*The grammatical description as a collection of form-meaning-pairs Sebastian Nordhoff

Max Planck Institute for Evolutionary Anthropology

Abstract

This paper analyzes the structure of books containing grammatical descriptions and builds up on work by Good (2004). It argues that the discussion of morphology, syntax, semantics, and intonation found in grammatical descriptions can be seen as a collection of interdependent form-meaning-pairs. These form-meaning-pairs form part of the larger structure of frontmatter, mainmatter and backmatter (Mosel 2006) and have themselves an internal structure which includes, among other things, linguistic examples as formalized by Bow et al. (2003).

1   Introduction

In this paper I will be concerned with the structure of a certain genre of texts, namely grammatical descriptions.[1] These texts have as an aim to store knowledge about the grammatical structure of a language, which may have a long literary tradition like French, or about which little may be known, as for Vedda, a small language in Sri Lanka. One thing which is important in the context of this paper is that I am dealing with *texts* and not with abstract entities like computational grammars which generate sentences (Maxwell this volume). Neither do I deal with the mental representations of grammatical knowledge. While I acknowledge the relevance of the latter two concepts, which form an interesting topic for formalization in themselves, in this paper I will concentrate on texts which are used to communicate grammatical information about a language from a knowledgeable person (the describer) to a person wishing to know more about the language (the reader). This is to say, I treat the "grammatical description as a communicative act" (Payne 2006).

2   Content-based and form-based structures

In an ideal world, all grammatical descriptions would conform to the same schema. Once this schema is established and applied to all grammars, the reader will be able to navigate a new description very easily. An approach pursuing the unification of the description of grammatical knowledge is for example the Crosslinguistic Reference Grammar project (Peterson 2002, Zaefferer 2006, Black & Black this volume). This approach presents a more or less elaborate apparatus for filling in grammatical information in predefined fields. The value of these approaches is heavily dependent on the quality of the underlying apparatus. In case the language in question does not exhibit a required phenomenon,[2] or it shows phenomena which were not known at the time when the apparatus was designed, the description cannot be implemented. Keeping

---

[1] I would like to thank Jeff Good, Nick Thieberger, and Michael Cysouw for comments on earlier versions of this paper.

[2] For a non-technical overview of how little we can assume about language structure, see Evans & Levinson (2009).

track with theoretical developments (Upward-compatibility) seems to be a reasonable expectation for a formalization of grammatical description,[3] but it is unclear how this can be done with a rigid formalization of 'what grammar is like' at the foundation of the apparatus. Furthermore, an apparatus based on the predefined categories necessarily relies on the cross-linguistic applicability of these categories, but it is still subject to debate whether it is even possible to formulate crosslinguistically valid categories at all (Haspelmath 2007). Moreover, language describers who do field work in remote places often have a strong personality with a disliking for being told how to describe 'their' language. The feeling that 'their' language is unique and cannot be pressed into a one-size-fits-all approach is widespread (Weber 2006). Even without the fundamental problems alluded to above, it is unclear whether a 'universal schema' is actually in line with the wishes of the describers. Finally, a universal schema is difficult to retrofit on already existing descriptions, which might be desirable if one wants to broaden the domain of semantic searches.

The 'universalist' or content-based schemas mentioned above have as an implicit aim to structure what grammar is like, with 'grammar' being used in its sense of 'mental representation'. Following Good (2004) and Nordhoff (2008), I take a slightly different approach in structuring what *grammatical descriptions* are like. This approach can be called 'form-based'. In the content-based approach, the elements of the structure are phonemes, words, suffixes etc, while the form-based approach has paragraphs, examples, and glosses as its elements. No claim is made about the universal applicability of the grammatical terms in the descriptions. Every author can describe the language how they see fit. However, the hypothesis is that grammar authors will make use of a number of typical textual elements like 'linguistic examples', 'paradigm tables', 'word-gloss-pairs' etc (Good 2004). Good (2004) surveys a sample of grammars and lists recurring structural elements, which are nested in a certain way. The highest structural element he recognizes is the 'annotation', which can contain 'exemplars', 'prose', 'references' and 'links to ontologies', and further annotations in a recursive fashion. He formalizes this nested structure in a DTD. The linguistic example itself is at a lower level in the structure. Theoretical discussions of its structure can be found in Drude (2002), formalizations are given in Peterson (2002) or Bow et al. (2003). In this paper, I want to concentrate on the higher elements in the structure, i.e. chapters and sections, while also refining Good's analysis of the 'annotation'. I use lower level elements occasionally where necessary, but often omit them to keep the presentation visually appealing and free of clutter. My basic approach has as an aim to be compatible with Bow et al. (2003) on the low level and Good (2004) on the mid-level.

3 The sample

I will exemplify my claims in this paper by a sample of descriptive grammars which includes all traditions and publication forms and covers different areas of the globe. The sample consists of

- Bloomfield (1962), a grammar of the North American language Menomini

---

[3]See Mosel (2006) for the necessity to update theoretical analyses.

- Seiler (1985), a grammar of the North American language Cahuilla *

- Li & Thompson (1981), a grammar of Mandarin Chinese

- Epps (2008), a grammar of the Amazonian language Hup

- Buechel (1939), a grammar of the North American language Lakota *

- Haspelmath (1993), a grammar of the Caucasian language Lezgian *

- Frohnmeyer (1889), a grammar of the Indian language Malayalam

- Newman (2000), a grammar of the Subsaharan language Hausa *

While all books were consulted, a general discussion will take up too much space, which is why the general findings are discussed based only on the starred items in the list above. In the remainder of this paper, I will follow Good (2004) in claiming that grammatical descriptions are semi-structured texts. When they are annotated for structure, semantic searches become possible, which is a useful resource for typologists. Figure 1 from Haspelmath (1993) illustrates a typical layout of a descriptive grammar.

In Figure 1, the discussion is made up of prose, which is found before and after examples in a particular format which are use to illustrate the topic at hand, distributive numerals in this case. As a first step in structuring the text, we can separate the examples from the prose which discusses them (cf. Good 2004). The internal structure of the examples can then be worked out, as done for instance in Bow et al. (2003). The prose part has received less attention overall, which is why I will focus on this aspect here, next to the whole overarching structure of the book.

4 The structure of grammatical descriptions: an overview

Taking a look at the table of contents of the books in the sample, we find a certain recurrent ordering in the topics discussed. I take these findings to be uncontroversial, so I give an abbreviated XML-notation right away.[4]

```
(1)  <book>
       <frontmatter>
         <tableofcontents/ >
         <listoftables/ >
         <listoffigures/ >
         <listofabbreviations/ >
         <acknowledgments> ...  </acknowledgments>
       </frontmatter>
       <mainmatter>
         <background> ... </background>
```

[4]The actual ordering of the chapters in the mainmatter may differ (Mosel 2006), but what it is important here is that they are discussed as part of the mainmatter; the actual internal ordering is less relevant, as will be discussed below.

13.1.7. Distributive numerals

Distributive numerals are formed by reduplication. The stress is on the first instance of the numeral. **prose**

(599)	sá-sa(d)	'one each'
	q'wé-q'we(d)	'two each'
	púd-pud	'three each'
	c'uwad-c'uwad	'fifteen each', etc.

examples

In complex numerals, only the last component is reduplicated (Gajdarov 1987:63). **prose**

(600)	wiš-ni ǯan-ni wad-wad	'125 each'
	q'ud wiš-ni c'urugud-c'urugud	'416 each'

examples

If the last component is wiš '100', aǧzur '1000', or million/milliard, the component that precedes it is reduplicated. **prose**

(601)	aǧzur-ni q'ud-q'ud wiš	'1400 each'
	ǯan-ni irid-irid million	'27 000 000 each'

examples

Examples for the use of distributive numerals: **prose**

(602) a. Ca-z   q'we=q'we ič   ła-na. (G54:155)
        we-DAT two=two   apple become-AOR

   'We received two apples each.' **examples**

   b. Fejzillah sa=sa xürünwi.di-n wil-er.i-z kilig-na. (HQ89:8)
        Fejzillah   one=one villager-GEN   eye-PL-DAT look-AOR

   'Fejzillah looked into the eyes of the villagers, one (villager) at a time.'

   c. Emirmet.a muhman-ar acuq'ar-na. Axpa sa=sada-waj
        Emirmet(ERC) guest-PL   make sit-AOR   then   one=one-ADEL

   žuzun-ar awu-na. (Q81:112)
   question-PL do-AOR

   'Emirmet made the guests sit down. Then he asked them questions, one (guest) at a time.'

Figure 1: The discussion of the distributive numerals in Haspelmath (1993)

4

```
        <phonology> ... </phonology>
        <morphology> ... </morphology>
        <syntax> ... </syntax>
        <semantics> ... </semantics>
    </mainmatter>
    <backmatter>
     <references/>
     <wordlist/>
     <texts>
        <text id="story1"> ... </text>
        <text id="recipe3"> ... </text>
     </texts>
    </backmatter>
    </book>
```

The nature, relevance and functions of front- and backmatter are similar to what we find in other kinds of scientific books (Mosel 2006) so that the need to discuss these parts in this paper is less urgent. I will focus on the parts of the mainmatter then.

5 The structure of the mainnmatter

5.1 Background

Grammatical descriptions typically contain a chapter on the sociohistorical background of the language (Lehmann 2002). In that chapter, the history and current sociological and political situation of the speech community is discussed. Topics covered are genealogical affiliation, geographical and political distribution, demographic factors such as ethnicity, religion, occupation and institutional representation. This part of the mainmatter does not seem to exhibit particular strong recurring structure and it does not seem wise to impose too tight a skeleton on it, so that I will treat it as unstructured data here.

5.2 Morphosyntax: form-meaning-pairs ordered according to form

Departing from the order normally found of books, I will now first discuss morphosyntax before coming back to phonology – which is normally the first thing to be discussed – in a minute. Depending on the language at hand, the division between morphology and syntax can be clear or rather subtle. While in Latin, the distinction is easy in most cases, other languages, like Tamil for instance, present challenges to the analyzer when they have to decide whether a given item is a suffix, an enclitic or an independent particle. There are ongoing theoretical discussions in particular languages about whether there is a division between morphology and syntax, and where it would be located (See Culicover & Jackendoff (2005) for an overview for the English facts, Lehmann (2002) for the consequences for language description). In light of these facts, it does not seem wise to impose a division in the schema until the differences are sorted out. What morphological and syntactical analyses have in common is that there is a certain

form X which is said to have a certain function F. Whether X is treated as belonging to the morphological domain or to the syntactic domain is not substantial here. As an example, we can take the English possessive marker *'s*. No matter the analysis, we can say that *'s* is used to encode possession, widely construed. This is to say we are dealing with a *form-meaning-pair*. The form *'s* of the English language is paired with the meaning *possession*. In this paper, I propose that most of the morphosyntax of a language can be treated as discussion of form-meaning-pairs (henceforth abbreviated as 'fomp').[5] Form-meaning-pairs consist of a topic of the discussion, which is at the same time the lemma (the name if you want) of the discussion. This topic is discussed with the help of illustrative examples[6] and surrounding prose.[7] We can illustrate this with the following fragment:

(2)
```
<fomp type="form-to-function" lemma="'s">
   <prose>
     The phrasal affix <form> 's" </form>  is used to
     code <meaning> possession </meaning>.
   </prose>
   <examples>
     <example> My friend's car </example>
   </examples>
    <prose>
      As we see in the example above, the affix
      attaches to the right edge of an NP, in this
      case <objectlanguage> My friend </objectlanguage>.
    </prose>
</fomp>
```

As a consequence of the bipartite nature of the form-meaning-pair, discussion of the sign can focus on the form part (signifiant) or on the meaning part (signifié) (Lehmann 2004b). Figure 2 from Seiler (1985) shows a neat separation between the discussion of formal properties and the discussion of functional uses. In the formal part on top, marked in red, allomorphs, a purely formal phenomenon, are discussed, while in the functional part on the bottom, marked in blue, the communicative situations where this morpheme can be used are explicated, in this case WISHES.

The division in a discussion of formal properties and functional properties can be squared with the alternation between prose and examples. Figure 3 shows such a more complex configuration.

---

[5]A 'fomp' is a subset of the 'Annotation' element proposed by Good (2004). 'Annotations' are broader and could be used for other domains as well. It is a topic for further research to determine the relations between the general superordinate 'annotation' type and the subset of form-meaning-pairs on the one hand and other types of description used in grammars (e.g. phonology) on the other hand.

[6]It seems sensible to have a container to contain a collection of individual examples illustrating aspects of the same phenomenon each. Good (2004) calls this container `<exSet>`. In line with the general use of full words in markup in this paper, I use `<examples>`, but the two terms can be subsituted for each other.

[7]The prose has an internal structure as well, consisting of running text interspersed with some special elements like references, *word* 'gloss'-pairs, technical terms and references. Specialized markup for these elements and links to ontologies enhance the computational usability of the grammatical description (Farrar & Langendoen 2003, Good 2004). For reasons of space and in order not to clutter the examples with tags, I

```
2.1.3.2.4.   'Possibility': -pulu

                                                          form

      The suffix shows two variants, one with initia…      prose
consonant /-pulu/, one with initial vowel /-alu/. Their
distribution follows the regularities formulated for the
VC/CV - suffixes in the Morphophonemics, I, 2.2.

      Meaning: The possibility of an event is underlined.  prose
One frequent contextual variant is 'wish':

(164)        hem-píi-pulu       'I wish they would go.'    example

      For further examples see Fuchs, p. 31.               prose

                                                          function
```

Figure 2: The discussion of the morpheme *pulu-* in Seiler (1985)

This structure of the text can be represented in semantic markup (3).

```
(3)   <fomp>
        <formaldescription>
          <prose>
            <form> XYX </form> has the following properties
          </prose>
          <examples>
            <example> ...</example>
          </examples>
        </formaldescription>
        <functionaldescription>
          <prose>
            <form> XYX </form> is used for <meaning> Function
            FGH </meaning>
          </prose>
          <examples>
            <example> ...</example>
          </examples>
        </functionaldescription>
      </fomp>
```

The kind of discussions we find in morphology have similarities to what we find in lexicography (Schultze-Berndt 1998, Mosel 2006, Weber 2006). First the forms are enumerated, then the possible meanings are given; additional information about domain, register or etymology may also be provided. In light of the similarity to the

---

omit the markup around the mentioned elements.

1. AGENT (ma-...-ī)

1.1. Form

Nouns of agent, which are comparable to words with the -er ending in English, have three forms depending on gender and number. (Many words formed according to this derivation function also as adjectives, see below, §1.7). All agent nouns use the same H-tone ma- prefix. In addition, masculine singulars add a suffix -ī)^{LH}, which results in an H-(L)-(L)-L-H tone pattern. Feminine singulars use the suffix -íyā)^{JHLH}. The suffix for plural agents is -ā)^{LH} with the same tone melody used with the masculine singulars. Examples:

	masculine	feminine	plural
quarrelsome psn	mafàɗàcī	mafàɗàcíyā	mafàɗàtā
parent	mahàifī	mahaifíyā	mahàifā
beggar, praise singer	marōƙī	marōƙíyā	marōƙā
coward	matsòràcī	matsòraciyā	matsòràtā

The plural formation as such is entirely regular. A few frozen/lexicalized agent nouns, however, employ other plurals. e.g., magājiyā / magájiyóyī 'a madam' f./pl.; magūdíyā / màsu gūɗā = 'yan gūɗā 'woman who ululates during festivities' f./pl.; macìjī / màcîzai 'snake' (lit. one who bites) m./pl.

°AN: Derivationally, macìjī is formed from the verb cīzā 'bite' even though in Hausa, snakes do not bite but rather slash, e.g., macìjī yā sārè ta 'A snake bit (lit. slashed) her'.

Monosyllabic verbs employ an epenthetic /y/ between the verb root and the suffixal ending, e.g.,

bi	follow	mabìyī / mabìyìyā / mabìyā	a follower of s.o. (m./f./pl.)
ƙi	dislike	maƙìyī / maƙìyìyā / maƙìyā	enemy (m./f./pl.)
shā	drink	mashàyī / mashàyìyā / mashàyā	drinker, alcoholic (m./f./pl.)

In the above feminine forms, the sequence /iyìyā/ is usually pronounced without the /i/ between the two /y/'s. The floating L tone attaches to the preceding syllable to produce a fall, i.e., mabìyìyā → [mabìyyā], maƙìyìyā → [maƙîyyā].

1.2. Verb stems with -TA

[...]

1.3. Meaning

The basic meaning of an agent noun is someone who customarily does the action of the underlying verb, commonly as a profession, e.g., maɗinkī 'tailor' (< ɗinkà 'to sew'). The semantic connection between the agent nouns and their source words is generally evident, e.g., ma'àskī barber < askè 'shave'. In some cases, however, these words have a lexicalized meaning that is more specialized and restricted than that of the related verb. Examples:

mabìyī	a follower (esp. religious); younger brother or sister < bì follow
macìyī	voracious < ci eat
mafàshī (usu. ɗan fashì)	robber < fasà break, shatter; commit robbery
makàɗàicī	unique (referring to God) < kàɗaità sit apart; acknowledge the unity of God
manìyyàcī	an intending pilgrim < *niyyata < nìyyā intention, wish
mariƙī	guardian, foster parent < riƙè grasp, hold
matàshī	adolescent, youth < tāshì rise, grow up

In a couple of special cases, the agent does not denote the doer of the action but rather the one affected by the action. The word ma'àikī (< àikā 'send') is used in the designation ma'àikin Allàh 'the Prophet Muhammad, i.e., the one who was sent by God', cf. Allàh ma'àikì 'God (lit. God the sender)'. The dictionaries also give the feminine agent word makulliyā with the meaning 'slave-concubine' (i.e., one who is locked up) < kullè 'lock'.

Figure 3: The discussion of the agentive morpheme *ma-* in Newman (2000) shows a division of formal and functional properties, and a further subdivision in prose parts and example parts.

lexicon, I will follow a suggestion by Lehmann & Maslova (2004) and call the space where these things are discussed *Morphemicon*.

5.3 Collections of fomps

In grammtical descriptions, as in other books, related phenomena are often grouped together. Sentences which cover related ideas are grouped into a paragraph, related paragraphs into sections, and related sections into chapters, with possibly some intervening levels (Good 2004). The conceptual unity of a discussion within a grammatical description is typically reflected in typography by white space. Tight coherence is mirrored by little space (e.g. between sentences), while more loose coherence is expressed by blank lines between paragraphs or blank pages between chapters. These organizational blocks thus reflect the semantic structure of the grammatical description. In a book, they are necessarily ordered linearly. However, when discussing the members of a set of morphemes, there is no inherent order. To take the French question words *qui* 'who', *quand* 'when', *quoi* 'what', *comment* 'how', *où* 'where', *pourquoi* 'why', no clear order of discussion suggests itself. The discussion of *qui* is pretty much independent of the discussion of *quand*, and both are independent of the discussion of *où*. The gist of the description does not change if you discuss *qui* before *quoi* or the other way round. The fact that in grammatical descriptions they are found in the sections numbered X.1, X.2, X.3 etc is simply a reflex of the requirement of the linear structure for printing. These numberings correctly indicate the subordination of these concepts to a higher complex of 'question words' (X in the example above), but they incorrectly suggest an inherent order among these items.[8] When creating the semantic markup for grammatical descriptions, we should not be fooled by incidental side effects of printing. However, the hierarchical structure of grammatical descriptions must be recognized. Some phenomena need to be discussed at an abstract level.

To take an analogy from classical zoologic taxonomy, in the family of *Felinae* we find the genera *Lynx, Leopardus, Puma,* and *Felis*, among others. There is surely no inherent order in discussing these genera, but some characteristics are shared among all members, e.g quadripedal, carnivore diet or moustaches. It would be redundant to state these facts at every individual level. They can better be discussed at the superordinate level of *genus proximum*. The same is true of linguistics. A semantic markup of grammatical description must provide for the possibility to state generalizations and sub/superordination. This is not a trivial problem. Here I would like to propose that this can be done by a general description followed by an enumeration of the members of the class with links to more detailed descriptions of the particular members. The XML-structure would be as follows:[9]

(4) `<fomp type="formlist" name="Question words">`
         `<overview>`

---

[8]Good (2004) concurs with the non-linear structure but remarks that the logical independence of these sections may be forfeited for a gain in didactic usefulness. To remain within the French example, the discussion of *quoi* should probably take place before *pourquoi* because the former is a component of the latter.

[9]For reasons of simplicity, I omit the representation of 'illustrative examples/paradigms' which are sometimes used in overview sections (Good 2004).

```
    <prose>
        Question normally start with the string "Wh".
        An exception is <form> how </form>. They are
        used to express <meaning> requests for
        information </meaning>. Question words normally
        trigger <form> do-support </form>.
    </prose>
</overview>
<ul>
  <li><form> Who </form></li>
  <li><form> What </form></li>
  <li><form> When </form></li>
  <li><form> Why </form></li>
  <li><form> Where </form></li>
  <li><form> How </form></li>
</ul>
</fomp>
```

This list structure can be recursive (Good 2004), so that deeper levels of subordination can be represented (free words>Nouns>Common Nouns>Count Nouns). Furthermore, multiple inheritance would also be possible.

As far as the treatment of formal (or semasiological) aspects is concerned, we thus have to distinguish two types of *fomps*: a kind of terminal node of the type "form-to-function" and a superordinate node of the type "formlist". The latter can include links to instances of the former.

The structure of the morphemicon can then be represented as in (5). Note that the linear order of the elements is a coincidence here. The morphemicon is an *un*ordered list, as discussed above.[10]

```
(5)  <morphemicon>
        <fomp type="formlist" name="Question words">
        ...
        </fomp>
        <fomp type="form-to-function" name="who">
        ...
        </fomp>
```

---

[10]As an illustration of the unordered nature of fomps, we can take Newman (2000), who lists them in alphabetical order, the following is an excerpt from the table of contents:

- Prepositions
- Pro-Verb *yi*
- Pronouns
- Questions
- Reason and Purpose
- Reduplication

```
<fomp type="form-to-function" name="what">
...
</fomp>
<fomp type="form-to-function" name="where">
...
</fomp>
<fomp type="form-to-function" name="why">
...
</fomp>
...
...
<fomp type="formlist" name="Demonstratives">
...
</fomp>
<fomp type="form-to-function" name="this">
...
</fomp>
<fomp type="form-to-function" name="that">
...
</fomp>
</morphemicon>
```

5.4 The treatment of examples: Bow, Hughes and Bird

Besides the structure of the higher level elements and descriptive prose, the linguistic example is obviously central to the discussion of semantic markup. This area has received a seizable amount of research (Drude 2002, Peterson 2002, Bow et al. 2003), which cannot be fully reviewed here. For the purposes of this paper, I adopt the XML-schema proposed by Bow et al. (2003), given for reference below.

```
(6) <interlinear-text>
      <item type="title"> The Title</item>
      <phrases>
        <phrase>
          <item type="gls"> A phrasal translation</item>
          <words>
            <word>
              <item type="txt"> Word</item>
              <morphemes>
                <morph>
                  <item type="txt"> Morph</item>
                  <item type="gls"> Gloss</item>
                </morph>
                <morph>
                  <item type="txt"> Morph</item>
                  <item type="gls"> Gloss</item>
```

```
            </morph>
          </morphemes>
        </word>
      </words>
    <phrase>
  </phrases>
</interlinear-text>
```

This 'raw' example can be further enhanced by information on meta-data (source, links to media files) and additional didactic annotations (constituency, highlighting of important aspects Good 2004).

5.5 Extending formal description: beyond the morpheme

In the paragraphs above I have discussed how morphemes can be linked to functions. However, morphemes are not the only meaning-bearing units in language. There are also constructions like *VERB the TIME away* e.g. *dance/waltz/chat the night/evening away* or more concrete *kick the bucket* (Culicover & Jackendoff 2005). The particular meaning of these constructions is more than what is present in their morphemic parts, so that we must assume some meaning stemming from the construction itself (Fillmore & Kay 1993, Goldberg 1995, Croft 2001). Another example is the difference between *John has come* and *Has John come?*. In this case, the relative order of auxiliary and subject indicates whether we are dealing with an assertion or a question. The meaning-bearing units of a language are thus not exclusively atomic, but they can be complex as well (Lehmann 1993). Furthermore, they are not always concrete as in the case of morphemes or idioms but they can also be schematic as in the case of the inversion questions. All this warrants the creation of a space where to discuss the meanings carried by these constructions. Following Goldberg (1995) I call this space *constructicon*.

The morphemicon deals with atomic and concrete elements, while the constructicon deals with schematic elements, which may be abstract or concrete. Both have in common that they deal with segmental material. Yet another complex bearing meaning in language is intonation, which is suprasegmental. The change of falling to rising intonation in the pair *Jim's mother has come.* vs. *Jim's mother has come?* has a predictable correspondence on the meaning side, the change of an assertion into a question. It seems best to treat intonation separated both from morphemes and constructions in a *contouricon* although there are of course some relations (WH-words trigger question intonation etc). The final structure of the morphosyntactic part is then

```
(7)  <forms>
      <morphemicon>
       <fomp type="morpheme" lemma="XX"> ... </fomp>
       <fomp type="morpheme"  lemma="YY"> ... </fomp>
       ...
      </morphemicon>
      <constructicon>
       <fomp type="construction" lemma="A B-C"> ... </fomp>
```

```
    <fomp type="construction" lemma="DE=F H G"> ... </fomp>
    ...
   </constructicon>
   <contouricon>
    <fomp type="intonation" lemma="HHL"> ... </fomp>
    <fomp type="intonation" lemma="HLH"> ... </fomp>
    ...
   </contouricon>
  </forms>
```

5.6 Phonology

The phonological part of grammatical description is normally structured as follows.

```
(8)  <phonology>
       <segments>
         <phonemechart/ >
         <vowels> ... </vowels>
         <consonants> ... </consonants>
       </segments>
       <phonotactics> ... </phonotactics>
       <stress> ... </stress>
       <!-- <intonation> ... </intonation> -->
```

As discussed above, I propose to treat intonation as something which does not merely distinguish meaning (like phonemes) but which carries a meaning of its own, more like morphemes (cf. Mosel 2006). Therefore, it can be meaningfully treated in the context of form-meaning-pairs, and there is no need to repeat the information in the phonological parts (although there should obviously be links between the two). This is why this element is commented out in (8).

The remaining content in the phonological domain belongs to the domain of 'distinguishing meaning'. This cannot be discussed in the context of form-meaning-pairs. The schmematization of this part will be left as a topic for future research.

5.7 Semantics: form-meaning-pairs ordered according to function

Above, I have discussed the structures we find in form-meaning-pairs based on morphemes and other forms. This approach is called the form-to-function or semasiological approach. Let's call this perspective form-based form-meaning-pairs, or *fo-fomps* for short. It is possible to take the converse approach, i.e. function-to-form or onomasiological (von der Gabelentz 1891, Jespersen 1924). This approach is the one which is generally relevant in typological work, although it is less prevalent in extant grammatical descriptions (Lehmann 1980, Comrie 1998, Lehmann 1998, 2004b, Schultze-Berndt 1998, Cristofaro 2006, Mosel 2006, Payne 2006, Zaefferer 2006), a notable example being (Willett 1991). Figure 4 shows the table of contents of Willet's description of Southeastern Tepehuan

Figure 4: Excerpt from the table of contents of Willett (1991). For reasons of space, only selected subsections are shown.

This is in many ways the mirror image of the former approach. Instead of taking a form and looking at the meanings it can express or the functions it can fulfill, we take a function and look at the forms it can be instantiated by. Let's call this perspective function-based form-meaning-pairs or *fu-fomp*. The division in prose parts and lists of examples is the same as above.

An illustrative example of the structure of a fu-fomp is given below in (9). Note the similarity to example (2) above.

```
(9)  <fomp type="function-to-form" lemma="Comparison">
       <prose>
         The <function> comparative degree </function> of
         adjectives can be expressed by the suffix <form>
         -er </form> or by the particle <form> more </more>
       </prose>
       <examples>
         <example> Mary is taller than John </example>
         <example> Mary is more intelligent than
         John </example>
       </examples>
       <prose>
         As we see in the example above, short adjectives
         form the comparative with <objectlanguage>
         -er </objectlanguage> while longer adjectives take
         the particle <objectlanguage> more </objectlanguage>.
       </prose>
     </fomp>
```

A more real-life example is given in Figure 5, which shows the discussion of the function of comparison of adjectives in Lakota. This function can be instantiated by three different constructions. After an initial overview of what this section is about, three different constructions which can be used to convey this meaning are discussed. These are a) an adverb meaning 'more', b) several other types of words with a rough meaning of 'surpass' and finally c) a contrastive juxtaposition of the type 'X is good and Y is bad'.

Given that the readers of grammatical descriptions are normally expected to have a basic knowledge of the world, the introductory portions of fu-fomps tend to be short. There is no need to belabour the intended meaning of the function called 'comparative degree' or 'temporal reference prior to speech act' as this is pretty much self-evident from the naming. In some more involved domains involving less familiar concepts like e.g. the paucal, the introduction can be longer and illustrate the functional domain at hand in more detail. This is especially true for semantic distinctions presumed unfamiliar to the average reader, like alienability, evidentials, or the paucal mentioned above.

In a function-to-form approach, additional contrastive information can be provided about shades of meaning, frequency or register. This information is less commonly found in form-to-function approaches. Figure 6 from Willett (1991) shows the description of a function and two formal instantiations thereof. Crucially, the two formal strategies come from different domains: the first one is a prefix (morphology) while the second one is a particle (syntax). In a form-to-function approach, the functional relatedness of these forms would have been difficult to convey. Furthermore, the two strategies can now be compared as to their felicity in different contexts, and their overall frequency. Without the functional *tertium comparationis* of 'Intention' spelled out, this information would have been very difficult to find in the grammar.

5.8   Extending functional description

I have shown above that different types of fo-fomps exist, namely those which belong to the morphemicon, the constructicon, and the contouricon. In what concerns fu-fomps, a similar division exists. We can distinguish meaning components which are purely semantic and relate to the communicated content. Examples are participants, events, space, and time as in *John ate a cake at the party at midnight*. These meaning components can be grouped in a *semanticon*. This purely semantic information is different from meaning components like topic and focus, which do not belong to the propositional content. An example would be the difference between *John came* and *It was John who came*. These sentences communicate the same semantic content and are truth-equivalent. Yet, there is a difference in information structure in that in the second sentence, the hearer is already expected to know that an act of coming occurred, which is not the case in the first sentence. These components of meaning which belong to information structure or discourse pragmatics can be discussed in a *discoursicon*. Speech acts are yet another type of function which is outside of both semantics proper and discourse. An example would be a request like *Please do come, John*. I will collect this interpersonal type of information in a *pragmaticon*. This division mirrors the layered structure of the clause as found in a number of contemporary grammatical theories like

COMPARISON OF ADJECTIVES

Introduction

While in English we have two ways of comparing, that is, by inflection, adding "er" and "est" to form the comparative and superlative, and by phrasal comparison, using the adverbs "more" and "most", in Lakota we have only the latter method.

#61 **The Comparative Degree**

instantiation

prose

a) The comparative is formed by placing the adverb "sanpa", more (which is usually shortened into sam or sanp), before the adjective, as in

examples

sam ksapa, more wise, wiser
sam sica, more bad, worse

prose

When, however, the object with which another is contrasted is mentioned, the inseparable preposition "i", with reference to, is prefixed to "sam". This composite adverb follows the noun or pronoun with which comparison is made, as in

examples

Hoksila kin atkuku kin isan hanska.
(boy the his father the more than tall)

The boy is taller than his father.

prose

When the pronoun is a personal pronoun, the abbreviated personal pronoun of the Third Class of verbs (cf #49) is used and prefixed to "isan", as in

examples

Hoksila kin misan hanska. The boy is taller than I.
Hoksila kin nisan hanska. The boy is taller than thou.
Hoksila kin unkisanpapi hanska. The boy is taller than we.
Hoksila kin wicisam hanska. The boy is taller than they.

prose

N.B. Note that in "unkisanpapi", the plural "pi" does not belong to "sanpa" but to the personal pronoun "unk".

instantiation

prose

b) The comparative is expressed quite often by employing other adverbs or verbs meaning that one thing surpasses or is above another, as in

examples

Wowase un kin he itancan kin he iwankabtu sni.
(servant the his lord the is above him not)

The servant is not greater than his lord.

Waniyetu yamni omakihe.
(winters three he follows me)

He is three years younger than I.

prose

This, however, is not a comparative in the true sense.

instantiation

prose

c) The comparative is expressed also by using two contrasting clauses, one with a positive, the other with a negative adjective or verb, as in

examples

Mastincala kin waste, tka sintehla kin sice.
(rabbit the is good but rattlesnake the is bad)

The rabbit is better than the rattlesnake.

Figure 5: The discussion of the function 'Comparison of Adjectives' in Lakota in Buechel (1939), with three forms instantiating this function.

9.31. Intention, deliberate action, and objective.

...

INTENTION is the agent's premeditated purpose to perform an action. In Southeastern Tepehuan, two types of intention are expressed, both of which are always in the first person.

instantiation

The less common form of intention is expressed by the use of the imperative prefix *xi-* with the future tense. This same prefix is used with the present tense for commands (§9.14) and with the past tense for deliberate action, as discussed below. The expression of intention is always in the first-person singular, as in (475) and (476).

(475) *Na-ñ   pai'dyuc   mu-jimi-a'   mu-ja'c   Juc-tɨr,   xi-ja-'ɨgui'ñ-dya-'-iñ*
SUB-1s   when   AWY-go-FUT   AWY-DIR   pine-in   INT-3p-tell-APL-FUT-1s
na   jax   chu-tu'm-cam   ya'   Mejic.
SUB   how   EXT-appear-like   here   Mexico
When I go (back) to Pine Grove, I'm going to tell (everyone) what it's like here in Mexico City.

(476) *Va-sava'ñi-añ-ich   dyo.   Xi-gámo-'iñ   ya'-ñi*
RLZ-buy-1s-PRF   PE   INT-put^in^pocket-FUT-1s   here-PRE
jiñ-'arpus   a'm.
1s-bag   into
I just bought it. I'm going to put it here in my bag.

⋮   ⋮   ⋮

instantiation

Another more common means of expressing intention is to preface the statement by one of two forms of the anticipation interjection (§13.3). In the imperative mode, this interjection is used with second person in the present tense to express immediate command, and with the future tense to indicate immediate request (§9.14). In the indicative mode, when the speaker himself intends to do something IMMEDIATELY following the time of speech, he normally prefaces the statement of his IMMEDIATE INTENTION with *ea* (or *eco-* plus subject enclitic). This statement is always in the future tense, and always in the first person, as it is with intention that is not immediate. But unlike nonimmediate intention, the expression of immediate intention does not require the use of the imperative prefix *xi-*, nor is it restricted to the singular of the first person. The examples given in (478)–(480) illustrate these differences.

(478) *Ea   na-ñ   cúpa-'   dyi   puerta.*
ANT   SUB-1s   close-FUT   ART   door
I'm going to close this door (now).

(479) *Eco-ñ   mo   mummu   xi-m-do'ñcho'   qui'ñgov.*
ANT-1s   DSC   there   INT-2s-drop   doorway
I'll show you out to the doorway.

(480) *Ea   na-ch   guio   bai'-p   vaidya-'.*
ANT   SUB-1p   again   TWD-RBP   invite-FUT
(It's agreed). Let's invite him back.

Figure 6: A function-to-form description of 'Intention' in Tepehuan (Willett 1991).

17

Role & Reference Grammar (Foley & Van Valin 1984, Van Valin & Lapolla 1997) or Functional Grammar (Hengeveld 1989, Hengeveld & Mackenzie 2008). We can add this structure to the general outline of grammatical descriptions (10).

```
(10)  <forms>
       <morphemicon>
        <fomp type="morpheme" lemma="ab"> ... </fomp>
        <fomp type="morpheme"  lemma="def"> ... </fomp>
       </morphemicon>
       <constructicon>
        <fomp type="construction" lemma="A B-C"> ... </fomp>
        <fomp type="construction" lemma="DE=F H G"> ... </fomp>
       </constructicon>
       <contouricon>
        <fomp type="intonation" lemma="HHL"> ... </fomp>
        <fomp type="intonation" lemma="HLH"> ... </fomp>
       </contouricon>
      </forms>
      <functions>
       <semanticon>
        <fomp type="semantics" lemma="space"> ... </fomp>
        <fomp type="semantics"  lemma="kin"> ... </fomp>
       </semanticon>
       <discoursicon>
        <fomp type="discourse" lemma="argument focus"> ... </fomp>
        <fomp type="discourse" lemma="new topic"> ... </fomp>
       </discoursicon>
       <pragmaticon>
        <fomp type="pragmatics" lemma="requests"> ... </fomp>
        <fomp type="pragmatics" lemma="insults"> ... </fomp>
       </contouricon>
      <pragmaticon>
```

A more extensive subdivision into 13 subcategories of meaning can be found in Lehmann (2004a).[11] These subcategories can be partitioned among the semanticon, discoursicon and pragmaticon, which will not be formalized here.

5.9 Collections of fu-fomps

Like fo-fomps in (4), fu-fomps can also be arranged hierarchically, for instance the hierarchy

(11) Expressing time > Expressing internal temporal structure > Expressing imperfective aspect.

---

[11] Apprehension and nomination, concept modification, quantification, reference, possession, space construction, predication, design of situations, temporal orientation, illocution and modality, contrasting, nexion, articulation of discourse

One can state general observations at the higher levels of the hierarchy ("Tense and aspect are nearly always expressed by prefixes") and more particular observations lower down in the hierarchy ("Progressive aspect can be expressed by *papu-* or by *pipo-*")

```
(12)   <fomp type="funclist" name="Expressing time">
         <overview>
             <prose>
                 Expressions of time cover lexical solutions
                 like <objectlanguage> aujourd'hui </objectlanguage>
                 <gloss> today </gloss> or <objectlanguage>
                 hier </objectlanguage> <gloss> yesterday </gloss>.
                 When time is expressed by bound morphemes, these
                 are normally <form> suffixes<form>. This is true
                 for both tense and aspect.
             </prose>
         </overview>
         <ul>
           <li><meaning> Expressing tense  </meaning></li>
           <li><meaning> Expressing aspect </meaning></li>
         </ul>
       </fomp>
```

What has been said above about collections of fo-fomps applies *mutatis mutandis* to collections of fu-fomps as well.

5.10   Interaction of fo-fomp and fu-fomp

Mosel (2006) notes that an ideal grammatical description could actually be required to state everything twice: once from a formal perspective and again from a functional perspective. A recent grammar which does precisely that is Nordhoff (2009). The table of contents for the formal and the functional part are given in figure 7.

The diligent reader will have noticed that the fo-fomp in example (2) and the fu-fomp in example (9) contain some additional markup. This markup can be used to link the form-to-function (semasiological) description with the function-to-form (onomasiological) description, a common desideratum for electronic grammars (Comrie 1998, Lehmann 1998, Zaefferer 1998b, Mosel 2006, Nordhoff 2008). I will illustrate this with a fragment of the grammar of French, namely question formation.

In French, there exist three principal ways to request information from the hearer: rising intonation contour, a question formative *est-ce que*,[12] and inversion. These three patterns are illustrated in (14). (13) gives the declarative sentence for comparison.

---

[12]For the ease of discussion, I essentially treat *est-ce que* as unanalyzable here. It is true that this introducer can be segmented into *est* 'is', *ce* 'this' and *que* 'that', but the construction has grammaticalized to such an extent that there is little awareness of the internal constituency. This can also be seen from the fact that an authortitative reference grammar of French (Grevisse & Goosse 1995) treats this construction as basically monomorphemic. The term 'introducer' ('introducteur') is also taken from this work.

Verbs
Nouns
Adjectives
Adverbs
Copula
Personal pronouns
Interrogative pronouns
Deictics
Quantifiers
Numerals
Interjections
Modal particles
Negative particles
Other particles
Conjunctions
Classifiers
Affixes
Simple clitics
Bound words
Nominal and verbal morphology
Verbal predicates
Existential predicate
Modal predicate
Nominal predicates
Circumstantial predicate
Adjectival predicate
Noun phrases
Postpositional phrases
Main clauses
Relative clause
Conjunctive participle clause
Purposive clauses
Subordinate interrogative clauses
Supraordination
The position of adjuncts
Reported speech
Agreement

Particpants
Participants of different entity orders
.   Participant roles
.   Unknown participants
.   Modifying participants
Predication
.   States
.   Events
.   Causation
Modification
Space
.   Figure-ground relations
.   Indicating spatial orientation
Time
.   Figure and ground
.   Phasal information
.   Aspectual structure
Quantification
Modality
Conditionals
Gradation
Comparison
Possession
.   Assertion of the possessee
.   Possession of abstract concepts
.   Assertion of the possessor
Negation
Kin
Referents and reference
Topic
Presupposition and assertion
Canceling implicatures
Parsing
Speech acts
Blending in the social tissue

Figure 7: The formal (left) and functional (right) table of contents of Nordhoff's grammar of Upcountry Sri Lanka Malay.

(13)   *Elle danse.*
       She dances
       'She dances/is dancing'

(14)   a. $\widetilde{\text{Elle}}$ $\overset{\diagup}{\text{danse}}$?
       b. Est-ce qu'elle danse?
       c. Danse-t-elle?
           'Does she dance/Is she dancing?'

We see that there is a many-to-one relation form form to function, and a one-to many relation from form to function. We can illustrate this as in Figure 8.
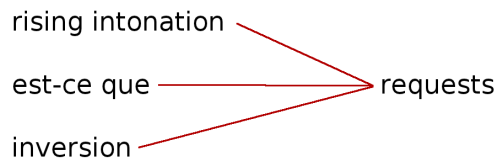

Figure 8: Many-to-one relations of form-meaning pairs

At closer scrutiny, we find that inversion is used for other functions as well, for example with coordination as in (15).

(15)   *..., aussi chante-t-elle bien.*
        also   sing-link-she well
       '..., and also does she sing.'

There is thus a many-to-many relation between form and function in language (Noonan 2006) as shown in Figure 9.
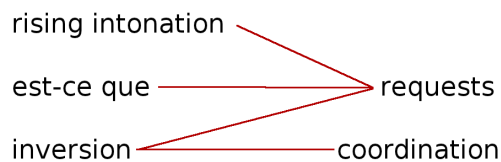

Figure 9: Many-to-one relations of form-meaning pairs

This relation can be expressed by a set of fo-fomps and fu-fomps as follows.[13]

---

[13]In order to not complicate the example further, I dispense with the difference between universal conceptual categories like 'question' and cross-linguistically common instantiations thereof, i.e. 'interrogative sentence' (but see Lehmann 1993). This distinction is important and should be reflected in markup used

```
(16)  <fomp type="morpheme" lemma="est-ce que">
        <formaldescription>
          <prose>
            The introducer <form> est-ce que </form>  with
            the literal meaning <gloss> Is it so
            that ... ? </gloss>. In front of a following
            vowel, the form is <objectlanguage> est-ce
            qu'</objectlanguage>. Both forms are shown in
            the following examples.
          </prose>
          <examples>
            <example> [example with est-ce que]</example>
            <example> [example with est-ce qu']</example>
          </examples>
        </formaldescription>
        <functionaldescription>
          <form> Est-ce que </form> is used for <meaning>
          questions </meaning>
        </functionaldescription>
      </fomp>

      <fomp type="contour" lemma="H%">
        <formaldescription>
          <prose>
            The rising contour has a high tone target on the
            last syllable.
          </prose>
          <examples>
            <example> [example with high tone target]</example>
          </examples>
        </formaldescription>
        <functionaldescription>
          This contour is used for <meaning> question
          formation </meaning>.
        </functionaldescription>
      </fomp>

      <fomp type="construction" lemma="Inversion">
        <formaldescription>
          <prose>
            In the <form> Inversion Construction</form>, the
            subject is repeated after the verb. Nominal subjects
            remain in front of the verb but pronominal subjects
```

---

for grammatical descriptions. However, in the context of this paper, it would incur too much theoretical overhead and would obscure the main line of argumentation. Future publications will explicate the relation in more detail than what can be covered here.

```
      are deleted.
    </prose>
    <examples>
      <example> Marie danse-t-elle?</example>
      <example> (*Elle) Danse-t-elle?</example>
    </examples>
  </formaldescription>
  <functionaldescription>
    <prose>
      The <form> Inversion Construction</form> is used
      for <meaning> question formation </meaning> and
      for <meaning> coordination </meaning>.
    </prose>
    <examples>
      <example> Chante-t-elle? </example>
      <example> Aussi chante-t-elle bien </example>
    </examples>
  </functionaldescription>
</fomp>

<fomp type="Speechacts" lemma="Requests">
  <overview>
    <prose>
      A <meaning> request </meaning> is used to elicit
      information from the addressee.
    </prose>
  </overview>
  <instantiations>
    <prose>
      Three strategies
      can be used to form requests. These are <form>
      rising intonation </form>, <form> preposing the
      introducer <objectlanguage> est-ce
      que </objectlanguage></form>, and <form>
      inversion</form>.
    </prose>
    <examples>
      <example>  Elle danse?</example>
      <example>  Est-ce qu'elle danse danse?</example>
      <example>  Danse-t-elle?  </example>
    </examples>
    <prose>
      The first example is the least formal, the middle
      one is quite neutral, while the third one is decidedly
      formal and pertains to the written language.
    </prose>
```

```
        </instantiations>
    </fomp>
```

This example models the many-to-many relations between form and function in a transparent way. The most relevant parts in the context of this discussion are form /form and meaning /meaning, which can be made to point to the page where the relevant formal or functional phenomenon is discussed in more detail. The reader might have noticed that the text between the tags varies and is not drawn from a restricted vocabulary. While there might be a possibility to avoid this arbitrariness in future descriptions, this is not possible when retrofitting the schema on extant descriptions. Therefore, the precise target of the form-links and the meaning-links has to be specified. So instead of

(17) ```<form> preposing the introducer <objectlanguage> est-ce que </objectlanguage></form>```

we should have something like

(18) ```<form target="est-ce que"> preposing the introducer <objectlanguage> est-ce que </objectlanguage></form>```

Note that the target `"est-ce que"` matches the lemma tag of `<fomp type="morpheme" lemma="est-ce que">`.

In the same vein, we can rewrite

(19) ```A <meaning> request </meaning> is used to elicit
      information from the addressee.```

as

(20) ```A <meaning target="Requests"> request </meaning> is used
      to elicit information from the addressee.```

These targets are ideally linked to an ontology to make the references clear and consistent and facilitate cross-linguistic searches (Farrar & Langendoen 2003). This will not be pursued here for reasons of space, but see Good (2004, this volume) for ideas how this can be done.

6 Interaction with the user

As Weber (2006) remarks, grammatical descriptions are never finished. New insights are continuously gained.[14] When a grammatical description is made available electronically, the findings can be updated. Nordhoff (2008) discusses the advantages of and requirements for electronic grammar writing. An aspect not discussed in Nordhoff (2008) is the possibility for users to add tags to pages of an electronic description. These

---

[14]See Comrie (1998), Cristofaro (2006), Mosel (2006), Payne (2006), Rice (2006) and Zaefferer (2006) for similar observations.

tags can be either arbitrary like `Compound`, `Important`, `SimilarToWarlpiri`, `Grammaticalization` or `V-Movement`. This kind of tag would have to be distinguished from a set of tags drawn from a closed restricted vocabulary. One possibility would be to rely on established schemas like the LDS questionnaire, so that tags like `LDS_2.3.4` would have a clear and defined meaning. Another obvious provider for a restricted and controlled set of vocabulary would be the GOLD ontology (Farrar & Langendoen 2003). If grammatical descriptions manage to draw a critical mass of tagging users, tag clouds can give a quick overview of the aspects of a certain page which the majority of the users find particularly relevant (cf Bouda & Cysouw this vol.).

7   Schematization

I have discussed the overall structure of a grammatical description above, including frontmatter, mainmatter and backmatter. The mainmatter was analyzed as consisting of a background part, a part for segmental phonology and two interdependent collections of form-meaning-pairs ('fomps'). The first one is based on the form-to-function or semasiological approach to grammatical analysis, while the second takes the converse onomasiological approach, function-to-form. The form-based and function-based fomps show similar structure. Both consist of alternating parts of prose and examples (Figure 10). These findings can be described in the RelaxNG schema given below (parts irrelevant in the context of this paper are treated as unstructured "texts" to keep the size of the schema within bounds).
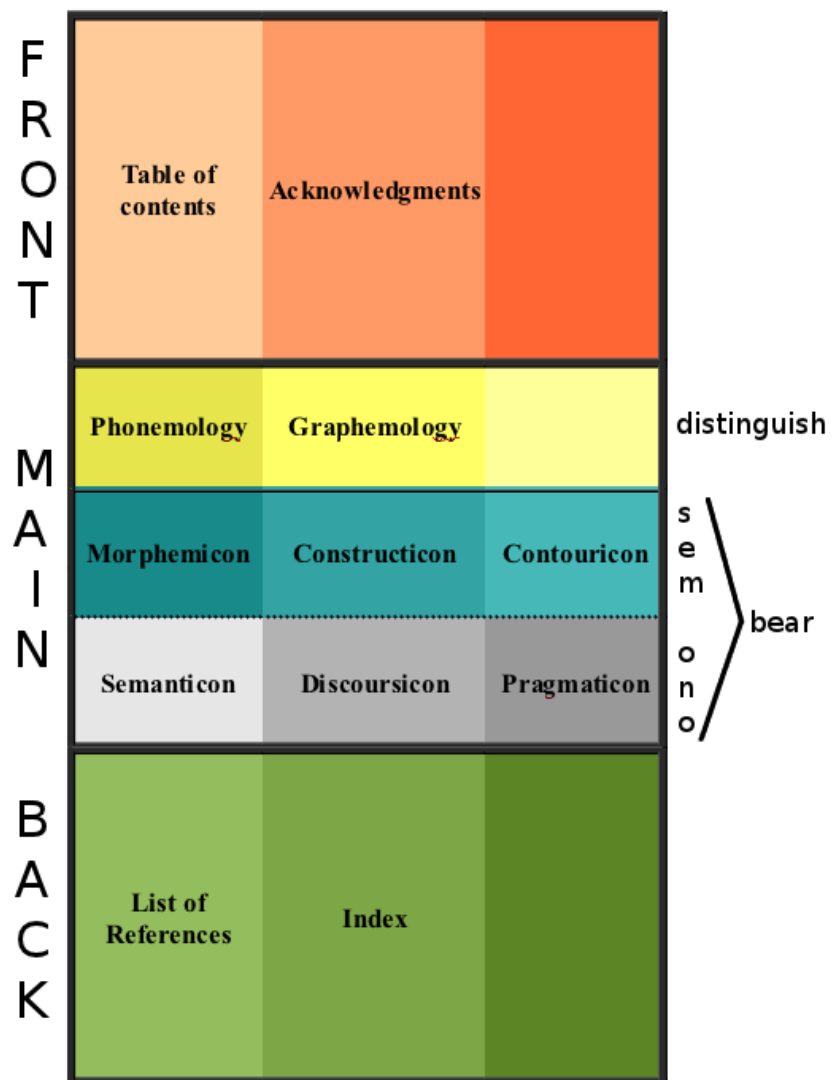
Figure 10: Visual illustration of the schema of grammatical descriptions.

```
(21)  GD =  element gd { Frontmatter,Mainmatter,Backmatter }
      Frontmatter = element frontmatter { TOC, LOF, LOT, LOA, Acknowledgments }
      Backmatter = element backmatter { References,Index }
      Mainmatter = element mainmatter { Phonemology, Semasiology, Onomasiology }

      Phonemology = element phonemology { Phonemicon }
      Semasiology = element semasiology { Contouricon, Morphemicon, Constructicon }
      Onomasiology = element onomasiology { Semanticon, Discoursicon, Pragmaticon }

      Phonemicon = element phonemicon  { text }
      Contouricon = element contouricon  { Fo-Part }
      Morphemicon = element morphemicon  { Fo-Part }
      Constructicon = element constructicon  { Fo-Part }

      Semanticon = element semanticon  { Fu-Part }
      Discoursicon = element discoursicon { Fu-Part }
      Pragmaticon = element pragmaticon { Fu-Part }

      Fo-Part = element fo-collection { (Fo-Collection|Fo-Fomp)* }
      Fo-Collection = element fo-list { Tags,  Prose, Examples, Formlinklist }

      Fu-Part = element fu-collection { (Fu-Collection|Fu-Fomp)* }
      Fu-Collection = element fu-list { Tags,  Prose, Examples, Funclinklist }

      Examples = element examples { Example+ }

      Fo-Fomp = element fo-fomp { Tags,  Overview, Formaldescription, Functionaldescription }
      Formaldescription = element formaldescription { (Prose|Example)* }
      Functionaldescription = element functionaldescription { (Prose|Example)* }

      Fu-Fomp = element fu-fomp { Tags,  Overview, Instantiations }
      Instantiations = element instantiations { (Prose|Example)* }

      Overview = element overview { text }
      Prose = element prose { text }

      Example = element example { Tags,  Bowhughesbird }
      Bowhughesbird = element bowhughesbird { text }

      Formlinklist = element formlinklist  { Formlink+ }
      Funclinklist = element funclinklist  { Funclink+ }

      Formlink = Link
      Funclink = Link

      Link = element link { attribute name { text }, attribute target { text } }
      Tags = element tag { attribute name { text } }*

      TOT = element tableofcontents { text }
      LOF = element listoffigures { text }
      LOT = element listoftables { text }
      LOA = element listofabbreviations { text }
      Acknowledgments = element acknowledgments { text }
      References = element references { text }
      Index = element index { text }
```

8   Conclusion and outlook

This paper has analyzed the semantic structure of grammatical descriptions and shown that in the domain of form-meaning pairs, the interaction between the semasiological and the onomasiological approach can be formalized in a RelaxNG schema. Grammars structured along this schema have a number of advantages. First, the schema encourages encapsulation of the descriptive content. The descriptive content in each fomp should be independent of the surrounding fomps. If the schema is adhered to, the constraint of linearity disappears. The elements are self-contained, which allows for addition and modification of elements without affecting the overall structure (terminological consistency remains an issue of course). This means that grammars can be written and published in an incremental heap-like way, making new insights available to the general public as they are gained (cf. Weber 2006, Good this volume). Furthermore, the basic advantages of structured text obtain, e.g. semantic searches, extraction, modification, differential presentation (Maxwell this volume).

The schema proposed here is designed to be compatible with recent structuring proposals in other domains of grammar, namely Bow et al. (2003) and Good (2004). Further work in analyzing the structure of grammatical descriptions needs to be done. Issues for further theoretical work are: the structure of phonological descriptions, the nature of tags and links, and the implementation of a controlled vocabulary for certain fields through an ontology. As far as practical applications are concerned, the schema will have to be measured against the actual requirements of future and past grammars. Is it possible to use this schema when writing a grammar, and is it possible to retrofit this schema on an existing grammar? As for the former question, first results are positive. Nordhoff (2009) is a descriptive grammar of a previously undescribed language, Sri Lanka Malay. While this grammar is not in XML-format yet, it was designed with the application of the above schema in mind. As such it contains a formal part and a functional part, which are roughly structured as outlined above. Furthermore, the individual sections in the two parts are parallel to fo-fomps and fu-fomps. The conversion process of the manuscript to XML is currently under way and looks promising. Retrofitting the schema on legacy descriptions is a more difficult task. The book has to be split into independent fomps. Depending on whether the author adhered to a strict separation of formal and functional discussion (e.g. Seiler 1985), the task is more or less easy. Resolving interparagraph dependencies (*as demonstrated in the last paragraph, as shown below, contrary to what was said in the preceding section* etc) will probably be a problematic issue in splitting the grammatical description into independent chunks on which the schema can be applied. In a first step, retrofitting will be done manually, but in a second step, semi-automatic analysis of the structure of grammatical descriptions remains a goal. Extraction tools like Lewis (2006) are probably a worthwhile domain to investigate for these prospects.

The ultimate goal would be to have an online repository of all existing grammatical descriptions which are converted to XML. These could be queried in a semantic fashion. The query would yield all the descriptive content of the selected grammars about a particular domain, a very useful feature for large sample typology. While there is still a very long way to go, the GALOES platform (Nordhoff 2007b,c,a) is aimed at supporting language describers in writing XML-based grammars. Descriptive departments in

several European countries have expressed interest in collaboration. In the long run, this should assure that future grammatical descriptions comply with the schema. As for legacy descriptions, the next step will be an analysis of the 10,000 electronic grammars collected by Harald Hammarstöm to see how far automatic extraction procedures can get us. This topic will be treated in future papers.

References

Ameka, Felix K., Alan Dench & Nicholas Evans (eds.). 2006. *Catching language – The Standing Challenge of Grammar Writing*. Berlin, New York: Mouton de Gruyter.

Black, Cheryl A. & H. Andrew Black. this volume. Grammars for the people, by the people, made easier using PAWS and XLingPaper. In Sebastian Nordhoff (ed.), *Electronic grammaticography*, 103–28. Manoa: University of Hawai'i Press.

Bloomfield, Leonard. 1962. *The Menomini Language*. New Haven, London: Yale University Press.

Bouda, Peter & Michael Cysouw. this vol. Treating Dictionaries as a Linked-Data Corpus. In Christian Chiarcos, Sebastian Nordhoff & Sebastian Hellmann (eds.), *Linked data in linguistics. representing language data and metadata*, 15–23. Heidelberg: Springer.

Bow, Cathy, Baden Hughes & Steven Bird. 2003. Towards a general model for interlinear text. *Proceedings of the EMELD Language Digitization Project Conference* http://www.linguistlist.org/emeld/workshop/2003/bowbadenBird-paper.pdf.

Buechel, Eugene. 1939. *A Grammar of Lakota*. Rosebud: Rosebud Educational Society.

Comrie, Bernard. 1998. Ein Strukturrahmen für deskriptive Grammatiken: Allgemeine Bemerkungen. In Zaefferer (1998a) 7–16.

Cristofaro, Sonia. 2006. The organization of reference grammars: a typologist user's point of view. In Ameka et al. (2006) 137–170.

Croft, William. 2001. *Radical Construction Grammar*. Oxford: OUP.

Culicover, Peter & Ray Jackendoff. 2005. *Simpler Syntax*. Oxford: OUP.

Drude, Sebastian. 2002. Advanced glossing – A language documentation format and its implementation with Shoebox. In Peter Austin, Helen Dry & Peter Wittenburg (eds.), *Proceedings of the international lrec workshop on resources and tools in field linguistics*, .

Epps, Patience. 2008. *A Grammar of Hup*. Berlin, New York: Mouton de Gruyter.

Evans, Nick & Stephen Levinson. 2009. The myth of language universals. *Behavioral and Brain Sciences* 32. 429–492.

Farrar, Scott & Terry Langendoen. 2003. A linguistic ontology for the semantic web. *GLOT International* 7. 200–203.

Fillmore, Charles J. & Paul Kay. 1993. *Construction grammar coursebook : chapters 1 thru 11*. Berkeley: University of California.

Foley, William A. & Robert D. Van Valin. 1984. *Functional syntax and universal grammar*. Cambridge: CUP.

Frohnmeyer, L. J. 1889. *A progressive grammar of the Malayalam language*.

Goldberg, Adele E. 1995. *Constructions : a construction grammar approach to argument structure*. Chicago: The University of Chicago Press.

Good, Jeff. 2004. The descriptive grammar as a (meta)database. Paper presented at the EMELD Language Digitization Project Conference 2004. http://linguistlist.org/emeld/ workshop/2004/jcgood-paper.html.

Good, Jeff. this volume. Deconstructing descriptive grammars. In Sebastian Nordhoff (ed.), *Electronic grammaticography*, 2–32. Manoa: University of Hawai'i Press.

Grevisse, Maurice & André Goosse. 1995. *Nouvelle Grammaire Franҫise*. Brussels: De Boeck & Larcier 3rd edn.

Haspelmath, Martin. 1993. *A Grammar of Lezgian*. Berlin, New York: Mouton de Gruyter.

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

Hengeveld, Kees. 1989. Layers and operators in Functional Grammar. *Journal of Linguistics* 25(1). 127–157.

Hengeveld, Kees & Lachlan Mackenzie. 2008. *Functional Discourse Grammar*. Oxford: Oxford University Press.

Jespersen, Otto. 1924. *The Philosophy of Grammar*. London: Allen & Unwin.

Lehmann, Christian. 1980. Aufbau einer Grammatik zwischen Sprachtypologie und Universalienforschung. In Hansjakob Seiler, Gunter Brettschneider & Christian Lehmann (eds.), *Wege zur universalienforschung*, 29–37. Tübingen: Narr.

Lehmann, Christian. 1993. *On the system of semasiological grammar*, vol. 1 Allgemein-Vergleichende Grammatik. Bielefeld: Universität Bielefeld, Universität München.

Lehmann, Christian. 1998. Ein Strukturrahmen für deskriptive Grammatiken. In Zaefferer (1998a) 39–52.

Lehmann, Christian. 2002. Structure of a comprehensive presentation of a language. In Tasaku Tsunoda (ed.), *Basic materials in minority languages*, 5–33. Osaka: Osaka Gakuin University.

Lehmann, Christian. 2004a. Documentation of grammar. In Osamu Sakiyama, Fubito Endo, Honore Watanabe & Fumiko Sasama (eds.), *Lectures on endangered languages: 4. from kyoto conference 2001*, 61–74. Osaka: Osaka Gakuin University.

Lehmann, Christian. 2004b. Funktionale Grammatikographie. In Waldfried Premper (ed.), *Dimensionen und kontinua. beiträge zu hansjakob seilers universalienforschung.*, 147–165. Bochum: N. Brockmeyer.

Lehmann, Christian & Elena Maslova. 2004. Grammaticography. In Geert Booij, Christian Lehmann, Joachim Mugdan & Stavros Skopeteas (eds.), *Morphologie. ein handbuch zur flexion und wortbildung*, vol. 2, Berlin, New York: de Gruyter.

Lewis, William D. 2006. ODIN: A Modle for Adapting and Enriching Legacy Infrastructure. Paper presented at the e-Humanities workshop at e-Science 2006.

Li, Charles N. & Sandra A. Thompson. 1981. *Mandarin Chinese – A functional reference grammar*. Berkeley: University of California Press.

Maxwell, Mike. this volume. Electronic Grammars and Reproducible Research. In Sebastian Nordhoff (ed.), *Electronic grammaticography*, 207–234. Manoa: University of Hawai'i Press.

Mosel, Urlike. 2006. Grammaticography: The art and craft of writing grammars. In Ameka et al. (2006).

Newman, Paul. 2000. *The Hausa Language – An encyclopedic reference grammar*.

New Haven, London: Yale University Press.

Noonan, Michael. 2006. Grammar writing for a grammar-reading audience. *Studies in Language* 30(2). 351–365.

Nordhoff, Sebastian. 2007a. The grammar authoring system GALOES. Paper presented at the workshop "Wikifying research" at the MPI Leipzig.

Nordhoff, Sebastian. 2007b. Grammar writing in the Electronic Age. Paper presented at the ALT VII conference in Paris.

Nordhoff, Sebastian. 2007c. Growing a grammar with GALOES. Paper presented at the Dobes workshop at the MPI Nijmegen.

Nordhoff, Sebastian. 2008. Electronic reference grammars for typology – challenges and solutions. *Journal for Language Documentation & Conservation* 2(2). 296–324.

Nordhoff, Sebastian. 2009. *A Grammar of Upcountry Sri Lanka Malay*: University of Amsterdam dissertation.

Payne, Thomas. 2006. A grammar as a communicative act or What does a grammatical description really describe? *Studies in Language* 30(2). 367–383.

Peterson, John. 2002. *Cross-Linguistic Reference Grammar (Final report)*. München: Centrum für Informations- und Sprachverarbeitung.

Rice, Keren. 2006. A typology of good grammars. *Studies in Language* 30(2). 385–415.

Schultze-Berndt, Eva. 1998. Zur Interaktion von semasiologischer und onomasiologischer Grammatik: Der Verbkomplex im Jaminjung. In Zaefferer (1998a) 149–176.

Seiler, Walter. 1985. *Imonda, a Papuan language*. Canberra: Department of Linguistics.

Van Valin, Robert D. & Randy Lapolla. 1997. *Syntax – Structure, meaning and function*. Cambridge: Cambridge University Press.

von der Gabelentz, Georg. 1891. Die Sprachwissenschaft. Ihre Aufgaben, Methoden und bisherigen Ergebnisse, Leipzig.

Weber, David. 2006. Thoughts on growing a grammar. *Studies in Language* 30(2). 417–444.

Willett, Thomas Leslie. 1991. *Southeastern Tepehuan*. Dallas: SIL.

Zaefferer, Dietmar (ed.). 1998a. *Deskriptive Grammatik und allgemeiner Sprachvergleich*. Tübingen: Niemeyer.

Zaefferer, Dietmar. 1998b. Ein Strukturrahmen für deskriptive Grammatiken: Die Beschreibung sprachlicher Funktionen. In Zaefferer (1998a) 29–38.

Zaefferer, Dietmar. 2006. Realizing Humboldt's dream: Cross-linguistic grammaticography. In Ameka et al. (2006) 113–136.