

DO NOT CIRCULATE

Replicability in comparative linguistics, with particular reference to syntactic typology

Jan Rijkhoff (Aarhus University, Denmark)

1. Introduction

Due to idiosyncratic or opaque methodology, major studies in comparative linguistics (especially in the area of syntactic typology) cannot be REPLICATED (or perhaps rather: REPRODUCED; cf. Drummond 2009). Consequently, the results of these studies remain basically unverified.

Apart from sampling issues, the main problem concerns the way members of some linguistic domain within and across languages ('morphosyntactic variants') are identified and selected:

- (1a) even though researchers have claimed that they (basically) employ semantic criteria to identify 'variants', they commonly use additional formal and/or functional criteria to make the semantic domain (if it is that) 'manageable';
- (1b) the real criteria used by the researcher to identify and select variants for an investigation are often left implicit;
- (1c) the (implicit or explicit) employment of so-called 'mixed criteria' in a single characterization or definition does not allow others to determine exactly how variants were selected (or why).

All these factors make it difficult or even impossible (a) to properly evaluate or interpret the results, and (b) to replicate a cross-linguistic investigation (esp. in the area of constituent order).

- (2) "Failure to provide a precise and explicit language-independent definition of the domain at hand will not only be confusing to the reader. Even worse, it will also, in all probability, lumber the researcher with a database that is unwieldy and internally contradictory, not to say chaotic. As a result, the researcher will run the serious risk of comparing data that are essentially incomparable, thereby damaging the overall value of his or her conclusions."
(Stassen 2010: 90-91)

Over the years, it has become clear that

- (3a) by themselves, semantic criteria are often too vague: semantically defined categories usually include variants that have too little in common for a useful cross-linguistic comparison (esp. in morphosyntactic typology);
- (3b) due to the combined, simultaneous and *ad hoc* use of different kinds of criteria in the same definition (if provided at all), the reader doesn't know exactly which variants are ultimately included the domain of investigation (or why) and which variants have been excluded (or why) to make the domain 'manageable' (see below);
- (3c) in morphosyntactic typology, the situation is aggravated by the vague criteria that are used to select only 'basic' or 'dominant' variants for the investigation.¹

The main goal of this paper is to show (a) how current research methods in comparative linguistics are often flawed, (b) how these methods can be improved, leading to studies that (in principle) can be replicated. More specifically, in this paper I will argue that

- (4a) different kinds of criteria are commonly required to define a cross-linguistic domain whose members have enough in common for a comparative investigation (this is uncontroversial);
- (4b) the use of 'mixed definitions' or 'comparative concepts' is not helpful if this means that functional or formal criteria remain unspecified or if they are contained ('mixed') in a single definition, as this obscures the way variants within and across languages are identified and subsequently 'managed' for the actual investigation. In other words, the various kinds of criteria should be applied separately, so as to allow others to determine exactly which variants are included (and why) and which variants are subsequently excluded from the investigation (and why).

¹ Recall that it is one of the central goals of typology 'to describe and explain the structural diversity of the world's languages'. This goal cannot be reached if certain variants are consistently ignored.

(4c) for various reasons (cognitive, linguistic, methodological) functional criteria/categories are usually the best starting point for a cross-linguistic investigation.

(5) Structure of this talk:

2. Some methodological issues in comparative linguistics
3. A case in point: the cross-linguistic category ‘Adjective’
4. The many-to-many relationship between Form and Function
5. Towards an explicit method of domain specification (with particular reference to constituent order)

2. Some methodological issues in comparative linguistics

Do comparative linguists / ‘Greenbergian’ typologists really use purely/basically semantic criteria? If so, what are they?

2.1. ‘Purely/basically semantic criteria’?

In morphosyntactic typology, researchers usually claim they have employed ‘basically/purely’ semantic criteria to identify variants:

(6) “It is here assumed, among other things, that all languages have subject-predicate constructions, differentiated word classes, and genitive constructions, to mention but a few. I fully realize in identifying such phenomena in languages of different structure, **one is basically employing semantic criteria ...**” (Greenberg 1966: 74)

(7) “Greenberg’s point of view has, explicitly or tacitly, been adopted by all authors of major typological studies in the last three decades, and has been canonized in textbooks such as Croft (2003: 13-19).” (Stassen 2010: 92; see also e.g. Haspelmath 2007, 2010)

(8) “In the final analysis, **semantic criteria will suffice** to make the cross-linguistic equation.” (Hawkins 1983: 12)

(9) - “The standard practice in word order typology since Greenberg 1963 has been to identify different categories **largely on the basis of semantic criteria. Underlying this practice is an implicit assumption either that semantic categories like ‘numeral’ and ‘demonstrative’ correspond to universal grammatical categories or that any differences in the way in which these semantic categories are realized in particular grammars are somehow irrelevant to the correlations.**” (Dryer 1992: 120)

- “Many linguists have used **purely semantic criteria** in identifying constructions, ...”. (Dryer 2007a: 80-1):

JR: This seems to be an overstatement: (a) the use of mixed criteria is the rule rather than the exception, (b) if/when (basically/purely) semantic criteria were used, they are hardly mentioned, (c) the set of cross-linguistic variants ‘identified’ by these (basically/purely) semantic criteria is usually not described in any detail.

In sum, the reader generally does not know exactly

- (10) a. how cross-linguistic domains are defined,
- b. how variants are identified within and across languages, and (in the ‘data management phase’),
- c. which variants are excluded (or why) and which variants are ultimately deemed suitable for the cross-linguistic investigation (or why).

Due to their opaque methodology these studies cannot be replicated. How reliable are the results of studies that have been conducted in this fashion?

2.2. ‘Non-formal’ criteria ⇒ functional (or ‘external’) criteria

Apparently semantically criteria do **not** “suffice to make the cross-linguistic equation” (cf. Hawkins’s statement above).²

- (11) Semantic criteria are often not explicitly mentioned (see, e.g. Greenberg 1963, Hawkins 1983, Dryer 1992).
- (12) If ‘semantic criteria’ are used, it often turns out that they are confused with functional criteria, which refer to the way members of a certain domain are used in actual discourse. Semantic criteria are probably better regarded as an inherent component of the functional characterization of a cross-linguistic domain.³

For example, members of the cross-linguistic semantic category Adjective have been characterized as words denoting a ‘descriptive property’ or a ‘property concept’ (Thompson 1988: 168), but it is clear that in practice membership of cross-linguistic domain Adjective is restricted to variants that serve as **adnominal modifiers** (see below, section 3). Sometimes this is explicitly mentioned as part of the definition (as in Haspelmath (2010: 670):

- (13) “An adjective is a lexeme that denotes a descriptive property and that **can be used to narrow the reference of a noun**”,

but often the extra functional criterion has to be inferred from the context. It is probably for this reason (i.e. the addition of functional criteria) that Stassen (2010: 92, 95) also characterized typological investigations in the tradition of Greenberg as ‘**non-formal**’ or ‘**functional**’ approaches to cross-linguistic identification.

As we will see below (section 4), a semantic characterization of the cross-linguistic domain Adjective (such as: denoting a ‘property concept’ or ‘quality’) can be included in the **functional domain specification**. For example, in the domain label ‘qualifying modifier’, ‘qualifying’ mostly relates to the meaning component (further specifying a more or less inherent property or ‘quality’ of an entity), where ‘modifier’ specifies the overall function (as opposed to e.g. predicating or referring).

2.3. Semantic/non-formal/functional criteria and formal criteria.

The criteria used to identify members of some domain within and across languages (‘morpho-syntactic variants’) are hardly ever made fully explicit in comparative linguistics. For example, we often have to guess which semantic criteria were used to identify members of such categories as Subject, Object, Adjective, Relative Clause or Genitive in major typological investigations.

But whichever semantic/non-formal/functional (aka ‘external’) criteria were employed (tacitly or explicitly), it appears that, as a rule, additional formal criteria are often applied (tacitly or explicitly) to make the cross-linguistic domain of investigation ‘**manageable**’, allowing for the researcher “**to weed out ‘concomitant’ factors, i.e. instances of cross-linguistic variation that are not considered to be essential to the domain under study**” (Stassen 2010: 95-96).⁴

² Cf. Bateson (1979: 15): “Without context, words and actions have no meaning at all. This is true not only of human communication in words but also of all communication whatsoever, of all mental process, of all mind, including that which tells the sea anemone how to grow and the amoeba what he should do next.” See also Akman and Bazzanella (2003).

³ Thus ‘semantic criteria’ should not be equated with ‘functional criteria’. Semantic criteria alone are usually not enough to define a cross-linguistic domain (Rijkhoff 2009; see also fn. 2): they are too vague or too abstract to define a set of variants that share enough grammatical features to make them comparable. **In other words, we need to know how a variant is used in actual discourse so as to be able to keep variants with irrelevant meanings out of the cross-linguistic domain (see below). The same probably holds for variants with irrelevant formal properties, which is why functional criteria should normally take priority over semantic and formal criteria (section 5).**

⁴ Both Dryer’s and Haspelmath’s characterization of the cross-linguistic category Adjective also include a formal criterion: adjectives must be ‘words’ (Dryer 1988, 2104a) or ‘lexemes’ (Haspelmath 2010; notice that his ‘comparative concepts’ can also include formal components). More on this in section 3.2.

Sometimes the additional formal criterion to eliminate ‘non-essential’ variants is explicitly stated.⁵ Thus, Greenberg (1963: 88) wrote that he could only formulate *Universal 22* about comparative constructions by ignoring variants in his semantic domain that involves a verb with a general meaning ‘to surpass’.⁶ Almost two decades later, Stassen (1984: 15) also indicated that his semantic definition of the comparative construction needed to be supplemented with formal criteria to exclude variants in which the two compared objects are not expressed by noun phrases (cf. Stassen 2010: 96).

The employment of additional formal (‘internal’) criteria is, however, not always (clearly) indicated. As in the case of non-formal (‘external’) criteria, this means that the reader does not know (i) which variants the researcher excluded from the database on formal grounds so as to make the domain ‘manageable’, (ii) what formally defines the variants that the researcher ultimately deemed suitable for the cross-linguistic investigation.

3. A case in point: the cross-linguistic category ‘Adjective’

Problems with the current approach to categorization and domain specification in comparative research can be illustrated on the basis of Dryer’s work, esp. regarding the cross-linguistic category Adjective. This is not because Dryer’s work is better or worse compared to the work of others, but rather because he has produced a large body of work in the area of morphosyntactic typology, especially with regard to ‘Greenbergian word order correlations’ (e.g. Dryer 1988, 1992, 2013).

3.1. The cross linguistic domain Adjective: semantic criteria

This is, for example, how the cross-linguistic category ‘Adjective’ is characterized in the *World Atlas of Language Structures* (WALS):

(14) “For the purposes of this map, the term *adjective* should be interpreted in a semantic sense, as a word denoting a descriptive property, with meanings such as ‘big’, ‘good’, or ‘red’. It does not include non-descriptive words that commonly modify nouns, such as demonstratives (like *this* in ‘*this* dog’) ..., numerals (as in ‘*two* dogs’) ..., or words meaning ‘other’ (as in ‘the *other* dog’). In some languages, like English, adjectives form a distinct word class. In other languages, however, adjectives do not form a distinct word class and are verbs or nouns (...). For example, in Eastern Ojibwa (Algonquian; eastern Canada and United States), words expressing adjectival meaning are just like verbs morphologically and syntactically.” (Dryer 2104a)

A definition of the cross-linguistic category ‘Adjective’ can also be found in Dryer’s (1988) investigation of the relation between Object-Verb order and Adjective-Noun order.

(15) “In collecting the data for his paper, I have followed the implicit practice of Greenberg (1963) in employing essentially semantic criteria in identifying adjectives. Namely, a word is identified as an adjective if it expresses the kind of meaning associated with descriptive adjectives in English. It is clear, however, that in many languages, the words in question are simply verbs; in other languages they are nouns. The use of the term *adjective* in word order studies is thus potentially misleading, and the expression *property term* might more accurately convey the fact that it is a semantic category rather than a grammatical category.” (Dryer 1988: 197)

⁵ To ignore ‘non-essential’ variants altogether would go against the basic goal of linguistic typology: to investigate the full range of cross-linguistic variation. It might be interesting to see which variants have been treated this way in the history of comparative linguistics (see e.g. the Mouton volumes on *rara* and *rarissima*). Notice also that there is the danger that ‘non-essential’ variants could sometimes be used as a euphemism for variants that are somehow problematic; this could not happen in an explicit, transparent procedure (section 5).

⁶ “*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). See also e.g. Greenberg’s (1966: 108) remarks on Burmese adjectives in his Appendix I.

Makwe (Devos 2008: 136): nominal expression of adjectival meaning

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

Hausa (Newman 1987: 721): nominal expression of adjectival meaning

- (21) *ríjìyáa màì zúrfi*
well 'owner, possessor of' depth
'a deep well'

- (22) **Thus, even if we exclude non-modifying variants, there are still many variants of the semantic category Adjective within and across languages that are not taken into account** (e.g. 16b-h). In 'basic word order' studies certain variants are consistently excluded, because they are not considered to be 'basic' or 'dominant' (Recall fn. 1 about the central goal of typology).

Below (section 5): since semantic criteria, definitions or concepts are generally rather vague and abstract, they tend to include too many variants (Rijkhoff 2009a; see also Malt et al. *in press*). Therefore it seems better to formulate the property denoting part of the characterization in terms of a more general functional definition, one that includes the property-denoting function and the modifying function (sections 2.2 and 4).

Native speaker bias: As to the criterion that the word in question should translate as a descriptive adjective in English, Rijkhoff (2009a; 98) already mentioned the danger of using one's own language as the measure of all things linguistic.⁸ Whereas a 'Western' linguist could be biased toward variants that resemble members of the English word class Adjective, a linguist from Kiribata might have a strong preference for 'property denoting expressions' that resemble Kiribatese 18b, similar to English (16)g, i.e. the variant with the adnominal relative clause ('a house that was rather big/small/expensive'). This is another reason why formal (language-specific) criteria should (as a rule) not take precedence over functional criteria.

3.2. Formal criteria

In (14) and (15), members of the semantic, cross-linguistic category 'Adjective' are also formally characterized as single 'words'. See also Dryer (1992: 109):

(23) "Another problem that arises in identifying adjectives in some languages is that in many languages the meanings in question are expressed by **words** that belong either to the class of verbs in the language or to the class of nouns We thus encounter again the question of to what extent the categories assumed in word order typology are semantic and to what extent they are motivated as categories within each language. **We follow here the general practice in word order typology of assuming a semantic notion of adjective, so that we include words that in some languages belong to the class of nouns, in others to the class of verbs**"

? Perhaps the 'word' problem would not have occurred, if instead the label 'constituent order typology' had been used.

However, in various places Dryer (1992: 96 fn. 12) noted that "what I call adjectives are really verbs, and 'adjectives' modifying nouns are **really just a kind of relative clause**" (see also Dryer 1988: 192, 197; 1992: 102, 109-111; 2007a: 103, 2009; Dryer 2014a-b).

This remark points to a serious problem in comparative linguistics: **the many-to-many relationship between the form and function of constituents** (section 4). This is one of the reasons

⁸ Cf. also Dryer (2007a: 80-1): "Many linguists have used purely semantic criteria in identifying constructions, but the semantic criteria employed are often strongly influenced by English translations and there is the danger of imposing English categories on languages to which they do not apply."

why we should distinguish between formal and functional criteria in characterizing the members of some cross-linguistic domain.

If adjectives can take the form of a relative clause, it becomes difficult, if not impossible, to distinguish between members of the cross-linguistic domain Adjective and members of the cross-linguistic domain Relative Clause, which, by definition must be clauses (cf. Dryer 2104b).

- (24) “A construction is considered a relative clause for the purposes of this map if it is a clause which, either alone or in combination with a noun, denotes something and if the thing denoted has a semantic role within the relative clause.”

Dryer (1992: 120, 122) is aware of the problematic, heterogenous character of the members of some of his domains (see also Dryer 2009) and suggests we would need to consider “more specific grammatical properties” so as to be able to account for the morphosyntactic behaviour of members of cross-linguistic ‘semantic’ categories such as demonstratives, numerals, adjectives (also e.g. ‘negative words’, ‘tense/aspect words’):

- (25) “The situation with demonstratives is thus not unlike what we observed for numerals. While they may constitute a fairly well-defined category from a semantic point of view, their grammatical properties seem to vary from language to language, ...” (Dryer 1992: 122)⁹

Another Form-Function problem: What about languages in which relative clauses cannot be distinguished from adverbial and other constructions (Hale 1976: 78-9) or from ‘Generalized Noun Modifying Clause Constructions’ (or: GNMCCs; Matsumoto 1997) Comrie et al. 2013).¹⁰

- (26) [*gakuseiga kat-ta*] *hon* (27) [*gakuseiga hon o kat-ta*] *zizitu*
student NOM buy-PST book student NOM book ACC buy-PST fact
‘the book that the student bought’ ‘the fact that the student bought the book’

- (28) [*sakana o yak-u*] *nioi*
fish ACC grill-PRS smell
‘the smell of (someone) grilling fish’

(29) It seems there is only one real solution to the basic problem of category membership in comparative linguistics: to seriously take into account the fact that there is a many-to-many relationship between the form and the function of constituents: the same form (e.g. adjectival lexeme, genitival phrase, relative clause) can have different functions and, vice versa, the same function (e.g. qualifying modifier) can be performed by members of different form classes.

3.3. ‘Basic word order’

We saw above that it is often unclear which variants are ultimately deemed ‘comparable’ for the cross-linguistic study at hand. One may assume that the total set of variants deemed comparable provides the pool from which the subset of variants is selected with members that only occur in some basic pattern or environment (as is often the case in ‘Greenbergian’ studies about word order correlations). Obviously, if we do not have a clear picture of the complete set of (basic and non-

⁹ Cf. also Dryer (2007a: 80): “With many pairs of elements, a pervasive problem is that of to what extent one should use purely semantic criteria in identifying constructions and to what extent specific syntactic properties of the construction in the language are relevant.” See Rijkhoff (2004: 278-91) for a more detailed discussion of membership of Dryer’s semantic categories.

¹⁰ Not to mention differences in constituent structure, e.g. appositional vs. non-appositional constructions (or: *configurational* vs. *non-configurational*, *tight* vs. *scrambled*, *hierarchical* vs. *flat* constructions) as frequently attested in certain Australian (Heath 1984: 500) or Tibeto-Burman languages (Herring 1991: 55); cf. Rijkhoff 2004: 19-23).

basic) variants that make up a cross-linguistic domain, how do we know which variants count as ‘basic’ and which as ‘non-basic’ - and why?

‘**What is ‘basic’?**’ There is no consensus among linguists what counts as ‘basic’ (i.e. individual researchers can take different decisions; also, they are often dependent on their theory or the quality of the written grammar).¹¹To mention some possibilities (which are not mutually exclusive): frequency (the most frequently used variant), complexity (the least complex variant), markedness (the least marked variant, with regard to meaning, form or use, incl. distribution and style/register), the order found in main (declarative) clauses, order found in subordinate clauses.

Siewierska (1988: 8) on the basic order of S, V and O: “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.” But such sentences (e.g. ‘The old man feeds the hungry dog’ hardly ever occur in spoken languages; also, some linguists ‘merge’ lexical NPs and pronouns, even though they often have different distributional properties.

Kroeger (2005: 198-9): FREQUENCY (?style, gender, spoken, written?), NEUTRAL SEMANTICS / PRAGMATICS (e.g. no special emphasis), AVOID PRONOUNS (*but lexical S and O in same clause are rare in spoken language!*), SUBORDINATE CLAUSES TAKE PRIORITY, DISTRIBUTION (IN MOST CONTEXTS)

Dryer (2007a: 74, 77): “frequency has been the primary criterion in word order typology”. Other criteria: distribution (e.g. all adjectives AN, except ‘big_A’: NA; or nearly only prepositions, but one postposition); simplicity, pragmatics, ... Dryer writes “in most cases, frequency coincides with other criteria”, but what about style/register/etc, the infrequent ‘The old man feeds the hungry dog’ etc.

4. The many-to-many relationship between Form and Function

The title of this talk is ‘Replicability in comparative linguistics’, but perhaps ‘Reproducibility’ would have been a better choice. Drummond (2009) has argued that that ‘reproducibility requires changes; replicability avoids them’. Apart from the general requirement that research procedures should be transparent, some weaknesses that should be AVOIDED in comparative linguistics are:

- a. the confusion between functional and formal criteria/categories (Rijkhoff 2009a)
- b. the suggestion that there is a one-to-one relationship between form and function.

As to a. (confusion between functional and formal criteria/categories): this is clearly demonstrated in the way some basically/purely ‘semantic’ categories are treated in Hawkins (1983) and Dryer (1992).

- In **Dryer**’s account of ‘Greenbergian word order correlations’, the cross-linguistic domain of Adjectives is first a semantic category, but in the Branching Direction Theory members of the adjectival domain are basically treated as a non-branching elements, i.e. a purely formal category, because adjectives are basically identified as ‘words’ (as we saw above, in various places Dryer also admits that some of his ‘adjectives’ are actually relative clauses, which leads to various modifications and adjustments; see also Dryer 2009);

- **Hawkins** claims that that his cross-linguistic category ‘Adjective’ (like all other domains in his word order book) were identified on the basis of semantic criteria. But when he attempts to account for word order phenomena with his *Heaviness Serialization Principle* (HSP), members the semantic category Adjective are (tacitly) reclassified as members of the relatively ‘light’ formal (lexical, syntactic) category ‘Adjective’.

(30) *Heaviness Serialization Principle* (HSP): Rel \geq_R Gen \geq_R A \geq_R Dem/Num

In the HSP, one constituent is considered ‘heavier’ than another, if one or more of the following

¹¹ We should also mention the role of grammatical theories (see also Dryer 2013 on this topic). For example, a ‘generative grammar’ of some language (e.g. Afar) may contain previous little information about actual, surface forms and constructions.

factors imposes an ordering between them (Hawkins 1983: 90): (i) length and quantity of morphemes, (ii) quantity of words, (iii) syntactic depth of branching nodes, (iv) inclusion of dominated constituents. Notice also that some semantic adjective formally speaking Relative Clauses, so that we also see here (see above, Dryer) that it becomes impossible to distinguish between the cross-linguistic domains Adjective and Relative Clause.

This kind of categorial confusion seems to occur when the researcher fails to

- clearly distinguish between functional and formal criteria;
- take into account the fact the same linguistic form can be used for different communicative or discourse functions and, vice versa, that members of different form classes (e.g. adjective, PP, relative clause) can be used for the same communicative or discourse function.

Preliminaries: Form & Function.

(31) Some formal categories: A, Rel.Cl, PP, ...; complement, adjunct, determiner, ...

(32) Some functional modifier categories: (i) classifying, (ii) qualifying, (iii) quantifying, (iv) localizing or anchoring, and (v) discourse-referential modifiers (e.g. Rijkhoff 2008a, 2008c, 2009); on attitudinal modification, see Rijkhoff 2010).¹²

- ☞ The notion ‘function’ should be interpreted in the sense of Prague School linguists, who were “seeking to understand what jobs the various components were doing [...]” (Sampson 1980: 104).
- ☞ The same five FUNCTIONAL MODIFIER CATEGORIES can be used to analyze NPs and clauses
- ☞ Functional modifier categories include both grammatical and lexical modifiers, i.e. ‘modifier’ covers all dependents that are not arguments or complements.
- ☞ Modifiers are distributed over ‘nested’ layers around the head constituent, reflecting differences in SEMANTIC SCOPE. Conventionally, grammatical modifier categories (‘operators’) like *Demonstratives* or *Tense* (both: LOCALIZING/ANCHORING MODIFIERS) are represented before the head constituent in a linear F(D)G representation, whereas lexical modifiers (‘satellites’), such as adjectives, adverb(ial)s, PPs or relative clauses, are represented after the head constituent.
- ☞ The layered model of representation applies to languages with a hierarchically organized NPs and clauses (fn.11)

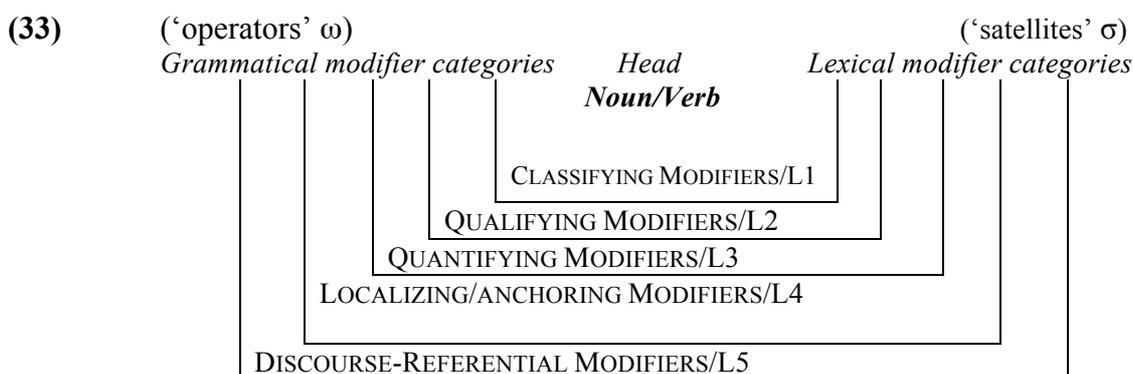


Figure 1. Functional modifier categories in a ‘layered’ representation of NP/clause structure.

- ☞ **Recall:** the same linguistic form or construction can often be used in more than one function (Fig. 2), and vice versa, different linguistic forms or constructions can be used in the same modifier function (Fig.3). On the ‘act of modification’, see Rijkhoff (2014).
- ☞ **Hierarchical organization:** modifiers of the outermost layer L5 have the widest scope, whereas the scope of modifiers represented at the innermost layer L1 (classifying modifiers) only further specify the *kind* of entity denoted by the head noun (L0).

¹² Clauses have at least two more layers of modification, accommodating propositional (modal) and illocutionary modifiers. Van de Velde (2007, 2012) has argued that a complete representation of NP-structure should also allow for illocutionary and modal modifier categories.

☞ NP structure developed on the basis of NPs that are used to refer to FIRST ORDER (SPATIAL) ENTITIES (concrete objects or ‘things’; so no nominalizations, for example).

(34) So far: five functional modifier categories in NPs and clauses - but cf. fn 12 (with examples of English **adnominal** modifiers) :

- a. CLASSIFYING MODIFIERS further specify what KIND of entity (which subcategory) is denoted by the head constituent (as in e.g. ‘SOLAR energy’, ‘a VIRAL infection’, ‘a man OF THE STAGE’);
- b. QUALIFYING MODIFIERS further 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 locatable/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’).

* On names for linguistic categories, see García Velasco Rijkhoff (2008c: 14-16).

=====

Simplified formal representation of NP structure (reflecting scope relations among functional modifier categories)

(35) NP_i: [ω₅ [ω₄ [ω₃ [- [ω₁ Noun(x_i) σ₁] σ₂] σ₃] σ₄] σ₅] (*simplified version*)

x = referent variable (symbolizes the referent of the NP);

ω = grammatical modifier category (‘operator’): ω₁ = classifying operator, ω₃ = quantifying operator, ω₄ = localizing operator, ω₅ = discourse-referential operator). Classifying operator ω₁ (such as a nominal aspect marker; cf. e.g. Rijkhoff 2008b: 738-9) only has the head noun in its scope, whereas discourse-referential operator ω₅ has the widest scope and is concerned with interpersonal/pragmatic properties of the referent;

σ = lexical adnominal modifier (‘satellite’): σ₁ = classifying satellite, σ₂ = qualifying satellite, σ₃ = quantifying satellite, σ₄ = localizing satellite, σ₅ = discourse-referential satellite).

(36) The underlying structure of the noun phrase (Rijkhoff 2008a-b-c):

NP_i: [ω₅ L₄[ω₄ L₃[ω₃ L₂[- L₁[ω₁ L₀[Noun(f_i)](x_i)]L₀ σ₁(L₀)]L₁ σ₂(L₁)]L₂ σ₃(L₂)]L₃ σ₄(L₃)]L₄ σ₅(L₄)]

For example (slightly simplified): ‘those three famous nuclear physicists across the table’

(37) NP_i: [Def L₄[Dem_{Rem} L₃[3 L₂[- L₁[ω₁ L₀[physicist_N(f_i)](x_i)]L₀ nuclear_A(L₀)]L₁ famous_A(L₁)]L₂ σ₃(L₂)]L₃ across the table(L₃)]L₄ σ₅(L₄)]

=====

4.1. Functional modifier categories I: Same Form - Different Functions

☞ Example 1: Form = Adnominal PP with *van* ‘of’ in Dutch (‘genitives’)

According to Dryer (2007a: 86-7) “there is relatively little cross-linguistic difficulty in identifying genitive constructions”. This may be true from a purely formal perspective (e.g. when ‘genitives’ are defined as ‘constructions carrying a genitive case marker’), but it is generally known that ‘genitive constructions’ are extremely versatile from a function/meaning-perspective (Williams 1991: 89), both within and across languages. The paperback edition of *Longman’s Dictionary of Contemporary English* alone lists twenty-five different senses of the possessive preposition ‘of’. This is at least partly due to the fact that the wide ‘belonging to’ relation covered by ‘genitives’ plays a central and fundamental role in human cognition (Seiler 1983, Heine 1997, Rijkhoff 2009b; Fig. 2 differs in some respects from the table given in Rijkhoff 2009b).

In Samoan (Mosel & Hovdhaugen 1992: 318), numerals occur as the head of a special kind of relative clause introduced by the general tense-aspect-mood marker [GENR] *e*; this particular construction is only used as a modifier in a specific NP:

- (42) *Sa fau=sia e Tagaloaalagi fale e tolu ...*
 PAST build=ES ERG Tagaloaalagi house GENR three ...
 ‘Tagaloaalagi built three houses ...’

Rel.CI = Localizing modifier (providing a referential ‘anchor’ to make the referent identifiable) Lehman (1984: 402, 405): “Mit einem Relativsatz kann man leicht einen bestimmten Gegenstand durch Spezifikation der Situation, an der er teilhat, identifizieren Das Adjektiv dient mehr der Begriffsbildung, der Relativsatz mehr der Gegenstandsidentifikation”

- (43) Phil Bronstein got the first interviews with *the man who actually fired the bullets that killed Osama bin Laden*. (*Osama bin Laden* is the referential anchor)

4.2. Functional modifier categories II: Same Function - Different Forms

☞ **Example: Function = qualifying modifier**

The fact that the same function (here ‘qualifying modification’) can be served by members of different form classes is shown in Fig. 3:

| | | | | |
|--------------------|--|-------------------------|-----------------------|--------------------------------|
| (44) | One Function: Qualifying Modifier here in NP (not in clause), i.e. variants that further specify a more or less inherent property of a participant or prop that is mentioned as part of a speech act. | | | |
| ↓ Many Forms ↓ | | | | |
| Adjective | <i>big N</i> | <i>expensive N</i> | <i>young N</i> | <i>red N</i> |
| PP | <i>N of enormous size</i> | <i>N of great value</i> | <i>N under age 16</i> | <i>N of incredible redness</i> |
| Relative clause | <i>N that was rather big</i> - also (18)b & (19)b | ... | ... | ... |
| Genitive/Possessor | - also (20) & (21) | ... | ... | ... |
| ... | ... | ... | ... | ... |

Figure 3. **ONE FUNCTION - MANY FORMS:** Formal categories serving as QUALIFYING MODIFIERS (N = head noun)

(45) **Recall:** “The standard practice in word order typology since Greenberg 1963 has been to identify different categories largely on the basis of semantic criteria. Underlying this practice is an **implicit assumption** either that semantic categories like ‘numeral’ and ‘demonstrative’ correspond to universal grammatical categories or that any differences in the way in which these semantic categories are realized in particular grammars are somehow irrelevant to the correlations.” (Dryer 1992: 120)

Fig. 3 shows that this ‘implicit assumption’ is not correct.¹⁴

- (46) **‘Form’ matters:** it is important to distinguish between formal properties of variants within the same functional domain, as they have different grammatical properties. For example,
 - (within the functional domain of qualifying modifiers) members of the English adjectival word class precede the noun, whereas PPs and Relative Clauses follow the noun in English;
 - cross-linguistically, heavy/complex material tends to appear later in the NP (Hawkins’s HSP).

In sum:

- some syntactic typologists appear to have also confused formal and functional criteria, thus strengthening the suspicion that categorization is sometimes biased by the native language of the speaker (e.g. English adjectives are the best examples of the cross-linguistic semantic domain ‘Adjective’);
- members of the same formal category (e.g. PP; Fig. 2) can have different properties (due to

¹⁴ See also Dryer (1992: 12): “The situation with demonstratives is thus not unlike what we observed for numerals. While they may constitute a fairly well-defined category from a semantic point of view, their grammatical properties seem to vary from language to language, ...”; see also fn. 8 above.

different functional specifications);

- members of the same functional category (e.g. qualifying modifier; Fig. 3) can have different grammatical properties (due to different semantic or formal properties).

The many-to-many relationship between linguistic forms and functions is often insufficiently taken into account in comparative linguistics, especially in the area of linguistic typology.¹⁵

=> **Form & Function I: do not confuse functional and formal criteria; keep them separate in domain specification)**

=> **Form & Function II: account for many-to-many relationship between Form and Function**

5. Towards an explicit method of domain specification (with particular reference to constituent order)

☞ REPLICABILITY/REPRODUCIBILITY of results is essential to the advancement of comparative linguistics

The results of major studies in syntactic typology cannot be replicated or reproduced (avoiding the methodological problems mentioned above), because the research methods (sampling, identification and selection of variants) are idiosyncratic or unclear. Consequently the results of these studies remain unverified (nevertheless their validity is often taken for granted).

One of the problems highlighted in this talk concerns the fact that formal and functional criteria are often left implicit and, if specified, are applied in such a way that the reader does not know exactly how variants were identified or selected. How can we avoid the various kinds of weaknesses in comparative research and make them replicable/reproducible?

5.1 Requirements for REPLICABILITY/REPRODUCIBILITY:

- (47) a. a general (non-idiosyncratic), explicit sampling method (e.g. Rijkhoff et al. 1993, Rijkhoff & Bakker 1998);
- b. language-independent definitions of cross-linguistic domains (this is probably only possible in the case of functional or 'external' criteria), that make it clear how members of a certain cross-linguistic domain were identified and selected;
- c. a distinction between functional and formal criteria, which should not be applied simultaneously (section 5.1);
- d. an account of way members of some cross-linguistic domain were identified 'in the first round' (*identification*), i.e. before certain variants are excluded to make the domain 'manageable' for the investigation at hand (*selection*); possible also how the origination definition/criteria had to be modified during the identification process;
- e. a typological overview of the various kinds of members that were identified 'in the first round' (e.g. a division into functional and formal (sub)types);
- f. a discussion of the way the researcher dealt with problematic cases (i.e. reasons to include or exclude borderline cases as members of the cross-linguistic domain); in other words, it should be clear which variants are included 'in the first round' (and why) and, when relevant, which certain variants were subsequently deemed unsuitable for the investigation at hand - and why (selection).

If formal and functional criteria are applied separately, the question is: which should be used first and which should be employed later (if necessary) to eliminate 'unsuitable' variants? Here are some reasons why, as a rule, Function should precede Form.

5.1. Functional categories have a much wider cross-linguistic applicability than formal categories

There are several ways to approach linguistic functions. For example, **Jakobson** (1960) proposed the following six more or less stylistic functions: referential, expressive, conative, poetic, phatic, and metalingual. **Halliday** (1975) recognizes seven functional linguistic categories for children in

¹⁵ On the false idea there should normally be a one-to-one relationship between form and function, see also e.g. McCawley's (1985: 675) review of Newmeyer (1983): "... functional accounts need not recognize such a thing as THE function of anything".

their early years. Four categories to satisfy physical, emotional and social needs: the instrumental, regulatory, interactional, and the personal function. The remaining three functions are: heuristic, imaginative, and representational. The *categorizing* function of language has been discussed in the psycholinguistic and anthropological linguistic literature (Rosch 1978, Berlin 1978).

At a more pragmatic (interpersonal) / grammatical level, the basic discourse functions of linguistic material can be divided into three ‘actional’ categories (Searle 1976, Croft 1990, Rijkhoff 2014): predicating (used to ascribe properties or relations; associated with lexemes, esp. verbs), referring (used to ‘pinpoint’ to entities; associated with nouns, or rather NPs) and modifying (used to ‘enrich’ or ‘supplement’; often associated with adjectives and adverbs, but may other forms and constructions can be serve as a modifier).

These functions can be said to be relevant for all the languages across the globe. It seems unlikely that there are languages that do not allow their speakers to predicate / refer / modify or (from an informative perspective) to differentiate between focal and topical information.¹⁶ Thus, it seems much easier to find common ground among the world’s languages with regard to functional criteria/categories.

By contrast, due to the enormous degree of language specific variety in linguistic forms and structures, it seems much more difficult to compare all languages in terms of formal criteria or categories (cf. Evans and Levinson (2009: 430-1) on the absence of “uncontroversial facts about substantive universals”).

It would be counterproductive to start with formally defined cross-linguistic domains and then apply functional criteria to make create a manageable set. Apart from the fact that it is unlikely that we could apply the same set of formal criteria in languages across the globe (in the worst case, potential variants in each language need to be identified with a set of different formal criteria), the variants would not form a coherent domain to begin with (see 5.2 below). Also that formal categories such as NP or PP are inherently non-relational; the full range of grammatical properties only become visible when forms or constructions are used in an actual utterance (Fig. 2): 5.2.

5.2. Functional categories are relational, formal categories aren’t

A formal category such as PP can be characterized in terms of certain intrinsic properties (e.g. P + NP), but many if not most of its grammatical properties can only be investigated when we know how a member of a formal category is used, e.g. as

- (48) a. a locative argument (‘She put the book on the shelf’)
b. a localizing adjunct (of an event): ‘They kissed under the mistletoe’
c. a classifying modifier /‘complement’: ‘a student of physics’
d. a localizing modifier (of a participant/prop): ‘the house across the road’

In all these examples, we find a construction with the formal category label PP, but this does not tell us anything about the relation between the PP and the construction in which it is used (Dik 1997: 25).

Only when we investigate how the PP is used (‘in context’), can we determine its non-intrinsic properties. For example, whether it is optional or obligatory, or whether its position is fixed or not. This is also shown in Fig. 2 (Dutch PPs with *van* ‘of’), where the grammatical properties of a member of the same formal category (PP) co-vary with its function in the larger construction: classifying, qualifying or localizing qualifier.

Functional categories (such as those listed in (33)), on the other hand, are inherently relational, as they are characterized in terms of their communicative function an utterance. Membership of a functional category informs us about the relations between a constituent and the construction in which it occurs and this determines to a large extent its grammatical properties (see Figs. 1 and 2.)

¹⁶ One could argue that all the linguistic elements of a speech act have a communicative function (Rijkhoff 2014). See also Halliday’s ‘principle of exhaustiveness’ (2004: 60): “everything in the wording has some function at every rank (cf. Halliday 1961, 1966).”

5.3. Psycholinguistic arguments why Function precedes Form

Levelt's psycholinguistic model of speech production (1989, 1995) deals with questions such as 'How do you get from some communicative intention to the overt articulation of speech?' and 'What are the processing components involved in the generation of fluent speech and how do they interact in real time?'

Roughly speaking, speakers first deal with matters of content (including communicative intentions), and only later with matters of form. What Levelt calls the Conceptualizer is primarily concerned with choice of content: "Conceptualizing is primarily deciding on what to express, given the present intention" (1995: 15). **The choice of content is functionally motivated: presumably content is only included because it has a communicative function in a given context.** Subsequently the message is linguistically (morphosyntactically, phonologically) encoded in the Formulator. In its turn, the choice of formal material is determined by the kind of content prepared in the Conceptualizer (incl. speech act, style, register, politeness, sarcasm etc.).

Thus it appears that functional (pragmatic, semantic) considerations like 'what should be the content of my message in this situation (addressees, setting)?' precede formal operations like 'which morphosyntactic or phonological means/tools are appropriate to express this content?'

5.4. It is not 'trivial' whether formal or functional criteria are applied first

Stassen (2010: 97) writes that there is no principle reason why formal criteria should not be applied first, i.e. why formal criteria cannot have 'wider scope' than functional criteria. But **(a)** see sections 5.1-2-3 above, **(b)** the choice (whether formal or functional criteria are applied first) is not 'trivial', especially when it is not clear to the reader in which order the criteria were applied. From a purely logical point of view it may look as if the same set of variants is selected, but to make the identification & selection procedure transparent and replicable, the reader needs to know which variants were excluded from the initial domain (and why). Cf. Dik's (1997: 135) discussion of the difference between 'Buddhist Japanese' and 'Japanese Buddhists' (also Dahl 1971). The only possible exception concerns the forms that have just one function (aka 'one-trick-ponies'; Rijkhoff 2010).

Selected references

- Comrie, Bernard et al. 2013. GNMCCs in Bezhta and Hinuq (Nakh-Dagestanian). Talk given at Association for Linguistic Typology (ALT) 10, 15-18 August 2013, Leipzig.
- Drummond, Chris. 2009. Replicability is not reproducibility: nor is it good science. *Proceedings of the Evaluation Methods for Machine Learning Workshop* at the 26th ICML, Montreal, Canada. National Research Council of Canada.
- Dryer, Matthew S. 1988. Object-verb order and adjective-noun order: Dispelling a myth. *Lingua* 74, 77-109.
- Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68-1, 81-138. Talk given at Association for Linguistic Typology 10, 15-18 August 2013, Leipzig
- Dryer, Matthew S. 2007a. Word order. In Timothy Shopen (ed.), *Language Typology and Syntactic Description: Clause Structure (Second Edition), Volume I*, 61-131. Cambridge: Cambridge University Press.
- Dryer, Matthew S. 2007b. Noun phrase structure. In Timothy Shopen (ed.), *Language Typology and Syntactic Description: Complex Constructions (Second Edition), Volume II*, 151-236. Cambridge: Cambridge University Press.
- Dryer, Matthew S. 2009. The Branching Direction Theory of word order correlations revisited. In S. Scalise et al. (eds.), *Universals of Language Today*, 185-208. Springer, Berlin.
- Dryer, Matthew S. 2014a. Order of Adjective and Noun. In M. Dryer and M. Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info/chapter/87>, Accessed on 2014-02-14.)
- Dryer, Matthew S. 2014b. Order of Relative Clause and Noun. In M. Dryer and M. Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info/chapter/90>, Accessed on 2014-02-26.)
- Evans, Nicholas and Stephen C. Levinson. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32-5, 429-448.
- Greenberg, Joseph H. 1966. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of Language* (2nd edition), 73-113. Cambridge: MIT.
- Halliday, Michael A. K. and Matthiessen, Christian M.I.M. 2004. *An introduction to Functional Grammar* (3rd ed.). London: Arnold.
- Haspelmath, Martin. 2007. Pre-established categories don't exist: Consequences for language description and typology. *Linguistic Typology* 11, 119-132.
- Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86-3, 663-87.
- Hawkins, John A. (1983). *Word order universals*. New York: Academic Press.
- Heine, Bernd. 1997. *Possession: Cognitive sources, forces, and grammaticalization*. Cambridge: Cambridge University Press
- Koptjevskaja-Tamm, Maria. 2002. Adnominal possession in the European languages: form and function. *Sprachtypologie und Universalienforschung* (STUF) 55 (2), 141-172.
- Koptjevskaja-Tamm, Maria. 2003. A woman of sin, a man of duty, and a hell of a mess: non-determiner genitives in Swedish. In *Noun phrase structure in the languages of Europe*, F. Plank (ed.), 515-558. Berlin: Mouton de Gruyter.
- Malchukov, A.L. 2000. Dependency reversal in noun-attributive constructions: towards a typology, Lincom, München.
- Malt, B. C., S. Gennari, M. Imai, E. Ameal, N. Saji, and S. Majid. (in press). Where are the concepts? What words can and can't reveal. In E. Margolis and S. Laurence (eds.), *Concepts: New directions*. Cambridge, MA: MIT Press.
- Matsumoto, Yoshiko. 1997. *Noun-modifying constructions in Japanese: a frame semantics approach*. Amsterdam: Benjamins.

- McCawley, James D. 1985. Review of Frederick J. Newmeyer (1983), *Grammatical theory: Its limits and possibilities*. *Language* 61-3, 668-679
- Newmeyer, Frederick J. 2010. On comparative concepts and descriptive categories: A reply to Haspelmath. *Language* 86, 688-95.
- Rijkhoff, J. 2004. *The Noun Phrase*. Oxford: Oxford University Press [expanded Pb of the 2002 Hb edition].
- Rijkhoff, Jan. 2008a. Synchronic and diachronic evidence for parallels between noun phrases and sentences. In F. Josephson and I. Söhrman (eds.), *Interdependence of Diachronic and Synchronic Analyses*, 13-42. Amsterdam and Philadelphia: Benjamins.
- Rijkhoff, Jan. 2008b. On flexible and rigid nouns. *Studies in Language* 32-3, 727-752.
- Rijkhoff, Jan. 2008c. Layers, levels and contexts in Functional Discourse Grammar. In D. García Velasco and J. Rijkhoff (eds.), *The Noun Phrase in Functional Discourse Grammar*, 63-115. Berlin and New York: Mouton de Gruyter.
- Rijkhoff, Jan. 2009a. On the (un)suitability of semantic categories. *Linguistic Typology* 13-1, 95-104.
- Rijkhoff, Jan. 2009b. On the co-variation between form and function of adnominal possessive modifiers in Dutch and English. In W. B. McGregor (ed.), *The Expression of Possession (The Expression of Cognitive Categories [ECC] 2)*, 51-106. Berlin and New York: Mouton de Gruyter.
- Rijkhoff, Jan. 2010. Functional categories in the noun phrase: on jacks-of-all-trades and one-trick-ponies in Danish, Dutch and German. *Deutsche Sprache* 2010 (Jg.38), nr. 2 (Special issue: *Modifikation im Deutschen: Kontrastive Untersuchungen zur Nominalphrase*), 97-123. Online: <http://pub.ids-mannheim.de/laufend/deusprach/ds10.html>
- Rijkhoff, Jan. (to appear 2014). Modification as a propositional act. In M. de los Ángeles Gómez González, et al. (eds.), *Theory and Practice in Functional-Cognitive Space..* Amsterdam: Benjamins.
- Rijkhoff, J., Eva van Lier (eds.). 2013. *Flexible Word Classes: typological studies of underspecified parts of speech*, Oxford University Press, Oxford.
- Rijkhoff, Jan and Dik Bakker. 1998. Language sampling. *Linguistic Typology* 2-3, 263-314.
- Ross, M., 1998. Proto-Oceanic adjectival categories and their morphosyntax. *Oceanic Linguistics* 37-1, 85-119.
- Stassen, Leon. 2010. The problem of cross-linguistic identification. In Jae Jung Song (ed.), *The Oxford Handbook of Linguistic Typology*, 90-99. Oxford: Oxford University Press.
- Thompson, Sandra. 1988. A discourse approach to the cross-linguistic category “adjective”. In J. A. Hawkins (Ed.), *Explaining Language Universals*, 167- 5. Oxford: Blackwell.
- Velde, Freek Van de. 2007. Interpersonal modification in the English noun phrase. *Functions of Language* 14-2, 203-230.
- Velde, Freek Van de. 2012. A structural-functional account of NP-internal mood. *Lingua* 122, 1-23.
- Williams, Edward. 1981. Argument structure and morphology. *The Linguistic Review* 1, 81-114.

Appendix 1: The cross-linguistic category **Article** (Dryer)

Consider also the way articles are defined (Dryer 2007a: 94): “In European languages, the term ‘article’ is used to denote words that code definiteness or indefiniteness and which in some languages vary with respect to other grammatical features of the noun phrase as well, such as case, gender, or number. Some languages elsewhere in the world employ words that do not vary for definiteness but which resemble articles in European languages in that they are words that are very common in noun phrases and which vary for grammatical features of the noun phrase (including number, case, gender), even if this does not include definiteness.” **Here it seems we the only real criterion is a formal one: being a very common constituent of the noun phrase (‘number, case, gender’ can also be expressed on other constituents of the NP).**

See also Rijkhoff (2004: 285): “In his study Dryer (1989a: 83) used two criteria to identify what he calls articles, but in my view the two groups of elements thus identified have very little in common. He counted as an article not only a word indicating (in)definiteness “or some related discourse notion”, but also a word serving as a noun phrase marker “in the sense that noun phrases in that language [...] typically occur with one of the words in question”. This resulted in a group of elements that are formally identical or similar to a demonstrative modifier or the numeral ‘one’ and another group whose members are quite distinct from a demonstrative modifier or the numeral ‘one’. The latter subgroup includes such varied items as noun (phrase) markers (as in the Fijian example below), noun classifiers (so that both *hune’* and *no’* count as articles in the Jacaltepec example below), pronouns (such as *rá* in the example from Jicaltepec Mixtec) and items such as the “polyvalent lexeme” *cái* in Vietnamese (Nguyễn Đình Hoà 1997: 180), which (it is true) some have called an article (or even “super article”; Bulteau 1953: 21) but whose exact meaning is far from clear. For example, it seems to serve as a classifier, as an emphasizing, individualizing, and particularizing element, as a modifier “to explain more clearly or to reinforce the meaning of the noun” or “to attract the reader’s attention to the noun” and it can be used to give the NP a pejorative connotation (Nguyễn Đình Hoà 1997: 179-180; cf. also Vũ Duy-Tù’ 1983: 30, 32).

Boumaa Fijian (Dryer 1989a: 84; Dixon 1988: 243)

(1) *sa tau-ra a drano o Boumaa*
 ASP hold ART lake ART Boumaa
 ‘Boumaa held the lake’

Jacaltepec (Dryer 1989a: 86; Craig 1977: 137)

(2) *hune’ no’ txitam*
 a CLF pig
 ‘a pig’

Jicaltepec Mixtec (Dryer 1989a: 93; Bradley 1970: 79)

(3) *číka ča?a ña sa?ma čì?ì rá ahili*
 thing-that give she clothes to he angel
 ‘That’s why she gave the clothes to the angel’

Vietnamese (Nguyễn Đình Hoà 1997: 180; Lê Văn Lý 1960: 213)

(4) *Cái con ngựa a ấy chạy nhanh thật*
 CÁI CLF horse that run fast real
 ‘That horse over there runs really fast’