides us with another example of the relevance of semantics for constituent order. In this language, color adjectives must precede the noun, whereas adjectives belonging to other semantic categories can occur on either side of the head, although some (e.g. adjectives denoting a physical property) tend to precede the noun and others (e.g. those denoting 'value') usually follow the noun. To give another example, in Maybrat (Bird's Head Peninsula, New Guinea) the genitive precedes the noun (GN) in the case of inalienable possession, but follows it (NG) in the case of alienable possession.

Formal factors clearly play a role when, for example, constituent order is determined by the way a constituent is formally expressed, which in its turn is connected with the internal complexity of a constituent (see above on 'morphosyntactic weight' or 'heaviness'). Whether a constituent takes the form of a single word (adjective, adverb), a phrase (NP, PP) or a clause (relative clause) can (co-)determine its position in a linguistic expression (see section 5.1 on Domain Integrity). The role of structural complexity was already captured in 1932 by Behaghel's *Gesetz der wachsenden Glieder* ('law of growing members'), according to which constituents are arranged in their order of relative increasing complexity (cf. 'heavy shift' in transformational grammar and the Principle of Increasing Complexity in Dik's Functional Grammar). Compare

- (14) a. ?I gave *the paintings that my uncle left to me as part of his inheritance* to her.
b. I gave to her *the paintings that my uncle left to me as part of his inheritance*.

The version with the long constituent in clause-final position (14b) is preferred over the one with the short constituent ('to her') at the end (14a). The same holds for heavy constituents in the NP: the short *of*-phrase in (15a) can appear immediately after the head noun, but the preferred position of the longer *of*-phrase is at the end of the NP, as in (15b).

- (15) a. the painting *of him* that she found
b. the painting that she found *of her grandfather wearing that strange jacket*

In sum, it is necessary to distinguish between meaning-based and form-based categories, as both play their own distinct role in structuring discourse through constituent order. For a more complete and satisfactory account of constituent order, however, the discourse function of a constituent in actual speech also needs to be taken into consideration; this is discussed in section 4.

4. Form and function in constituent order

An important reason why 'semantic' and formal categories have been mixed up in some constituent order studies is that these studies did not distinguish between semantics and pragmatics, i.e. between the coded meaning of a constituent and the discourse function of a constituent in an actual utterance (section 3.3.3). This means that functional categorization is

essentially context-dependent (see e.g. section 1 about ‘Subject-before-Object’). Furthermore, not all constituent order studies took into account that there is often no direct correspondence between the form and the function of a constituent. The same constituent (e.g. a relative clause) can have different functions in interpersonal communication (section 4.2), and vice versa, constituents belonging to different formal categories (e.g. adjectives and relative clauses) can be used in the same function (section 4.3). In either case we see that the discourse function of a constituent in actual speech may put certain constraints on its ordering possibilities.

4.1. Functional modifier categories

The functional, instrumental approach to categorization can be said to have its roots in the Prague School of Linguistics, which has greatly influenced current functional theories of grammar. Functional categorization is not directly concerned with the formal or semantic properties of a constituent, but rather with the actual job of a linguistic form or construction in the process of verbal communication. Here we will only deal with some FUNCTIONAL MODIFIER CATEGORIES that are relevant for both NPs and clauses (for details, see Rijkhoff, 2008a, 2008c). They are briefly characterized in (16a-e), where they are listed according to scope increase. Thus, a CLASSIFYING MODIFIER such as ‘presidential’ in ‘the presidential election’ only has the head constituent (‘election’) in its scope, whereas a DISCOURSE-REFERENTIAL MODIFIER like a definite or indefinite article has all other constituents of an NP in its scope (see also section 4.3 on the Scope Principle).

(16) Five functional modifier categories in NPs and clauses (with examples of English adnominal modifiers)

- a. CLASSIFYING MODIFIERS further specify what KIND of entity (subcategory) is denoted by the head constituent (as in e.g. ‘SOLAR energy’, ‘a VIRAL infection’, ‘a man OF THE STAGE’);
- b. QUALIFYING MODIFIERS specify more or less inherent properties (‘qualities’) of an entity (e.g. ‘a BLUE hat’, ‘children UNDER AGE 16’, ‘a house THAT WAS RATHER BIG’);
- c. QUANTIFYING MODIFIERS specify quantitative properties (quantity, number, cardinality) of an entity (e.g. ‘TWO cars’, ‘SEVERAL books’);
- d. LOCALIZING or ANCHORING MODIFIERS specify locative properties of an entity in spatial, temporal or cognitive dimension. By ‘anchoring’ an entity in the world of discourse, the speaker makes the referent identifiable for the hearer (e.g. ‘THIS car’, ‘the house ON THE CORNER’, ‘That’s the man WHO STOLE MY BIKE’);
- e. DISCOURSE-REFERENTIAL MODIFIERS specify discourse-referential (pragmatic) properties of an entity (e.g. ‘THE/A car’).

Since there is often no one-to-one correspondence between the form and the function of a constituent, the same constituent can have different functions in actual speech. Thus, an adnominal demonstrative is a LOCALIZING MODIFIER in its deictic function, but a DISCOURSE-REFERENTIAL MODIFIER when it merely marks definiteness (section 4.2). Similarly, a relative clause is a QUALIFYING MODIFIER when it serves to express a more or less inherent property (example (6b), but a LOCALIZING/ANCHORING MODIFIER when it is used to make the referent of an NP identifiable (as in example (16d)). Conversely, constituents that belong to different formal categories (e.g. adjectives and relative clauses) can be used in the same function (e.g. QUALIFYING MODIFIER - see Table 4 below). The many-to-many correspondence between the form and the function of constituents and its relevance for constituent order is discussed in sections 4.2 and 4.3.

4.2. Same constituent, different functions

Membership of formal categories such as Adjective, Relative Clause or Complement Clause is generally not a very reliable indicator of the discourse function of a constituent in actual speech. For example, the adjective ‘civil’ is a CLASSIFYING MODIFIER in ‘civil rights’ (implying e.g. that it cannot be modified by a degree adverb: *very civil rights), whereas ‘civil’ in ‘a civil reply’ serves as a QUALIFYING MODIFIER (which means it can be modified by a degree adverb: ‘a very civil reply’) (Halliday, 2004: 320-322; Bisetto, 2010).

There are many examples of constituents whose actual function in an utterance is marked by constituent order. Berbice Dutch Creole is one of the languages in which the same constituent can be used as a definite article or as an adnominal demonstrative modifier, but as an article it precedes the noun and when it has a deictic (localizing) function it follows the noun. When a numeral appears before the noun in Russian (num N) it expresses cardinality (*dvadcat’ knig* ‘twenty books’), but when the order is reversed (N num) the numeral has an approximative meaning (*knig dvadcat’* ‘approximately twenty books’). In Turkish and many other languages the same constituent can occur as an indefinite article or as the numeral ‘one’ (in Turkish *bir*). The position of *bir* in the NP is determined by its actual modifier function. Similar things can be said about adnominal possessor phrases or genitives and relative clauses. For example, in the Nepalese language Kham a qualifying relative clause appears immediately before the noun, whereas a localizing relative clause normally precedes the numeral and clusters with other localizing modifiers in the order (see section 4.1 on functional modifier categories): [dem Rel num Rel N].

4.3. Different constituents, same function

The fact that the same function can be fulfilled by constituents that belong to different form classes is seen in Table 4, which shows that English adjectives, prepositional phrases and relative clauses can all be used as a descriptive or QUALIFYING MODIFIER, which specifies a more or less inherent property of an entity.

Functional Category: QUALIFYING MODIFIER (in the noun phrase)			
Semantic Categories	Formal Categories		
	ADJECTIVE	PREP. PHRASE	RELATIVE CLAUSE
SIZE	<i>a big N</i>	<i>a N of enormous size</i>	<i>a N that is rather big</i>
VALUE / QUALITY	<i>an expensive N</i>	<i>a N of great value</i>	<i>a N that is pretty cheap</i>
AGE	<i>a young N</i>	<i>a N under age 16</i>	<i>a N that is much too young</i>
COLOR	<i>a black N</i>	<i>a N of blackness</i>	<i>a N that is pitch black</i>

Table 4. The same function (QUALIFYING MODIFIER) expressed by members of different formal categories (Adjective, Prepositional Phrase, Relative Clause)

According to the Scope Principle (section 5.3), constituents that belong to the same functional modifier category occupy the same position before or after the head constituent. This was shown to be the case in the Nepalese language Kham (section 4.2), where qualifying relative clauses (but not localizing relative clauses) occupy the same position as true adjectives in other languages (cf. the ordering patterns in (31)).

When a language has alternative forms of expression for the same function (Table 4), the choice is often determined by linguistic genre, pragmatic considerations or internal complexity (section 3.3.4). This can be seen in (18), where the ‘heavy’ postnominal qualifying relative clause contains coordinated constituents (both examples from the Internet).

(17) *I am looking for a RATHER BIG_A house*

(18) *Indeed there was a house THAT WAS RATHER BIG AND VERY PRETTY*_{REL}.

5. General tendencies in constituent ordering

Globally observed constituent ordering tendencies are generally regarded as manifestations of iconicity: non-arbitrary relations between ‘form’ and ‘content’ (Haiman ed., 1985). Iconicity in grammar concerns any formal manifestation of the (largely unconscious) effort to maintain a close correspondence between, on the one hand, non-linear, multidimensional MENTAL representations of entities (as well as relations between these entities) and, on the other hand, the way these representations are mapped onto a linear, one-dimensional structure in speech. Constituent order is considered iconic whenever it reflects aspects of the conceptual or semantic content, which typically manifests itself as ‘iconicity of distance’, where conceptual or semantic distance matches with spatial distance. Especially among functional linguists, it is widely assumed that there is direct causal connection between iconicity in grammar and increased communicative efficiency.

Under the narrow view of iconicity, the notion refers to the tendency of linguistic expressions to mirror the temporal or logical sequential order of events. This can be seen in (19a-b), where the coordinated clauses in the a-sentence (but not in the b-sentence) reflect the actual order of events.

- (19) a. They met in Konstanz and married in Amsterdam. (unmarked)
b. ?They married in Amsterdam and met in Konstanz. (marked)

Another instance of iconicity in the narrow sense is the order of clauses in conditional sentences (e.g. ‘If he comes, I’ll leave’), where the condition (*protasis*) precedes the consequence (*apodosis*) as the unmarked order in all languages (Universal 14 – Greenberg, 1963: 84).

In its broad sense, iconicity also refers to any correspondence between linguistic expressions and “generally accepted perceptions of dominance and salience and/or semantic dependencies” (Siewierska, 2006: 646). Defined this way, iconicity accounts for several of the general ordering tendencies mentioned earlier, such as the strong universal preference for Subject arguments to precede Object arguments (subjects tend to be agentive, definite, more salient and higher on the person and animacy hierarchy than objects) and even the tendency to place long and complex constituents after relatively short, less complex constituents. The use of informationally and structurally ‘light’ constituents correlates with topicality and high identifiability: placing light constituents before heavy constituents would iconically reflect the preferred way of processing information.

Behagel’s famous first law (1932: 4) already captured the insight that constituent order is often an iconic reflection of the way we perceive or conceptualize entities or relations between entities to facilitate language processing.

(20) Behagel’s First Law

Das oberste Gesetz ist dieses, daß das geistig eng Zusammengehörige auch eng zusammengestellt wird (‘The principal law is this: that what belongs together mentally is also placed close together’).

It is possible to formulate more precise iconic sub-principles on the basis of Behagel’s first law: (i) the Principle of Domain Integrity, (ii) the Principle of Head Proximity and (iii) the Principle of Scope (Rijkhoff, 2004, 2008b). These principles are briefly discussed below, mostly illustrated with examples involving NP-internal constituents, as the effects of the various principles are better seen at the level of the phrase than in the larger clausal domain.

5.1 Domain Integrity

The Principle of Domain Integrity (PDI) can be formulated as follows:

- (21) Constituents prefer to remain within the boundaries of their proper domain; embedded domains (subordinate clauses or phrasal modifiers such as adnominal possessive NPs) prefer to be in the periphery of their embedding domain (e.g. main clause, argument NP), so as to avoid possible discontinuity among other constituents of the embedding domain.

The notion ‘domain’ stands for an integral, hierarchically organized linguistic unit and can be regarded as the functional equivalent of ‘constituency’ in syntactocentric theories of grammar. The PDI is perhaps the most obvious interpretation of Behaghel’s first law and accounts for the fact that discontinuity is a marked phenomenon in languages across the globe. This principle sums up ideas that have been proposed in various forms, namely that constituents belonging to the same conceptual or semantic unit (for instance, because they help to describe or refer to the same object or event) are processed faster and easier when they appear in a continuous, uninterrupted sequence. Some languages, such as the western Austronesian languages, do not allow discontinuous phrases at all, but when a language does allow a constituent to appear outside the boundaries of its proper linguistic domain, this is commonly due to some pragmatic factor (e.g. contrastive focus) or complexity (‘heaviness’). For example, in Ngiti (Zaire) highly emphasized numerals may appear in clause-initial position (Kutsch Lojenga, 1994: 355).

Ngiti

- (22) [*árɛ*]_{num} *ma m-ɔ̀kèrè* [*itsu*]_{NP}
eight 1.SG SC-cut:PERF.PRES tree
‘I have cut EIGHT trees’

In Dutch ‘heavy’ adnominal modifiers, such as relative clauses, may appear in clause-final position:

Dutch

- (23) *Ik heb net [de man]_{NP} gezien [die gisteren mijn fiets gestolen heeft]_{Rel}*
I have just the man seen who yesterday my bike stolen has
‘I just saw the man who stole my bike yesterday’

Since appositional constituents are not part of an integral domain, it is less unusual for them to appear discontinuously (section 1). It appears that loose, appositional relationships between constituents that together form a conceptual or semantic unit historically precede the employment of hierarchically organized linguistic domains (Van de Velde 2009).

5.2 Head Proximity

The Principle of Head Proximity (PHP) basically states that languages prefer to have a short distance between lexical heads of hierarchically ordered domains (e.g. between head ‘N’ of subordinate domain ‘NP’ and head ‘V’ of superordinate domain ‘clause’).

- (24) Constituent ordering rules conspire to keep the heads of different, hierarchically ordered domains as closely together as possible.

This general ordering principle is based on the processing strategy which basically holds that domain identification (e.g. of a clause or NP) plays a central role in language processing. One of the most reliable strategies to identify a linguistic domain is to find the constituent that is the head of the domain, because the head is typically (i) the only obligatorily constituent in a domain and (ii) the most informative constituent of a domain. In languages that have distinct classes of verbs and nouns, verbs typically serve as the head of a clausal domain and nouns as the head of a phrase (incl. prepositional phrase; notice that adpositions are not regarded as heads in e.g. Functional (Discourse) Grammar).

The PHP could account for several ordering phenomena (Rijkhoff, 2004: 261-312), including some of the constituent order correlations observed in Greenberg (1963) and Dryer (1992), such as (notice that, due to the pragmatic status of Subject, SVO languages are comparable to V-initial languages regarding head proximity; see Rijkhoff (1986: 1-12) for details):

- the tendency to avoid embedded domains (like relative clauses or genitives) appearing between the head of the clause V and the head of the phrase N (cf. Table 1 and Dryer's *Branching Direction Theory*; dots symbolize adnominal modifiers);

(25) a. V [N....]_{NP} (verb-initial)
 b. [...N]_{NP} V (verb-final)

- Greenberg's Universal 16 (see section 3.2.):

(26) a. Aux V (verb-initial)
 b. V Aux (verb-final)

- Behaghel's (1932) *Gesetz der wachsenden Glieder* ('law of growing members'), according to which long constituents in the clause appear at a greater distance from the head of the clause V than short constituents, which would result in a better overall ratings in terms of 'head proximity'.

(27) Aux V [...]_{NP} [....]_{NP} (verb-initial)

Notice, however, that in terms of head proximity the long-before-short order is preferred in verb-final languages:

(28) [....]_{NP} [...]_{NP} V Aux (verb-final)

While it is true that some verb-final languages have a preference for 'long-before-short' (Hawkins, 1994: 152), which is preferred in terms of head proximity, the structure in (28) also shows that in verb-final languages the PHP is in conflict with Behaghel's *Gesetz der wachsenden Glieder*, which predicts a general preference for short-before-long. This fundamental conflict between the two ordering principles could explain why so many verb-final languages are also characterized as 'free word order' languages, as there is often no clear statistical preference for a 'basic word order' in these languages. This conflict may also be the reason why some verb-final languages allow constituents to appear after the verb (also known as 'leaking').

The constituent orders exemplified in (27) and (28), i.e. short-before-long in verb-initial and SVO languages and long-before-short in verb-final languages are also predicted as the preferred patterns by Hawkins's *Principle of Early Immediate Constituents*, which says that

structures with shorter recognition domains are easier to process (a detailed comparison between the *Principle of Head Proximity* (PHP) the *Principle of Early Immediate Constituents* and the *Branching Direction Theory* can be found in Rijkhoff (2004: 261-312); see also Song 2009 and Song 2012: 234-303). Hawkins's *Principle of Early Immediate Constituents* is similar in spirit to the PHP with regard to ordering preferences among short and long constituents in a sentence, but the PHP seems to cover more ground in that it could also account for certain seemingly disparate grammatical phenomena, in addition to the ordering tendencies mentioned above. For example, one could regard the two following phenomena as strategies that reduce the overall ratio of head_N-to-head_V proximity in a clause (Rijkhoff, 2004: 311-312):

- (i) keep the number of NPs in a clause to a minimum: the fewer the number of NPs in a clause, the better will be the overall head proximity ratio. This can be seen in Cayuga and other polysynthetic languages, where ordinary discourse can consist largely of verbs.
- (ii) increase the number of verbs in a clause: the more lexical verbs a clause contains, the shorter will be the overall distance between heads V and N. This is exemplified in the Papuan language Kobon: if a sentence would contain many arguments and adjuncts, speakers have a strong tendency to divide them over two or more clauses (which may contain an identical predicate) rather than including them all in one sentence.

5.3 Scope

The Principle of Scope (PoS) states that differences in scope among modifiers in a certain domain is iconically reflected in the actual linguistic expression:

- (29) Modifiers tend to occur next to the part of the expression that they have in their scope.

Scopal relationships between modifiers in a domain can be captured in meaning-based, layered representations of linguistic expressions as employed in, for example, Van Valin's (1993) *Role and Reference Grammar*, Dik's (1997) *Theory of Functional Grammar* and its successor *Functional Discourse Grammar* (Hengeveld and Mackenzie, 2008). In a layered representation it is clearly indicated which part ('layer') of a domain is within the scope of members of a certain functional modifier category. As was shown in (16), at least the following functional modifier categories can be used to analyze both NPs and clauses (Rijkhoff, 2008a, 2008c): (i) CLASSIFYING MODIFIERS, (ii) QUALIFYING MODIFIERS, (iii) QUANTIFYING MODIFIERS, (iv) LOCALIZING or ANCHORING MODIFIERS, and (v) DISCOURSE-REFERENTIAL MODIFIERS.

Modifier categories include both grammatical and lexical elements (cf. (16)), covering all the 'dependents' in a domain that are not arguments or complements. In a formal, meaning-based representation, these modifiers are distributed over 'nested' layers around the head constituent, reflecting differences in semantic scope. Conventionally, grammatical modifier categories (LOCALIZING/ANCHORING MODIFIERS such as demonstratives or tense markers) are represented on the left, whereas lexical modifiers (e.g. QUALIFYING MODIFIERS in the form of an adjective, a prepositional phrase or a relative clause; see Table 4) are represented on the right. The various layers are shown in a strongly simplified form in Figure 1 (notice that clauses have at least two more outer layers of modification to accommodate propositional and illocutionary modifiers; see Van de Velde 2012 on NP internal mood).

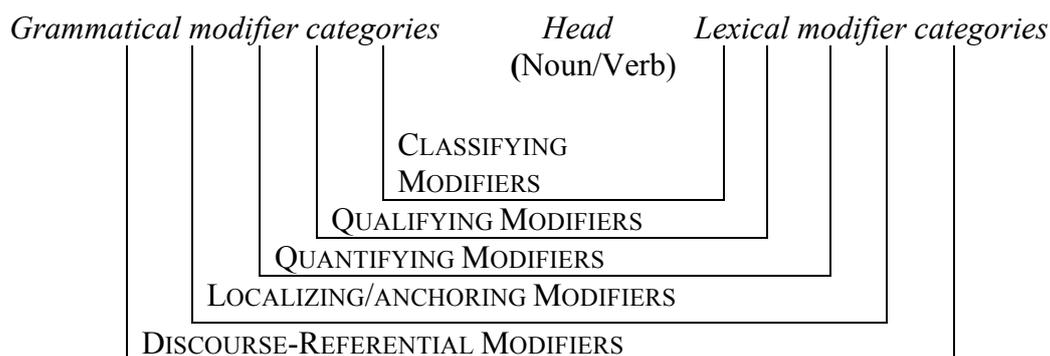


Figure 1. Some functional modifier categories in a layered representation of NP/clause structure reflecting differences in scope (Rijkhoff, 2008a, 2008c).

The iconic relationship between the distance of a functional modifier category relative to the head in the layered representation of a linguistic expression and constituent order is most clearly seen in the case of simple (non-phrasal) modifiers in the NP, because here the PoS does not compete with Domain Integrity (section 5.1), which says that the preferred position of an embedded domain (e.g. Genitive, Relative Clause) is in the periphery of its embedding domain.⁵ Furthermore, embedded domains tend to be more versatile in functional terms than non-embedded domains (section 4.2; Rijkhoff, 2009b). On the assumption that deictic adnominal demonstratives (dem) typically serve as LOCALIZING/ANCHORING MODIFIERS, cardinal numerals (num) as QUANTIFYING MODIFIERS and genuine, non-derived adjectives (A) as QUALIFYING MODIFIERS, there are 24 logically possible combinations in an NP where the noun is modified by a member of all three modifier categories (as in English ‘those three black boxes’). Only eight of these ordering patterns iconically reflect scopal relations among the three modifiers and it appears that each of these ‘iconic’ constituent orders is attested (Rijkhoff, 2008b: 802):

- | | | |
|------|-----------------------|------------------------------------|
| (30) | dem num A N (Dutch) | num A N dem (Berbice Dutch Creole) |
| | dem num N A (Spanish) | num N A dem (Hmong Njua) |
| | dem A N num (Zande) | A N num dem (Sango) |
| | dem N A num (Bambara) | N A num dem (Oromo) |

In each ordering pattern in (30) the adjective is adjacent with the noun and the demonstrative is in the periphery, mirroring the layered representation in Figure 1. It has been argued that claims about the occurrence of non-iconic patterns only constitute apparent counterexamples, because in such cases at least one of the following statements is true (Rijkhoff, 2004: 272–273; 2008b):

- (31)
- (a) constituents treated as semantic numerals or adjectives are formally speaking embedded domains headed by stative verbs or abstract nouns (see section 5.1 on Domain Integrity and heaviness);
 - (b) constituents are expressed as bound rather than free elements, which means their expression is a matter of morphology rather than syntax (section 2);
 - (c) constituents are in apposition (rather than fully integrated constituents; section 1).

⁵ Behaghel’s (1932) *Gesetz der wachsenden Glieder* (‘law of growing members’) is another competing ordering principle that is relevant here (see above).

- (d) constituents are assigned a special pragmatic function like Focus, indicating we are dealing with a marked pattern (see, for example, the marked position of adjectives in Hungarian NPs).

Recently Cinque (2005) has argued that 14 orders are attested: the eight iconic patterns listed in (31) plus the six non-iconic ones in (32). Interestingly, the non-iconic orderings only involve postnominal modifiers (cf. section 1 on apposition):

- | | | | |
|------|-------------|-------------|-------------|
| (32) | dem N num A | A N dem num | N dem A num |
| | N dem num A | N A dem num | N num A dem |

However, since Cinque (2005: 315) employed semantic criteria to identify constituents (section 2.3.2), he also included constituents that are characterized by one or more of the properties listed in (31). For example,

- in Arbore (Cinque: N dem num A) demonstratives are suffixes and numerals are regarded as a subset of nouns;
- in Sampur (Cinque: dem N num A) adjectives are basically verbs heading a relative construction.

In sum, when we exclude apparent counterexamples like those mentioned under (31), it remains to be seen how many non-iconic ordering patterns in (32) actually occur (in fact, there are just a handful of languages with non-iconic orders in Cinque's study).

5.4. Arbitrary patterning

Notice finally that even non-iconic or 'arbitrary' ordering patterns can aid in language processing. Any consistent form of linear organization can be used to structure discourse. For example, speakers of the Peruvian language Iquito mark the distinction between REALIS and IRREALIS clauses solely by means of word order constraints (Beier et al., 2011). It is easier to identify constituents and process the linguistic expression they are part of when constituents with the same functional specification occupy the same position in a linear structure. Thus, whether a language has OV or VO order is essentially arbitrary, but knowing that one's language normally has the object argument before the verb provides the language user with interpretation or decoding strategies. Similarly, whether a language has AN or NA order is essentially arbitrary, but knowing which order is used in a language facilitates information processing.

Abbreviations

1 = first person, 2 = second person, 3 = third person, A = adjective, ACC = accusative, ART = article, AUX = auxiliary verb, COMP = complementizer, DAT = dative, DT = dative-transformative, dem = demonstrative, ERG = ergative, FEM = feminine, G/GEN = genitive, possessor phrase, HABIT = habitual, N = noun, NC1 = noun class 1, NC7 = noun class 7, NP = noun phrase, num = cardinal numeral, O = object, OBL = oblique, PT = past tense, PERF = perfective, PL = plural, PP1 = class 1 pronominal prefix, PREP. PHRASE = prepositional phrase, PRES = present tense, REFL reflexive, Rel = relative clause, S = subject, SC = subject concord, SG = singular, V = verb, V-1 = verb-initial.

See also: Iconicity in linguistics; Linguistic typology; Syntax; Grammatical relations; Markedness in linguistics; Dependency in language; Functional approaches to grammar; Information structure in linguistics; Word classes/parts of speech; Comparative method in linguistics; Linguistic fieldwork.

List of relevant web pages

- The Universals Archive: <http://typo.uni-konstanz.de/archive/intro/>
- Das Grammatische Raritätenkabinett - a leisurely collection to entertain and instruct: <http://typo.uni-konstanz.de/rara/intro/>
- The World Atlas of Language Structures (WALS): <http://wals.info/>

References

- Behaghel, O., 1932. *Deutsche Syntax / 4. Wortstellung, Periodenbau*. Winter, Heidelberg.
- Beier, C., Hansen, C., Lai, I-W., Michael, L., 2011. Exploiting word order to express an inflectional category: reality status in Iquito. *Linguistic Typology* 15, 65-99.
- Bisetto, A., 2010. Relational adjectives crosslinguistically. *Lingue e linguaggio* 9-1, 65-85.
- Blake, B.J., 1983. Structure and word order in Kalkatungu: the anatomy of a flat language. *Australian Journal of Linguistics* 3-2, 143-75.
- Blake, B.J., 2006. Australian languages. In Keith Brown (Ed.), *Encyclopedia of Language & Linguistics (Second Edition)*, 585-93. Elsevier, Oxford.
- Cinque, G., 2005. Deriving Greenberg's Universal 20 and its exceptions. *Linguistic Inquiry* 36-3, 315-33.
- Devos, M., 2008. *A Grammar of Makwe*, Lincom Europa, München.
- Dik, S.C., 1997. *The Theory of Functional Grammar - Part 1: The structure of the clause* (2nd revised edition, edited by Kees Hengeveld). Mouton de Gruyter, Berlin.
- Dryer, M.S., 1992. The Greenbergian word order correlations. *Language* 68-1, 81-138.
- Dryer, M.S., 2005. Order of Subject, Object and Verb, in Dryer, M.S., Haspelmath, M. (Eds.), *The World Atlas of Language Structures Online*. Max Planck Digital Library, Munich, chapter 81. Available online at <http://wals.info/chapter/81>, accessed on 6 March 2013.
- Dryer, M.S., 2009. The Branching Direction Theory of word order correlations revisited, in Scalise, S., Magni, E., Bisetto, A. (Eds.), *Universals of Language Today*, Springer, Berlin, pp. 185-208.
- Foley, W.A., 1991. *The Yimas Language of New Guinea*. Stanford University Press, Stanford CA.
- Greenberg, J.H., 1963. Some universals of grammar with particular reference to the order of meaningful elements, in Greenberg, J.H. (Ed.), *Universals of Language*, MIT, Cambridge MA, pp. 73-113.
- Haiman, J. (Ed.), 1985. *Iconicity in syntax. Proceedings of a Symposium on Iconicity in Syntax, Stanford, June 24-6, 1983* (Typological Studies in Language 6), Benjamins, Amsterdam.
- Hale, K.L., 1976. The adjoined relative clause in Australia, in Dixon, R.M.W. (Ed.), *Grammatical Categories in Australian Languages*, Australian Institute of Aboriginal Studies, Canberra, pp. 78-105.
- Halliday, M.A.K., 2004. *An Introduction to Functional Grammar* (3rd edition), Arnold, London.
- Hawkins, J.A., 1983. *Word Order Universals: quantitative analyses of linguistic structure*, Academic Press, New York.
- Hawkins, J.A., 1994. *A Performance Theory of Order and Constituency*, Cambridge University Press, Cambridge.
- Heine, B., König, C., 2010. On the linear order of ditransitive objects. *Language Sciences* 32, 87-131.
- Hengeveld, K., Mackenzie, J.L., 2008. *Functional Discourse Grammar*, Oxford University Press, Oxford.

- Kittilä, S., 2006. Object-, animacy- and role-based strategies: a typology of object marking. *Studies in Language* 30-1, 1-32.
- Kutsch Lojenga, C., 1994. *Ngiti: a Central-Sudanic language of Zaire*. Köppe, Köln.
- Lehmann, W.P., 1973. A structural principle and its implications. *Language* 49-1, 42-66.
- Malchukov, A.L., 2000. Dependency reversal in noun-attributive constructions: towards a typology, Lincom, München.
- Mattissen, J., 2002. Dependent-head synthesis in Nivkh – with an outlook on polysynthesis in the Far Northeast, in Evans, N., Sasse, H.-J. (Eds.), *Problems of Polysynthesis* (Studia Typologica 4), Akademie Verlag, Berlin, pp. 135-166.
- Perkins, R.D., 2001. Sampling procedures and statistical methods. In Haspelmath, M., König, E., Oesterreicher, W., Raible, R. (Eds.), *Language Typology and Linguistic Universals: an international handbook. Volume 1*, Walter de Gruyter, Berlin, pp. 419-434.
- Rijkhoff, J., 1986. Word order universals revisited: the Principle of Head Proximity. *Belgian Journal of Linguistics* 1, 95–125.
- Rijkhoff, J., 2004. *The Noun Phrase*, Oxford University Press, Oxford [expanded Pb of the 2002 Hb edition].
- Rijkhoff, J., 2008a. Layers, levels and contexts in Functional Discourse Grammar, in García Velasco, D., Rijkhoff, J. (Eds.), *The Noun Phrase in Functional Discourse Grammar*, Mouton de Gruyter, Berlin, pp. 63-115.
- Rijkhoff, J., 2008b. Descriptive and discourse-referential modifiers in a layered model of the noun phrase. *Linguistics* 46-4, 789–829.
- Rijkhoff, J., 2008c. Synchronic and diachronic evidence for parallels between noun phrases and sentences, in Josephson, F., Söhrman, I. (Eds.), *Interdependence of Diachronic and Synchronic Analyses* [Studies in Language Companion Series 103], Amsterdam and Philadelphia: Benjamins, pp. 13-42.
- Rijkhoff, J., 2009a. On the (un)suitability of semantic categories. *Linguistic Typology* 13-1, 95-104.
- Rijkhoff, J., 2009b. On the co-variation between form and function of adnominal possessive modifiers in Dutch and English, in McGregor, W.B. (Ed.), *The Expression of Possession*, Mouton de Gruyter, Berlin, pp. 51-106.
- Rijkhoff, J., van Lier, E. (Eds.), 2013. *Flexible Word Classes: typological studies of underspecified parts of speech*, Oxford University Press, Oxford.
- Ross, M., 1998. Proto-Oceanic adjectival categories and their morphosyntax. *Oceanic Linguistics* 37-1, 85-119.
- Siewierska, A., 1988. *Word Order Rules*, Croom Helm, London.
- Siewierska, A., 2006. Word order and linearization, in Brown, K. (Ed.), *Encyclopedia of Language & Linguistics* (2nd edition), Elsevier, Oxford, pp. 642-649.
- Song, J.J., 2009. Word order patterns and principles: an overview. *Language and Linguistics Compass* 3-5, 1328–1341.
- Song, J.J., 2012. *Word Order*, Cambridge University Press, Cambridge.
- Stassen, L., 1984. The comparative compared. *Journal of Semantics* 3-1/2, 143–182.
- Testelec, Y.G., 1997. Word order variation in some SOV languages of Europe, in Siewierska, A. (Ed.), *Constituent Order in the Languages of Europe*, Mouton de Gruyter, Berlin, pp. 649-679.
- Tomlin, R., 1986. *Basic Word Order: functional principles*, Croom Helm, London.
- Van de Velde, F., 2009. The emergence of modification patterns in the Dutch noun phrase. *Linguistics* 47-4, 1021-1049.
- Van de Velde, F., 2012. A structural-functional account of NP-internal mood. *Lingua* 122, 1-23.

- Van Valin, R.D., Jr. (Ed.), 1993. *Advances in Role and Reference Grammar*, Benjamins, Amsterdam.
- Vennemann, T., 1975. Analogy in generative grammar: The origin of word order, in Heilmann, L. (Ed.), *Proceedings of the Eleventh International Congress of Linguists, Bologna-Florence, Aug. 28 - Sept. 2, 1972*, vol. II, Il Mulino, Bologna, pp. 79-83.
- Weil, H., 1844. *De L'ordre des Mots dans les Langues Anciennes Comparées aux Langues Modernes*. Translated by C.W. Super as *The Order of Words in the Ancient Languages Compared with that of the Modern Languages* (1978), Benjamins, Amsterdam.
- Weir, E.M.H., 1994, Nadëb, in Kahrel, P., van den Berg, R. (Eds.), *Typological Studies in Negation*, Benjamins, Amsterdam, pp. 291-323.

DRAFT