

## Lexical and grammatical categories in RRG

John Peterson, University of Kiel<sup>1</sup>

This chapter deals with the status of lexical and grammatical categories in RRG, including parts of speech and categories such as TAM and negation. While RRG allows for parts of speech such as “noun”, “verb”, “adjective”, etc., it does not view them as universal. Instead, it assumes functionally motivated, non-endocentric syntactic categories such as nucleus, referential expressions and modifying phrases. Although these syntactic slots are typically realized by verbs, nouns and adjectives/adverbs, respectively, this is not required, as many languages, including English, allow for non-verbal predicates, non-adjectival modifying phrases, etc., while other languages show little or no evidence for lexical classes or allow for clausal referents, modifiers and predicates. And even languages which do possess the traditionally assumed parts of speech often show “category squishes” in which membership to a particular lexical class is at best arbitrary. Finally, this chapter provides a description of the “operators”, i.e., grammatical categories such as tense, aspect, deontic and epistemic mood, etc., which “ground” or modify a clause, core or nucleus and which play a pivotal role in the determination of nexus and juncture relations in complex sentences in RRG.

### 1. Introduction

The present chapter deals with the status of lexical and grammatical categories in RRG, including part-of-speech systems and grammatical categories such as TAM and negation, which are referred to in RRG as “operators”. With respect to **lexical categories**, RRG differs from many other theories of language which assume a small, closed set of universal parts of speech such as “noun”, “verb”, “adjective”, etc., and which consider these to be the “heads” of corresponding syntactic categories, i.e., NP, VP and AdjP. Instead, RRG does not assume that such categories are universal, in line with an increasing amount of typologically oriented research questioning the universality of these categories. Instead, RRG assumes functionally motivated, non-endocentric syntactic categories such as nucleus or “NUC” (containing the predicate), (potentially) referential phrases or “RPs”, and modifying phrases or “MPs”. Although these syntactic slots are typically realized by verbs, nouns and adjectives/adverbs, respectively, this is not required by the theory: Many languages, including English, allow for non-verbal predicates, non-adjectival modifying phrases, etc., while others show little or no evidence for lexical categories such as “noun” and “verb” or allow for clausal referents, modifiers and predicates. And even those languages which do possess traditional parts of speech often show so-called “category squishes”, in which assignment to a particular category is at best an arbitrary decision. Hence, categories such as NUC, RP and MP are not universally linked to particular lexical categories *per se*, although individual languages will generally have language-specific restrictions as to which elements may occupy these slots.

The present chapter also provides a description of the **grammatical categories** referred to in RRG as “operators”, i.e., those categories which “ground” a clause, core or nucleus and which are closely linked to finiteness, a topic which has been receiving increasing attention in typological research in the past few years. These categories include tense, aspect, deontic mood (“modality” in RRG terms), epistemic mood (“status”), evidentials, etc. These units all have scope over a particular level of the sentence structure (nucleus, core, clause) and play an important role in the determination of nexus relations in complex sentences within the theory. This also includes categories which are primarily concerned with questions of reference, such as definiteness, deixis, etc., which ground the RP.

---

<sup>1</sup> I would like to thank Utz Maas for his comments on an earlier version of the present study. Needless to say any remaining errors and misconceptions are my own.

## 2. Lexical and syntactic categories in RRG

### 2.1 General issues

A basic tenet of descriptive linguistics is that linguistic categories are only valid to the extent that they are empirically justified, and the view that there is no such thing as a universal grammatical category is gaining acceptance.<sup>2</sup> This holds for all categories, including **lexical categories** such as **noun**, **verb** and **adjective** (to name just three) as well as the corresponding endocentric **syntactic categories**, i.e., **NP**, **AdjP** and **VP**, even though these categories – or at least the lexical categories – are generally considered universal.

Although there are considerable differences of opinion with respect to the “correct” approach to this topic, most researchers would probably agree with an approach which views lexical categories as language-specific categories based on feature bundles, in which “prototypical” members of a particular class possess all of the features potentially associated with the particular class.<sup>3</sup> For example, “prototypical” nouns in English would denote “persons, places or things”, mark for the plural (usually with *-s*), be compatible with definite and indefinite articles, etc., while “less prototypical” nouns may, for example, denote “things” but may not be compatible with plural marking or the definite article. Problematic for this approach, however, is determining which features are more important than others, if indeed a hierarchy can be established at all, as well as how many features are “enough” to be assigned to a particular category. We will return to this topic in the following pages.

As Himmelmann (2008: 260) notes, there is little doubt that “the inventory of function words is highly language-specific.” Rather, “[w]hen there is a controversy regarding the number and kind of syntactic categories in a given language, it pertains to the linguistic classification of content words.” Himmelmann assumes two grammatical levels for the discussion of parts of speech, 1. “the level of terminal syntactic categories where lexical items are categorised according to their phrase-structural properties”, and 2. “the level of lexical categories proper where lexical items are categorised according to those grammatical features which are not directly relevant for phrase structure.” (Himmelmann, 2008: 263). To illustrate these two levels,<sup>4</sup> Himmelmann proposes a hypothetical example in a language L, in which the most basic phrasal categories always consist of an overt function word X and an overt content word Y, with virtually all content words being compatible with all functions words. This can be schematically portrayed as in (1). Here the status of the function word X distinguishes the phrase from other phrase types, hence “XP”.

(1) [XP [X] [Y]]

---

<sup>2</sup> Cf. e.g. Haspelmath (2007).

<sup>3</sup> A notable exception to this is Croft (e.g., 2000, 2005) who approaches the topic from a cross-linguistic perspective and hence begins with semantics instead of language-specific morphosyntactic categories, assuming prototypical nouns (= objects), verbs (= actions) and adjectives (= attributes) and mapping language-specific forms/constructions against these. As what is at issue for our discussion at this point is the necessity (or rather, the lack thereof) for assuming universally valid CATEGORIES for language-specific feature bundles, we follow here the approach laid out in Himmelmann (2007). Further approaches to the topic abound in the literature; to cite just a few of the more recent typological works on this topic: Evans & Osada, 2005a and the responses to this article (Peterson, 2005; Croft, 2005; Hengeveld & Rijkhoff, 2005); Luuk, 2010; and the many contributions in Rijkhoff & van Lier (eds.), 2013.

<sup>4</sup> In the following, we will assume that there is a clear distinction between “word level” and “syntactic level”, a prerequisite for this analysis, although it is well known (cf. e.g. Haspelmath, 2011) that the boundary between the two levels is often not as clear as generally assumed.

These two levels, i.e., lexical and syntactic, need not correspond to one another directly, although this will often be the case, e.g., with Ns as heads of NPs, Adj's as heads of AdjPs and Vs as heads of VPs in many languages. Himmelmann (2008: 264) notes five logical possibilities with respect to these two levels and their correspondence to one another, given in Table 1. With respect to category I, both levels possess distinct categories but these do not directly correlate in Ia ("≠"), while they do in Ib ("=").

|     | LEXICAL CATEGORIES |   | TERMINAL SYNTACTIC CATEGORIES |
|-----|--------------------|---|-------------------------------|
| Ia  | distinct           | ≠ | distinct                      |
| Ib  | distinct           | = | distinct                      |
| II  | indistinct         |   | distinct                      |
| III | distinct           |   | indistinct                    |
| IV  | indistinct         |   | indistinct                    |

Table 1: Possible alignments between lexical and syntactic categories (Himmelmann, 2008: 264)

Returning to our example above, if following Himmelmann we imagine that L has a further restriction such that not all content words Y behave similarly, as some Ys take plural suffixes, others take plural prefixes, while the members of a third class do not allow pluralisation, then we would have case Ia, with distinct lexical categories based on the compatibility of Y with plural markers and also distinct terminal syntactic categories, but with no direct correspondence between the two levels, as the compatibility of plural affixes does not correlate with the compatibility of a particular Y with a particular X.

To these two descriptive levels we can add a third level, that of function, here primarily reference, predication and attribution. This can involve a direct correlation between terminal syntactic categories and functional categories, i.e., XP<sub>1</sub> = reference, XP<sub>2</sub> = predication, etc., or there can be no direct correlation, so that XP<sub>1</sub>, XP<sub>2</sub>, etc., can both be used in reference, predication and modification. At one extreme we would then have endocentric categories specialized for certain sentence-level functions, e.g., N = NP = reference, V = VP = predication, etc., and at the other extreme – at least putatively – no correlations whatsoever, although it is questionable whether such a language is possible.

Such mismatches are not only a theoretical possibility; many researchers have claimed that the languages they are describing either do not in fact possess a certain category such as “N” or “NP”, etc., or that the respective language can at least best be described without recourse to such categories. Even in languages such as English which possess relatively clearly defined lexical classes such as noun, verb, and adjective, it has long been known that there is no one-to-one relationship between part of speech and function. Consider the examples in (1a)-(1g), from Van Valin (2008: 165).

- |       |  |                                 |
|-------|--|---------------------------------|
| (1)a. | Chris will [NUC [PRED see]] the movie.       | PRED in NUC = V                 |
| b.    | Chris [NUC is [PRED a very good detective]]. | PRED in NUC = NP                |
| c.    | Pat [NUC is [PRED exceedingly tall]].        | PRED in NUC = ADJP              |
| d.    | Pat [NUC is [PRED in the house]].            | PRED in NUC = PP                |
| e.    | Chris [NUC [PRED wiped]]                     | PRED in NUC <sub>1</sub> = V,   |
|       | the table [NUC [PRED squeaky clean]]         | PRED in NUC <sub>2</sub> = ADJP |
| f.    | Pat [NUC [PRED pushed]]                      | PRED in NUC <sub>1</sub> = V,   |
|       | the table [NUC [PRED out the door]]          | PRED in NUC <sub>2</sub> = PP   |

g. Chris [<sub>NUC</sub> was [<sub>PRED</sub> elected]]  
 [<sub>NUC</sub> [<sub>PRED</sub> president of the club]].

PRED in NUC<sub>1</sub> = V,  
 PRED in NUC<sub>2</sub> = NP

As the examples in (1) show, the predicate of a nucleus in English need not be verbal: While the predicate is verbal in (1a), in (1b) it is an NP, in (1c) an AdjP and in (1d) a PP. Figure 1 presents the constituent projections for (1a) (from this author), and (1b) and (1d) (from Van Valin, 2008: 166).<sup>5</sup>

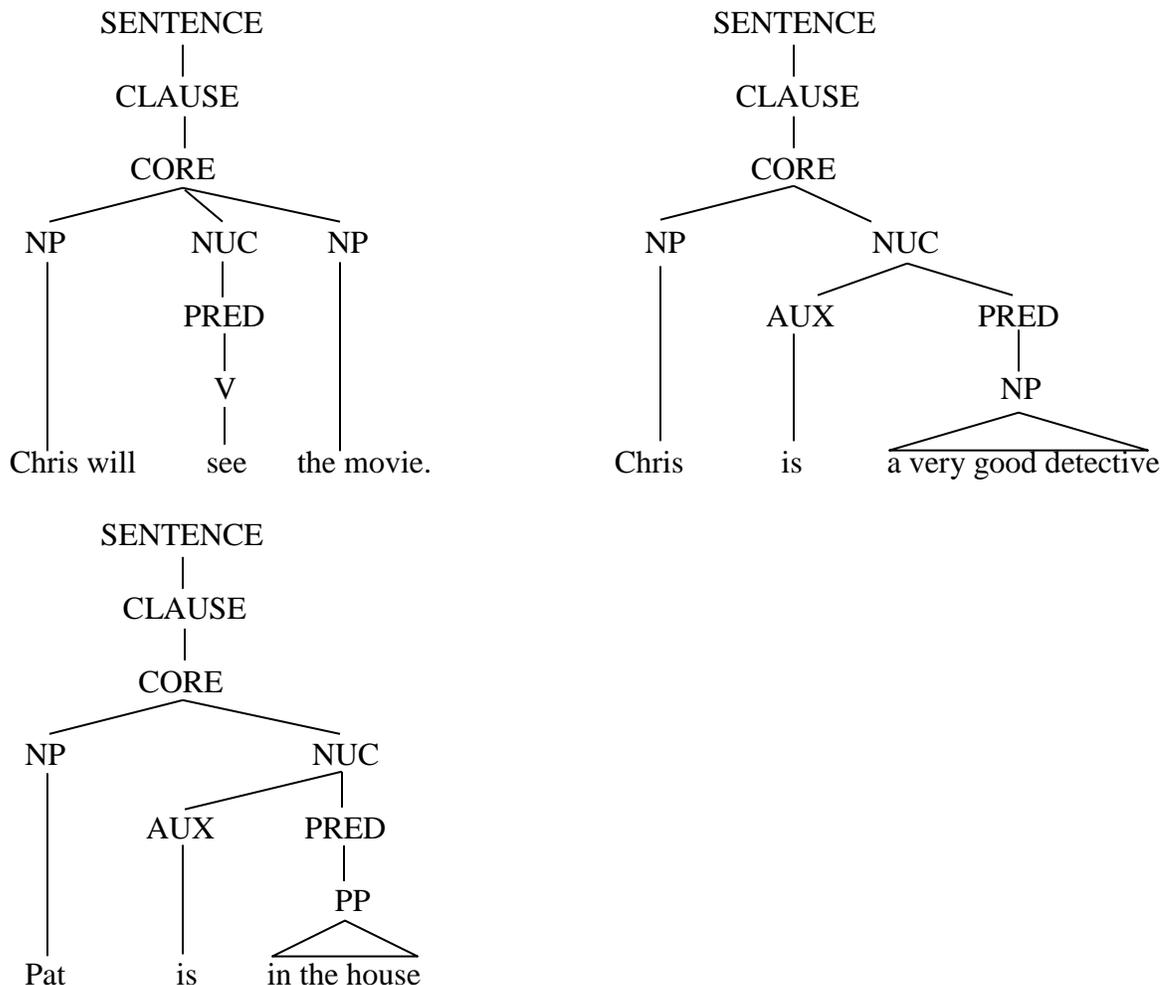


Figure 1: Constituent projections for (1a), (1b) and (1d)

In English, as in many other languages, there is a general requirement that clauses in which the predicate is not verbal must contain a kind of “auxiliary” such as the copula *is* in order for the clause to be grammatical. As *a very good detective* in (1b), *exceedingly tall* in (1c) and *in the house* in (1d) are not verbs, the copula *is* is required, however it is important to stress two things here: First, that while *is* is a verb, it is not the predicate; this function is filled by the NP in (1b), by the AdjP in (1c), and by the PP in (1d). Second, this is a language-specific criterion which is not found in all languages or which is only found with certain interpretations. For example, one

<sup>5</sup> The operator projections have been omitted here as these are not discussed until Section 3 below.

quite common situation is found in Russian (Eastern Slavic), where a clause whose predicate is non-verbal does not require a copula when there is a present or atemporal interpretation, as in (2).

### Russian

- (2) *eto stol*  
this table  
'This [is a] table.'

In other languages, the element which is often considered a copula may not even be verbal, as in Maltese (Semitic, Malta), where the “copula” is in fact a pronoun agreeing with the subject with respect to person, number and gender (in the third person, singular) with a present or atemporal interpretation, as in (3).

### Maltese

- (3) *Dan il-professur huwa Malti.*  
this.M.SG DEF-professor 3SG.M(PRO) Maltese.M.SG  
'This professor is Maltese.' (Ambros, 1998: 69)

As Van Valin (2008: 166) notes, non-verbal predicates do not pose a problem in RRG, since the clause is not an endocentric category: “Indeed, there is no head at all. The nucleus cannot be considered the “head” of the core or the clause, because it is not a lexical category, on the one hand, and is often phrasal, on the other. The notion of “head” is of no relevance to the layered structure of the clause.” Instead, all universal categories assumed in RRG are semantically motivated, so that in RRG a clause minimally consists of a core, which minimally consists of a nucleus which contains the predicate, and possibly one or more arguments.

Instead of assuming “universal” lexical categories which are assumed to be the heads of corresponding phrasal categories – cross-linguistically an even more problematic assumption – mainstream RRG has assumed since Van Valin (2008) that the core contains a **predicate**, the functional / semantic base of the nucleus, and possibly one or more “referential phrases”, referred to as **RPs**. Just as the predicate need not necessarily be verbal, so too does an RP not need to be nominal. In addition, RRG assumes that these units may be modified by modifying phrases or **MPs**, which can be specialized so as to modify only an RP, as with prototypical adjectives, or only predicative elements, as with prototypical adverbs, although again this is not necessary, as there are many languages with either no or only very few adjectives and/or adverbs, where these functions are realized by other categories, such as nouns or verbs.

There is one potential pit-fall here: The RP is not necessarily referential, since the same structural units which are typically used referentially can also be used as expletives or “dummy” subjects in many languages. Consider *it* in the two examples in (4).

- (4) a. *Where's the book? It's over there, on the table.* Referential  
b. *It looks like it's going to rain today.* Non-referential, “dummy” subjects

The non-referential use of RPs is not limited to pronominals, however, as *a very good detective* example (1b) shows, although this very same form can of course also be used referentially, as in (5).

- (5) *A very good detective just arrived and is waiting for you in your office.*

Consequently, RPs are defined as POTENTIALLY referential expressions whose DEFAULT interpretation is referential, thereby capturing the structural similarities between referential and non-referential forms, while at the same time recognizing the fact that the default use of these forms is for reference (Van Valin, 2008: 170).

As mentioned above, modifying phrases or MPs in many languages have as their default value either adjectives or adverbs, depending on what is being modified, i.e., a noun (*the tall tree*) or a verb (*the mouse ran quickly into the closet*). These appear in the peripheries of RPs and clauses (Van Valin, 2008: 172).<sup>6</sup> As Van Valin notes, MPs also have a layered structure, for two reasons: First, many languages allow modifier phrases with an adjectival nucleus to take a core argument, such as the German *der auf seinen Sohn stolze Vater* [the of his son proud father]. Secondly, the modifiers can themselves be modified, “which means that they must have a periphery to house the modifying MP, e.g., *the very quickly extinguished fire*.” (Van Valin, 2008: 172).

In addition to adjectives and adverbs many languages, including English, also allow other, much more complex categories to serve as modifiers. Consider the following examples, from the same source.

- (6) a. The *Charles and Di* syndrome is no longer relevant.
- b. The *God is dead* philosophers are mostly dead.
- c. My grandson likes to give me the *who’s the boss now, silly old grandpa* wink frequently.

In (6a) the MP contains a conjoined RP in its nucleus, that in (6b) contains a clausal nucleus, and the nucleus of the MP in (6c) contains a constituent question as well as a vocative in its right-detached position. Figure 2 gives the constituent projection of (6c).

Even the seemingly inconspicuous category of pre-/postpositional phrases is not without problems. As Van Valin (2008: 170) notes, these are normally thought of as endocentric projections of a (pre- or postpositional) lexical head in predicative adpositional phrases, such as the underlined element in (7), which has a prepositional nucleus consisting of the preposition *in*, which functions as a predicate and thus licenses the object *the library*.

- (7) *Chris saw Pat in the library.*

The status of this unit in (7) is markedly different from that of the underlined units in (8), from the same source, referred to as non-predicative PPs as these units do not predicate but are licensed by the verb and therefore do not license an object.

- (8) a. *Chris showed the photo to Pat.*
- b. *Chris stole the photo from Pat.*
- c. *Christ presented Pat with the photo.*

As the underlined units in (8) are non-predicative, they do not contain a PRED and hence neither a NUC<sub>p</sub> (“nucleus of the PP”) nor a CORE<sub>p</sub>. Rather, they consist merely of a pre- or postposition and an RP. Figure 3 contrasts the internal structure of these two PP types.

---

<sup>6</sup> For an example of this, cf. the three peripheries in Figure 2 below.

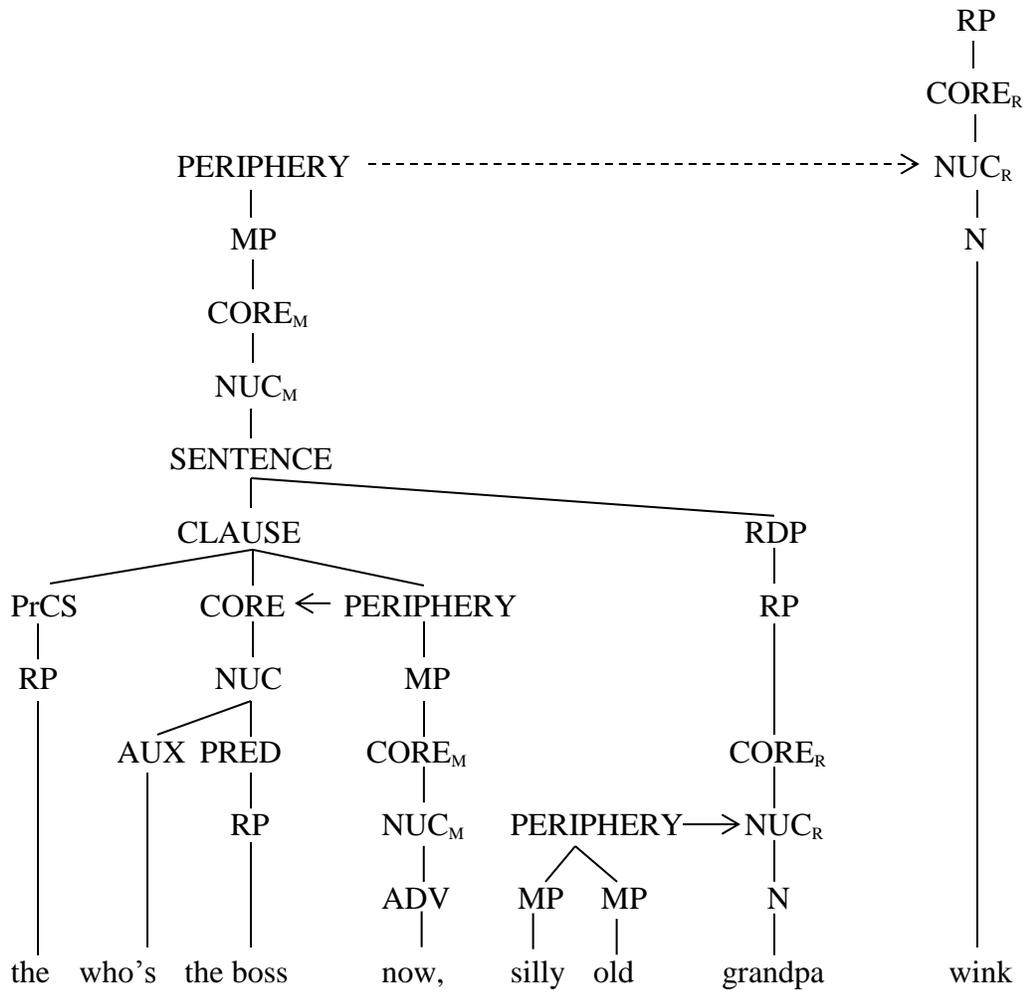


Figure 2: Structure of RP with MP containing sentential nucleus (Van Valin, 2008: 173)

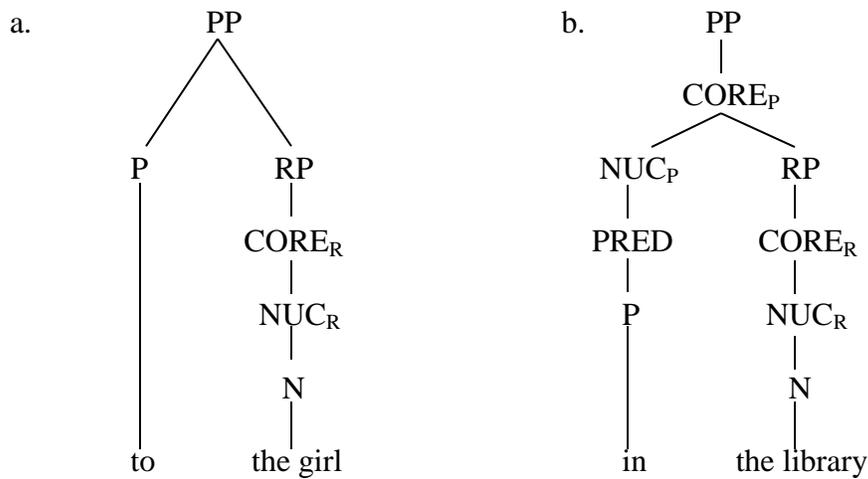


Figure 3: Non-predicative (a) and predicative (b) prepositional phrases in English (Van Valin, 2008: 171)

With respect to these two PP types, Van Valin (2008: 171) notes that predicative PPs most closely correspond to the X-bar notion of an endocentric category, as the nucleus here is always a

P and always a “head”. He also notes that a modifier like *right* in *right under the table* can be handled in terms of a modifier in the periphery within the predicative PP, while with a non-predicative PP there is no prepositional nucleus, the only nucleus being that of the RP. As such, they can arguably be analyzed as “*exocentric* PPs, i.e., PPs without a prepositional nucleus. Hence there is no layered structure, and this predicts that the admittedly limited set of PP-internal modifiers should not occur with non-predicative PP, which seems to be the case: \**Chris showed the photo right to Pat*/\**Chris stole the photo right from Pat*/\**Chris presented Pat right with the photo*.”

## 2.2 Gradience in lexical categories

Any definition of lexical classes in terms of feature bundles will eventually run into problems as there will always be at least some content morphemes which show many but not all of the “prototypical” features associated with a certain lexical category. For example, although most nouns in English and many other languages can appear in the plural, this is problematic with others, such as *peace*. Similarly, while most nouns e.g. in German are compatible with definite and indefinite articles, such as *der / ein Mann* ‘the / a man’, others, such as *Verlaub* ‘leave; permission’ are compatible with neither, in this case being restricted to its occurrence in the fixed (prepositional) phrase *mit Verlaub* with the preposition *mit* ‘with’ and the meaning ‘with respect; if you will pardon my saying so’. Although *Verlaub* is listed in many dictionaries with the definite article *der* (masculine, singular), it is in fact never found with this article in actual speech, although it is considered a noun for other reasons, e.g., its compatibility with the preposition *mit*, which presupposes an NP (which is assumed to have a nominal head!) and its perceived inherent (masculine) gender, as inherent gender is restricted to nouns in German.

Similar problems are encountered with what Ross (1972) refers to as “category squishes” or sequences of forms whose feature bundles gradually change so that it is often difficult to decide which category they should belong to. Van Valin (2008: 174) cites the examples in (9).

- (9) a. The savant quickly proved the theorem  
 b. That the savant quickly proved the theorem (astonished the professors)  
 c. the savant/\*the quickly proving the theorem (astonished the professors)  
 c'. the savant having quickly proved the theorem  
 c''. (I am aware of) the theorem having quickly been proved by the savant.  
 d. the savant's/\*the quickly proving the theorem (astonished the professors)  
 d'. the savant's having quickly proved the theorem  
 d''. the theorem's quickly being proved by the savant  
 e. the savant's/the quick proving of the theorem (astonished the professors)  
 e'. \*the savant's quick having proved of the theorem  
 e''. the quick proving of the theorem by the savant  
 e'''. \*the theorem's quick proving by the savant  
 f. the savant's/the quick proof of the theorem (astonished the professors)  
 f'. the quick proof of the theorem by the savant  
 f''. the savant's/the quick proof (astonished the professors)

Although *proved* in example (a) is clearly verbal and *proof* in the (f) examples is clearly nominal, the status of the forms ending in *-ing* is much less clear: While some instances of *proving* such as those in (c) and (d) take a direct object, suggesting that this form is a verb, the

forms in (e) and (e’)) require an “object” marked by the preposition *of*, suggesting that this is a noun. Furthermore, the “subject” NP in (c) appears as a bare NP, while in (d) – as in the more “nominal” form in (e) or the clearly nominal form *proof* in the (f) examples – it appears in the genitive. These and other factors such as adjectival modification, optional complements, tense, aspect and voice are summarized in tabular form in Van Valin (2008: 175).<sup>7</sup>

As we see in (9), even in languages such as English, which otherwise seem to have quite clear lexical classes, it is often not feasible to assume that each and every word form unambiguously belongs to a particular lexical category. RRG thus assumes that it is the INDIVIDUAL FEATURES which are relevant and not the labels themselves: “What is relevant to the grammar is the features themselves, not any category labels that might be overlaid on them. So “noun” is just a useful descriptive label for a certain pattern of lexical item distributions in a language, just like “subject” is a useful descriptive label for a certain consistent pattern of restrictive neutralizations in a language. Analogous to grammatical relations, some languages have well-defined and sharply differentiated lexical categories, while others do not.” (Van Valin, 2008: 176)

In the following two subsections, we discuss concrete examples from two other languages, Kharia (Munda, India) and Wari’ (Chapakuran, Brazil and Bolivia), where assuming clear-cut lexical classes and strictly endocentric syntactic categories is even more difficult, and we will show how these “problematic” elements are in fact dealt with straight-forwardly in RRG.

## 2.3 “Problematic” Categories

### 2.3.1 Parts of speech in Kharia

In the South Munda language Kharia, spoken in eastern-central India, Peterson (e.g., 2011a; b; 2013) argues that the familiar lexical categories of “noun”, “verb”, “adjective” etc. are not only unnecessary for an adequate grammatical description of the language, assuming their presence in fact considerably complicates the discussion.

Peterson divides the Kharia lexicon into two major classes: The vast majority of content morphemes, i.e., morphemes denoting referents, attributes and events/states, may be used freely in referential, attributive and predicative functions with no overt derivational marking, “light verbs”, etc. This class is open as new morphemes (e.g., modern technology and loan words in general) are constantly being added to its ranks. There is also a closed class which can be further divided into two subclasses: “Proforms / Deictics”, which may be used freely in referential and predicative function, but may only be used attributively if marked for the genitive, and “grammatical morphemes”, which may never be used referentially, predicatively or attributively. This is illustrated in Figure 4.

(10)-(11) present two simple examples demonstrating the flexibility of content morphemes.

- (10) a. *lebu del=ki.*                      b. *bhagwan lebu=ki*                      *ro del=ki.*  
       man come=MID.PST                      God                      man=MID.PST                      and                      come=MID.PST  
       ‘The / a man came.’                      ‘God became man [= Jesus] and came [to earth].’  
       [(1b) adapted from Malhotra, 1982:136]

<sup>7</sup> Cf. Maas (2004) for similar data from German (p. 368) and Portuguese (p. 374).

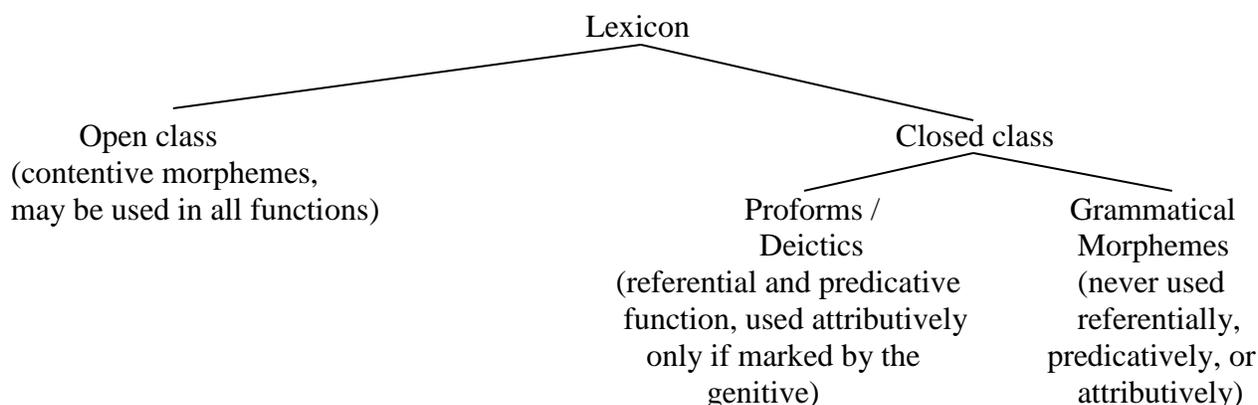


Figure 4: The Kharia lexicon (Peterson, 2011a: 78)

(11) (in a play about me and you, in which both of us will be taking part):

*“natak=te in=ga ho=kar=na=in ro am=ga in=na=m.” “umbo?*  
 play=OBL 1SG=FOC that=SG.HUM=MID.IRR=1SG and 2SG=FOC 1SG=MID.IRR=2SG no

*am=na um=in pal=e. direktar sen=ga? in=te ho=kar=o?*  
 2SG=INF NEG=1SG be.able=ACT.IRR director early=FOC 1SG=OBL that=SG.HUM=ACT.PST

*am=ga am=na=m.”*  
 2SG=FOC 2SG=MID.IRR=2SG

“In the play I will be him and you will be me.” “No. I can’t be you. The director already made me him. You will be you.”“

This process is entirely productive, given a proper context. Cf. e.g. (12)-(21), from Peterson (2011a: 75-92), which gives some idea of the pervasiveness of this flexibility throughout the lexicon.

(12) Interrogatives: *i* ‘what; which; do what?’

(13) Indefinites: *jahā* ‘something; some (attribution); do something’

(14) Quantifiers: *moṇ* ‘one (referential/attributive); become one’

(15) Properties: *rusuṇ* ‘red (one); red (attributive); become red’, *maha* ‘big (one); big; grow, become big’

(16) Proper names: *a?ghrom* ‘Aghrom (name of a town) (referential/attributive); come to be called “Aghrom” (middle voice), name [something] “Aghrom” (active voice)’

(17) Status and Role: *ayo* ‘mother; become a mother (middle voice), accept someone as a mother (active voice)’

(18) Deictics and proforms: *iḍa?* ‘yesterday; become yesterday (middle), turn (e.g. today) into yesterday (active, e.g., with God as subject)’

(19) Physical objects and animate entities: *kaḍoṇ* ‘fish; become a fish (middle), turn into a fish (active)’

(20) Locative: *tobluy* ‘top, rise (middle), raise (active)’

(21) Activities: *silo?* ‘plowing (n.); plowed; plow’

In addition to this “precategoryality”, what appear to be entire NPs can also denote an event. Consider examples (22)-(23), both of which have the same propositional content.

(22) *ho rocho<sup>2</sup>b=te col=ki=n*  
 that side=OBL go=MID.PST=1SG  
 ‘I went to that side’

(23) *ho rocho<sup>2</sup>b=ki=n*  
 that side=MID.PST=1SG  
 ‘I went to that side’ (lit.: “I that-side-d”)

Example (23) shows that an analysis of Kharia as possessing an N/V distinction is problematic: Even if we were to treat *rocho<sup>2</sup>b* ‘side’ in (23) as a (presumably zero-derived) verb, we would then have a verb modified by the demonstrative *ho* ‘that’. As (24)-(25) show, the construction in (23), with an apparent NP as the semantic base of the predicate, is entirely productive and can also contain both quantifiers and genitive attributes, in addition to demonstratives. This is especially problematic for assuming “rampant zero conversion” (Evans & Osada, 2005), as this would entail productive zero derivation of a verb, i.e., a lexical stem, from a full-fledged NP, i.e., from a complex syntactic unit.

(24) *ubar rocho<sup>2</sup>b=ki=n*  
 two side=MID.PST=1SG  
 ‘I moved to both sides (i.e., this way and then that).’

(25) *o<sup>2</sup>=ya<sup>2</sup> telon=ki.*                      *o<sup>2</sup>=ya<sup>2</sup> telon=o<sup>2</sup>=ki.*  
 house=GEN roof=MID.PST                      house=GEN roof=ACT.PST=PL  
 ‘The house’s roof was thatched.’    ‘They thatched the house’s roof.’

On the basis of data such as these, Peterson (2013) assumes two structurally defined categories, the “Case-syntagma” and the “TAM/Person-syntagma”, both of which have the same potential structure for the semantic base and which differ only with respect to their functional marking, i.e., case or TAM/voice and person/number/honorific marking, motivated by similar structures assumed for Tongan by Broschart (1997). The structure of the semantic base is given in (26). None of these elements is obligatory as long as some non-enclitic unit is present. The Kleene star (\*) denotes that potentially any number of content morphemes (LEX) is possible, including zero.

(26) (LEX=GEN) (DEM) (QUANT (CLASS)) (LEX=GEN) (LEX\*) (=POSS) (=NUM)

The maximal potential structure of the Case-syntagma is given in (27), where “X” is equal to the structure of the semantic base in (26).

(27) X=CASE

“CASE” in (27) refers to postpositions such as *buy* ‘with; INST’, *sey* ‘before’, etc., as well as the two cases given in (28), so that any postposition, =*te* or the lack of any overt marking signals that the unit is a Case-syntagma.

(28) Case: Direct (zero marking)<sup>8</sup> – the case of subjects and indefinite direct objects;  
Oblique (marked by =*te*) – marks definite direct objects, “indirect objects” and  
 adverbials

<sup>8</sup> Zero marking is restricted to the Case-syntagma, as the TAM/Person-syntagma is always overtly marked as such. Thus, the lack of any marking is unambiguous as it is only found with Case-syntagmas in the function of subject or indefinite direct object.

The genitive is not a case in this sense: Whereas the direct and oblique cases are relevant at the clause level to mark the relation of a constituent to the predicate (roughly: subject / non-subject), the genitive is only relevant within the semantic base and serves to integrate one semantic base into a larger semantic base. The oblique case also cannot appear in a TAM/Person-syntagma, as (29) shows, whereas the genitive can (30).

(29) \*sahar=te=ki=jn.  
 city=OBL=MID.PST=1SG  
 ‘I went to the city.’

(30) ayo=yaʔ=yoʔ  
 mother=GEN=ACT.PST.1SG  
 ‘he or she made [it] mother’s’ (lit.: “he or she mother’s-ed [it]”)

The structure of the TAM/Person-syntagma is given in (31), where “X” again refers to the semantic base given in (26). “(v2)” or “vector verbs”, as they are referred to in South Asian studies, refers to phonological and morphosyntactic words which mark Aktionsart or the passive. VOICE refers to “basic voice”, i.e., the active and middle (Klaiman, 1991).<sup>9</sup>

(31) X (V2\*) (=PERF)=TAM/VOICE=PERSON/NUM/HON

Examples (32)-(35) illustrate this structure with a number of TAM/Person-syntagmas in Kharia. The semantic base is underlined.

(32) col=ki=may  
 go=MID.PST=3PL  
 ‘they went’

(33) boksel=nom      goʔq=ki  
 sister.in.law=2POSS    C:TEL=MID.PST  
 ‘she became your sister-in-law’ (= ‘she “your sister-in-law-ed”’)

(34) ho=jeʔ      u=jeʔ=ki      goʔq=ki  
 that=SG.NHUM    this=SG.NHUM=PL    C:TEL=MID.PST  
 ‘that became these’ (= ‘that “these-d”’)

(35) ro tama am=pe u nāw kuʔum=te=ga sadi biha kerson=na=pe ...  
 and now 2=2PL    this nine    family=OBL=FOC    marry marry marry=MID.IRR=2PL  
 ‘And now you will marry in only these nine families ...’ [AK, 1:69]

Examples (36)-(37) show a Case-syntagma and a TAM/Person-syntagma with the same semantic base (underlined), differing only with respect to their functional marking. Their basic structures are illustrated in Figure 5.<sup>10</sup>

<sup>9</sup> As the passive is marked by a V2, not by an obligatory fusional TAM/VOICE marker, and always occurs together with the middle voice, it has a different status than the active and middle, which can be considered “basic”.

<sup>10</sup> The triangles in Figure 5 indicate that there is further internal structure which we cannot deal with here but which does not affect the discussion at hand. For a fuller discussion, cf. Peterson (2013).

(36) *bides=a? lebu=ki=ya? rupraŋ=te*  
 abroad=GEN person=PL=GEN appearance=OBL  
 ‘the appearance of foreigners’ (e.g., as the object of the predicate)

(37) *bides=a? lebu=ki=ya? rupraŋ=ki=may.*  
 abroad=GEN person=PL=GEN appearance=MID.PST=3PL  
 ‘They took on the appearance of foreigners (e.g. by living abroad so long).’

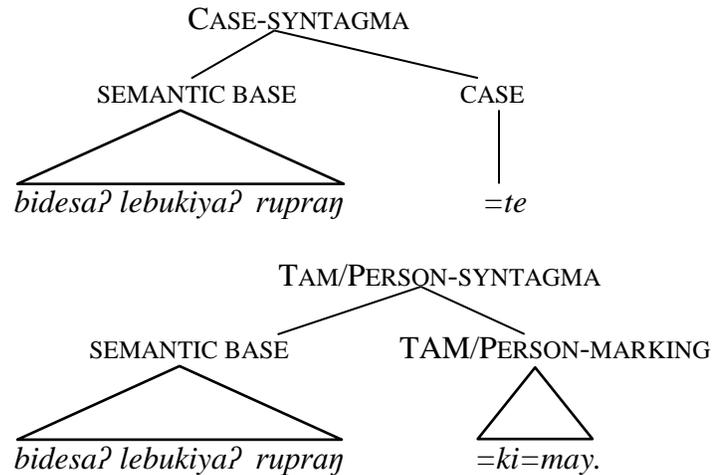


Figure 5: TAM/Person- and Case-syntagms with the same semantic base

These two syntactic units are defined purely structurally and both can appear in attributive, predicative and referential functions, as the following examples from Peterson (2013: 137) show.

Case-syntagma in attributive function

(38) *kuda kolon daru* (Note: *kuda* modifies *kolon* and the two together modify *daru*.)<sup>11</sup>  
 millet bread tree  
 ‘a millet bread tree’ (from a children’s story)

TAM/Person-syntagma in attributive function

(39) *yo=yo?j lebu col=ki.*  
 see=ACT.PST.1SG man go=MID.PST  
 ‘The man I saw left’

Case-syntagma in predicative function

(40) ... *ro u=ga ho jinis=a? komaŋ.* [AK 1:57]  
 and this=FOC that animal=GEN meat  
 ‘...and this [is] that animal’s meat.’

TAM/Person-syntagma in referential function

(41) *kunqab aw=ki tomlin khariya gam dom=na la?=ki=may ina no*  
 behind QUAL=MID.PST milk Kharia say PASS=INF IPFV=MID.PST=3PL because

<sup>11</sup> In Peterson (2006, section 4.6.2) it is shown that *kuda kolon daru* is not a compound.

*u=ki tomlin u<sup>2</sup>d=ga del=ki=may.* [MT, 1:180]  
 this=PL milk drink=FOC come=MID.PST=3PL  
 ‘[Those who] were in the rear (= literally: ‘they were behind’) were called “Milk Kharia” because they came drinking milk.’

Finally, an analysis of the forms given in the preceding pages as consisting of nominals and “hidden verbs” is also not tenable: As noted above, it is not possible in Kharia to combine TAM/Person marking with case marking or with postpositions, so that a construction such as the one in (42), which ends in a postposition, is not grammatical if directly followed by TAM/Person-marking, as (43) shows.<sup>12</sup>

(42) *iku<sup>2</sup>d jughay du?kho bun*  
 very much sorrow INST  
 ‘very sad’, literally: ‘with very much sorrow’

(43) \**iku<sup>2</sup>d jughay du?kho bun go<sup>2</sup>d=ki.*  
 very much sorrow INST C:TEL=MID.PST  
 ‘(He) became very sad’ (= ‘with very much sorrow’)

However, if the semantic part of this unit is re-ordered as in (44), with two constituents instead of just the one in (42), the unit which is not marked by a postposition may serve as the semantic base of a TAM/Person-syntagma. This is a common predicate type in Kharia (Peterson, 2011a: 220).

(44) *gupa lebu du?kho bun iku<sup>2</sup>d jughay go<sup>2</sup>d=ki.*  
 watch person sorrow INST very much C:TEL=MID.PST  
 ‘The shepherd (= watch person) became very sad.’ [RD, 1:18]

At issue here is that the unit ending in the postposition may not be directly followed by a TAM marker, whereas the unit which does not end in a postposition/case marker can. Furthermore, if a “light verb” such as *hoy* ‘become’ (borrowed from Kharia's Indo-Aryan neighbour Sadri) is inserted between *bun* ‘INST’ and *go<sup>2</sup>d=ki* in (43), this results in a grammatical predicate, consisting of a Case-syntagma and a marker of qualitative predication ( $\approx$  copula).<sup>13</sup>

(45) *iku<sup>2</sup>d jughay du?kho bun hoy go<sup>2</sup>d=ki.*  
 very much sorrow INST become C:TEL=MID.PST  
 ‘(He) became very sad.’

Thus “hidden verbs” in Kharia would have a very different distribution than the “light verb” *hoy* ‘become’ or any in fact other supposed verb: They could only be assumed to occur after “predicate nominals” which are not case marked (e.g., for the structure in (33), i.e., those units we consider to be the semantic base of either a TAM/Person- or of a Case-syntagma), but not in structures such as (43), so that they would also not be able to occur where the overt “light verb” *hoy* does (compare (43) with (45)). Thus, hidden verbs do not simplify the analysis but rather

<sup>12</sup> The following discussion and the corresponding examples are from Peterson (2013: 145-146).

<sup>13</sup> This discussion and the accompanying examples have been taken from Peterson (2011a: 111).

complicate it considerably, as they would not only be non-overt but would also have a unique distribution.

Example (46) presents a simplified example of an intransitive sentence in Kharia and Figure 6 presents its constituent projection in the present analysis.

- (46) *u=kiyar taŋ el=a? sori=ga nog=e uq=e=kiyar.*  
 this=DU now 1PL.EXCL=GEN COM=FOC eat=ACT.IRR drink=ACT.IRR=DU  
 'They two will now eat and drink with us.' [adapted from Roy & Roy, 1937: 180f.]

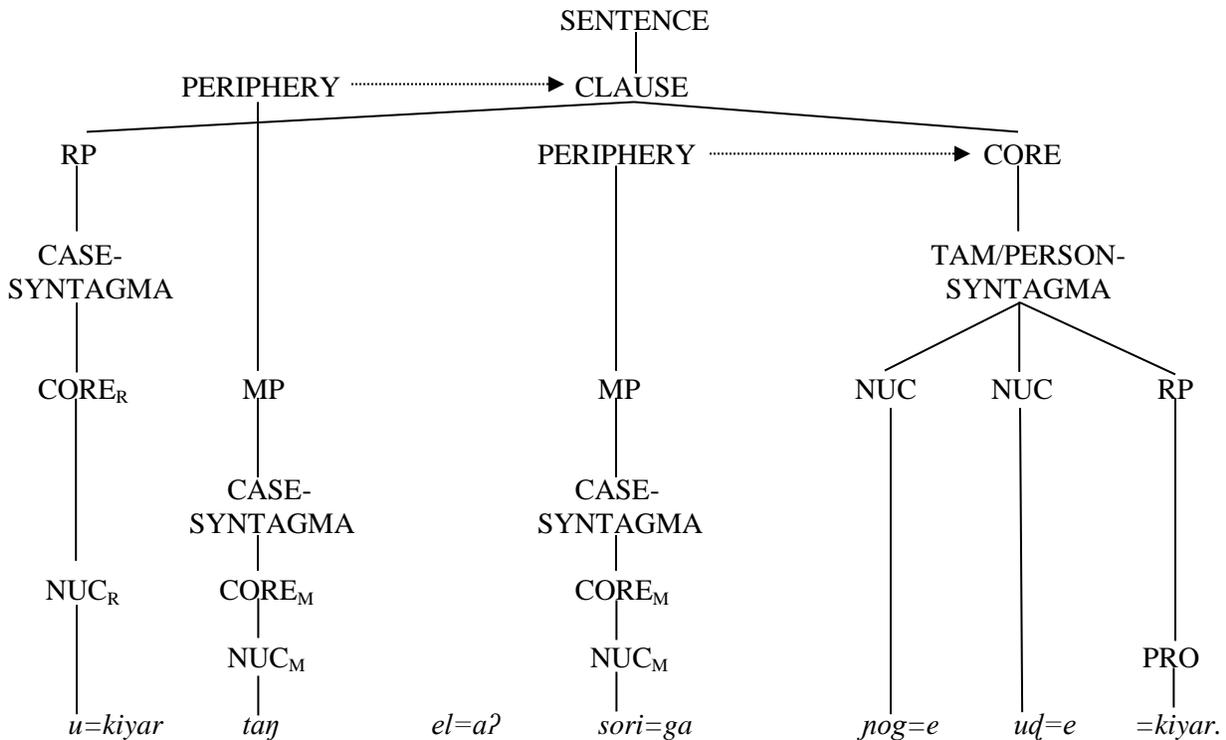


Figure 6: The constituent projection of (46)

Whether or not this analysis of parts of speech in Kharia is the “correct” or “best” analysis is not what is at issue here. Rather, the present discussion is merely intended to show how RRG can account for analyses of individual languages which do not assume the existence of such *a priori* “universal” categories.

Finally, there is a rather marginal type of TAM/Person-syntagma in Kharia, quotatives, in which the semantic base consists of an entire sentence, as in example (47).

- (47) *idib tunbo? “kerson=e la! kersonə la!” lo?=na=kiyar.*  
 night daytime marry=ACT.IRR VOC REP CONT=MID.IRR=DU  
 ‘Day and night they both will keep on [say]ing “Marry! Marry!”’ [Kerkeṭfā, 1991:31]

Although Peterson (2013) offers no final analysis of this form, as it is unique in his corpus, in the next section we will encounter very similar predicates in the Chapakuran language Wari’, in which these structures appear to be much more common.

### 2.3.2 The “Internal State Constructions” in Wari’

A similar construction to the rather marginal Kharia quotative predicate discussed in the previous section is the much more productive “Internal State Constructions” (ISC) in Wari’, a member of the Chapakuran family of Brazil and Bolivia, to which we now turn.

As Everett (2008: 383) notes, in Wari’, a predicate-initial language (VOS), in verb-initial sentences the verb is immediately followed by the “VIC” or “Verbal Inflectional Clitic”, which marks for tense, voice, person, number and gender and agrees with both subject and object. Cf. examples (48)-(49), where the VIC is underlined.

(48) *Quep* *na-in* *xirim* *te* *pane* *ta*.  
do 3SG.REAL.PST/PRS-3N house father.1SG REM.PST EMPH  
‘My father made a house long ago.’ (Everett, 2008: 383)

(49) *Ten* *ta* *wao*’.  
weave PASS.3SG type of basket  
‘Baskets are woven.’ (Everett, 2008: 384)

In sentences in which the verb is not the first sentence constituent, e.g., with constituent questions, this sentence-initial unit is followed by a different class of clitics which mark tense and which agree with the gender of the item in sentence-initial position, regardless of its function in the sentence, as in (50).

(50) *Ma*’ *co* *tomi*’ *na*?  
that.PROX.HEARER M/F.REAL.PST/PRS speak 3SG.REAL.PST/PRS  
‘Who is speaking?’ (Everett, 2008: 385)

As Everett (2008: 386) writes, “Many Amazonian languages report on others’ thoughts, character, reactions, and other results of intentional states by means of quotatives, i.e., literally putting words in people’s mouths.” Wari’ also makes use of quotatives for these purposes, but their use is apparently much more common in that language, having also spread to other uses. What is interesting is that in Wari’, similar to the Kharia example in (47), there is no morpheme denoting ‘say’. (51) provides an example of this construction in Wari’, where the indices “i”, “j” and “k” show the respective congruence relationships.

(51) *Ma*’ *co* *mao* *na* *-in<sub>i</sub>* *Guajará<sub>i</sub>*  
that.PROX.HEARER M/F.REAL.PST/PRS go.SG 3SG.REAL.PST/PRS-3N Guajará (Brazilian city)  
  
*na<sub>j</sub>* *-nam<sub>k</sub>* *’oro* *narima<sub>k</sub>*’ *taramaxicon<sub>j</sub>*.  
3SG.REAL.PST/PRS-3PL.F collective woman chief  
‘“Who went to Guajará?” (said) the chief to the women.’ (Everett, 2008: 386)

Everett (2008: 391-392) notes a number of similarities between the the ISC predicator and simple verbs, suggesting that the clausal semantic base of the ISC occupies the same functional slot as verbs otherwise do, i.e., the predicate, although it consists of an entire clause. These include:<sup>14</sup>

<sup>14</sup> For reasons of space, Everett’s argumentation has been somewhat simplified here.

- the ISC predicate appears in the clausal position otherwise occupied exclusively by the verb;
- the last syllable of the ISC predicate is stressed, as with the verb in other sentence types. This suggests that the ISC predicate is considered a single unit;
- the predicate of an ISC may be modified just like a verb;
- the predicate of an ISC may undergo compounding, like all verbs.

If the ISC predicate is considered the predicate, just like verbs in verb-initial sentences, the following generalizations may be made for constituent order in Wari' (Everett, 2008: 392):

- Wari' sentences always begin with a verb, ISC predicate, or preverbal mood marker;
- the “VIC” appears directly after the predicate, whether this is clausal or a simple verb;
- tense appears in sentence-second position.

Everett (2008: 399) argues that Wari' inflectional clitics – i.e., those which follow the predicate of the clause – follow the NUC in an RRG analysis. As such, they follow the ISC predicate as well, which then occupies the NUC slot. Figure 7, from Everett (2008: 398),<sup>15</sup> shows the structural analysis of the structure which he assumes for example (51), where the entire “quote” is viewed as a nucleus.<sup>16</sup> The structure in Figure 7 captures the fact that the nucleus of the clause, which contains the predicate, consists of a form which, despite its role as the predicate, also has many of the trappings of a clause, as it consists of a precore unit and a core, which itself contains a nucleus (predicate) and two arguments. Nevertheless, it is clearly the predicate of the larger clause. This then, together with the two NPs *narima'* and *taramaxicon*, forms the core of the higher-level clause.

### 3. Grammatical categories – the operators

A great deal of research has traditionally gone into modeling an appropriate representation of the argument structure of predicates – both in RRG and other syntax theories. As discussed in Chapter XX of the present volume, in RRG this information is stored in the lexicon in the form of logical structures, which contain information on the basic meaning of the predicate, its Aktionsart class(es), and the number of arguments and their thematic relations. The predicate and its arguments provide us with what can be termed the “sentence base” (cf. Maas, 2004), i.e., the basic propositional information, but without any information relating to the specifics of the utterance, i.e., is the information a command, a request, a statement or a question? Does it refer to a past, present, future, iterative or hypothetical event/state? Are the referents identifiable, etc.? Information of this type is referred to as “sentence modality” or “semantic finiteness” in Maas (2004), to distinguish it from the overt marking of this information, which he refers to as “morphological finiteness”. This information is also often referred to as the “grounding” or “anchoring” of the sentence. This can be portrayed as in Figure 8.

---

<sup>15</sup> Here and in the following, minor changes will be made to analyses and figures, such as the use of “RP” as opposed to “NP”, as the concept of the RP was not fully justified until Van Valin (2008). Furthermore, minor changes with respect to glosses, etc., have silently been made and will not be commented on in the following figures.

<sup>16</sup> The representation of the operators has been left out of Figure 7, as these are not dealt with until Section 3 below.

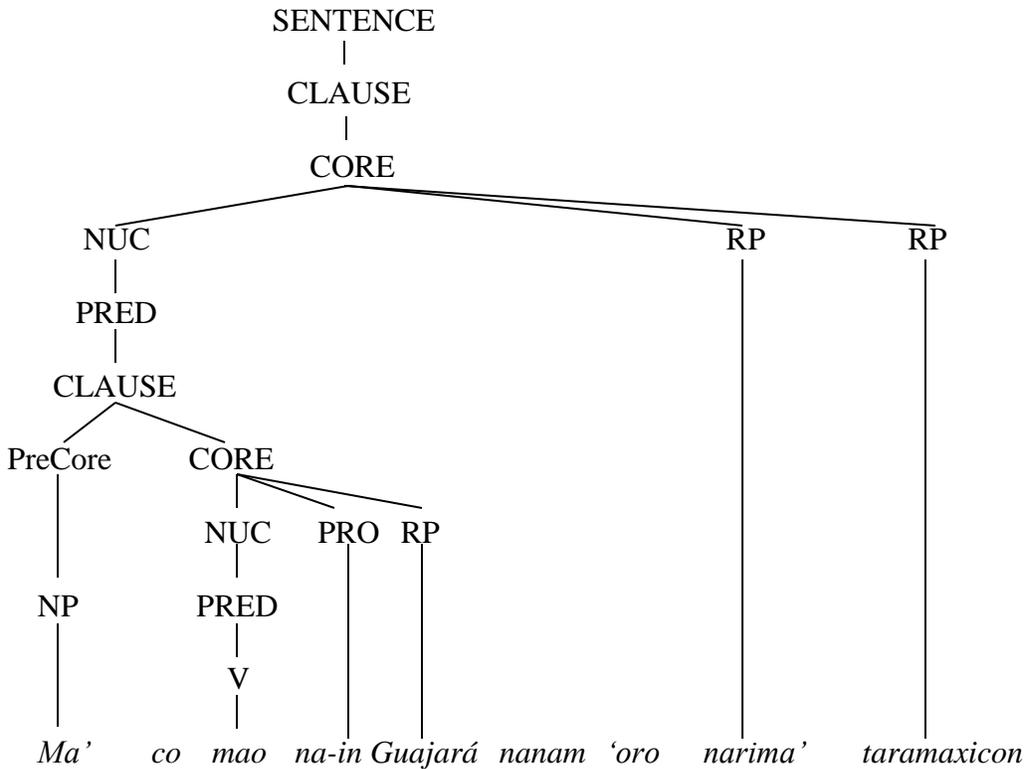


Figure 7: Constituent projection of (51) (Everett, 2008: 398)

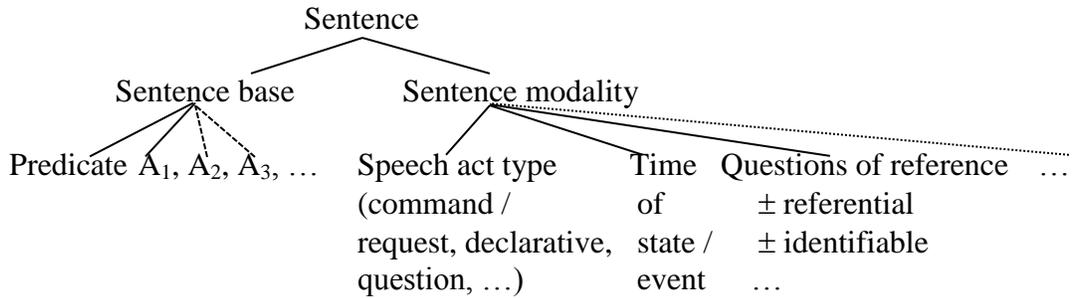


Figure 8: The sentence base and sentence modality

This grounding corresponds closely to Tomasello’s (2008) three kinds of “syntax” i.e., “simple”, “serious” and “fancy syntax”, which accompany what he views as the three basic evolutionary foundations of human cooperative communication:

“Thus, since requesting prototypically involves only you and me in the here and now and the action I want you to perform, combinations of natural gestures and / or linguistic conventions require no real syntactic marking but only a kind of “simple syntax” in a grammar of requesting [...] But when we produce utterances designed to inform others of things helpfully, this often involves all kinds of events and participants displaced in time and space, and this creates functional pressure for doing such things as marking participant roles and speech act functions with “serious syntax” in a grammar of informing. Finally, when we want to share with others in the narrative mode about a complex series of events with multiple

participants playing different roles in different events, we need even more complex syntactic devices to relate the events to one another and to track the participant across them, which leads to the conventionalization of “fancy syntax” in a grammar of sharing and narrative.” (Tomasello, 2008: 244-245).<sup>17</sup>

In RRG, this information is divided into two broad categories: The first group we will refer to as “proposition-grounding”, as these categories ground the event or state with respect to time, modality, speech act, etc. The second group we will refer to as “referent-grounding”, as it refers to referential identifiability.

### 3.1 Proposition-grounding operators

We begin with a discussion of propositional grounding, which is expressed in individual languages by language-specific grammatical categories, such as “present tense”, “past tense”, “hearsay”, “mirative”, etc. The grounding information expressed by these language-specific categories is divided into three different groups according to which level of the layered structure of the clause (LSC) they ground, i.e., the nucleus, the core or the clause. RRG assumes that individual languages will differ as to which grammatical categories express which of these notions and there is no requirement that every language have all of the possible operators. In fact, most languages do not have grammatical categories expressing all of these semantic categories, although it is assumed that all languages will have grammatical means to express at least two of these categories – negation and illocutionary force (Van Valin, 2005: 9). These operators are summarized in Table 2.

Nuclear operators such as aspect or directionals modify the state/event itself without reference to the participants, while the the core level operators modify the relation between the arguments and the event, most notably directionals, (deontic) modality (‘be able’, ‘must’, etc.) and narrow scope negation, but also event quantification. Finally, operators which have scope over the entire clause include “status”, i.e., epistemic modality and external negation, as well as tense, evidentials and illocutionary force.

---

<sup>17</sup> Of these three functions, the last two, i.e., helpful informing and the narrative mode, seem to be species-specific and are not found even with our closest evolutionary relatives, the non-human primates. E.g., Tomasello (2008: 38) cites studies on “linguistic apes”, i.e., human-raised apes which could communicate through signs, in which 96-98% of all signed productions were imperatives, with the remaining 2-4% having no clear functional interpretation but with no clear cases of helpful information or narratives. Similarly, the data cited in Håkansson & Westander (2013) contain no clear cases of non-human animal communication where helpful information is being offered (other than (largely involuntary) alarm signals) or narratives are being told.

Hockett (1960 [1982]: 6) already notes that his design-feature of “displacement” or “being able to talk about things that are remote in space or time (or both) from where the talking goes on” is found not only in human communication but also in bee-dancing. The referential communication of bee-dancing, however, arguably does not correspond to what Tomasello (2008) refers to as “inform[ing] others of things helpfully” (and certainly not to the “narrative mode”) but presumably to a prompt or command to other members of the hive to go the place being described. With that, the last two functions in Tomasello’s (2008) analysis appear to be unique to humans. But even if bee-dancing is viewed as “helpful informing”, the narrative mode, and with that Tomasello’s “fancy syntax”, would still be uniquely human.

---

**Operators in the layered structure of the clause**

---

**Nuclear operators:**

Aspect

Negation

Directionals (only those modifying the orientation of an action or event without reference to the participants)

**Core operators:**

Directionals (expressing the orientation or motion of one participant with reference to another participant or to the speaker)

Event quantification

Modality (root modals, e.g. ability, permission, obligation)

Internal (narrow scope) negation

**Clause operators:**

Status (epistemic modals, external negation)

Tense

Evidentials

Illocutionary force

---

Table 2: Operators in the layered structure of the clause, from Van Valin (2005: 9)

Operators are represented by an operator projection, given below the respective language example. This projection is joined to the constituent projection through the nucleus. Figures 9 and 10 illustrate this for English and Japanese.

Note that in the left-hand example in Figure 9 the information for the aspect operator comes from two different positions: the auxiliary *be* preceding *leaving* and the suffix *-ing* on this form. In fact, this information can potentially appear anywhere in the sentence. The only claim that RRG makes with respect to the placement of operators is “that the ordering of the morphemes expressing operators with respect to the verb indicates their relative scopes. That is, taking the nucleus as the reference point, the morphemes realizing nuclear operators should be closer to the nucleus than those expressing core operators, and those manifesting clausal operators should be outside of those signalling nuclear and core operators.” (Van Valin, 2005: 11) All evidence so far suggests that this is true of language in general: Although there may be language-specific variation for the ordering of the operators within a particular level, cross-linguistic data for operator marking appears to substantiate this claim for those markers for different levels which appear together on the same side of the nucleus. Cf. e.g. examples (52)-(56), from Van Valin (2005: 10).<sup>18</sup>

---

<sup>18</sup> For the original references, see the respective examples in Van Valin (2005: 10).

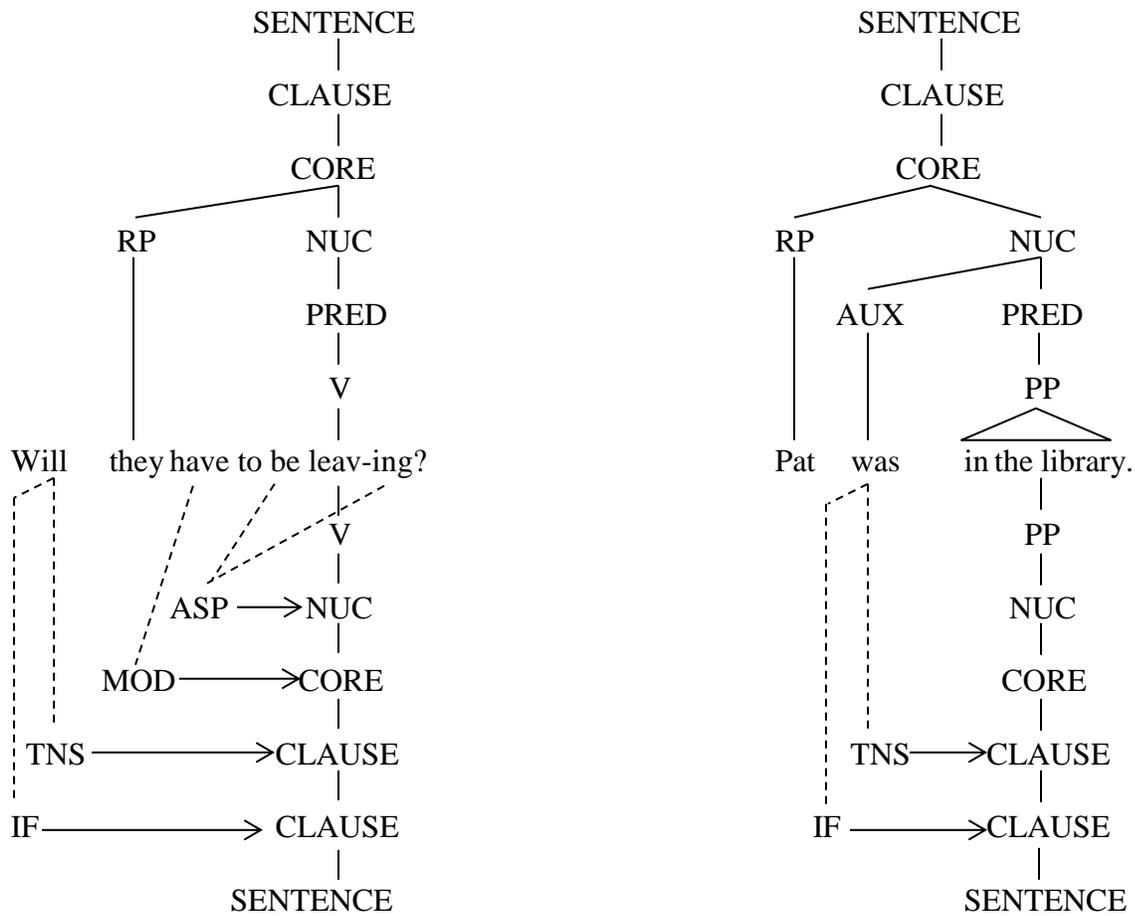


Figure 9: Constituent and operator projections for two English sentences (Van Valin, 2005: 14)<sup>19</sup>

<sup>19</sup> "... in English, illocutionary force is indicated by the position of the tense marker in the main clause: interrogative by core-initial tense, declarative by core-internal tense, and imperative by no tense." (Van Valin, 2005: 10, fn. 2).



Korean

(54) *An-tul-li-wu-(si)-lswuiss-cianh-ass-keys-up-nita.*

NEG-hear-CAUS-PASS-(SH)-ABLE-NEG-PST-PRESUM-POL-DEC

‘(I) guess that (he) [HON] might not be heard.’

(NEG-V-MOD-NEG-TNS-EVID-IF)

English

(55) a. *He may be leaving soon.*

(IF/TNS-STA-ASP-V)<sup>20</sup>

b. *She was able to see them.*

(IF/TNS-MOD-V)

c. *Will they have to be leaving?*

(IF/TNS-MOD-ASP-V)

Tiwi

(56) *ŋə-ru-un̩tiŋ-apa.*

1SG-PST-PROG-eat

‘I was eating.’

(TNS-ASP-V)

Cross-linguistic data on the position of operators (or markers of finiteness in general)<sup>21</sup> within the sentence also suggests that these operators can theoretically appear at any position in the sentence and are not restricted to any particular position (e.g., adjacent to the verb, sentence-final, or sentence-second (“Wackernagel”) position). For example, in his cross-linguistic study of finiteness, Maas (2004) suggests the typology of morphological finiteness given below in Table 3. Types I, II and III in Table 3 denote simple predicates which consist of a single predicating unit (generally a simple verb), while Types Ia, IIa, IIb and IIc are complex predicate types which consist of at least two parts (whether verbal or not) which together express a single predicate. “E” refers here to the complement of a predicate, which may be either an argument or an adjunct, “P” denotes the predicate and “f” refers to markers of (personal) finiteness. “+” refers to a loose concatenation of words and “-” to a narrow morpheme concatenation.

| Type       | Scheme                                     |  |
|------------|--|--|
| <b>I</b>   | E + P                                      | No (verbal) morphology   |
| <b>Ia</b>  | E + P (Modifying + Modified)               | No (verbal) morphology   |
| <b>II</b>  | E + P-finite                               | Morphologically finite predicate   |
| <b>IIa</b> | E + P (Modifying + Modified-finite)        | Modified element (= “lexical verb”) is finite, modifying (or “situating”) element is invariable (e.g., “particle”) |
| <b>IIb</b> | E + P (Modifying-finite + Modified)        | Modifying element is finite, modified element is non-finite (= auxiliary-participle type)                          |
| <b>IIc</b> | E + P (Modifying-finite + Modified-finite) | Both elements are finite   |
| <b>III</b> | (E + P)-finite                             | Movable affixes  |

Table 3: Predicate types with respect to “finiteness” marking (Maas, 2004: 379, modified)

<sup>20</sup> See footnote 19.

<sup>21</sup> Although the notion of “finiteness” is closely related to operators in RRG, the two are not entirely identical, as e.g., RRG does not view markers for honorificity (“politeness”, “deferentiality”) as operators, although these are generally viewed as markers of finiteness in languages where these are found, e.g. Korean or in many South Asian languages. For further discussion of finiteness from a typological perspective, cf. Maas (2004), Bisang (2001) and the studies in Nicolaeva (ed.), (2007).







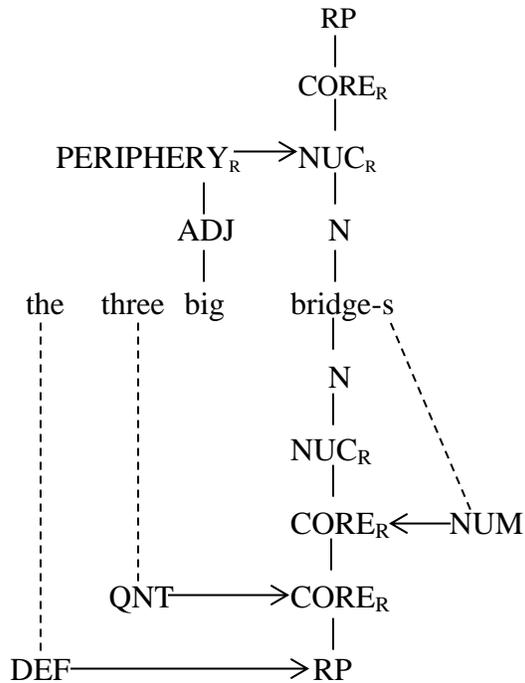


Figure 12: Layered structure of the RP in English (Van Valin, 2005: 25)

To give an example of the positional variability of these operators, consider the different orders in the Dyirbal examples in (64). Dyirbal is spoken in northern Australia and has exceptionally free word order. Note that all three variants in (64) are all grammatical and have the same meaning. (Van Valin, 2005: 28) Note also the position of the determiners with respect to the nouns which these modify.

- (64) a. *Bayi bargan baŋul yaɾa-ŋgu ɟurɟa-ŋu gambi-ɾa.*  
 DET.ABS wallaby.ABS DET.ERG man-ERG spear-TNS mountains-LOC  
 ‘The man speared the wallaby in the mountains.’
- b. *Baŋul gambi-ɾa yaɾa-ŋgu bayi ɟurɟa-ŋu bargan.*  
 DET.ERG mountains-LOC man.ERG DET.ABS spear-TNS wallaby.ABS
- c. *Yaɾa-ŋgu ɟurɟa-ŋu gambi-ɾa bargan baŋul bayi.*  
 man.ERG spear-TNS mountains-LOC wallaby.ABS DET.ERG DET.ABS

The operator projection of the various orders in (64) is straightforward and is given in Figure 13, which presents the projection of the structures in (64)b.

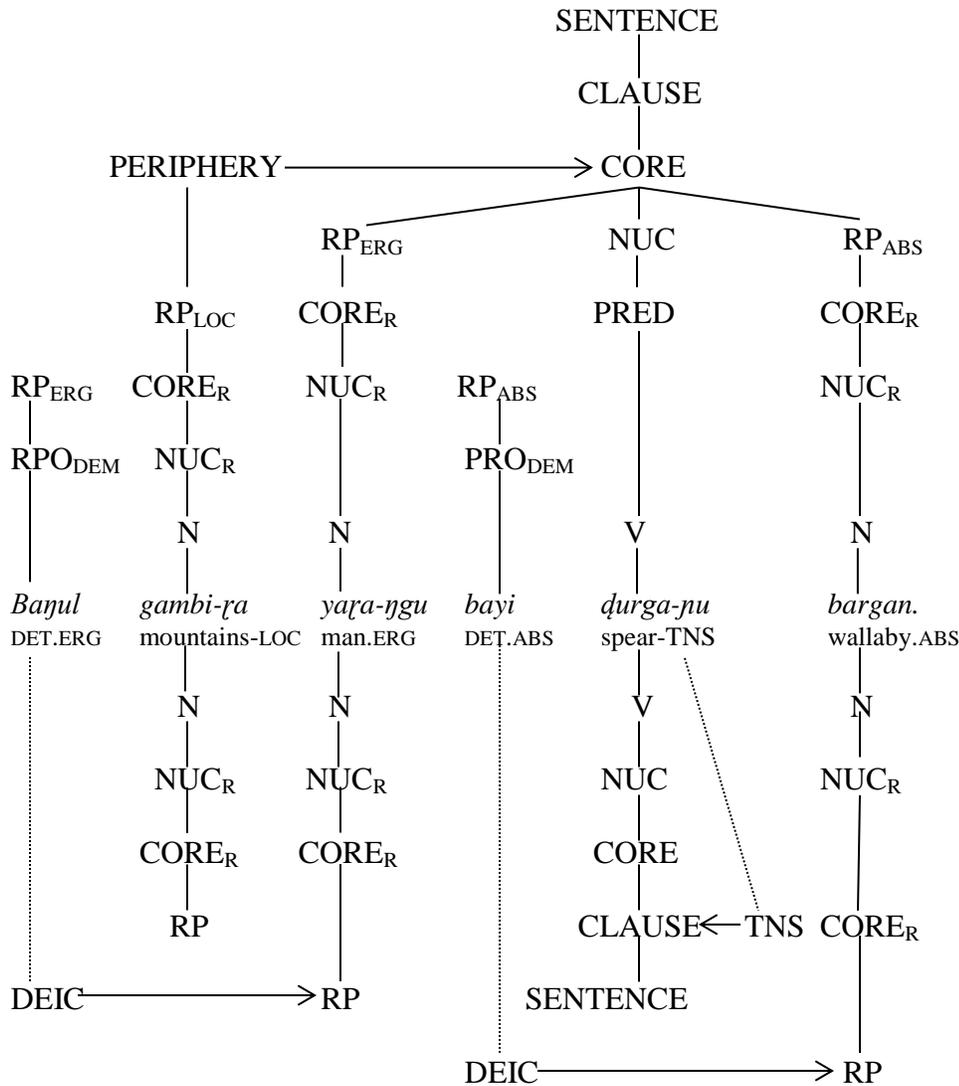


Figure 13: The structure of the Dyrbal sentence in (64)b (Van Valin, 2005: 29)

Similar to the proposition-grounding operators in the semantic representation of the sentence, the reference-grounding operators are represented in the semantic representation of the RP. An example is given in (65), from Van Valin (2005: 52), assuming that these are the same for the RP as those previously assumed for the NP. The conventions used here (i.e., angled brackets denoting scope, italicized small caps, etc.) are the same as those used to represent the proposition-grounding operators. “ $\exists$ ” is the “existential operator”, roughly meaning “there is (at least one)”.

(65)  $\langle_{\text{DEIC}} \textit{PROX} \langle_{\text{DEF}} + \langle_{\text{NEG}} \emptyset \langle_{\text{QNT}} \exists \langle_{\text{NUM}} \textit{SG} \langle_{\text{NASP}} \textit{COUNT} \langle_{\text{N}} \rangle \rangle \rangle \rangle \rangle \rangle$

The logical structure for a phrase like *the scarf* would then be represented as in (66), also from Van Valin (2005: 52).

(66)  $\langle_{\text{DEF}} + \langle_{\text{NEG}} \emptyset \langle_{\text{QNT}} \exists \langle_{\text{NUM}} \textit{SG} \langle_{\text{NASP}} \textit{COUNT} \langle \langle \textit{scarf} \rangle \rangle \rangle \rangle \rangle \rangle$

Although in principle every RP in a sentence has such a semantic representation, for practical reasons it is seldom feasible to include all of this information in the full semantic representation of the sentence, hence it is generally omitted. An example of a full structure is given in (67) for the simple intransitive question *Has the tall man been crying?* (from Van Valin, 2005: 53). The underlining of *man* in the semantic representation of the RP denotes that it is the head of the RP.<sup>22</sup>

(67)  $\langle_{\text{IF}} \text{INT} \langle_{\text{TNS}} \text{PRES} \langle_{\text{ASP}} \text{PERF PROG} \langle \text{do}' (x, [\text{cry}' (x)]) \rangle \rangle \rangle \rangle$   
 $\langle_{\text{DEF}} + \langle_{\text{QNT}} \exists \langle_{\text{NUM}} \text{SG} \langle_{\text{NASP}} \text{COUNT} \langle \text{be}' (\underline{\text{man}} (x), [\text{tall}']) \rangle \rangle \rangle \rangle \rangle$

#### 4. Questions for future research

Rauh (2010: 378-383) raises a number of important questions with respect to lexical and syntactic categories in RRG which need to be addressed and to which we now turn. These questions can be divided into two broad categories: The first are questions concerning the status of lexical and syntactic categories and how these relate to one another in RRG. For example, Rauh (2010: 380) raises a question with respect to examples (1b)-(1d) above, repeated here in (68).

- |         |  |                    |
|---------|--|--------------------|
| (68) b. | Chris [ <sub>NUC</sub> is [ <sub>PRED</sub> a very good detective]]. | PRED in NUC = NP   |
| c.      | Pat [ <sub>NUC</sub> is [ <sub>PRED</sub> exceedingly tall]].        | PRED in NUC = ADJP |
| d.      | Pat [ <sub>NUC</sub> is [ <sub>PRED</sub> in the house]].            | PRED in NUC = PP   |

As Rauh notes, *is* in these examples is not viewed as the predicate in RRG but rather the “NP”, “AdjP” and “PP”, respectively. As such, the question arises as to how *is* is introduced into the sentence: The usual answer would be to assume that English, like many languages, requires a “dummy verb” such as the copula when the predicate is not verbal. However, as Rauh rightly notes, if labels such as “noun” and “verb” are merely useful descriptive labels for certain language-specific feature-bundle patterns and not universally valid categories, this is not an option.

As has been argued in the preceding pages, in RRG there is no need to specify that predicates which are traditionally viewed as consisting of “NPs”, “AdjPs”, “PPs”, etc., require a “dummy verb” – in English or elsewhere. Recall from Section 2.2 above that it is the FEATURES which are relevant to the grammar, not these convenient categorial labels (Van Valin, 2008: 176). It would therefore be more correct to say that a copula is required with stative predication in English when the distributional properties of the content morpheme of the predicate are not compatible with TAM/person marking (*\*Pat (exceedingly) tall(s).*): No reference is necessary to categories such as “nouns”, “adjectives”, “verbs”, etc. Rather, the theory recognizes that individual content morphemes can have language-specific distributional properties, which can only be determined through tests and not through “logical” conjecture, based e.g. on whether they denote an action, object, etc. These language-specific restrictions include the (non-)compatibility of a particular content morpheme with TAM or case markers, determiners, etc. However, as we saw in Section 2.3.1 for Kharia, this is not equivalent to saying that these content morphemes actually ARE “nouns” or “verbs”. Also, in view of the problems with “category squishes” described in Section

<sup>22</sup> On the use of this notation, see Chapter XX (this volume) and Van Valin (2005: 52).

2.2, where only some of the supposedly defining features of a particular category are present, direct reference to specific features as opposed to (at best more-or-less predictable) feature bundles provides us with a much more exact description of the grammatical operations involved.

Another area mentioned by Rauh, qualia structure (Pustejovsky, 1995), is of a very different nature and raises more serious questions, which we however can only touch upon here. Qualia structure is one of four interpretive levels, the other three being argument structure, event structure and the interaction of semantic levels. This level consists of four basic roles (Pustejovsky, 1998:76):

- “CONSTITUTIVE: the relation between an object and its constituent parts;
- FORMAL: that which distinguishes the object within a larger domain;
- TELIC: its purpose and function;
- AGENTIVE: factors involved in the origin or “bringing it about”.”

Rauh refers to qualia structure on two occasions with respect to open questions in RRG, highlighting what is in her view the importance of lexical categories in constructing the logical structure of the sentence from the lexicon. For example, “Lexically, the NPs should be represented by items which are described on the basis of qualia structures in the lexicon, since such a description identifies them as Ns and indicates their compatibility with those peripheries and operators that are N-specific rather than CLAUSE-specific.” (Rauh, 2010: 380) Also, “To begin with, in order to construct the semantic representation of predicates and their arguments, a specification of nouns (items with a qualia structure) is required.” (Rauh, 2010: 382) As just noted, we believe that there is no need to refer to lexical categories in describing the compatibility which Rauh is referring to, if such restrictions are described in terms of individual features and not lexical categories.

The status of qualia structure, however, requires some discussion. To begin with, the presence of qualia structure is not restricted to “nouns”, as Rauh’s comments would seem to imply, since verbs also have qualia structure (cf. e.g., Section 6.2.5 in Pustejovsky, 1995). Nor is event structure restricted to “verbs”, since nominalizations or nouns such as *war* in English also have event structures (cf. e.g. Pustejovsky, 1998, §§8.4 and 8.5). Of course, the structure and type of information contained at these levels depend among other things on whether an action or an object is being described, but it is worth recalling here that saying that a content morpheme denotes an action is not equivalent to saying that it is a “verb” with a predictable set of distributional (and other) features, and to my knowledge the lexical structures given in Pustejovsky (1995) do not contain any direct reference to such lexical categories, although they do, of course, contain information relating to semantic features, such as “**physobj**”, “**process**”, “**state**”, etc.

This issue requires further study, and it is indeed likely that we will have to re-evaluate a number of assumptions with respect to semantic structure in order to be able to do away with lexical categories entirely. Furthermore, languages such as Kharia, discussed in Section 2.3.1, where virtually all content morphemes can be used predicatively, referentially and attributively will undoubtedly present us with new challenges with respect to structuring the lexicon. However, as lexical categories do not play a very prominent role in Pustejovsky’s (1995) formalism, this does not appear to be an insurmountable task, although Rauh (2010: 378-383) is certainly correct in noting that this will require assuming a much richer lexicon in RRG, containing considerably more language-specific information on individual morphemes than is currently the case. At any rate, languages such as Kharia with its almost total lack of

distributional restrictions on content morphemes, or Wari’ with its clausal predicates in the Internal State Constructions, discussed in Section 2.3.2, as well as “category squishes” even in languages where we have come to accept the presence of clear-cut lexical categories as self-evident, force us to accept that these highly intuitive lexical categories cannot play a role in a theory of language with claims to universality.

In sum, while much work remains to be done, it seems clear that traditional “categorical thinking” cannot serve as the basis for a universal theory of language and is in fact at the heart of many descriptive problems which have plagued linguistic theory for decades. The view taken here is that these problems can only be overcome by the wholesale rejection of such categories in language theory, granting them nothing more than the status of convenient labels in informal discussion.

#### Abbreviations

|                              |                                   |   |
|------------------------------|-----------------------------------|---|
| 1, 2, 3 – person markers     | FOC – restrictive focus           | PL – plural                             |
| ABL – ablative               | FUT – future                      | POL - politeness                        |
| ABLE – ability               | GEN – genitive                    | POSS – inalienable possession           |
| ABS – absolutive             | HON – honorific                   | PRESUM – presumptive evidential         |
| ACT – active voice           | HS - hearsay                      | PROX.HEARER – proximate to hearer       |
| ACC – accusative             | HUM – human                       | PROG – progressive                      |
| AdjP – adjective phrase      | INF – infinitive                  | PRS – present                           |
| AOR – aorist                 | INST – instrumental               | PSBL - possibility                      |
| CAUS – causative             | INT – interrogative               | PST – past                              |
| CL – classifier              | IPFV – imperfective               | Q – interrogative                       |
| C:TEL – culminatory telic V2 | IRR – irrealis                    | QUAL – qualitative predication (≈ ‘be’) |
| CLASS – classifier           | ISC – Internal State Construction | REAL – realis                           |
| COM - comitative             | LOC – locative                    | REM – remote                            |
| CONT – continuous V2         | M – masculine                     | REP – repetition                        |
| COP – copula                 | MID – middle voice                | SG – singular                           |
| DEC – declarative            | N – neuter                        | SH – subject honorific                  |
| DEF – definite               | NEG – negative morpheme           | TAM – tense, aspect, mood               |
| DEM – demonstrative          | NHUM – non-human                  | THEM – theme                            |
| DET – determiner             | N – neutral                       | TNS – tense                             |
| DU – dual                    | OBL – oblique case                | V2 – unit denoting aspect/aktionsart    |
| EMPH – emphatic              | OBLIG – obligation                | VIC – verbal inflectional clitic        |
| ERG – ergative               | OPT – optative                    | VOC – vocative                          |
| EXCL – exclusive             | PASS – passive                    |   |
| F – feminine                 | PERF – perfect                    |   |
| FAC – factive                | PFV – perfective                  |   |

#### References

- Ambros, Arne A. 1998. *Bongornu, kif int? Einführung in die maltesische Sprache*. Wiesbaden: Reichert Verlag.
- Bisang, Walter. 2001. “Finite vs. non finite languages. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals*, Vol. 2. Berlin: Mouton de Gruyter. 1400—1413.
- Borer, H. 2005. *Structuring sense. The normal course of events*. Oxford: Oxford University Press. (Oxford Linguistics).
- Broschart, Jürgen. 1997. “Why Tongan does it differently: Categorical distinctions in a language without nouns and verbs.” *Linguistic Typology* 1/2: 123-165
- Croft, William. 2000. “Parts of speech as language universals and as language-particular categories.” Petra M. Vogel & Bernard Comrie (eds.), *Approaches to the typology of word classes*. Berlin. (Empirical Approaches to Language Typology, 23). 65-102.

- Croft, William. 2005. "Word classes, parts of speech, and syntactic argumentation." *Linguistic Typology* 9/3: 431-441.
- Evans, Nicholas & Toshiki Osada. 2005. "Mundari: The myth of a language without word classes." *Linguistic Typology* 9/3: 351-390.
- Everett, Daniel L. 2008. "Wari' Intentional State Constructions." In Van Valin (ed.), 381-410.
- Håkansson, Gisela & Jennie Westander. 2013. *Communication in humans and other animals*. Amsterdam: Benjamins. (Advances in Interaction Studies, 4).
- Haspelmath, Martin. 2007. "Preestablished categories don't exist: Consequences for language description and typology." *Linguistic Typology* 11/1: 119-132.
- Haspelmath, Martin. 2011. "The indeterminacy of word segmentation and the nature of morphology and syntax." *Folia linguistica* 45/1: 31-80.
- Hengeveld, Kees & Jan Rijkhoff. 2005. "Mundari as a flexible language." *Linguistic Typology* 9/3: 406-431.
- Himmelmann, Nikolaus P. 2008. "Lexical categories and voice in Tagalog." In Peter K. Austin & Simon Musgrave (eds.), *Voice and grammatical relations in Austronesian languages*. Stanford, CA: CSLI Publications. 247-293. (Studies in Constraint-Based Lexicalism).
- Hockett, Charles F. 1960. "The Origin of Speech." *Scientific American* 203: 88-111. [Reprinted in Wang, William S.-Y. 1982. *Human communication: Language and its psychobiological bases*. *Scientific American*. 4-12].
- Kerkeṭṭa, K.P. 1990. *jujhair dāṛ (kharīyā nāṭak)*. Ranchi: Janjātīya Bhāṣā Akādemī, Bihār Sarkār.
- Klaiman, M.H. 1991. *Grammatical voice*. Cambridge: Cambridge University Press (Cambridge Studies in Linguistics, 59).
- Luuk, Erkki. 2010. "Nouns, verbs and flexibles: implications for typologies of word classes." *Language Sciences*, 32/3: 349-365.
- Maas, Utz. 2004. "Finite" and "nonfinite" from a typological perspective. *Linguistics* 42/2: 359-385.
- Malhotra, V. (1982). *The structure of Kharia: A study of linguistic typology and language change*. Unpublished Ph.D. dissertation. New Delhi: Jawaharlal Nehru University.
- Nikolaeva, Irina (ed.). 2007. *Finiteness. Theoretical and empirical foundations*. Oxford: Oxford University Press.
- Peterson, John. 2005. "There's a grain of truth in every "myth", or, Why the discussion of lexical classes in Mundari isn't quite over yet." *Linguistic Typology* 9/3: 391-405.
- Peterson, John. 2007. "Languages without nouns and verbs? An alternative to lexical classes in Kharia." In Colin Masica (ed.), *Grammatical and semantic perspectives on South Asian and other languages. Papers growing out of the Fifth International Conference on South Asian Linguistics (ICOSAL-5), held at Moscow, Russia in July 2003*. Delhi: Motilal Banarsidass. 274-303.
- Peterson, John. 2011a. *Kharia. A South Munda language*. Leiden / Boston: Brill. (Brill's Studies in the Languages of South and Southwest Asia, 1).
- Peterson, John. 2011b. "Aspects of Kharia grammar – A Role and Reference Grammar (RRG) approach." Rajendra Singh & Ghanshyam Sharma (eds.), *Annual Review of South Asian Languages and Linguistics*. Berlin: Mouton de Gruyter. 81-124.
- Peterson, John. 2013. "Parts of speech in Kharia: a formal account." Rijkhoff & van Lier (eds.), 131-168.
- Pustejovsky, James J. 1995. *The generative lexicon*. Cambridge, MA: MIT Press.
- Rauh, Gisa. 2010. *Syntactic categories. Their identification and description in linguistic theories*. Oxford: Oxford University Press. (Oxford Surveys in Syntax and Morphology, 7).

- Rijkhoff, Jan & Eva van Lier. 2013. *Flexible word classes. Typological studies of underspecified parts of speech*. Oxford: Oxford University Press.
- Ross, John Robert. 1972. "The Category Squish: Endstation Haputwort." *Papers from the Eighth Regional Meeting. Chicago Linguistic Society. April 14-16, 1972*. 316-328.
- Roy, Sarat Chandra & Ramesh Chandra Roy. 1937. *The Khārīās*. Ranchi: Man in India.
- Sasse, Hans-Jürgen. 1993. "Das Nomen – eine universale Kategorie?" *Sprachtypologie und Universalienforschung (STUF)* 46/3: 187-221.
- Tomasello, Michael. 2008. *Origins of human communication*. Cambridge, MA: MIT Press. (The Jean Nicod Lectures).
- Van Valin, Robert D., Jr. 2005. *Exploring the syntax-semantics interface*. Cambridge: Cambridge University Press.
- Van Valin, Robert D., Jr. 2008. "RPs and the nature of lexical and syntactic categories in Role and Reference Grammar." In Van Valin (ed.), 161-178.
- Van Valin, Robert D., Jr. (ed.). 2008. *Investigations of the syntax-semantics-pragmatics interface*. Amsterdam: Benjamins. (Studies in Language Companion Series, SLCS, 105).