

## CLIPP

**Christiani Lehmanni inedita, publicanda, publicata**

### titulus

Language description and general comparative  
grammar

### huius textus situs retis mundialis

[http://www.christianlehmann.eu/publ/lg\\_descr.pdf](http://www.christianlehmann.eu/publ/lg_descr.pdf)

### dies manuscripti postremum modificati

14.12.2012

### occasio orationis habitae

11<sup>th</sup> International Congress of Linguists, 13.08.1987, Berlin

### volumen publicationem continens

Graustein, Gottfried & Leitner, Gerhard (eds.), Reference  
grammars and modern linguistic theory. Tübingen:  
Niemeyer (Linguistische Arbeiten, 226)

### annus publicationis

1989

### paginae

133-162

# **Language description and general comparative grammar**

**Christian Lehmann**

## **Abstract**

The following postulates are formulated with respect to a scientific description of a language:

1. It is usable for whatever linguistic purpose.
2. It comprises an account of the linguistic system, a lexicon, a text corpus and a statement of the historical situation of the language.
3. The description of the language system accounts not only for core, but also for peripheral subsystems.
4. The linguistic system and the lexicon are presented both in a synthetic and in an analytic form.
5. The description brings out the dynamic character of the language.

These postulates can be complied with if the description of the language system instantiates a general comparative grammar. This in itself obeys the postulates. Specific proposals for the implementation of a general comparative grammar, esp. with respect to postulate 4, are made.

## 1. Introduction

The form of a model of the syntactic structure of a language has been the subject of extensive discussion in the confines of certain theories of grammar. The question, however, of how a **comprehensive description of a language** is to be devised has, to my knowledge, received surprisingly little attention from linguists, either in formal terms or at a general level.<sup>1</sup> This is the basic problem of the **theory of linguistic description** (whose central component is the theory of grammar).

However, the question is not just a theoretical one. Answers to it have to be inspired empirically by actual language descriptions. In fact, the general problem of a language description is at least as much one of **general comparative linguistics** as it is one of the theory of linguistic description. In what follows, the position will be advocated that the elaboration of a model of linguistic description is bound up with the elaboration of a general comparative grammar.

The presentation is organized as follows. In §2, a set of requirements to be met by any language description is derived from a definition of 'language description' and a thesis on its composition. These are put forward in the form of postulates (numbered consecutively with a capital P). In §3, the scope is broadened to include further postulates and theses (with a capital T) which connected with the requirement that a language description should be comparable with those of other languages and therefore be based on a general comparative grammar. This is discussed partly on the basis of a critical evaluation of the *Lingua Descriptive Studies Questionnaire*. In §4, the organization of the general comparative and the language-specific grammar according to the synthetic and analytic viewpoints is explained.

## 2. Demands on a language description

In §2, I will formulate and argue for a set of postulates and theses which circumscribe the contents and form of a language description.

### 2.1. Aim

Definition: A **language description** is an encyclopedic description of a language the contents and form of which are defined with respect to linguistics as a science.<sup>2</sup>

The linguistic description here envisaged scientifically describes one natural language, the

---

<sup>1</sup> Among the exceptions that I am aware of are Gabelentz 1901, *Zweites Buch*, Seiler 1969, Lehmann 1980 and Mosel 1987.

<sup>2</sup> *Science* is, of course, to be taken in the sense of 'Wissenschaft', not in the sense of 'Naturwissenschaft'.

**object language**, in terms of another natural language, the **background language** (also called metalanguage).<sup>3</sup> It is one kind of language description.

Other kinds are, e.g., a pedagogic description or a general purpose description for an encyclopedia. Disregarding these for the rest of the discussion, I will follow the above definition and henceforth call the linguistic description of a language simply a language description. It must be usable by any linguist for whatever his specific interests may be. This implies, among other things,

- a. that no knowledge of the language or any aspect of it (e.g. its writing system) is presupposed;
- b. that the general state of the art in linguistics is presupposed, so that a layman is not expected to be able to use the description;
- c. that the description will not be framed in terms of some formal model, but in terms intelligible to any linguist, and that practical considerations of usability shape it to some extent;
- d. that didactic considerations play a minor role in the presentation, since the description must be usable as a manual.<sup>4</sup>

## 2.2. Composition

P1. A language description is a comprehensive presentation of a language under all its aspects.

The term 'comprehensive presentation' is intended to include not only a **scientific analysis** or its result, a model, but also a **documentation** of representative specimens of its subject matter. This leads to the following thesis:

T1. A language description consists of four parts:

1. description of the language system,
2. lexicon,
3. text corpus,
4. description of the historical situation of the language.

Let us now elaborate on these four parts in turn.

---

<sup>3</sup> I call it background language rather than metalanguage because it is the language in terms of which we understand the object language. It is also used in the dictionary (cf. §2.2 ad 2), where it would not normally be said to have the status of a metalanguage in the scientific sense.

<sup>4</sup> Most grammars do indeed present information by stepwise building up complex structures from simpler ones. The *Lingua Descriptive Studies Questionnaire*, to be discussed in §3.2, and the grammars based on it represent the most blatant and remarkable violation of this tendency, by starting with sentence types and subordination.

- Ad 1. The description of the **language system** comprises
- a. the phonology with its interfaces to phonetics and orthography,
  - b. the grammar *stricto sensu*, i.e. morphology and syntax,
  - c. the semantics with its interfaces to pragmatics and stylistics.<sup>5</sup>

The interfaces are described to the degree that they are systematic.

Given that, according to the definition in §2.1, a language description is constituted inside the science of linguistics, signs of the language should be given in phonological representation. Here, however, the practical considerations also mentioned there come into play. Provided that the orthographic representation preserves the identity of the sign, it is to be preferred. On the other hand, if the orthography is Non-Roman, a transliteration is the best solution for an alphabetic script, a phonological representation, for a non-alphabetic script.

Ad 2. Again given the definition in §2.1, the **lexicon** should take the form of a lexicological description. However, on the one hand, the lexicon by definition contains what is non-systematic about the signs of the language; insofar the list form is adequate to it. And on the other, for ease of reference, the lexicological description would have to be supplemented by an index, anyway. Practical considerations therefore speak in favor of representing the lexicon in the form of a "bilingual" dictionary, which associates the object language with the background language. However, such a bilingual arrangement also has some properly linguistic motivation, for which see §3.3.

Each entry contains information of the following kinds (cf. the lists of T1 and ad 1 above):

- a. phonological, incl. phonetic and orthographic,
- b. grammatical (i.e. morphological and syntactic),
- c. semantic, incl. pragmatic and stylistic,
- d. historical, incl. all the rubrics mentioned ad 4.
- e. examples, with reference to the text collection.

Again mainly for practical reasons, the lemma is given in orthographic representation, which entails that phonological and phonetic information (item a) will be specified only to the extent that it is idiosyncratic with respect to the orthography. Analogous considerations apply to the pragmatic and stylistic information.

Ad 3. There are various reasons why a language description should include a **text corpus**.

---

<sup>5</sup> Pragmatics is here taken to include all kinds of non-linguistic knowledge which determines the use and interpretation of utterances, in particular socio-cultural conventions as they have been described variously in sociolinguistics, the ethnography of communication or the theory of actions. Thus, pragmatics is not part of the linguistic system (or its description). It is distinct from functional sentence perspective (which pragmatics reduces to in some terminologies), which is accommodated in the present model in §4.1.

First, the texts confirm or falsify the statements of the description. Second, our techniques of analysis at levels above the sentence are presently so little advanced that it is safe to supplement the efforts in that area by simple ostension. Third, while grammar and lexicon present the system, the texts represent the norm, or various norms, in the sense of Coseriu 1952. Especially for the latter reason, it is essential that the texts illustrate different forms of communication (text sorts); cf. ad 4, b. They should not all be narratives; there should be speeches, instructions, rituals, plays, jokes and, above all, dialogues.

The constitutive parts of each text are the following:

- a. orthographic representation,
- b. interlinear glossing, providing morphological and possibly syntactic information,
- c. translation into the background language,
- d. commentary.

These kinds of information bear some correspondence to the items ad 2. As in the other parts of the description, orthographic representation is chosen for practical reasons. Also, it is normally closer to the morphophonemics than a purely phonological representation and thus facilitates the interlinear glossing. The latter should be done according to the guidelines set out in Lehmann 1983. The translation itself can then be quite idiomatic. The commentary accounts for any aspects of the texts whose understanding is not derivable from the other three parts of the description.

Ad 4. A language could be regarded as a system established over an inventory of signs from which texts may be constructed. Insofar, it would be accounted for by the first three parts of the description. However, an adequate notion of a natural language is more comprehensive than that. A language is bound up with the life and culture of its speech community; it is a historical phenomenon. As such, it is not adequately grasped by a purely structural description. The global historical situation of the language has to be described as well. This account may be subdivided into the inner and outer situation as follows:

- a. inner: dialectal and sociolectal varieties, history, genetic affiliation of the language;
- b. outer: ethnographic aspects of the speech community itself, role of the language in the society (e.g. in a multilingual situation), conventional forms of communication, literature.

The description of the language system and the lexicon together form the **core description**. All the components of the language description are tightly linked to each other. It is general practice that the description of the language system (especially, the grammar) and the lexicon complement and therefore constantly refer to each other. The same is true of the other parts. In particular, the core description refers to the text corpus for further examples. The lexicon also refers to the statement of the historical situation for points of fact. The texts of the corpus

illustrate, and the commentaries accompanying it refer to, the statement of the historical situation. F1 shows the composition of the language description schematically.

### F1. *Composition of a language description*

Core description			
1. Language system [global subdivision]	2. Lexicon [subdivision of entry]	3. Text corpus [representation levels]	4. Historical situation [global subdivision]
stylistics	pragmatics		outer situation
semantics		translation	
grammar syntax morphology		interlinear glossing	inner situation
phonology			
phonetics	orthography	orthographic representation	

## 2.3. Completeness

P2. The description of the language system is complete in its account of the - central and peripheral - subsystems.

Every linguistic category and subsystem has the status of a prototype. This means that it has a **center** and a **periphery** (cf. Daneš 1966). Even the linguistic system as a whole has peripheral subsystems. What these are should be subject to empirical research. However, there have been inveterate preconceptions in linguistics which regard certain subsystems such as the phoneme system, parts of speech, morphological categories, syntactic relations and clause structure as central.<sup>6</sup> Numerous grammars, among them even some generally regarded as good specimina of their kind, limit themselves to an account of these. One often misses word formation, very frequently complex sentences, almost always particles (modal particles, interjections, ideophones etc.). I know of no linguistic theory which proves that these areas are peripheral to the linguistic system, although they may be.

Given that a language is not systematic throughout and, even as a system, is open in every direction, it cannot be described exhaustively. However, the theory of linguistic description has to guarantee that all the subsystems of a language are accounted for to the degree of their

---

<sup>6</sup> Some of these preconceptions may be based on theories which confuse centrality with regularity. Obviously, the regular parts of a language are more easily described than the irregular ones.

relevance to its functioning.

In the last years, general comparative linguistics has drawn our attention to phenomena which used to be neglected in earlier language descriptions. Consequently, linguistic comparison also serves as a heuristic tool to guarantee the completeness of the description.

## **2.4. Hermeneutics and dynamicity**

P3. A language description renders the object language intelligible.

Language is a human activity whose immediate goal it is to make sense. This goal is common to speaker and hearer. The linguist has to respond to it. His description has to bring out the sense that is hidden in linguistic structure. To this extent, his task is an interpretative or hermeneutic one.

In the present framework, the hermeneutic quality of a description takes the place of what is termed explanation in other frameworks. Instead of trying to explain an instance of language by positing laws for it, we should try to understand it or, rather, to show how it is to be understood (cf. Lehmann 1987, §4.2 for some examples).

To understand a human act (e.g., an utterance) means to know its goal and the conditions under which one might do or have done that act oneself (if one had the ability). Analogously, to understand an activity (in particular, a culture-bound activity such as a language) means to know the circumstances under which this activity has to fulfill its goal, so that one might engage in it (if one had the ability).

Knowledge of the conditions of an act includes knowledge of the available alternatives. A hermeneutic linguistic description therefore accounts for variation. Knowledge of the circumstances of a social activity includes knowledge of its direction in time, of the sense in which it is currently developing. Consider, for illustration, the projection of a movie. It is normally impossible to fully understand a still picture if one has no knowledge of the segments immediately preceding and, perhaps, following it. An "ideal" synchronic linguistic description in the sense of a pure momentary cross-section therefore is not a hermeneutic description. This leads to the following postulate:

P4. A language description represents the operational and evolutive dynamism of the language.

As an activity, language is dynamic. For the synchrony, this entails that a language is not reduced to an inventory of categories and relations,<sup>7</sup> but that it consists equally of operations

---

<sup>7</sup> This reduction is one of the few things in Langacker's (1987) cognitive grammar that are incompatible with the present framework.

which associate functions with structures, which select items from their paradigmatic class and combine them with their syntagmatic context. For the diachrony, the dynamicity of language entails that at any given moment some components (grammatical concepts, expression devices and operational strategies) are declining, others gaining ground. This is the presence of diachrony in synchrony or what Sapir called the 'drift' of a language. Cf. also Coseriu's (1987:46f) 'principle of dynamic description'.

Linguistic description accounts for this by ordering grammatical concepts and expression devices on continua. Grammaticalization scales play a prominent role here. Since they incorporate both expression and content of the language sign, they can serve as an organizational principle both in the synthetic and in the analytic part of the description. A description which accounts for the dynamicity of language acquires diachronic depth without mingling different historical stages. It represents the diachrony in the synchrony.

### 3. General comparative and language-specific grammar

There is a mutual dependency between general comparative linguistics and descriptive linguistics, which proves sometimes painful, more often fruitful. On the one hand, progress in general comparative linguistics depends on the availability of good language descriptions. On the other hand, progress in descriptive linguistics depends on the availability of good theories of linguistic description, the central part of which is expected from general comparative linguistics.

The kind of relationship that I assume between general comparative and specific grammar is formulated in P5, which is meant as a postulate to be followed by the descriptive linguist.

P5. Describe your language in such a way that the maxim of your description could serve, at the same time, as the principle of a general comparative grammar - and, thus, as the maxim of the description of any other language.

P5 is the **categorical imperative of language description**. It naturally leads to the following thesis:

T2. The description of a specific language is a concretization of general comparative grammar.

The following subsections will establish T2 and show how P5 can be complied with.

#### 3.1. General comparative grammar

Most linguists, I presume, will find it desirable that a language description conform to the definition and postulates put forward in §§2.1 - 2.3 and, possibly, 2.4. Everybody, not only the general comparative linguist, wants a language description to be comprehensive. There

are very few, if any, language descriptions which are complete in the sense explained in §2.2f. Most of the time one has to gather the information wanted from different and independent sources. In part, the completeness of a language description depends on the resources available to the analyst. For another part, however, it depends on the availability of information on the contents and form of a comprehensive language description and how it is to be done. As long as such information is not available, no linguist can blame any other for the unsatisfactory state of affairs.

The general comparative linguist has another demand to pose upon a language description which we may formulate in the following postulate:

P6. A language description is comparable with the descriptions of other languages.

If every analyst organizes his language description according to a personal scheme, his description may be as comprehensive as may be; but it will be difficult to compare information drawn from it with information drawn from the description of another language. Again, as long as no generally applicable schema is available, the descriptive linguist can hardly do any better.

Requirements both of comprehensiveness and of commensurability of language descriptions thus converge in the call for a general schema on which the description of any language can be based. This is what used to be called a **general grammar**. In the heyday of rationalism, a general grammar was conceived as a deductive enterprise. As such, it would have to be grounded in a general theory of language. However, to the degree that such a deductive basis was not in fact available, the general grammars of those times were largely based on principles of logic and of the normative grammar of certain classic Indo-European languages. With the advent of linguistics as an empirical science, this pastime was abandoned. We witness its partial resurrection in our day in the form of so-called universal grammar.

In the last decades, much comparison of languages has been done on an empirical basis. The Dobbs Ferry Conference on universals of language (cf. Greenberg (ed.) 1963) has stimulated large amounts of research which strive for contributions to general grammar by accumulating inductive generalizations over the languages compared. This approach, in turn, has been criticized for being atheoretical, for putting out sets of largely isolated observations and hypotheses whose relevance to a general theory of language is not clear.

From this account, it becomes obvious that the general grammar that we want will have to combine a deductive and an inductive approach. This is why I call it **general comparative grammar** (GCG). It does not yet exist, but important contributions towards it have forthcome both from the deductive and from the inductive approaches.

Each of the four parts of language description presented in §2.2 should be the subject matter of general comparative linguistics. Thus, there should be a general comparative discipline

occupying itself with text corpora;<sup>8</sup> and there should be one dedicated to the historical contexts of languages. These will be neglected here. A general comparative model of language structure will comprise the core description. GCG *stricto sensu* refers only to the central portion of this model, the grammar as opposed to semantics, phonology and lexicon. In what follows, I will restrict my attention to GCG in this narrow sense (for contributions to general comparative lexicology, see Talmy 1985 and Lehmann 1988[P]).

### 3.2. The Lingua Descriptive Studies Questionnaire

*Lingua Descriptive Studies* (1979-1982, since 1984 *Croom Helm Descriptive Grammars*) is a series of grammars of different, mostly exotic, languages which are all organized according to the same schema. The schema was published in 1977 by the editors of the series, B. Comrie and N. Smith, in the form of a questionnaire. The Lingua Descriptive Studies Questionnaire (LDSQ) does not pretend to be a GCG. It is just meant to be a systematic catalogue of questions the answering of which would secure the completeness and comparability of the descriptions. According to its introduction (p. 5), "it is important that the general framework be sufficiently flexible to enable any arbitrary language to be described within this framework". However, to the degree that the LDSQ approaches this aim, it does represent an important contribution towards GCG - incidentally one reflecting predominantly the empirical-inductive approach. I will briefly discuss its overall organization, so that the requirements to be fulfilled by a GCG become more apparent.<sup>9</sup>

The main subdivision of the LDSQ is in

1. Syntax
2. Morphology
3. Phonology
4. Ideophones and interjections
5. Lexicon.

Part 4 just asks for the ideophones and interjections of the language and is not subdivided further. The lexical part asks for the items of a couple of semantic fields and for 207 items of basic vocabulary. Thus, the bulk of the conceptual work has been invested into parts 1 - 3.

The main section headings of part 1 are:

---

<sup>8</sup> Apparently, next to nothing is known about how a representative text collection of a language is composed.

<sup>9</sup> I am aware of one discussion of the LDSQ in the literature, namely Uhlenbeck 1980. This is essentially concerned with the objection that the questionnaire should not presuppose the universal existence of conceptual categories to be variously expressed in languages.

1. General questions [i.e. sentence types and subordination]
2. Structural questions [i.e. clause and phrase structure]
3. Coordination
4. Negation
5. Anaphora
6. Reflexives
7. Reciprocals
8. Comparison
9. Equatives
10. Possession
11. Emphasis
12. Topic
13. Heavy shift
14. Other movement processes
15. Minor sentence types
16. Operational definitions of word-classes.

The main section headings of part 2, with headings of section 2.1 shown for clarity's sake, are:

1. Inflection
  - 1.1. Noun inflection
  - 1.2. Pronouns
  - 1.3. Verb morphology
  - 1.4. Adjectives
  - 1.5. Prepositions/postpositions
  - 1.6. Numerals/quantifiers
  - 1.7. Adverbs
  - 1.8. Clitics
2. Derivational morphology.

The main section headings of part 3 are:

1. Phonological units (segmental)
2. Phonotactics
3. Suprasegmentals
4. Morphophonology (segmental)
5. Morphophonology (suprasegmental).

So much should suffice to give an idea of the general disposition. According to the introduction (p. 8), "the general direction of description within the questionnaire is from function to form." However, this principle is not at all observed consistently in the LDSQ. Its

first violation lies in the subdivision between syntax and morphology. Comrie & Smith, just as most of us, draw the boundary between these two with respect to the grammatical level of the word. Now this is clearly a structural criterion. Consequently, all the function-form oriented questions posed inside the morphology chapter presuppose that any language to be described will fulfill the corresponding functions at a certain structural level. For instance, the question for definiteness/indefiniteness marking is asked in the section on noun inflection. What then shall we say of languages such as English, which express this distinction by separate words (after all, articles are not part of noun inflection) or, worse, of languages such as Russian which express a similar distinction by word order?

The same problem repeats itself at the lower hierarchical levels of the LDSQ. To resume the last example, there are languages (such as Hungarian) which mark definiteness by verb agreement. For their description, the question as to the expression of definiteness would properly belong into the section on verb inflection. Similarly, the question as to the expression of the syntactic functions of noun phrases is also posed in the noun inflection section, although it might as well be asked in the verb inflection section. The curious consequence of this organization of the LDSQ is that a large part of the verbal morphology of several languages described in the *Lingua Descriptive Studies* series (e.g. of Abkhaz, cf. Hewitt 1979) is presented in the section on noun inflection.

Obviously, such problems could be solved within the LDSQ framework. It is not necessary that a question concerning definiteness presuppose noun inflection; it might be asked appropriately in a chapter on deixis and reference. Similarly, syntactic functions of noun phrases will naturally emerge from a discussion of the functional domain of participation, i.e. the articulation of an event in terms of core and participants. A purely function-form oriented GCG would be feasible. What, then, if the LDSQ were organized consistently in such a way?

A purely function-form oriented approach inevitably leads to the consequence that the different uses of a polysemous or multifunctional form are dispersed over the various chapters. For instance, in a description of English, the verb *to be* would have to be treated in ch. 1.2 (clause structure, namely copular sentences), 1.10 (possession), 1.11 (emphasis, namely clefting), 2.1.3 (verb morphology). Comrie & Smith acknowledge this problem and ask contributors to counteract it, chiefly by extensive cross-referencing.

Two objections may be raised against this procedure. First, whatever the analyst can do within the predominantly function-based framework in order to alleviate the above problem, will remain patchwork, since it is not systematically provided for in the conception. One can understand a language linguistically only if one has a coherent picture of its functioning in its own terms. This much, at least, appears to be acceptable of the justification for the purely form-based American structuralist grammars.

Second, there is no reason why a grammar - a GCG or a specific grammar - should be preferably function-based rather than form-based. Comrie & Smith assume that general comparative linguists will use *Lingua Descriptive Studies* essentially with such questions in mind as 'How is the concept of definiteness, or of possession, or whatever, expressed in this language?' This is doubtless a frequent and important kind of question asked by such linguists. However, there are also questions such as 'What does verbal prefixation, or vowel alternation, or front shift of constituents, express in this language?' The grammars of this series answer such questions only insofar as they, or the LDSQ, are really not purely function-based and, thus, inconsistent.

This discussion of a well-known and meritorious contribution to GCG leads us to P7.

### 3.3. Analytic and synthetic viewpoints

P7. In a language description, those items have to be treated together which are similar in the object language.

P7 is a basic requirement to be met by any language description. It is based both on grammar-theoretical reasons of adequacy and on practical reasons of usability. Certain directions of American structuralism have inferred from P7 that the description has to follow exclusively a principle which results from the structure of the language in question. This conclusion proved to be undesirable, as such a presentation renders the description unusable for non-specialists in the language. Moreover, it does not really follow from P7.

A weaker thesis, however, does follow from it, namely that the peculiar structure of the object language should be brought out by the description. This follows also from P3. Now this thesis immediately seems to conflict with P5 and T2, since the latter amount to the requirement that every language should be described according to one general schema. §§3.4 and 4 will be devoted to the resolution of what has proved to be a basic dilemma in descriptive linguistics. Here we may observe that a very similar dilemma arises already inside the particular language description which tries to obey P7, quite regardless of any requirements of general grammar.

All elements of a language which relate both to expression and to content may be similar in either of two respects: they may be functionally similar or structurally similar.<sup>10</sup> The association of function with structure in any language is partly motivated, partly arbitrary. To the degree that it is motivated, functional similarity correlates with structural similarity. To the degree that it is arbitrary, functionally similar elements are structurally dissimilar, and vice

---

<sup>10</sup> Although the most general terms for the two sides of the language sign are expression (significans) and content/meaning (significatum), it is customary to speak, with respect to expression and content within grammar, of form/structure and function, instead.

versa. If this is accepted, P7 entails the following thesis:

T3. The core description (cf. F1) is carried out according to two complementary viewpoints, the analytic and the synthetic.

In one part, the principle of the disposition of the material is a formal-structural one. One starts from the structures, interprets these and thus arrives at the functions. The corresponding part of the lexicon leads from the object language to the background language. This corresponds to the viewpoint of the hearer or of a user who is confronted with a text in the language and wants to understand it. This is the form-function oriented or **analytic part** of the description.

In the other part, the principle of disposition is a functional-semantic one. One starts from the functions, looks for their realization and thus arrives at the structures. The corresponding part of the lexicon leads from the background language to the object language. This corresponds to the viewpoint of the speaker or of a user who wonders how a given function is fulfilled in this language. This is the function-form oriented or **synthetic part** of the description.

Not only the lexicon and the grammar, but each of the three sections of the language system as shown in §2.2, ad 1, are described according to the two complementary viewpoints. This is shown in F2.<sup>11</sup>

---

<sup>11</sup> This conception originates with Gabelentz 1901:84-125. A similar one is reflected in Sapir's (1921, ch. IVf) distinction between grammatical processes and grammatical concepts, which itself goes back to F. Boas. It may be useful to recall that trends in the recent history of linguistics partly differ in their preference for one of the directions of description. Thus, early American structuralism worked "from phoneme to utterance", while generative semantics worked from semantics to phonology. Cf. also the directionality debate of the late sixties in generative grammar. Another pair of terms belonging into this context is 'recognition grammar vs. production grammar'. For further discussion, cf. Jespersen 1924:39-46, Lehmann 1980, Mosel 1987.



which is an essential constituent of language and, therefore, represented in every language. A theory of language universals is a central part of a theory of language. Structural properties of languages have, in principle, the theoretical status of variants;<sup>13</sup> as such, they are generally not language universals. Consequently, while structural properties of language systems are represented in a GCG, they are generally not represented in a theory of language universals. Consequently, a GCG is clearly distinct from a theory of language universals. The expression 'universal grammar' occasionally found in contemporary general linguistics is a misnomer for either of these two things.<sup>14</sup>

In §3.3, it was postulated (P7) that a language description must treat together what is similar in the object language. In §3, it was maintained (T2) that a language description be based on a general comparative grammar. Now a language is not just an eclectic accumulation of items from the superset of all possible language properties; it is a system "où tout se tient". Therefore, the question arises how the two propositions are reconcilable.

The problem whether a language description based on a non-language-specific grid can represent the spirit of the individual language is a very real one. Experience with a dozen of volumes of *Lingua Descriptive Studies* has shown that a general framework cannot possibly foresee all the fanciful associations of functions with structures that appear in the various languages.

However, we have to differentiate between **polysemy** and **homonymy**. At a cross-linguistic level, the distinction between the two can be reformulated as a distinction between a content-expression association recurring in unrelated languages and one occurring in only one language. Take two German examples: a) The word *selbst* functions both as an identifier ('self') and as an emphasizer ('even'). b) The suffix *-er* functions both as a plural marker on declinable words and as a comparative marker on adjectives. An internal analysis of German here clearly brings out the polysemy of case *a*, the homonymy of case *b*. Cross-linguistic evidence confirms this. Identity of elements functioning in identification and emphasis recurs in Portuguese (*mesmo*), Tamil (*taan*) and many other languages.<sup>15</sup> Identity of plural and comparative morphemes recurs only in languages genetically related to German.

Thus, one of the requirements to be put on a GCG will be that it be organized in such a way as to bring out the connection between identification and emphasis. And, in fact, for most of the cases, it will do this both in its synthetic and in its analytic portion. In the synthetic system, identification of an entity with itself and emphatic underscoring of it will be treated, in the functional dimension of reference, as two similar kinds of relationship of an entity with the universe of discourse. In the analytic system, the two uses of the morpheme in question

---

<sup>13</sup> This has been emphasized by H. Seiler since 1972.

<sup>14</sup> I have argued this repeatedly, e.g., with special reference to a couple of established strands of language universals research, in Lehmann 1982.

<sup>15</sup> Cf. Moravcsik 1972 and Edmondson & Plank 1978 for data and explanations.

will be treated in the same or at least in adjacent chapters to the degree that the distributions of the morpheme in the two cases are similar. German *selbst*, e.g., is in any case a sentence level particle with order properties describable by quantifier floating.

The conclusion is that the framework of GCG will allow to bring out the spirit of the function-form association of a specific language to the degree that the individual associations are functionally or structurally motivated. It will not provide a common denominator for cases of homonymy; but then there is none, in the first place.

## 4. Synthetic and analytic systems of GCG

The discussion in §3.2-4 was intended to support the thesis (T3) that a GCG, just like an individual grammar, has to be organized according to the analytic and the synthetic viewpoints. In the following sections, this proposal will be fleshed out.

### 4.1. The synthetic system

The synthetic part of the core description is organized independently of linguistic expression structures and, instead, grounded in human cognition and semiosis. The set of concepts and operations represented in human languages is structured in various ways. We can assume it to be articulated, at the highest hierarchical level, in what may be called cognitive domains.<sup>16</sup> Most of these are represented only in the lexica of languages. Examples are the domain comprising the physical constitution of the cosmos, the life cycle or the ingredients and operations of cooking. Some of them, however, reach over into the grammar; indeed, they provide the semantic subject matter of human grammars. To these belong the domains of possession, of spatial and temporal orientation and a limited set of others, to be reviewed presently.

There are two ways in which one might misunderstand this conception. One would be to assume that the lexicon, or lexical meanings, are "more universal" than the grammar, or grammatical meanings. What is meant, instead, is that all universal conceptual domains find their manifestation in the lexicon of every language, but some of them have no relevance for the grammatical structuring of languages. Another misunderstanding would be to conclude that the entire lexicon of every language is somehow contained in the set of universal cognitive domains, or reducible to universals. What the universal domains provide is only the framework of the structure of linguistic meanings.

In the UNITYP project of Cologne, those cognitive domains that manifest themselves in the grammars of languages have been analyzed at a cross-linguistic level and assumed to be universal. The list in F3 is not meant to be complete or definitive in any sense, but may suffice to give an impression of what is involved.

---

<sup>16</sup> Cf. Langacker 1987 for this conception.

### F3. *Functional domains of the synthetic system*

1. **Nomination:** an entity is named by a descriptive expression or a label (Seiler 1975).
2. **Apprehension:** an entity is grasped by categorizing and individualizing it (Seiler & Lehmann (eds.) 1982, Seiler & Stachowiak (eds.) 1982, Seiler 1986).
3. **Attribution:** a representation is modified so that the concept is enriched or the object is identified (Seiler 1978, Lehmann 1984).
4. **Possession:** the relation of an entity to another one is represented as inherent in one of them or established between them (Seiler 1983).
5. **Quantification:** the extent of the involvement of a set of entities in a predication is specified.
6. **Reference:** a representation is determined so that it can be related to and delimited within the universe of discourse (Seiler 1978).
7. **Participation:** a situation is articulated into an immaterial center and a set of participants and circumstances linked to it in various ways.
8. **Spatial orientation:** an entity is localized in the referential world, this is tied to the universe of discourse, and this is anchored in deixis.
9. **Temporal orientation:** a situation is designed with respect to its internal temporal structure, its temporal limits and relations at various levels.
10. **Intensification and comparison:** a concept or a thought is assessed qualitatively by explicit or implicit contrast with similar ones.
11. **Nexion:** a situation is expanded into a complex one, or several situations are linked together (Lehmann 1988[T]).
12. **Functional sentence perspective:** a thought is articulated into subject and predicate (cf. Sasse 1987, Himmelmann 1988), topic and comment, focus and background.
13. **Modality:** a thought is rendered relative to illocution and reality.

From the bibliographic references, it is apparent that only some of these have been worked out in the UNITYP context; the others are projected. A detailed account of the internal structure of the former may be found in the sources. Very little is as yet known about the set as a whole, what constitutes membership in it, how the set is structured, whether the domains differ only by being situated in different realms of cognitive space or also by logical or semiotic properties, etc.

Each of these domains has an internal structure. Within each of them, a number of techniques are available which translate the universal concepts and operations in specific languages. A couple of functional principles account for the ordering of these techniques. For example, in the domain of attribution, relative clauses, participials and adjectives of various kinds are ordered on a continuum whose poles are defined by the functions of identification of an entity vs. enrichment of a concept. By virtue of this gradient internal organization, the functional domains are the locus of synchronic and diachronic variation.<sup>17</sup>

---

<sup>17</sup> This is why they are called dimensions in the UNITYP framework.

The synthetic system of a GCG - and, consequently, of a specific grammar - is organized according to these domains and their internal structure. It thereby satisfies all the postulates P3 - P7.

## 4.2. The analytic system

The analytic part of the core description is organized independently of linguistic meaning structures and, instead, based on the structural forms possible in linguistic expressions. Such an approach has been taken in several schools of European and American structuralism, including transformational grammar. A number of alternative proposals have been put forward, but no generally recognized theory of linguistic structure has come forward so far. The following proposal must be regarded as tentative.

In every language, most of the meanings are expressed by complexes of phonological units which constitute the significantia of morphemes. These are inventorized in the lexicon. Some of them are grammatical formatives and therefore reappear in the grammar, but in a different perspective.

The analytic system is based on the paradigmatic and syntagmatic relations of significantia. The expression aspects of these are:

- paradigmatic: possibility of substituting a significantia in its position by another one or by zero; relationships of opposition, complementary distribution and free variation;
- syntagmatic: modification of a significantia by segmental alternation, intonation or accent; sequential relationship of a significantia to a neighbouring one, including permutability and bondedness (phonological ties, separability etc.).

However, it would be unwonted and impractical if these structural aspects were the primary principle of disposition in the grammar. Instead, they are used to define levels of linguistic structure, units occupying these levels and subclasses of such units. For these, then, the paradigmatic and syntagmatic relations are specified. This leads to the organization in F4.<sup>18</sup>

### F4. *Structural hierarchy of the analytic system*

1. Units of different **grammatical levels**  
(Word, syntagm [phrase], clause, sentence, paragraph)
2. [For each unit:] **Subclasses** of unit  
(Word classes, syntactic categories, clause types, sentence types, paragraph types)
3. [For each subclass of unit:] Internal **syntagmatic structure**, according to the following types of structural device:

---

<sup>18</sup> F4 follows both Heger 1976 and Coseriu 1987 in treating grammatical levels (ranks) and units as a primary structuring device. However, in neither of the two models, such concepts would be defined inside the analytic system.

- structural relations between the members of the syntagm [i.e. the units of the next lower level], in particular: distribution of the members, including positions for combination of peripheral elements with the head;
- morphological modification [of the flexional type];
- segmental phonological modification;
- prosodic modification.

4. [For each type of structural device:] **Paradigmatic relations** between members of the type.

For an example, let us take a particular path through this hierarchy. At the first level, we deal with the syntagm unit (where 'syntagm' is to be taken in the narrow sense of 'word group'). At the second level, we subclassify the syntagms and thus come to the noun phrase. At the third level, we display the internal structure of the noun phrase: we give its constituents, among them the noun, the adjective attribute, the article; we show the structural positions of these and the processes of modification which they undergo in the combination. At the fourth level, we discuss the difference between prenominal and postnominal position of an attribute, between intonation patterns of the noun phrase, etc.

The hierarchical character of the system in F4 is not equally pronounced at all levels. Certain grammatical processes (e.g. emphatic accent) may be the same for several of the levels. However, there are level-specific processes, such as infixation or vowel harmony. These testify to the essential hierarchical character of linguistic structure.

The fact that terms familiar from English grammar appear in F4 should mislead no one into thinking that here the structure of English is once again made the basis of a general grammar. Some languages, such as Dyirbal, do not have a noun phrase, although they may have nouns and adjectives and syntactic relations between the two. Other languages, such as Turkana and Guarani, have no or almost no adjectives. While the synthetic system cares for the question of how the function of the English noun phrase and adjective are fulfilled in these languages, the grid of the analytic systems of Dyirbal and Turkana will just remain unoccupied at the places occupied by the noun phrase and the adjective in English. Again, in the analytic grammars of the former two languages, the structural process of correlative morphological modification of a noun and its attribute, known as agreement, will play a prominent role, but not appear in English.

The terms appearing in F4 thus refer to prototypical elements of linguistic structure which tend to recur, under one form or another, in different languages. Given that expression devices vary in a gradient fashion, there is a continuum between the concepts of all the levels of F4. There is no clear boundary between phrase and clause, nor is there one between sentence types, between types of morphological and phonological modification or between the various paradigmatic relations. The analytic system of a GCG thus presents a framework of systematically varying possibilities which assume, in large part, the form of continua. It provides for the structural possibilities of natural languages in a systematic, but non-peremptory way and, thus, again satisfies postulates P3 - P7.

## 5. Conclusion

Throughout the history of general linguistics, two opposing views have been held as regards the form of a language description. The universalists have maintained that all languages are fundamentally alike, therefore there must be a general model of language structure which can be applied in the description of any language. The relativists have argued that every language is a homogeneous system, therefore a property of one language is never like a property of any other language, and consequently every language has to be described in its own terms. Both are right and wrong. There are universals of language, and consequently there can be a framework for the description of any language. However, there is no universal grammar, since the association of expression with content is done inside the specific language, not at a universal level.

The GCG as proposed here tries to solve both the theoretical problem of the above antinomy and the practical problem of guaranteeing complete and comparable language descriptions. It achieves this by virtue of two properties: First, it reflects in a systematic way the cross-linguistic variation and thus provides for the variants appearing in any one language. Second, like the description of the individual language system, it is organized in a synthetic and an analytic system. Associations of expression with content which are motivated by similarity relations of one or the other are therefore captured in a principled way.

Finally, the GCG provides a maximally language-independent and neutral framework which fits the concepts, categories, relations and processes of any language. It does this by ordering them on cross-linguistic continua which are the locus for synchronic and diachronic, interlingual and intralingual, variation.

## References

- Brettschneider, Gunter & Lehmann, Christian (eds.) 1980, *Wege zur Universalienforschung*. Sprachwissenschaftliche Beiträge zum 60. Geburtstag von Hansjakob Seiler. Tübingen: G. Narr.
- Comrie, Bernard & Smith, Norval 1977, "Lingua Descriptive Studies: questionnaire". *Lingua* 42:1-72.
- Coseriu, Eugenio 1987, *Formen und Funktionen*. Studien zur Grammatik. Tübingen: M. Niemeyer (Konzepte der Sprach- und Literaturwissenschaft, 33).
- Daneš, František 1966, "The relation of centre and periphery as a language universal". *TLP* 2:9-21.
- Dressler, Wolfgang U. 1967, "Wege der Sprachtypologie". *Die Sprache* 13:1-19.
- Edmondson, Jerrold A. & Plank, Frans 1978, "Great expectations: an intensive self analysis". *Linguistics and Philosophy* 2:373-413.
- Gabelentz, Georg von der 1901, *Die Sprachwissenschaft*. Ihre Aufgaben, Methoden und bisherigen Ergebnisse. 2. ed. Reimpr. 1972: Tübingen: G. Narr (TBL, 1).
- Greenberg, Joseph H. (ed.) 1963, *Universals of language*. Report of a conference held at

- Dobbs Ferry, New York, April 13-15, 1961. Cambridge, Mass.: MIT Press.
- Heger, Klaus 1976, *Monem, Wort, Satz und Text*. Tübingen: Niemeyer (Konzepte der Sprach- und Literaturwissenschaft, 8). 2. enlarged ed.
- Hewitt, B. George 1979, *Abkhaz*. Amsterdam: North-Holland (LDS, 2).
- Himmelman, Nikolaus 1987, *Morphosyntax und Morphology - Die Ausrichtungsauffixe im Tagalog*. München: W. Fink (Studien zur Theoretischen Linguistik, 8).
- Jespersen, Otto 1924, *The philosophy of grammar*. London: Allen & Unwin.
- Langacker, Ronald W. 1987, *Foundations of cognitive grammar*. I: Theoretical prerequisites. Stanford: Stanford UP.
- Lehmann, Christian 1980, "Aufbau einer Grammatik zwischen Sprachtypologie und Universalistik". Brettschneider & Lehmann (eds.) 1980:29-37.
- Lehmann, Christian 1982, "Some current views of the language universal". *LeSt* 17:91-111.
- Lehmann, Christian 1983, "Directions for interlinear morphemic translations". *FoL* 16, 1982:193-224.
- Lehmann, Christian 1984, *Der Relativsatz*. Typologie seiner Strukturen, Theorie seiner Funktionen, Kompendium seiner Grammatik. Tübingen: G. Narr (LUS, 3).
- Lehmann, Christian 1987, Theoretical implications of processes of grammaticalization. Conference on The Role of Theory in Language Description, Ocho Rios, Jamaica, Oct 31 - Nov 8, 1987.
- Lehmann, Christian 1988, "Predicate classes and participation". Lehmann, Ch., *Studies in general comparative linguistics*. Köln: Institut für Sprachwissenschaft der Universität (akup, 71); 33-77.
- Lehmann, Christian 1988, "Towards a typology of clause linkage". Haiman, John & Thompson, Sandra A. (eds.), *Clause combining in grammar and discourse*. Amsterdam: J. Benjamins; 181-225.
- Moravcsik, Edith A. 1972, "Some cross-linguistic generalizations about intensifier constructions". *CLS* 8:271-277.
- Mosel, Ulrike 1987, Inhalt und Aufbau deskriptiver Grammatiken (How to write a grammar). Köln: Institut für Sprachwissenschaft der Universität (Arbeitspapier Nr. 4 [N.F.]).
- Sapir, Edward 1921, *Language*. An introduction to the study of speech. New York: Harcourt, Brace & World (Harvest Books, 7).
- Sasse, Hans-Jürgen 1987, "The thematic/categorical distinction revisited". *Linguistics* 25:511-580.
- Seiler, Hansjakob 1969, "On the interrelation between text, translation, and grammar of an American Indian language". *Linguistische Berichte* 3:1-17.
- Seiler, Hansjakob 1972, "Universals of language". *Leuvense Bijdragen* 61: 371-393.
- Seiler, Hansjakob 1975, "Die Prinzipien der deskriptiven und der etikettierenden Benennung". Seiler, H. (ed.), *Linguistic workshop III*. Arbeiten des Kölner Universalienprojekts 1974. München: W. Fink (Structura, 9); 2-57.
- Seiler, Hansjakob 1978, "Determination: A functional dimension for inter-language comparison". Seiler, H. (ed.), *Language universals*. Papers from the Conference held at Gummersbach/Cologne, Germany, October 3-8, 1976. Tübingen: G. Narr (TBL, 111);

301-328.

- Seiler, Hansjakob 1983, *POSSESSION as an operational dimension of language*. Tübingen: G. Narr (LUS, 2).
- Seiler, Hansjakob 1986, *Apprehension. Language, object, and order*. Part III: The universal dimension of apprehension. Tübingen: G. Narr (LUS 1, III).
- Seiler, Hansjakob & Lehmann, Christian (eds.) 1982, *Apprehension. Das sprachliche Erfassen von Gegenständen*. Teil I: Bereich und Ordnung der Phänomene. Tübingen: G. Narr (LUS 1, I).
- Seiler, Hansjakob & Stachowiak, Franz Josef 1982, *Apprehension. Das sprachliche Erfassen von Gegenständen*. Teil II: Die Techniken und ihr Zusammenhang in den Einzelsprachen. Tübingen: G. Narr (LUS 1, II).
- Talmy, Leonard 1985, "Lexicalization patterns: semantic structure in lexical forms". Shopen, Timothy (ed.) 1985, *Language typology and syntactic description*. Cambridge etc.: Cambridge UP; vol.III:57-149.
- Uhlenbeck, Eugenius M. 1980, "Language universals, individual language structure and the Lingua Descriptive Series Project". Brettschneider & Lehmann (eds.) 1980:59-64.