Crosslinguistic categories in morphosyntactic typology: Problems and prospects

Jan Rijkhoff, Aarhus University (published in 2016 in: *Linguistic Typology* 20-2, 333-363)

Abstract: This article offers a new, transparent method to construe morphosyntactic categories for cross-linguistic research. It avoids the problem of categorial confusion attested in major post-Greenbergian studies in morphosyntactic typology, in particular in probabilistic typological investigations, which tend to mix up semantic and formal criteria and marginalize 'statistically insignificant' morphosyntactic variants. These and other problems are avoided by using functional criteria as the starting point in identifying comparable forms and constructions in different languages. Subsequently formal and semantic criteria are employed to arrive at a morphosyntactic category whose members are sufficiently similar in terms of function, form and meaning.

1. Introduction

Categorization is one of the most fundamental traits of human cognition (Harnad 2005), which makes it of central importance to the linguistic sciences, especially disciplines such as grammatical theory and morphosyntactic typology, both of which are concerned with linguistic forms and constructions used by speakers of different languages. Any discipline involved in cross-linguistic research is faced with the problem of cross-linguistic comparability (Song 2001: 10-15; Stassen 2011): how can we be sure that we are dealing with comparable linguistic units in the languages of the sample? The problem is not so much trying to identify linguistic units in the various languages whose grammatical properties are completely identical (no two linguistic forms or expressions are exactly the same grammatically, not even in a single language), but rather making sure that the units are similar enough to allow for a responsible cross-linguistic comparison.

Section 2 discusses some problems of traditional Greenbergian morphosyntactic typology, in particular problems that are due to the fact that morphosyntactic typology has used semantic criteria to identify and process members of formal, i.e. morphosyntactic categories, in a cross-linguistic investigation.[1] Since semantic categories contain members of different morphosyntactic categories (as illustrated in (1) and (2) below), studies in Greenbergian typology have suggested that the formal or structural properties of the members of a semantic category are basically irrelevant in morphosyntactic typology. This is a problematic assumption: as the name indicates, morphosyntactic typology concerns itself with formal rather than semantic properties of morphosyntactic units and members of different morphosyntactic categories (e.g. Adjective, Relative clause) usually come with their own distinct set of grammatical properties, also when they are deemed to belong to the same semantic category (see Table 2 in Section 4.2.2).

Section 3 argues that we need a third kind of category to do morphosyntactic typology, namely interpersonal ('communicative', 'discourse') functional categories, which should be used to make the first cut in a cross-linguistic investigation, as functional categories come closest to having universal applicability. It is difficult to image a language whose speakers do without the various kinds of speech acts (illocutionary acts, propositional acts etc.), each of

---

[1] The notions 'formal' and 'morphosyntactic' are often used more or less interchangeably. Here I will use 'formal' in connection with morphosyntactic, phonological and certain other CRITERIA (also including, for example, frequency) used to constrain membership of a crosslinguistic category. The qualification 'morphosyntactic' in 'morphosyntactic unit' or 'morphosyntactic category' is reserved for the linguistic unit that is the focus of the morphosyntactic investigation.

which results in a functional category in the current proposal. Since there is at present no list of functional categories, a relatively large part of this article will be devoted to a tentative typology of functional categories.[2]

Section 4 sketches a new method to set up morphosyntactic categories for cross-linguistic investigations, which involves the application of formal and semantic criteria within an all-inclusive functional category. According to this method, it is the interpersonal function of a member of a morphosyntactic category that serves as the basis in setting up a cross-linguistic category. The bottom line for reliable, transparent and reproducible morphosyntactic typological research is this: so as to make it possible to reproduce the result of a study in morphosyntactic typology based on data taken from a representative variety sample (see below on the importance of using language samples that are maximally diverse), it should be clear which functional, formal and semantic criteria were used in setting up the cross-linguistic category.

2. Categorization in morphosyntactic typology: semantic or formal categories?

In his famous study on universals of grammar, Greenberg (1966: 74) basically applied semantic criteria to identify members of cross-linguistic categories (such as Subject, Noun, Adjective, Genitive, Relative Clause) and the same procedure was followed by Hawkins (1983: 12) and Dryer (1992: 120) in their post-Greenbergian word order studies. This implies, for example, that the semantic category Adjective in these studies also included verbal and nominal expressions of adjectival notions (such as relative clauses and genitives), which are typically used by speakers of languages that lack a dedicated class of adjectives:

(1)   Kiribati (a.k.a. Gilbertese; Oceanic)
       te   uee   ae   e   tikiraoi   (relative clause)
       ART flower REL 3SG.S be.pretty
       'a pretty flower' (lit. 'a flower that pretties') (Ross 1998: 90)

(2)   Makwe (Bantu)
       muú-nu   w-á=ki-búúli   (genitive)
       NC1-person   PP1-GEN=NC7-silence
       'a silent person' (lit. 'a person of silence') (Devos 2008: 136)

Confusingly, Relative Clause and Genitive are also semantic categories in their own right in the three studies mentioned above and since the semantic criteria for category membership are not specified, we do not know when relative clauses and genitives were regarded as members of the semantic category Adjective or when they were categorized as members of the semantic categories Relative Clause or Genitive.

When subsequently Greenberg, Hawkins and Dryer put forward universals, principles or theories on the basis of the cross-linguistic data they collected, the fact that their meaning-based categories contained a variety of morphosyntactic variants (usually with their own set of grammatical properties) lead to a number of difficulties. Greenberg was not unaware of the problem, as shown in his discussion of Universal 15 on comparatives, which is concerned with the order of Standard, Marker and Adjective (e.g. English 'taller$_{Adjective}$ than$_{Marker}$ Sarah$_{Standard}$').[3] He wrote that he had to exclude certain members of the semantic category Comparative from his investigation on formal grounds, because in quite a few languages the

---

[2] Formal theories of grammar also employ what are called 'functional categories', but this is merely another label for grammatical (i.e. non-lexical) elements like affixes and particles (Muysken 2008).
[3] Notice that Greenberg (1966: 108) distinguished between 'participle of adjective-verb' and other expressions of adjectival notions in Burmese.

morphosyntactic properties of the comparative construction deviated so much from the English model, that he found it impossible to formulate a general statement about the order of Standard, Marker and Adjective for all the languages in his sample (Greenberg 1966: 88; also Rijkhoff 2009a).

The problem caused by mixing up categories is perhaps best illustrated in Hawkins's investigation of word order patterns. He, too, had claimed that semantic criteria 'suffice[d] to make the cross-linguistic equation' (Hawkins 1983: 12), but when he proposed his *Heaviness Serialization Principle*, the supposedly semantic category labels Adjective, Genitive and Relative Clause (but also e.g. Demonstrative and Numeral) appeared to have changed into labels for formal categories. Hawkins defined 'heaviness' in terms of such non-semantic criteria as (a) length and quantity of morphemes, (b) quantity of words, (c) syntactic depth of branching nodes, and (d) inclusion of dominated constituents (Hawkins 1983: 90-91).

(3)   *Heaviness Serialization Principle*: Rel $\geq_R$ Gen $\geq_R$ A $\geq_R$ Dem/Num

Thus a member of the (semantic? formal?) category Relative Clause is 'heavier' than a member of the (semantic? formal?) category Adjective. But Hawkins's semantic category Adjective must also have included members of the heavier formal categories Genitive and Relative Clause (see (1) and (2) above), so it is not clear if or how the original members of the single semantic category Adjective were later re-categorized and distributed over the formal categories Adjective, Genitive and Relative Clause (which also remained undefined) before Hawkins proposed his *Heaviness Serialization Principle*.

Dryer (1992: 120) also followed Greenberg's method of semantic categorization in his investigation of Greenbergian word order correlations. In this study, he proposed the Branching Direction Theory, which refers to a structural feature of the internal syntactic organization of a constituent, to explain why only some ordering combinations correlate with the order of object and verb. According to the Branching Direction Theory, only the position of members of phrasal or branching categories (such as relative clauses and genitives) correlates with OV or VO order (Dryer 1992: 107-111). Since adjectives are regarded as non-branching elements in the Branching Direction Theory, their position relative to the head noun should not correlate OV or VO order. In this case, too, the semantic category Adjective must also have included members of the formal categories Genitive and Relative Clause (see examples above). However, the author does not explain what happened to the branching members of the erstwhile(?) semantic category Adjective when this category became the formal (non-branching) category Adjective that is part of the Branching Direction Theory. It was only in a recent message on LingTyp (January 19, 2016) that the author wrote that he found that the formal realization of members of a semantic category (in this case Adjective) was 'overall irrelevant' in determining statistically significant word order correlations.[4]

It is important to point out here that what is statistically insignificant in probabilistic morphosyntactic typology, is highly relevant for a typology-based theory such as Functional

---

[4] Space limitations prevent me from addressing the question whether probabilistic typology, which requires that the sample is free of (genetic, geographic etc.) bias, is at all possible (cf. Maslova 2000). The general problem was formulated as follows in Rijkhoff & Bakker (1998: 265): "The requirement that the languages be independent units makes it rather difficult to construct a good probability sample. For example, even in a relatively small sample it is practically impossible to avoid the inclusion of languages that are not somehow genetically related or spoken in the same region. Several attempts have been made to deal with this problem (…), but basically there are only two ways out. Either a small sample is used which, however, is not quite representative with respect to the genetic, areal, and/or cultural diversity (…). Or a large sample is used and genetic, areal, and/or cultural relationships are manipulated so as to meet the requirements on statistical tests (e.g. Dryer 1992: 83). Essentially, however, there does not seem to be a real solution." See Perkins (2001: 432), for a critical assessment of the Dryer's use of statistical methods.

Discourse Grammar (Hengeveld & Mackenzie 2008), whose aim it is to describe and explain both the frequent and the less frequent grammatical phenomena in all languages (Rijkhoff 2002, 2010a). A theory that can account for both common and unusual phenomena is superior to a theory that can only handle frequently attested grammatical phenomena. This is, of course, also the reason that Rijkhoff et al. 1993 and Rijkhoff and Bakker 1998 developed a sampling method that is designed to produce a maximally diverse language sample or 'variety sample'. In a variety sample (as opposed to a probability sample) it is very important to have cases of the rarest type, since "exceptional types test the theory" (Perkins 1988: 367).

   In sum, traditional (in particular 'probabilistic') morphosyntactic typology has some important drawbacks:

(a) it suggests that semantic categories or criteria can be used to investigate morphosyntactic properties of constituents, even though a semantic category typically contains members of different formal categories, each of which has its own distinct set of grammatical properties;
(b) it often fails to explain how members of different formal categories are identified as members of the same semantic category and subsequently processed in order to formulate statistical statements about morphosyntactic properties such as constituent order;
(c) it does not always clearly distinguish between semantic and formal categories;
(d) its focus on dominant patterns or statistically significant combinations of ordering patterns has the undesirable consequence that the less frequent or statistically insignificant phenomena tend to be ignored in linguistic typology (for a notable exception, see Wohlgemuth & Cysouw (eds.) 2010a and 2010b).
(e) due to the opaque way members of a cross-linguistic category tend to be identified and processed in morphosyntactic typology, it is usually difficult if not impossible to reproduce the outcome of a typological investigation.

Notice, finally, that categorial confusion in morphosyntactic typology has also led to terminological confusion, as Greenberg's meaning-based cross-linguistic categories have subsequently been referred to as, for example, 'comparative concepts' (Haspelmath 2007: 126) and 'functional' or 'functional/semantic concepts' (Croft 2014; see also Croft 2001). I will show below that nothing is gained by conflating different kinds of categories or by making matters more abstract than is strictly necessary in typology, which is, after all, an empirical science. Linguistic typology needs clear categories for comparative purposes rather than vague concepts, especially since no one seems to know what a concept is (Machery 2009, Malt 2010). Morphosyntactic typology should use methods that allow us to identify and process actual linguistic forms and constructions that are similar enough in terms of function, meaning and form to make a responsible and reproducible cross-linguistic study.[5] I will propose such a procedure in Section 4. We shall see below that this also means that we should abandon the idea that we can compare everything in all languages.[6]

---

[5] Cf. Smith & Samuelson (1997: 190): "A successful theory of categories (...) might require that we give up timeless abstractions such as concepts."
[6] The problems discussed here are at least partly due to the fact that many typologists work in a theoretical vacuum. Charles Darwin already observed in a letter to Henry Fawcett (10 September 1861) that empirical research without a theoretical perspective is rather pointless. On this topic, see also e.g. Hale (1997: 28): "*I think you have to have a theory. You have to be operating within some kind of theoretical framework to ask any meaningful questions about language. Languages make sense, but only in terms of theory.*"

3. Cross-linguistic comparison requires functional, semantic and formal categories

In morphosyntactic typology, it is not enough that members of the same cross-linguistic category are similar enough with regard to their formal or structural properties. We also have to make sure that the members of the same morphosyntactic category (e.g. Adjective) share sufficient semantic and functional properties to make a fair and responsible comparison. The reason for this is, of course, that the same morphosyntactic unit (affix, word, phrase etc.) can have different meanings and functions in actual discourse, which can have an impact on its grammatical behavior (Section 3.1). Compare, for example, (a) 'Bob really is a poet' and (b) 'Really, Bob is a poet!' (Kaltenböck et al. 2011: 860). In the (a) example, the adverb 'really' is a modifier within a declarative illocution (Sections 3.2.1 and 3.2.3), whereas in the (b) example it serves as thetical constituent, i.e. it is not an integral constituent of the central illocution (see Section 3.2.2 for more details on theticals). Or compare the following examples of adnominal PPs with the preposition 'of': (a) 'the owner OF THE HOUSE', (b) 'statues OF THIS QUALITY', (c) 'a crown OF PURE GOLD', (d) 'a man OF MANY FACES', (e) 'a woman OF (GREAT) INFLUENCE', (f) 'a man OF THE WORLD' (I am ignoring several other variants here, such as 'the roof OF THE HOUSE', 'the City OF LONDON'; cf. Heine 1997 on the semantic and functional versatility of possessive constructions). Even though from a formal perspective we are dealing with the same kind of morphosyntactic unit (an adnominal prepositional phrase with 'of'), in each example it has different semantic and functional (incl. textual) properties; this is discussed in considerable detail in Rijkhoff 2009b.

The approach outlined in this article is based on the observation that the grammatical behavior of a morphosyntactic unit is basically determined by formal, semantic and functional (i.e. 'interpersonal', 'communicative' or 'discourse') factors. Morphosyntactic units are members of lexical and grammatical word classes, phrasal and clausal structures like noun phrases, subordinate clauses, complement clauses and members of morphological units such as prefixes and other affixes (phonological units are largely ignored here, but see e.g. Smith 2015 on phonological properties of the major word classes). Morphosyntactic units or categories are the subject matter of morphosyntactic typology. Depending on the research question, members of these morphosyntactic categories must be further specified for functional, formal and semantic properties so as to be able to create a cross-linguistic category whose members are similar to allow for a proper typological investigation.

The role of formal and semantic factors, and consequently the importance of employing form and meaning-based categories in cross-linguistic morphosyntactic investigations, is briefly illustrated in Section 3.1. Since there is no generally agreed upon list of interpersonal categories, I will propose such a list in Section 3.2. Section 4 briefly demonstrates how we can set up cross-linguistic categories suitable for morphosyntactic typology and typology-based approaches to grammar that aim to provide a model that can be used to describe and explain actual linguistic expressions in individual languages.

3.1. Formal and semantic factors

Morphosyntactic units are characterized by various kinds of formal properties, also including, for example, frequency of occurrence and degree of internal complexity, both of which can be relevant in a comparison of members of some morphosyntactic category in different languages. For example, the preferred or unmarked position of a constituent can depend on its 'morphosyntactic weight', as shown in the following examples.

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

The a. sentence, with the short phrase 'to her' before the longer, more complex constituent 'the paintings that my uncle left to me as part of his inheritance' is clearly preferred to the b. sentence. To explain this ordering preference, we only need to refer to a formal property: MORPHOSYNTACTIC COMPLEXITY, i.e. the relative degree of internal complexity of the linguistic unit (cf. Behaghel's *Gesetz der wachsenden Glieder* 'Law of Increasing Terms' already proposed in 1932; see also Hawkins' *Heaviness Serialization Principle* mentioned earlier and Dik's (1997 I: 411-413) *Language Independent Preferred Order of Constituents* or LIPOC). In other words, when investigating constituent order in the languages of the world, one should take into consideration their morphosyntactic weight so as to avoid comparing constituents of varying morphosyntactic complexity, which might otherwise compromise the reliability of the typological investigation.

Some examples of semantic factors relevant for a cross-linguistic morphosyntactic investigation are (a) semantic roles for participants and props that are part of an event (e.g. AGENT, PATIENT, RECIPIENT), (b) semantic features that are typically associated with certain entity types, such as COLOUR, SIZE, VOLUME, WEIGHT, ±COUNTABLE (i.e. ±SHAPE) or ±ANIMATE for first order entities ('concrete objects') and ±TELIC, ±CONTROL or ±DYNAMIC for second order entities ('events').

Let us have a closer look at a semantic category that has a significant influence on grammars of languages across the globe, viz. ANIMACY. The relevance of this semantic feature has been discussed in many places (e.g. Silverstein 1976; Comrie 1981: 178-193; Rijkhoff 2004: 112-114), showing the effects of (degree of) ANIMACY on such divers grammatical phenomena as differential object marking, split-intransitivity, number marking, and alignment systems (esp. in so-called active alignment systems). Below is an example from Kannauri, showing the effect of animacy on constituent order. When both objects (child, maternal aunt) are animate, as in (5a), the direct object precedes the indirect object. However, when the direct object is inanimate, as in (5b), the phrase referring to the animate entity (uncle) must precede the phrase referring to the inanimate entity (letter) in the unmarked order.

(5)   Kannauri (Tibeto-Burman)
  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' (Kittalä 2006: 2; taken from Sharma 1988: 79)

Membership of a semantic (sub)category is often more a matter of degree than a simple case of inclusion vs. exclusion. For example, the Navajo people rank entities according to the following hierarchy of animacy (Young & Morgan 1987: 65-66):

(6) Human > Infant/Big Animal > Medium-sized Animal > Small Animal > Natural Force > Abstraction

This ranking is observed in the ordering of sentence constituents (as well as verb morphology) in that the phrase referring to the more animate entity must precede the phrase referring to the entity that is lower in the animacy hierarchy. Hence (7b), with the 'direct-construction' (DIR), is considered ungrammatical, as 'girl' outranks 'bird' in animacy and should therefore come first, as in (7a), with the 'inverse-construction' (INV).

(7) Navajo (Na-Dene)
　　a. *At'ééd tsídii bi-shtąsh*　　　　b. **Tsídii at'ééd yi-shtąsh*
　　　 girl bird INV-pecked　　　　　　 bird girl DIR-pecked
　　　 'The girl was pecked by the bird'　　 'The bird pecked the girl'
　　　 (Young & Morgan 1987: 65-66)

The gradual nature of semantic properties can also be observed in other linguistic domains, showing that people speaking different languages may disagree on the criteria that determine category membership. For example, in the semantic domain of ±Realis (Elliott 2000) or ±Actual (Chung and Timberlake 1985: 241f.) we see that the same kind of event whose occurrence counts as ACTUAL (REALIS) in one language can be marked as NON-ACTUAL (IRREALIS) in another language, and vice versa.

　　The fact that people may take different decisions in matters of categorization was already anticipated by the British philosopher John Locke (1689/1825: 322). In 'An essay concerning human understanding' he concluded that all categories are the products of our cognitive system, i.e. without us humans, there would be no categories:

　　"Men determine the sorts of substances, which may be sorted variously. From what has been said, it is evident that men make sorts of things. For, it being different essences alone that make different species, it is plain that they who make those abstract ideas which are the nominal essences do thereby make the species, or sort. ...
　　… Nature makes many particular things, which do agree one with another in many sensible qualities, and probably too in their internal frame and constitution: but it is not this real essence that distinguishes them into species; it is men who, taking occasion from the qualities they find united in them, and wherein they observe often several individuals to agree, range them into sorts, in order to their naming, for the convenience of comprehensive signs; …"

In other words, since it is humans who do the categorization, we decide what is contained in the category ADJECTIVE or how we rank entities on a scale of animacy or actuality. How we define categories or which features count as 'relevant' or 'necessary' depends on one's theory, goals, method, data and, last but not least, cultural factors, so disagreement is an inherent aspect of the categorization enterprise. It is probably easier to agree on what counts as a member of the category BIRD (but see Bulmer 1967) than to agree on what counts as a member of the category ADJECTIVE - but this is just a matter of degree.[7]

3.2. Members of functional categories are the products of a speech act
Since morphosyntactic units such as clauses, phrases, words and free or bound morphemes can be characterized in terms of formal, semantic and functional properties, they can simultaneously belong to a formal, a semantic and a functional category.[8] Functional categorization is not directly concerned with the formal or semantic properties of a linguistic unit, but rather with the actual job of a form or construction in the process of verbal

---

[7] The same holds for members of formal categories. For instance, in the case of verbal nouns it is impossible to draw a hard and fast line between verbs and nouns (Ross 1972, Sasse 2001, Aarts 2004).
[8] The most obvious exceptions to this general rule are (a) zero expressions, which have a meaning and a discourse function, but no phonological form, and (b) empty forms, i.e. forms without a meaning or a discourse function. Empty forms can come in various guises, for example: (i) linking morphemes (e.g. 's' in *hunt-s-man*), (ii) so-called cranberry morphemes (named after the meaningless element 'cran' in 'cranberry', and (iii) retinue or servant words (cf. Haiman 2013 on meaningless, decorative morphology).

communication. Membership of an interpersonal functional category is determined by the way a linguistic unit is used in actual discourse (hence the alternative label 'discourse unit'). The functional approach to categorization has a long tradition, going back to the Prague School of Linguistics (if not earlier, cf. Weil 1844: 19), whose members were "seeking to understand what jobs the various components were doing […]" (Sampson 1980: 104). A considerable amount of literature has be devoted to pragmatic or functional processes, forces, factors and principles, but so far this has not resulted in a more or less comprehensive classification of the interpersonal ('communicative', 'discourse') functions that are served by the various kinds of linguistic units in languages across the globe. Hence a tentative list of functional categories is proposed below.

Functional categories are here defined as categories whose members are the product of an interpersonal speech act, i.e. a speech act that first and foremost concerns the role of a linguistic unit in the interaction between speaker and addressee from a communicative or discourse perspective. Speech acts are commonly divided into locutionary acts (the act of saying something), illocutionary acts (the act of providing an utterance with a communicative intention, e.g. an assertion, a question or a command), propositional acts (the acts of referring, predicating, and, as we shall see below, modifying), and perlocutionary acts (the actual effect of the speech act).[9] For our purposes we only need to concern ourselves with illocutionary acts and propositional acts, but we need two additional kinds of speech acts to account for functional categories across languages: thetical acts and pragmatic acts. Thus there are four main types of interpersonal functional categories, which I will call ILLOCUTIONS, THETICALS, PROPOSITIONALS and PRAGMATICALS for ease of reference. Each of these functional categories is briefly introduced below.

3.2.1. Functional categories I: linguistic units that are the products of an illocutionary act ('illocutions')

Illocutionary acts, introduced in Austin 1962 as part of his speech act theory, result in linguistic units which have a specific communicative function usually called 'illocutionary force or value', which is captured by labels such as Declarative, Interrogative or Imperative.

Each type of illocution comes with its own set of grammatical properties or 'mini-grammar', involving syntactic, morphological and/or prosodic features. For example, Nama Hottentot has dedicated particles to mark different kinds of illocutionary acts, whereas speakers of Greenlandic Eskimo use special affixes for this purpose:

(8)   Greenlandic Eskimo (Inuit)
        a. *Iga-voq*   b. *Iga-va*
           cook-DEC.3.SG   cook-INT.3.SG
           'He cooks'   'Does he cook?' (Sadock & Zwicky 1985: 167)

Notice that illocutions do not necessarily take the form of a complete clause, as in the case of the normal answer to a question such as 'Where is she?', where a simple 'In her office' suffices as a complete, declarative response. This implies that the formal expression of the illocution should be determined at an early stage (e.g. a complete clause or just the focal part, as in the example above) to make sure that the cross-linguistic category will contain comparable forms or structures (more on this below). For a recent overview of the grammatical properties of illocutions, I refer to König & Siemund 2007.

---

[9] The notion 'act' is in itself ambiguous between an activity and an abstract object. In this article, I focus on the activity interpretation, because it emphasizes the actional status of the speech act.

3.2.2. Functional categories II: linguistic units that are the products of a thetical act ('theticals')

Theticals cover more or less the same linguistic expressions as Parentheticals (Kaltenböck et al. 2011: 855-856) or what Dik (1997 II: 379-407) called Extra-Clausal Constituents in his *Theory of Functional Grammar* (now called Subsidiary Discourse Acts in its successor *Functional Discourse Grammar*; Hengeveld & Mackenzie 2008: 47, 53-56). A recent discussion of theticals is provided by Kaltenböck et al. (2011: 857), who list the following set of characteristic grammatical properties: (a) they are syntactically independent, (b) they are set off prosodically from the rest of the clause, (c) their meaning is "non-restrictive", (d) they tend to be positionally mobile, and (e) their internal structure is built on principles of sentence grammar but can be elliptic (see also Heine et al. 2014).

Several subtypes of theticals can be distinguished, such as Greetings and Leave-takings, Summonses and Addresses (Dik 1997: 384-5); some other subtypes are Motivation, Concession, Orientation, and Correction (Hengeveld & Mackenzie 2008: 53). Space limitations prevent me from offering a more detailed presentation, so I will end this section with some examples taken from Kaltenböck et al. 2011 and Dik 1997:

(9)   'The main point — WHY NOT HAVE A SEAT? — is outlined in the middle paragraph.'
(10)   'Is he a friend of yours, THAT GUY?'
(11)   'ANYWAY, what should we tell her when she comes back?'
(12)   'BETWEEN YOU AND ME, he failed the exam.'
(13)   'Mary — DON'T FORGET — is coming over to visit.'

3.2.3. Functional categories III: units that result from propositional acts ('propositionals')

Propositional acts (Searle 1969) come in three major subtypes, Acts of Predication, Acts of Reference and Acts of Modification, resulting in three types of propositionals: (i) verbal and non-verbal Predicates, (ii) Referential Phrases (such as NPs and complement clauses) and (iii) Modifiers.[10] Searle focused on acts of predication and reference; acts of modification are discussed in Croft 1990 and Rijkhoff 2014. Like other functional categories, Predicates, Referential Phrases, and Modifiers are characterized by their own grammatical properties. Here I will focus on Modifiers, because members of this functional category will be used to illustrate the new method to set up morphosyntactic categories for cross-linguistic purposes (Section 4). Modifiers enrich or supplement core linguistic material that is used in acts of predication or reference and are typically associated with formal categories like Adjective, Adverb(ia)l and other attributive, more or less optional linguistic forms and constructions in a noun phrase or a sentence.

Especially in the case of Modifiers, it is important to emphasize that there is usually no one-to-one relationship between FORM and FUNCTION (cf. Table 2): the same linguistic form or construction can often be used in more than one modifier function, and vice versa, different linguistic forms or constructions can be used in the same modifier function. This is discussed in considerable detail in Rijkhoff 2009b and 2010, basically showing that interpersonal functions come with their own set of grammatical properties or 'mini-grammar'.[11] Notice furthermore that the functional category Modifier includes both lexical and grammatical material. For example, both the English past tense suffix *-ed* and the adverb 'yesterday' serve

---

[10] Traditionally, only definite NPs are deemed to have referential force (Reboul 2001: 515; García and Rijkhoff 2008: 15). For this reason, Dik (1997 I: 130) distinguished between acts of constructing reference (resulting in indefinite NPs) and acts of identifying reference (definite NPs).

[11] So-called ONE-TRICK-PONIES, linguistic units that are invariably used in same interpersonal function ('one form, one function') are an important exception to this general rule (Rijkhoff 2010b).

as localizing modifiers: they inform the addressee where to locate an event in time (see below).

In a layered representation of noun phrases and clauses, members of functional modifier categories are distributed over nested layers around the head constituent (typically V in the clause and N in the noun phrase), reflecting differences in SEMANTIC SCOPE between the various subtypes of modifiers.[12] At least five modifier subcategories can be distinguished, which can be used to analyze both clauses and NPs, here ranked according to their scopal properties: CLASSIFYING MODIFIERS have the narrowest scope (only the head constituent) and DISCOURSE-REFERENTIAL MODIFIERS have the widest scope (more detailed presentations concerning the role of modifiers in a layered analysis of NPs and clauses can be found in, for example, Rijkhoff 2008b, 2010, 2014).

1. CLASSIFYING MODIFIERS specify a subtype of the kind of entity that is denoted by the head constituent (like 'presidential' in 'a PRESIDENTIAL election'; so-called stripped nouns (Miner 1986) are good examples of free classifying modifiers in the clause);
2. QUALIFYING MODIFIERS specify more or less inherent properties or 'qualities' (e.g. 'BLACK cars$_N$', 'walk$_V$ FAST');
3. QUANTIFYING MODIFIERS specify quantitative properties such as the number or cardinality of the thing or event (e.g. 'TWO cars$_N$' or 'fall$_V$ TWICE');
4. LOCALIZING OR ANCHORING MODIFIERS specify locative properties, thus making the referent (object or event) locatable and hence identifiable for the addressee (e.g. 'THAT house$_N$ ON THE CORNER', 'met$_V$ IN KONSTANZ');
5. DISCOURSE-REFERENTIAL MODIFIERS such as (in)definite articles specify the status of a thing or event as a discourse entity; (ir)realis markers are good examples of discourse-referential modifiers in the clause (Rijkhoff & Seibt 2005).

The layered model of the clause includes two additional layers for modal and illocutionary modifiers. Van de Velde (2007, 2012) has argued that a complete layered model of the NP should also provide slots for modal and illocutionary modifiers.

3.2.4. Functional categories IV: linguistic units that result from pragmatic acts ('pragmaticals')

Pragmaticals are linguistic units that are marked for their information value. Topic and Focus are the two main subtypes within this functional category. Dik (1997 I: 313, 327) only counts units as pragmaticals when they receive special grammatical treatment in a language, as when it gets
- a special form or is expressed in a special syntactic position;
- a special marker signaling its pragmatic status;
- a special prosodic contour;
or when it is expressed by some other special construction type.

For an overview of the various (sub)types of Topic and Focus, I refer to Dik (1997 I: 309-338) and Hengeveld & Mackenzie (2008: 89f.); for a discussion of the grammatical effects of information marking in general, see Foley & Van Valin 1985.

---

[12] A layered analysis for sentences was introduced in the context of Dik's *Functional Grammar* in Hengeveld 1989. Subsequently Rijkhoff 1990 proposed a unified analysis for a layered representation of noun phrases and sentences (see Rijkhoff (2004, 2008a, 2014) and Rijkhoff & Seibt 2005 for further developments).

3.3. Functional categories: summing up

There are four main types of functional categories, whose members consist of linguistic units that are the product of an interpersonal act:

Speech Act	Functional Category	Main Subtypes
Illocutionary Act	ILLOCUTIONS	DECLARATIVE, IMPERATIVE, INTERROGATIVE, …
Thetical Act	THETICALS	ADDRESS, SUMMONS, GREETING, LEAVE-TAKING, AFTERTHOUGHT, …
Propositional Act	PROPOSITIONALS	PREDICATE, REFERENTIAL PHRASE ('NOUN PHRASE', COMPLEMENT CLAUSE, ETC.), MODIFIER (ADJECTIVE, RELATIVE CLAUSE, GENITIVE, ADVERB, ADVERBIAL, PREPOSITIONAL PHRASE, ARTICLE, DEMONSTRATIVE, NUMERAL; ALSO TAM AFFIXES, NUMBER AFFIXES, …)
Pragmatic Act	PRAGMATICALS	TOPIC, FOCUS

Table 1. Speech acts, functional categories and some of their main subtypes.

Notice that these four functional categories basically cover all the linguistic material that is part of a complete utterance. Dedicated markers of interpersonal acts (i.e. special markers of the various kinds of functional categories) are regarded as part of the mini-grammar of the members of the functional categories associated with these acts. Some examples are
- distinct sets of illocutionary particles, as in e.g. Nama Hottentot (Hagman 1973: 257);
- dedicated markers of Acts of Predication ('predicate markers'), Reference ('noun phrase markers') and Modification ('ligatures', 'modification markers'), as attested in Austronesian languages (Foley 1976, Adelaar & Himmelmann eds. 2005):
- specialized markers of pragmaticals ('Topic markers', 'Focus markers'), as in Japanese.

4. How to set up a cross-linguistic category for morphosyntactic typology: outlining a new approach

The previous section showed that members of a morphosyntactic category can be categorized in terms of their functional, semantic and formal properties. This section sketches how functional, formal and semantic criteria can be used to establish a cross-linguistic category whose members are similar enough to allow for a responsible morphosyntactic investigation.

For the purpose of this demonstration, we shall be concerned with cross-linguistic categories for a typological investigation whose goal it is to investigate the order of modifiers in the noun phrase. This is not precise enough, of course, so we have to be more specific. For reasons explained below, we will first choose a subset of functional modifier categories, for example: only qualifying, quantifying and localizing modifiers (Section 3.2.3). But since each function can be expressed by various morphosyntactic variants and since we must end up with comparable forms and constructions, we will have to decide for each functional modifier category which particular morphosyntactic category will be selected for the investigation. The current paper is not the place to demonstrate in detail how one can establish valid cross-linguistic categories for each of the three functional categories mentioned above, so I will just concentrate on setting up a cross-linguistic category of Qualifying Modifier whose members are similar enough in function, form and meaning to allow for a responsible comparison in the area of morphosyntactic typology.

4.1. Stage I: getting an overview of morphosyntactic variation with a functional category
There are at least two reasons why a morphosyntactic investigation should start with the selection of a functional category (even with a particular form or construction already in mind; see below). Firstly, functional categories have the widest, perhaps even universal application (Sections 1 and 3). It is difficult to imagine a natural human language whose speakers do without illocutionary acts, thetical acts, propositional acts and pragmatic acts. The four main functional categories used in the method outlined here are the linguistic products of the speech acts mentioned above: (i) ILLOCUTIONS, (ii) THETICALS, (iii) PROPOSITIONALS (including PREDICATES, REFERENTIALS and MODIFIERS) and (iv) PRAGMATICALS. This should not imply that all languages make the same distinctions within these four main categories: languages may differ with regard to the number and kind of subtypes that are recognized within each of the four major functional categories.

The second reason why we should prioritize functional categories in morphosyntactic typology is that these categories give the typologist an indispensible, cross-linguistic overview of the various forms and constructions that can be used in the same communicative function. After all, it is one of the main goals of linguistic typology to investigate linguistic diversity. As we saw above, however, the less frequent (and consequently: 'statistically insignificant') variants are at best reduced to marginal linguistic phenomena in probabilistic typology. By contrast, the statistically insignificant forms and constructions are at least as important as the frequently attested variants in non-probabilistic approaches to typology and in typologically informed grammatical theories. Having a proper cross-linguistic overview of the morphosyntactic units that can be used in the same communicative function (e.g. Qualifying Modifier) before a particular morphosyntactic variant is chosen also prevents the typologist from selecting one of the familiar morphosyntactic variants as attested in the better studied European languages, thus implicitly introducing a Eurocentric bias in typological investigations. This does not necessarily imply that the researcher should start without a clear idea of the goal of the morphosyntactic investigation; it does mean, however, that the typologist should first find out more about the range of morphosyntactic variants within a certain functional category and report about the frequency and distribution of these variants before concentrating on one particular variant.

To illustrate the point of departure according to the new approach, we will concentrate on members of the functional category of Qualifying Modifier, which were characterized in Section 3.2.3 as morphosyntactic units that specify more or less inherent properties ('qualities') of the head constituent (e.g. 'BLACK cars$_N$', 'walk$_V$ FAST'; see also Section 4.2.4 on the role of entity types). Due to the fact that the function of Qualifying Modifier covers a great variety of forms and constructions both in the clause (e.g. adverbs, adverbials) and in the NP (e.g. adjectives, relative clauses, prepositional phrases; see below for some examples from Igbo and Kwaza), we must first limit the pool of potential members of the cross-linguistic category by adding a functional criterion. Since we are ultimately interested in the order of certain modifiers in the noun phrase in languages spoken across the globe, we will concentrate on Qualifying Modifiers used in a Referential Act (this should exclude modifiers at the level of the clause). This is not sufficient, as we shall see below, so in the next phase we will add semantic and formal criteria to further limit the kind of morphosyntactic variants that can be a member of the cross-linguistic category of Qualifying Modifier.

4.2. Stage 2: applying semantic and formal criteria
In this section we will apply additional criteria in establishing a proper cross-linguistic category for qualifying modifiers. We should point out that the order in which these criteria are applied may differ depending on the nature of the typological investigation. Sometimes it is important that certain criteria are applied before other criteria, in other cases this does not seem to make a real difference (cf. Stassen 2011: 97-98). In practice that can mean that the correct order of

application of the various kinds of formal and semantic criteria will be a matter of trial and error before the researcher has arrived at a group of with morphosyntactic units that allow for a fair and responsible cross-linguistic comparison.

4.2.1. Focusing on Qualifying Modifiers in integral (hierarchical) noun phrases

Now that we have set functional limits on potential members of the cross-linguistic category, we can make an initial inventory of the forms and constructions speakers of different languages use in the function of Qualifying Modifier as part of a Referential Act. Referential Acts, however, can formally manifest themselves in several ways, for example as a pronoun ('she'), a proper name ('Mette'), a noun phrase ('this old car'), a complement clause ('that he sold his house') or a headless relative clause ('what you just said'). For the purpose of the current demonstration, I will limit the typological investigation to qualifying modifiers in referential acts that formally result in a noun phrase, a kind of referential unit that can actually contain members of the category Qualifying Modifier. Notice that this formal criterion excludes languages that, strictly speaking, have no noun phrases, such as language without a clear verb-noun distinction (Hengeveld et al. 2004; Rijkhoff 2004: 12-14; Rijkhoff & van Lier eds. 2013). This is simply another reminder of the fact that we cannot compare everything in all languages.

We should also be careful to distinguish between languages with integral (i.e. hierarchically organized) NPs and languages where what would be a qualifying modifier in a hierarchically organized noun phrase (an 'integral NP') is formally expressed as an appositional element, which basically is in itself the product of a Referential Act (Blake 1983). Appositional constructions are rather common in the languages of Australia and Papua New Guinea and they have their own grammatical properties, but as far as I am aware nearly all typological investigations concerning the order of adnominal modifiers have ignored the distinction between modifiers in an integral, hierarchical NP and units in an appositional relationship (Rijkhoff 2004: 19-23, 327). Thus, so as to be able to exclude appositional constituents, we need to be even more specific in our formal characterization. This can be achieved by limiting the domain to qualifying modifiers used in a Referential Act that takes the form of a single, integral NP, which means that the modifier should appear in the same hierarchically organized phrase as the head of that phrase. A final formal constraint would be that the qualifying modifier should be a syntactically free unit, excluding bound manifestations of qualifying modifiers (as attested in e.g. polysynthetic languages; Rijkhoff 2015: 645). Obviously, mixing bound and free modifiers would not result in a cross-linguistic category that contains comparable morphosyntactic units.

4.2.2. A preliminary assessment

Next we will do a quick, preliminary investigation on the basis of the criteria we applied so far, just to get an impression of the kind of morphosyntactic variation we may encounter in the various languages. For this purpose of this outline, we will use a sample that contains only four languages: English (Europe), Kiribati (Pacific), Igbo (Africa) and Kwaza (South America).

If English would be the first language to investigate, we would find that there are at least three morphosyntactic units that can serve as a qualifying modifier in a referential act:[13]

---

[13] Obviously, all the morphosyntactic units should be defined in a morphosyntactic investigation. I won't attempt to do so here, since such definitions are inevitably influenced by theoretical considerations, among others. This part of the investigation is left to the individual researcher.

Functional Category: QUALIFYING MODIFIER used in a referential act (see 4.2.1)			
Semantic Categories	Formal Constraint/Category: the Referential Act should result in a hierarchically organized ('integral') noun phrase **Morphosyntactic Units:**		
	ADJECTIVE	PREP. PHRASE	RELATIVE CLAUSE
DIMENSION	*a big N*	*a N of enormous size*	*a N that was rather big*
VALUE / QUALITY	*a cheap N*	*a N of great value*	*a N that was pretty cheap*
AGE	*a young N*	*a N under age 16*	*a N that was much too young*
COLOR	*a black N*	*a N of blackness*	*a N that was pitch black*
WEIGHT	*a heavy N*	*a N of two kilos*	*a N that was rather heavy*

Table 2. Some semantic and formal categories (in English) that can be used in a referential act.

If we then add the relevant data from Kiribati (Austronesian, Oceanic) and Igbo (Niger-Conge, Kwa), our tentative cross-linguistic overview of morphosyntactic units that can serve as qualifying modifiers in a referential act would show that Kiribati only uses relative clauses (example 1) and that Igbo has only eight adjectives, constituting four pairs of antonyms (Emenanjo 1978): *úkwú* 'big', *ńtà* 'small'; *ójí'í* 'black, dark', *ócá* 'white, light'; *óhú'rú* 'new', *ócyè* 'old'; *ómá* 'good'; *ójó'ó* 'bad'. Notice furthermore that in Igbo a form such as *ógólógó* 'length, height' serves as the nominal head of an associative construction (*ógólógó ósísí* 'length of tree' (Welmers & Welmers 1969: 19); see Malchukov 2000 on Dependency Reversal). We already saw that Makwe also uses abstract nouns like 'silence' in a possessive-like construction (example 2) and the same goes for e.g. Hausa (Newman 1987: 721) and Somali (Banti 1988: 223). If we then include data from the Amazonian language Kwaza, we find that this language uses derived nominalized forms (van der Voort 2004: 94); as a matter of fact, it is claimed that in Kwaza all acts of modification (as part of a referential act) involve the juxtaposition of nouns (van der Voort 2006: 87).

With so many different morphosyntactic variants in a cross-linguistic category, it will be clear that it is impossible to compare the variants with regard to a single syntactic parameter such as the position of the qualifying modifier relative to the head noun. This can already be illustrated by just considering the three English morphosyntactic variants in Table 2: (simple) adjectives precede the noun, whereas prepositional phrases and relative clauses follow the noun. In other words, we need to apply additional semantic and formal criteria if we want to avoid comparing morphosyntactic apples and oranges.

Obviously many more languages should be investigated before one can appreciate the frequency and distribution of the chosen morphosyntactic variant (e.g. Adjective) that can serve as a qualifying modifier in the noun phrase, compared to the frequency and distribution of all the other variants (e.g. Relative Clause or Genitive). Such a broad investigation would allow the researcher to make an informed choice about the unit that is ultimately selected as the subject of the investigation. As was already mentioned above, researchers that fail to make such a cross-linguistic inventory at the outset may tend to opt for a familiar form or construction from e.g. English, which may lead to the undesirable situation that certain kinds of morphosyntactic variants (e.g. associative constructions in Igbo or nominalized forms in Kwaza - see above) are systematically ignored. This would go against the one of main goals of linguistic typology: to investigate how languages can differ by establishing the range of linguistic variation (Section 4.1).

Once we know which morphosyntactic variants are used as qualifying modifiers in languages across the globe, we can concentrate on the morphosyntactic category that will be subject of the cross-linguistic investigation. For the sake of this short sketch (and in spite of what I wrote about

Eurocentric bias), we will select a morphosyntactic category that is well attested in the European languages, namely adjectives.

4.2.3. Further restrictions on membership of the cross-linguistic category

The decision to make members of the morphosyntactic category Adjective the subject of the cross-linguistic investigation means that Kiribati (which uses relative clauses) and Kwaza (which uses juxtaposed nominal forms) will no longer be part of the investigation. But because Igbo only has eight adjectives, we must, strictly speaking, also reduce the number of semantic adjectival categories to four (Size, Value/Quality, Age and Color), because only adjectives that belong these four semantic categories can be compared in the two remaining languages of this pilot study: English and Igbo (Table 3). This could seem somewhat overcautious, but it may very well be the case that in some languages adjectives have different morphosyntactic properties, depending on the semantic category they belong to.

Functional Category: QUALIFYING MODIFIER used in a referential act				
Semantic Categories	Formal Constraint/Category: the Referential Act should result in a hierarchically organized ('integral') noun phrase **Selected Morphosyntactic Category: ADJECTIVE**			
	English	Igbo	Kiribati	Kwaza
SIZE	*big, small*	*úkwú* 'big', *ńtà* 'small'	Ø	Ø
VALUE / QUALITY	*good, bad*	*ómá* 'good', *ójó'ó* 'bad'	Ø	Ø
AGE	*old, new*	*ócyè* 'old', *óhú'rú* 'new'	Ø	Ø
COLOR	*dark, light*	*ójí'í* 'dark', *ócá* 'light'	Ø	Ø

Table 3. Members of the morphosyntactic category Adjective in English and Igbo that are sufficiently similar in terms of functional, semantic and formal properties to be used in a cross-linguistic comparison of QUALIFYING MODIFIERS.

If we had decided to focus on the Relative Clauses as the morphosyntactic category in their function as qualifying modifier in the noun phrase, Kiribati would have remained in the sample, in which case it would have been relevant to add that English relative clauses typically involve a degree modifier like 'rather' when they serve as qualifying modifiers.

Functional Category: QUALIFYING MODIFIER (here: used to enrich or supplement material used in reference to a concrete object)			
Semantic Categories	**Selected Morphosyntactic Category: RELATIVE CLAUSE**		
	English	Kiribati	Kwaza …
DIMENSION	*a N that was rather big*	(see ex. 1)	Ø
VALUE / QUALITY	*a N that was rather cheap*	…	Ø
AGE	*a N that was rather old*	…	Ø
COLOR	*a N that was pitch black*	…	Ø

Table 4. Members of the morphosyntactic category Relative Clause in English and Kiribati that are sufficiently similar in terms of functional, semantic and formal properties to be used in a cross-linguistic comparison of QUALIFYING MODIFIERS.

The fact that we have now limited the investigation to adjectives that belong to the four semantic categories listed in Table 3, also leads us to consider another semantic issue that could interfere with cross-linguistic comparability: entity types.

4.2.4. Entity types

The fact that Igbo only has adjectives that apply to first order entities (if not used in a metaphorical sense, that is), seems to exclude referentials that involve second and higher order nouns, which are modified by adjectives that tend to belong to different semantic categories. Only concrete spatial objects ('first order entities') have a size, a color, a weight, a shape etc. Second order nouns, which denote temporal entities, are typically specified for temporal features, as in 'a dynamic game' or 'a short meeting'. Notice that spatial notions such as 'short' and 'long' are used in a metaphorical sense when they can combine with second order nouns, expressing the duration of an event rather than length in the spatial dimension. Third order nouns are used to talk about entities beyond the spatial and temporal dimension (such as 'belief', 'thought', 'idea' or 'fact'), which tend to combine with their own qualifying modifiers, for example 'a possible fact', 'an interesting thought' or 'a hopeless idea' (on entity types, see Lyons 1977 and Hengeveld & Mackenzie 2008). Finally, one could ~~even~~ recognize even higher order nouns, which are used to refer speech acts, like 'a polite answer' or 'an inappropriate request'. So as to keep the semantic categories the same in the comparison, we should decide on the kind of entity that is involved in the referential act; let us here choose first order entities. In this particular case, this means there is no need to exclude NPs used to refer to second and higher order entities, due the kind of semantic categories covered by the limited number of adjectival semantic categories in Igbo.[14]

4.2.5. The final stage

At this point we have decided to investigate syntactic properties of members of the morphosyntactic category Adjective and we have applied certain semantic and formal criteria to make sure that the cross-linguistic category contains comparable linguistic units. Now we are also in a position to specify additional formal details of the selected morphosyntactic category. It will be recalled that, for the sake of this demonstration, we decided to focus on matters regarding word order. Our mini-investigation tells us that (simple) adjectives precede the noun in English, whereas they follow the noun in Igbo. But we could also have focused on morphological (e.g. agreement) or phonological (e.g. tone) properties. If we were to focus on phonology, we would find, for example, that Igbo is a tonal language, as opposed to English. If we had decided to select Relative Clause instead of Adjective, we could have added formal properties associated with relative clauses, such as clause internal syntactic properties (constituent order) or certain morphological aspects (e.g. finite vs. non-finite verb form).

For practical reasons, a cross-linguistic investigation in the area of morphosyntax will often have to abstract away from various kinds of more detailed formal or semantic parameters, but the extent to which these additional criteria are relevant is in itself not a trivial matter. Interestingly, Welmers & Welmers (1969: 322) end their article on noun modifiers in Igbo with the following observation: "The limitations and characteristics of the class of adjectives defined for Igbo in the foregoing suggests that, in other Niger-Congo languages as well, the apparently small groups of 'adjectives' should be investigated for special features of form, function, or meaning. The features may well be quite different in other languages, but they may be as significant and as unexpected as those here defined for Igbo adjectives."

---

[14] When speakers of different languages refer to first order entities (concrete physical objects), they do not necessarily use nouns (Rijkhoff and Van Lier eds. 2103). And if they use nouns to talk about first order entities like 'chair' or 'dog', they do not necessarily employ the same kind of noun. At least four different kinds of nouns (i.e. noun types defining different modes of being or *Seinsarten*) can be used to refer to a single object, each with its own set of grammatical properties: individual object nouns, set nouns, sort nouns, or general nouns (Rijkhoff 2004: 100-121; Rijkhoff 2009b). This is another semantic factor that may influence comparability of linguistic units in a typological investigation.

There are several ways to set up a cross-linguistic category for morphosyntactic purposes within a wider, all-encompassing functional category. The number and kind of formal and semantic criteria that are used to construct a cross-linguistic category within a larger functional category depends on various factors, including the phenomenon under investigation, the goal of the study, the level of detail pursued by the researcher and the theoretical framework within which the research question was formulated (see end of Section 3.1). As long as the researcher is fully explicit and transparent about the way the number of members of the functional supercategory is systematically reduced to a cross-linguistic category with members that are similar enough in terms of function, meaning and form/structure for the purpose of the investigation, the reader can decide for herself if she agrees with the final selection.

5. Summary and conclusion

It has been assumed for quite a while that we cannot simply compare morphosyntactic forms and constructions such as Relative Clause, Genitive or Adjective in different languages, because the formal properties of these constituents are deemed too language specific to be of any use in a meaningful cross-linguistic comparison. For this reason semantic criteria have been used to identify and process members of morphosyntactic categories, but this has lead to several serious problems, especially in the case of probabilistic morphosyntactic research (Section 2). In this article, I have outlined a method that aims to produce cross-linguistic categories whose members are similar enough to warrant an academically responsible comparison. Perhaps the most important feature of this method is that it consistently distinguishes between functional, semantic and formal criteria. Failing to do so in the past is probably the main reason why morphosyntactic typology has been struggling to come up with reliable, reproducible results.

The current approach requires that we first employ functional criteria, because they have the widest cross-linguistic applicability. This is to make sure that the linguistic units in the sample share the same interpersonal or communicative function. Prioritizing the functional criteria should also take care of the form-function problem: the same linguistic form or construction can be used in different functions, and vice versa, the same function can be fulfilled by different forms and constructions. Then, within the boundary of the functional category the researcher selects one particular morphosyntactic variant and subsequently makes sure that the members of the cross-linguistic morphosyntactic category are also similar enough from a formal and a semantic perspective. It is important that each step in the procedure is made explicit and, where necessary, motivated, so that others can decide for themselves whether the investigator has succeeded in setting up a valid cross-linguistic category.

The method outlined in the article implies that there will probably always be languages that have to be excluded on functional, semantic or formal grounds. This is shown in Rijkhoff's 2004 investigation of the relative order of Demonstrative - Numeral - Adjective - Noun in the simple noun phrase, where 50% of the languages in his sample had to be excluded, due to the fact that there was at least one problem with cross-linguistic comparability. To mention some of the reasons why languages were excluded (Rijkhoff 2004: 328):
- absence of the selected morphosyntactic variant (e.g. Hixkaryana has no attributive demonstratives);
- modifier is expressed as a bound form (rather than a free modifier);
- adjectival notions and cardinality are expressed by verbs or nouns (i.e. as relative clause or genitives);
- potential candidates for a modifier function were not part of an integral noun phrase, but

occurred as an appositional constituent.

If morphosyntactic typology is to be of any service to linguistics, it should be concerned with concrete members of specific morphosyntactic categories that are sufficiently similar in terms of function, meaning and form. In collecting and processing the data, the researcher should be fully explicit about the procedure, in particular with regard to the way functional, semantic and formal criteria are applied. Furthermore, data for a typological investigation should be obtained from reliable primary sources, such as fieldwork reports and detailed language descriptions, to be read from cover to cover.[15]

The upshot of this contribution is also that we cannot take the results from previous typological studies for granted and use them ~~to~~ as input for new typological investigations, as is often the case in the so-called 'big data approach' to typology. Since big data typology simply assumes that the labels for the various cross-linguistic categories used in previous studies cover the same linguistic forms and constructions, they ignore the fact that different investigators had different goals and theoretical assumptions and used different criteria to identify and process language specific forms and constructions as members of the various categories. This can be seen, for example, in Cysouw 2012, where categorial labels such as Adjective and NP were taken for granted, apparently on the assumption that these labels cover comparable linguistic forms and constructions in the various older studies (e.g. Dryer 1992, Cinque 2005, Rijkhoff 2004). This is a mistake, because the way Dryer 1992 collected and processed members of his cross-linguistic categories (to the extent that this can be determined; neither Dryer or Cinque are very explicit in explaining how they defined and constructed their categories) differs in some very important respects from the procedure that Rijkhoff 2004 used to establish the cross-linguistic categories for his investigation. So when Cysouw 2012 proposes a probabilistic model that claims to offer 'a new take on typological explanations', he has in all probability been comparing morphosyntactic apples and oranges. In other words, exercises in big data typology that blindly accept category labels and results from previous typological investigations (in particular probabilistic studies) to make new (probabilistic!) claims about the grammatical properties of linguistic units of different languages make matters only worse.

Correspondence address
Jan Rijkhoff, Linguistics, Aarhus University
Jens Chr. Skous Vej 2, Nobelparken building 1485-621
DK-8000 Aarhus C, Denmark
E-mail: linjr@cc.au.dk

Abbreviations: 1 = first person, 3 = third person, A = adjective, ACC = accusative, ART = article, DEC = declarative, DEM = demonstrative, DIR = direct, ERG = ergative, GEN = genitive, INT = interrogative, INV = inverse, N = noun, NP = noun phrase, NC1 = noun class 1, NC7 = noun class 7, NUM = numeral, PP1 = class 1 pronominal prefix, REFL = reflexive, REL = relative clause, S = subject, SG = singular.

---

[15] Due to lack of space, I have ignored the potential problem of the reliability of glosses for grammatical categories: do they represent the same meaning or function in different descriptions? It is my distinct impression that the problem is often rather exaggerated: if one reads the whole grammar (and not just the examples and the glosses in certain sections or subsections), one usually finds enough information about the way e.g. 'Perfective' or 'Dative' should be interpreted. Greenberg already pointed out that it is not good enough to just focus on specific passages in a grammatical description and ignore the rest. The following quotation is taken from Croft (2009: 174): "You know, you gotta muck around in grammars. You can't just focus on one specific thing and pick it out. You read around and you discover things you never would have thought of."

References

Aarts, Bas. 2004. Modelling linguistic gradience. *Studies in Language* 28(1). 1-49.

Adelaar, K. Alexander & Nikolaus P. Himmelmann (eds.). 2005. *The Austronesian languages of Asia and Madagascar*. London: Routledge.

Austin, John L. 1962. *How to do things with words: the William James Lectures delivered at Harvard University in 1955*. Oxford: Clarendon Press.

Banti, Giorgio. 1988. 'Adjectives' in East Cushitic. In Marianne Bechhaus-Gerst & Fritz Serzisko (eds.), *Cushitic – Oromo: papers from the International Symposium on Cushitic and Omotic languages, Cologne, January 6-9, 1986*, 205-259. Hamburg: Buske.

Behaghel, Otto. 1932. *Deutsche Syntax: eine geschichtliche Darstellung. Band IV: Wortstellung-Periodenbau*. Heidelberg: Carl Winter.

Blake, Barry J. 1983. Structure and word order in Kalkatungu: the anatomy of a flat language. *Australian Journal of Linguistics* 3(2). 143-175.

Bulmer, Ralph. 1967. Why is the cassowary not a bird? A problem of zoological taxonomy among the Karam of the New Guinea highlands. *Man* 2(1). 5-25.

Cinque, Guglielmo. 2005. Deriving Greenberg's universal 20 and its exceptions. *Linguistic Inquiry* 36. 315–332.

Comrie, Bernard. 1981. *Language universals and linguistic typology*. Oxford: Basil Blackwell.

Chung, Sandra & Alan Timberlake. 1985. Tense, aspect and mood. In Timothy Shopen (ed.), *Language typology and syntactic description* (Volume 3, first edition), 202-258. Cambridge: Cambridge University Press.

Croft, William. 1990. A conceptual framework for grammatical categories (or: a taxonomy of Propositional Acts). *Journal of Semantics* 7(3). 245-279.

Croft, William. 2001 Radical Construction Grammar. Syntactic theory in typological perspective. Oxford: Oxford University Press.

Croft, William. 2009. Joseph Harold Greenberg. *Bibliographic Memoirs 90*. 153-198. Washington, D.C.: The National Academies Press.

Croft, William. 2014. Comparing categories and constructions crosslinguistically (again): the diversity of ditransitives. *Linguistic Typology* 18(3). 533-551.

Cysouw, Michael. 2010. Dealing with diversity: towards an explanation of NP word order frequencies. *Linguistic Typology* 14(2). 253–287.

Dik, Simon C. 1997. *The theory of Functional Grammar* (2nd revised edition, edited by Kees Hengeveld). Volume I: *The structure of the clause*. Volume II: *Complex and derived constructions*. Berlin: Mouton de Gruyter.

Devos, Maud. 2008. *A grammar of Makwe*. München: Lincom Europa.

Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68(1). 81–138.

Elliott, Jennifer R. 2000. Realis and irrealis: forms and concepts of the grammaticalisation of reality. *Linguistic Typology* 4(1). 55-90.

Emenanjo, Nolue. 1976. *Elements of modern Igbo grammar*. Ibadan: Oxford University Press.

Foley, William A. 1976. Towards a universal typology of the noun phrase. *Studies in Language* 4(2). 171-199.

Foley, William A. & Robert D. Van Valin, Jr. 1985. Information packaging in the clause. In Timothy Shopen (ed.), *Language typology and syntactic description* (Volume 1, first edition), 282-364.

García Velasco, Daniel & Jan Rijkhoff. 2008. Introduction. In Daniel García Velasco & Jan Rijkhoff (eds.), *The noun phrase in Functional Discourse Grammar*, 1-42. Berlin: Mouton de Gruyter.

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.

Haiman, John. 2014. Decorative morphology in Khmer. In Jeffrey P. Williams (ed.), *The aesthetics of grammar: sound and meaning in the languages of mainland Southeast Asia*, 61-82. Cambridge: Cambridge University Press.

Hale, Kenneth. 1997. An interview with Professor Ken Hale, by Lisa Cheng & Rint Sybesma. *Glot International* 2(9/10). Interview begins on p.1 and continues on pp.26-28.

Harnad, Steven. 2005. To cognize is to categorize: cognition is categorization. In Henri Cohen & Claire Levebvre (eds.), *Handbook of categorization in cognitive science*, 20-44. Amsterdam: Elsevier.

Haspelmath, Martin, 2007. Pre-established categories don't exist: consequences for language description and typology. *Linguistic Typology* 11(1). 119-132.

Haspelmath, Martin, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.). 2001. *Language typology and language universals: an international handbook*, Vol. 1. Berlin: Walter de Gruyter.

Hawkins, John A. 1983. *Word order universals*. New York: Academic Press.

Heine, Bernd. 1997. *Possession: cognitive sources, forces, and grammaticalization*. Cambridge: Cambridge University Press.

Heine, Bernd, Tania Kuteva & Gunther Kaltenböck. 2014. Discourse Grammar, the dual process model, and brain lateralization: some correlations. *Language and Cognition* 6(1). 146-180.

Hengeveld, Kees. 1989. Layers and operators. *Journal of Linguistics* 25. 127-157.

Hengeveld, Kees, Jan Rijkhoff & Anna Siewierska. 2004. Parts-of-speech systems and word order. *Journal of Linguistics* 40(3). 527-570.

Hengeveld, Kees & J. Lachlan Mackenzie. 2008. *Functional Discourse Grammar: a typologically-based theory of language structure*. Oxford: Oxford University Press.

Kaltenböck, Gunther, Bernd Heine & Tania Kuteva. 2011. On thetical grammar. *Studies in Language* 35(4). 848-893.

Kittilä, Seppo. 2006. Object-, animacy- and role-based strategies: a typology of object marking. *Studies in Language* 30(1). 1-32.

König, Ekkehard & Peter Siemund. 2007. Speech act distinctions in grammar. In Timothy Shopen (ed.), *Language typology and syntactic description* (Volume 1, second edition), 276-324.

Locke, John. 1690 [1825]) *An essay concerning human understanding*. London: Tegg.

Lyons, John. 1977. *Semantics* (2 volumes). Cambridge: Cambridge University Press.

Machery, Edouard. 2009. *Doing without concepts*. Cambridge: Cambridge University Press.

Malchukov, Andrej L. 2000. *Dependency reversal in noun-attribute constructions: towards a typology*. Munich: Lincom Europa.

Malt, Barbara C. 2010. Why we should do without concepts. *Mind and Language* 25(2). 622-633.

Maslova, Elena. 2000. A dynamic approach to the verification of distributional universals. *Linguistic Typology* 4(3). 307-333.

Miner, Kenneth L. 1986. Noun stripping and loose incorporation in Zuni. *International Journal of American Linguistics* 52(3). 242–254.

Muysken, Pieter. 2008. *Functional categories*. Cambridge: Cambridge University Press.

Newman, Paul. 1990. *Nominal and verbal plurality in Chadic*. Dordrecht: Foris.

Perkins, Revere D. 1988. The covariation of culture and grammar. In Mark T. Hammond, Edith A. Moravcsik & Jessica R Wirth (eds,). *Studies in syntactic typology*, 359-378. Amsterdam: Benjamins.

Perkins, Revere D. 2001. Sampling procedures and statistical methods. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & gang Raible (eds.), 419-434.

Reboul, Anne. 2001. Foundations of reference and predication. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), 509-522.

Rijkhoff, Jan. 1990. Toward a unified analysis of terms and predications. In Jan Nuyts, A. Machtelt Bolkestein & Co Vet (eds.), *Layers and levels of representation in language theory: a functional view*, 165-191. Amsterdam: Benjamins.

Rijkhoff, Jan. 2002. On the interaction of Linguistic Typology and Functional Grammar. *Functions of Language* 9(2). 209–237.

Rijkhoff, Jan. 2004. *The noun phrase*. Oxford: Oxford University Press. [Expanded Pb of 2002 Hb edition]

Rijkhoff, Jan. 2008a. Synchronic and diachronic evidence for parallels between noun phrases and sentences. In Folke Josephson & Ingmar Söhrman (eds.), *Interdependence of diachronic and synchronic analyses*, 13-42. Amsterdam: Benjamins.

Rijkhoff, Jan. 2008b. On flexible and rigid nouns. *Studies in Language* 32(3). 727-752.

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 William B. McGregor (ed.), *The Expression of Possession*, 51-106. Berlin: Mouton de Gruyter.

Rijkhoff, Jan. 2010a. Rara and grammatical theory. In Jan Wohlgemuth & Michael Cysouw (eds.), *Rethinking universals: how rarities affect linguistic theory*, 223-240. Berlin: Mouton de Gruyter.

Rijkhoff, Jan. 2010b. Functional categories in the noun phrase: on jacks-of-all-trades and one-trick-ponies in Danish, Dutch and German. *Deutsche Sprache* 38(2/10). 97-123. Available online at: http://pub.ids-mannheim.de/laufend/deusprach/ds10-2.html

Rijkhoff, Jan. 2014. Modification as a propositional act. In María de los Ángeles Gómez González, Francisco José Ruiz de Mendoza Ibáñez & Francisco Gonzálvez-García (eds.), *Theory and practice in functional-cognitive space*, 129-150. Amsterdam: Benjamins.

Rijkhoff, Jan. 2015. Word order. In James D. Wright (editor-in-chief), *International encyclopedia of the social & behavioral sciences (Second edition)*, Vol. 25, 644–656. Oxford: Elsevier.

Rijkhoff, Jan & Dik Bakker. 1998. Language sampling. *Linguistic Typology* 2(3). 263–314.

Rijkhoff, Jan, Dik Bakker, Kees Hengeveld & Peter Kahrel. 1993. A method of language sampling. *Studies in Language* 17(1). 169-203.

Rijkhoff, Jan & Johanna Seibt. 2005. Mood, definiteness and specificity: a linguistic and a philosophical account of their similarities and differences. *Tidsskrift for Sprogforskning* 3(2). 85–132 (notice that the journal is now called *Skandinaviske Sprogstudier*). Available online at: http://ojs.statsbiblioteket.dk/index.php/tfs/article/view/89

Rijkhoff, Jan & Eva van Lier (eds.). 2013. *Flexible word classes: typological studies of underspecified parts of speech*. Oxford: Oxford University Press**.**

Ross, John Robert. 1972. The category squish: Endstation Hauptwort. *Chicago Linguistic Society* 8. 316–328.

Ross, Malcolm. 1998. Proto-Oceanic adjectival categories and their morphosyntax. *Oceanic Linguistics* 37(1). 85-119.

Sadock, Jerrold M. & Arnold M. Zwicky. 1985. Speech act distinctions in syntax. In Timothy Shopen (ed.), *Language typology and syntactic description* (Volume 1, first edition), 155-196.

Sampson, Geoffrey. 1980. *Schools of linguistics: competition and evolution*. London: Hutchinson.

Sharma, Devidatta D. 1988. *A descriptive grammar of Kinnauri*. Delhi: Mittal Publications.

Sasse, Hans-Jürgen. 2001. Scales of nouniness and verbiness. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), 495–509.

Searle, John. 1969. *Speech acts: an essay in the philosophy of language*. Cambridge: Cambridge University Press.

Shopen, Timothy (ed.). 1985. *Language typology and syntactic description* (3 volumes, first edition). Cambridge: Cambridge University Press.

Shopen, Timothy (ed.) 2007. *Language typology and syntactic description* (3 volumes, second edition). Cambridge: Cambridge University Press.

Silverstein, Michael. 1976. Hierarchy of features and ergativity. In Robert Dixon (ed.), *Grammatical categories in Australian languages*, 112-171. Canberra: Australian Institute of Aboriginal Studies.

Smith, Jennifer L. 2015. Parts of speech in phonology. 17th LIPP Symposium - 'Parts of speech across languages, in acquisition, mind and brain', held at Ludwig-Maximilians-Universität, München, July 6. Handout available at: http://www.unc.edu/~jlsmith/home/pdf/smith2012_PoS-in-phono_17LIPP-hdt.pdf

Smith, Linda B. & Larissa K. Samuelson. 1997. Perceiving and remembering: category stability, variability and development. In Koen Lamberts & David Shanks (eds.), *Knowledge, concepts and categories*, 161-195. Cambridge, MA: The MIT Press.

Song, Jae Jung. 2001. *Linguistic typology: morphology and syntax*. Harlow: Longman.

Stassen, Leon. 2011. The problem of cross-linguistic classification. In Jae Jung Song (ed.), *The Oxford handbook of linguistic typology*, 90-99. Oxford: Oxford University Press.

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.

Voort, Hein van der. 2004. *A grammar of Kwaza*. Berlin: Mouton de Gruyter.

Voort, Hein van der. 2006. Construções atributivas em Kwazá. *Boletim do Museu Paraense Emílio Goeldi* (Ciências humanas) 1(1). 87–104. Available online at http://marte.museu-goeldi.br/seer/index.php/boletimhumanas

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). Amsterdam: Benjamins.

Welmers, William E. & Beatrice F. Welmers. 1969. Noun modifiers in Igbo. *International Journal of American Linguistics* 35(4). 315-322.

Wohlgemuth, Jan & Michael Cysouw (eds.). 2010a. *Rethinking universals: how rarities affect linguistic theory*. Berlin: De Gruyter Mouton.

Wohlgemuth, Jan & Michael Cysouw (eds.). 2010b. *Rara & rarissima: documenting the fringes of linguistic diversity*. Berlin: De Gruyter Mouton.

Young, Robert W. & Morgan, William, Sr. 1987. *The Navajo language: a grammar and colloquial dictionary* (revised edition). Albuquerque: University of New Mexico Press.