Universität des Saarlandes – Philosophische Fakultät

Linguistic Modeling for Multilingual Machine Translation

Dissertation zur Erlangung des Grades
eines Doktors der Philosophie

vorgelegt von

Oliver Streiter

im April 1996

Oliver Streiter
Karcherstraße 6
66111 Saarbrücken

Acknowledgments

The thesis presented here summarizes the work I have done in the last 5 years at IAI. As a student of the University of Paris VII I established contact with Johann Haller who offered me practical training at IAI in summer 1990. Here I first came into contact with the CAT2 MT system. After a short stay in Torino and Palermo with interests other than MT I started work at IAI in January 1991 and from then on I have been developing and testing grammars for MT, first under the guidance of Randall Sharp and later, when he left for Mexico, without his help. Although now completely changed, the first large-scale grammars were built by Paul Schmidt and Johann Haller, who integrated many fruitful linguistic concepts developed in the Eurotra project into the CAT2 grammars. Linguistic programming is also done at present by Antje Schmidt-Wigger and Catherine Pease within the CAT2 framework, and Heinz-Dieter Maas who is responsible for the development of the morphological component called MPRO. Michael Carl programmed the interface between CAT2 and mpro. The implementation of the CAT2 formalism is entirely the responsibility of Randall Sharp. In addition, innumerable students from many countries have been to Saarbrücken to work with us for weeks or months, doing valuable work on various problems related to MT [Schmidt-Wigger91], [Steinberger92a], [Steinberger92b], [Raffy92], [Fakhour93], [Steinberger94], [Pease93].

Many people have contributed in some way to the existence of this thesis. First of all there are all those native speakers I have interrogated, or who have explained to me phenomena of their mother tongue. Among them are: Manon van Zuylen, Consolata Peyron, Nadia Mesli, Leonid Iomdin, Zhāng Xiăohóng and Li Jié. Assistance in LaTeX as well as for many technical problems was provided by Ute Hauck. She also took over the CAT2 system administration when Randall Sharp moved to Mexico. Valuable comments on earlier versions have been made by Michael Carl, Chris Chambers, Leonid Iomdin, Catherine Pease, Randall Sharp, Antje Schmidt-Wigger and Ralf Steinberger. The Bulgarian examples were checked by Ivan Derzhanski and Irina Nenova. Lĭ Jié checked the Chinese examples and gave important hints for the interpretation of the data. I thank all of them. But most of all I am indebted to Randall Sharp, my teacher and friend. It is to him that I dedicate this book.

Contents

transliteration
Cyrillic characters

Conventions Transcriptions Abbreviations

This section describes the conventions and abbreviations I have used. Normal text is written like this. Since most of what follows is about feature structures I will indicate them by the typewriter type style, e.g. “The `wh` feature is used to indicate whether a constituent undergoes wh-movement or not. Possible values are `wh=yes` and `wh=no`”. With bold characters I mark those **terms** I have defined in the text and which I use only in the sense of this definition.

For languages which do not use Latin characters, I used the transliteration and transcription system which I think is the most accepted one in linguistic circles.
For the transcription of Chinese I use the 58 symbol writing system developed in The People's Republic of China in 1955 and finally adopted in 1958, called ”pin ying”. This system uses monographs and digraphs for the representation of phonemes as well as diacritical signs for the representation of the distinctive tones (*bā, bá, bă, bà, bȧ*) (cf. [Huáng and Liào83], [Crystal87], [Zadoenko and Xuan93]).
For Russian I use the transliteration system described in [Bouvier77]. This system is valid for the transliteration of all Slavonic Cyrillic characters if not further specified by special transliteration rules as in the case of Bulgarian. The system used for Bulgarian deviates from the general transliteration of the Slavonic Cyrillic characters wrt the 26th and 27th character which are transliterated as št and â respectively.
Language examples from Serbo-Croatian are given for reasons of simplicity in the Croatian variant, which uses Latin characters with diacritical signs.

The following abbreviations are employed in the glosses:

Abbreviation	spelled out	example
1	first person	
2	second person	
3	third person	
ACC	accusative case	
ATT	attribute marker	*wǒ* **dė** *shū*
DAT	dative case	
DEF	definite inflection paradigma	*der groß***e** *Mann*
ERG	ergative case	
FEM	feminine gender	
GEN	genitive case	
INDEF	indefinite inflection paradigma	*ein groß***er** *Mann*
DISS INSTR	instrumental case	
MASC	masculine gender	
NEG	negation marker	*Il* **ne** *vient* **pas.**
NOM	nominative case	
PASS	passive marker	
PL	plural number	
PREF	separable prefix	*Es kommt* **vor.**
SG	singular number	
VERB	verb	
QU	question marker	*chī fàn lė* **må** *?*

Within the text I use the following abbreviations, all of which are standard in the related literature.

Abbreviation	spelled out
CEU	Commission of the European Union
CS	Constituent Structure
EBMT	Example-based MT
GB	Government and Binding Theory (cf. [Chomsky81])
GPSG	Generalized Phrase Structure Grammar (cf. [Gazdar et al.85], [Bennett95])
HPSG	Head Driven Phrase Structure Grammar (cf. [Pollard and Sag87], [Pollard and Sag94])
IS	Interface Structure
KBMT	Knowledge-based MT
LBMT	Linguistic-based MT
LF	Lexical Function (cf. [Mel'čuk74], [Mel'čuk84])
LFG	Lexical Functional Grammar (cf. [Bresnan (ed.)82])
MT	Machine Translation
NLP	Natural Language Processing
SBMT	Statistic-based MT
SL	Source Language
SS	Surface Structure
SV	Support Verb
SVC	Support Verb Construction
TG	Transformational Grammar
TL	Target Language

Background information to MT systems, linguistic concepts and other is given in a glossary at the end of this thesis. I have preferred this style of representation over the use of footnotes since it allows for the selectional reading of specific chapters without the necessity of repeating footnotes every time a concept is mentioned. Words for which an entry can be found in the glossary are marked by an asterisk, e.g. ETAP*.

Chapter 1

Introduction

1.1 About this Thesis

MT is said to have had its origins in the late forties. Since then many books have been published about MT. While some of them focus on the history of MT and give an overview of MT systems (e.g. [Hutchins86], [Kulagina89], [Schwanke91] and [Whitelock and Kilby95]), others try to describe the problems of MT and possible approaches to handling them (e.g. [Nirenburg (ed.)87], [Luckhardt87], [Lehrberger and Bourbeau88], [Luckhardt and Zimmermann91], [Laffling91], [Goddman and Nirenburg (eds.)92] and [Schütz94]). Still others describe concrete systems focusing on formal aspects and less on linguistic details (e.g. [Maas84], [Weisweber94] and [Sharp94]). Finally there are publications which specify linguistic details of MT systems (e.g. [Luckhardt and Maas83], [ERM90], [Apresjan et al.89], [Mehrjerdian92] and [Dorr93]), but only a few succeed in presenting a coherent linguistic theory of MT, which on the one hand is sufficiently abstract in order to allow knowledge transfer from the MT system they describe to other MT systems and on the other hand shows how such a theory can be implemented. With this thesis I try to overcome this gap in the literature in describing the linguistic foundations for MT and how they are realized in the CAT2 MT system. Although I concentrate on the linguistic aspects of MT, I do not assume that MT can be done exclusively based on linguistic approaches to analysis and generation, and first steps are currently being taken to extend the CAT2 system with components from alternative approaches such as SBMT, EBMT[1]. These alternative approaches, on the other hand, currently try to integrate linguistic knowledge in order to optimize the

[1] These approaches are discussed in [Brown et al.90], [Furuse and Iida92], [Brown et al.93], [Hutchins93], [Nirenburg93], [Isabelle93], [Jones95], [Fung and Wu95] and [Sumita and Iida95].

form
function
meaning
neutralization

output of such systems (cf. [Brown and Frederking95], [Collins96]). Therefore, I consider it to be my task to illustrate the linguistic foundations of MT, taking as an example the implementation work I have done during the last 5 years.

1.2 The Linguist's Contribution to MT

The linguist's contribution to the modeling of the translation process has mostly been understood as supplying a monolingual analysis according to the actual linguistic research paradigm[2] and to show how the collected information can be used for translation, e.g. [Kudo and Nomura86], [Sharp86], [Hauenschild and Busemann88]. This conception, interesting as it might be for the research paradigm, has never led to the construction of a running MT system. In order to achieve this, a contrastive analysis of languages should be the starting point for the modelling of MT if the contributions of the linguist are to have any impact. As Tsujii puts it: *The most crucial of all is that linguistics in LBMT have placed excessive importance on monolingual theories and largely ignored bilingual counterparts. As a result, their theories of MT become mere parasites of monolingual theories, while ideal theories of MT, to my mind, should center around a bilingual theory and reconstruct monolingual theories accordingly.*([Tsujii93], pg.97). To my knowledge, only few MT systems have been developed taking the contrastive view as a starting point for its modeling of MT, e.g. [Dorr93]. In order to achieve this in a systematic and valid way, a view on language has to be taken which involves the three following components of language description: **form, function** and **meaning.**

The **form** of a language describes those properties of a language for which no choice is possible: A speaker of German, for example, has no choice about whether to put the inflectional affix before or after an adjective stem; only the latter will result in a German word (e.g. **enschön* vs. *schönen*). The **form** of a language is thus the set of invariable constraints on the surface structure (henceforth SS) of a language[3]. To capture the hierarchical relations of these properties has been one of the main streams of linguistics in this century (cf. [Rombouts19], [Jakobson86], [Greenberg63], [Greenberg65]). The **form** does not only list structural constraints on the SS but also cases of neutralization, i.e. contexts in which a choice provided by the language is not possible, e.g. final devoicing in Dutch and German:

[2] Paradigms, as defined by Kuhn [Kuhn87], represent scientific performances which are taken by a group of scientists for a period of time as the basis of their scientific work.

[3] [Vinay and Darbelnet58] call this set of invariable constraints *ensemble de servitudes* (set of servitudes) (pg. 31), where a 'servitude' is defined as *Cas où le choix, la forme et l'orde des mots sont imposés par la langue* (Case where the choice, the form and the order of words are dictated by the language).

bidirectionality

(1)

The Form of German
no nasal vowels
no bilabial fricatives
+[obstr] ⇒ -[voice] / _#
no inflectional prefixes/infixes
articles precede the NP
nonfinite structures have no overt subject
...

Functions are involved whenever the speaker actually has a choice, say, between saying *speaker* and *speaking*, between *the dog eats the cat* and *the cat eats the dog.* Choices are only possible for a given language in a context which does not neutralize the possible choices. **Functions** relate a choice of possible SSs to **meaning** and vice versa. A **function** is identified by its name and the functional value, i.e. the choice that has been realized:

(2)

context	*The dog eats the cat*	*The cat eats the dog*
name of function	syntactic function	syntactic function
scope of function	*dog*	*dog*
functional value	subject	direct object
scope of meaning	*dog*	*dog*
meaning	agent	theme

For the contexts where MT is currently used, i.e. technical and scientific types of texts, it is reasonable to assume that translation is **meaning**-preserving, i.e. the same distinctions have to be made in SL and TL. These distinctions are responsible for the applications of the functions, i.e. they represent the **meaning** of the text[4]. Given the **form** of a language, the **meaning** determines completely the SS, i.e. **meaning** is the condition for the application of **functions**.

My central claim is that MT, especially in a multilingual architecture, has to be based on **meaning** instead of a set of heuristics which match **forms** and **functions** of SL and TL. The choice of the appropriate **function** in the TL depends on the interaction of the **form** and all constraints on **functions** of the TL. These constraints cannot be predicted from the SL.
The strategy followed by many MT systems to match **forms** and **functions** may have some success in bilingual MT systems involving cognate languages, but is impracticable in a multilingual environment.

The requirement of an MT system to use every language component both for

[4] Similar meaning based approach to translation can be found in [Mel'čuk74], [Kuz'min75], [Larson84], [Nitta86] and [Bateman89].

number marking

analysis and generation is supported by the proposed model. Bidirectional systems have to separate the descriptors of different languages, otherwise the generation part is underspecified wrt the description of the TL and cannot function as SL[5]. The oddness of such an approach becomes obvious when compared to the activity of a human translator. Human translators master in the best case SL and TL with the same degree of proficiency, otherwise they use the 'strong' language as the TL and the 'weak' one as SL. While the TL requires a complete and correct description, deficiencies in the SL can be supplemented by conceptual knowledge which helps to find out what is said. What monodirectional MT systems model, however, is a translator who can produce a text in a language she cannot understand, generating a structure according to certain heuristics, but not being able to control the output of these heuristics. In such systems A_{source} is mapped to A_{target} and B_{source} to B_{target}, but if A_{target} and B_{target} do not go together in the TL, the system will generate illegal output or fail. The process of translation, however, necessarily checks that the target structure makes the same distinctions (i.e. has the same **meaning**) as the source expression does. The system must have at its disposal information about the **form** and all constraints on the **functions** of the TL in order to decide which **functions** can be employed to express the **meaning**.

1.3 An Exemplary Analysis

The distinction of **form**, **function** and **meaning** is not always trivial. Let us take an example to illustrate how these components may be disentangled.

Many languages have a system for the marking of number of nouns. Thus, while the English nouns *woman* and *laundry* are singular, *women* and *clothes* are plural. This marking of number cannot be directly equated with **meaning**, as becomes apparent when we compare translationally equivalent expressions, as in the following German-Italian examples.

(3) German - Italian:
a. das Auto - la macchina
the car$_{SG}$ - the car$_{SG}$
'the car'
b. die Autos - le macchine
the car$_{PL}$ - the car$_{PL}$
'the cars'
c. die Nudeln - la pasta
the pasta$_{PL}$ - the pasta$_{SG}$
'the pasta'

[5] Systems using this architecture are MU (cf. [Nagao and Tsujii86] and [Sakamato et al.86], ETAP* and ETAP2* METAL (cf. [Thurmair90]), LOGOS (cf. [Scott89] [Scott92]), NTVECMT (cf. [Chen and Chen92]) and CHARON*.

pluralia tantum
countability

d. Nudeln - della pasta
pasta$_{PL}$ - of the pasta$_{SG}$
'pasta'

e. der Teig la pasta
the dough$_{SG}$ - the dough$_{SG}$
'the dough'

f. jedes Mal - tutte le volte
every time$_{SG}$ - all the times$_{PL}$
'all the times'

g. die Brille - gli occhiali
the glasses$_{SG}$ - the glasses$_{PL}$
'the pair of glasses'

h. die Brillen - gli occhiali
the glasses$_{PL}$ - the glasses$_{PL}$
'the pairs of glasses'

In addition, singular and plural marking can be observed on nouns which are the translations of parts of speech or syntactic constructions where no number marking is possible. In (4) the German verb *anschaffen* is translated into an English singular noun. Where can the singular marking come from if the marking is identified with the meaning itself?

(4) *German:*

Sie begrüßen es, daß ein Auto angeschafft wird.
they greet it, that a car acquired becomes

'They greet the acquisition of a car.'

In order to describe such phenomena correctly, one first has to isolate the form, i.e. constraints on functions, through the contrast of possible and impossible SSs. In our examples the constraints are of three kinds: First, German and Italian have only two types of number marking, singular and plural. Second, *occhiali* is a pluralia tantum, i.e. no singular is possible. Third, the number marking occurring with *jedes* and *tutte* is grammatically fixed as singular in German and plural in Italian, irrespective of their **meaning**. The opposition of singular and plural is thus neutralized. How these phenomena are to be handled may be a problem for implementation, but they do not contribute to the identification of the **functions** and the **meaning** related to number marking. They are excluded from the analysis of the phenomenon.

The remaining examples can be systematized when a second variable is integrated into the analysis. This variable is the 'countability' of nouns. The countability of a noun specifies whether plural and counting is possible or not.

With count nouns counting and plurals are possible, with mass nouns only singular is possible. The Italian *pasta* is a mass noun for which no plural is possible without a change in **meaning** from ‘pasta’ to ‘biscuits’. The German *Nudel* is a count noun since plural as well as counting of *Nudeln* are possible. We thus arrive at three classes of translational equivalence. The first class consists only of singular count nouns. Therefore singular count nouns can only be translated as singular count nouns (3a). The second class consists of plural count nouns and mass nouns, which may be translations of each other. This second class must be subdivided according to the values of the **function** determination, definite (3b) and indefinite (3c).

(5)	COUNT SG	COUNT PL DEF MASS DEF	COUNT PL INDEF MASS INDEF

The next step is to find the **meaning** which accounts for this pattern. In the proposed model, number is a **function** of count nouns which yields with the functional value ’singular’ the **meaning** ’quantized’. With the functional value ’plural’ the **meaning** depends on the **function** determination which yields the **meaning** ’cumulative’ or ’quantized’. For a definition of these terms cf. Section 7.3. Singular mass nouns equally refer according to the **function** determination ’cumulative’ or ’quantized’. The relations are summarized as follows, where the **meaning** is represented by three types of quantifications: cumulative, quantized ”X=1”, or quantized ”X>1”.

(6)	COUNT SG	COUNT PL DEF MASS DEF	COUNT PL INDEF MASS INDEF
	quantized		cumulative
	X=1	X>1	

The **meaning** so identified is applicable to other parts of speech. Events expressed through verbs may be quantized (*run 5 miles*) or cumulative (*run*). Adjectives may be lexically marked as to whether they refer to a quantized entity or a cumulative entity. *triangular* refers ’quantized’ while *angular* refers ’cumulatively’. This can be seen from the paraphrases with a nominal: *triangular = having 3 angles* (quantized) and *spiny = having thorns* (cumulative). Other types of quantification are found in the adjectives and adverbs like *weekly, monochrome* and *polychrome.*

The above deliberations can be summarized as follows: the treatment of a phenomenon in a multilingual setting requires (i) a contrastive listing of translational equivalences. (ii) the **form** has to be excluded from the analysis, since the **form** does not contribute to the identification of the **functions** and

meaning, although special attention has to be paid to its implementation. (iii) functional equivalences have to be identified and (iv) the **meaning** which underlies the different functional equivalences has to be identified. The representation of the **meaning** should be chosen in such a way as to be applicable to every language and to all **functions** related to all parts of speech which can express the **meaning**. The **meaning** is represented as a system of oppositions, where each opposition may be linked to zero, one or more **functions**. Although a parse of a sentence is already a possible **meaning** representation of that sentence[6], I derive from this representation a representation which is more adequate for the purpose of translation, i.e. I arrange that structure in such a way that all **functions** of all other languages can interact with that representation. meaning

1.4 Overview of the Thesis

As has already been stated, translation is meaning-preserving, at least in the contexts where MT is currently used. Therefore, the translation process, if modeled for the purposes of multilingual MT, can best be described as a mapping of SSs onto **meaning** and from this **meaning** to the SSs of the TL. **Form** and **meaning** are related via **functions**. Whenever a choice between different SSs is possible, this causes changes in the **meaning**. **Functions** map choices in SSs onto **meaning** (in analysis) and **meaning** to SSs (in synthesis). My central claim is that in a multilingual MT system **functions** cannot be translated into **functions**, which is what most bilingual systems do. The choice of the appropriate **function** in the TL depends on the interaction of the **form** and all constraints on the **functions** of the TL. These constraints cannot be predicted from the **functions** used in the SL.

Chapter 2 is dedicated to the CAT2 system. I shall describe the formalism and the basic design of this multilingual MT system, as all examples of implementation in the following chapters are taken from this system.

Chapter 3 collects some of the translation problems an MT system may encounter. The list of problems is, however, not exhaustive and the translation problems are presented in a descriptive rather than explanatory way. It will be shown that all phenomena underlying these translation problems may interact in a way which cannot be predicted from the SL.

Chapter 4 shows how most MT systems try to solve translation problems. My claim is that any attempt to handle divergences between languages by reference to the form or functions, instead of the meaning, must necessarily fail, as the transfer rules employed for this purpose are local rules which cannot take into

[6] cf. [Wittgenstein89] Lecture B XV.

account the totality of the constraints of the TL. Transfer rules which refer to the structure of a sentence instead of the meaning, may handle isolated phenomena of divergences between languages but cannot account for the full range of constraints and interactions of these phenomena in the TL.

Chapter 5 is dedicated to the units for translation. I shall argue that the correct choice of the size of the units for translation is crucial for translation to succeed and that the distinction of **form** and **function** helps with this choice. Although most MT systems rely on **concepts** as the basic unit for translation, I claim that **concepts** are language-specific segmentations of a meaning continuum and, as a consequence, cannot be directly mapped onto **concepts**, but transfer has to operate at a level below the **concept**, i.e. at the level of the **notional domain**.

Chapter 6 describes the syntactic treatment of function words, the most prototypical instance of a **function**.

In Chapter 7 I give an exemplary analysis of the **meaning** aimed at by one type of function words: the articles. Special attention is given to the independence of **meaning** of the part of speech.

Chapter 8 describes the Argument Structure as one of the most prominent **functions**. This function determines word order, case marking, the prepositional form and the part of speech for the constituents of a phrase. The mutual interaction of theses values with the morpho-syntactic properties of the predicate will be illustrated.

Chapter 9 offers an introduction to the treatment of pronouns in an MT system. Instances where the reconstruction of an unexpressed pronoun is necessary are discussed. The main objective of this chapter, however, is to prepare Chapter 10.

In Chapter 10 I analyse modifier structures as predicative structures where one argument has been realized externally to the syntactic projection of the predicate. Therefore, the internal argument position must be reconstructed by a pronoun. This strategy allows for a free translation of all types of modifiers such as relative clauses, prepositional phrases, adjectives and adverbs into one another according to the requirements of the TL.

Chapter 11 discusses some implementations of lexical functions (henceforth LF). Lexical functions are functions which are not realized through case marking, word order or the like, but through the choice of a specific lexeme. I shall show that the treatment by LFs allows a wide range of translation phenomena to be treated, for which otherwise unmotivated approaches would have to be taken.

A summary followed by a glossary, the bibliography and a subject index conclude the thesis.

Chapter 2

CAT2: Formalism and Design

CAT2 was first developed in 1987 as an alternative prototype to Eurotra*, the large-scale MT project sponsored by the CEU between 1983 and 1992. Within this project, the CEU developed in 1985 a first MT formalism, the so-called <C,A>,T framework (C=Constructors, A=Atoms, T=Translators)[1]. When the first inadequacies of the formalism became apparent, the CEU cancelled the development of this prototype and developed from 1987 on a second formalism, the so-called Engineering Framework[2]. In the meantime, a number of alternative MT prototypes had been developed in various Eurotra research centers in order to continue research into MT, among them the MiMo system[3], CLG[4], E-STAR[5] and the CAT2 system, the latter deriving from the <C,A>,T framework mentioned above.

The CAT2 system can be placed within a number of modern research paradigms, the most global of which is that of unification based NLP[6]. This paradigm may be traced back to the development of the Q-System and the PROLOG programming language by Colmerauer and his colleagues in the early 1970s[7] and has developed a sub-paradigm, that of unification based MT[8]. Secondly, CAT2 belongs to the rule-based branch of MT systems which all share to some extent a bundle of design features. "These include the well-known notions

[1] cf. [des Tombe et al.85], [Arnold et al.86], [Arnold and Sadler87]
[2] cf. [Bech and Nygaard88], [Bech90], [Bech91]
[3] cf. [Arnold and Sadler91], [van Noord et al.91]
[4] cf. [Balari et al.90]
[5] cf. [Allegranza and Soma93]
[6] cf. [Shieber86], [Carpenter et al.91], [Carpenter92]
[7] cf. [Pereira and M.87], [Melby89]
[8] cf. [Rohrer86], [Arnold and Sadler92], [Streiter et al.94]

feature structure

of linguistic rule-writing formalisms with software implemented independently of the linguistic procedures, stratificational analysis and generation, and an intermediate linguistically motivated representation which may or may not involve the direct application of contrastive linguistic knowledge" ([Somers90]). Within this approach CAT2 uses quite a number of modern and sophisticated methods in linguistics and compilations: HPSG-inspired concepts are combined with concepts coming from GB and the Meaning⇔Text Model, using the computational techniques of constraint-based unification and lazy evaluation.

2.1 The Formalism

The function of the formalism is to provide the linguist with a formal tool which functions independently of the hardware on which the system runs and independently of new releases of operating systems and programming languages. If adaptations become necessary, the formalism can be tested on its own without any effects on the lingware.
The second function of the formalism is to provide the linguist with formal means to describe easily those data structures and operations which are needed for linguistic modeling. As most modern linguistic theories are concerned with tree- and feature structures, these are the only data structures of the CAT2 formalism[9]. The operations on the feature- and tree structures are unification and transformations.

Feature structures are unordered sets of attribute-value pairs, represented within curly brackets. In the following example **lex**[10] and **head**[11] are attributes and mann and {**cat=n**} their values[12]. While mann is an atomic value, {**cat=n**} is a complex value itself consisting of a set of attribute-value pairs.

(7)
$$\left\{\begin{array}{l}\text{lex=mann,}\\ \text{head=}\left\{\text{cat=n}\right\}\end{array}\right\}$$

[9] More details about the formalism can be found in a number of publications [Sharp88], [Sharp91], [Sharp and Streiter92], [Haller93], [Sharp and Streiter95], reviews [Alshawi et al.91] and in the CAT2 Reference Manual [Sharp94]

[10] **lex** is the lexical unit of the entry and serves as the search key for the lexicon look up. It must be present in every lexical entry with an atomic value.

[11] **head** contains the semantic and syntactic features which are percolated up to the entry's phrasal projection. Non-head features may change their value at different levels of projection. Head features cannot change their values between the head daughter and the mother node. Our notion of "head" corresponds basically to that of GPSG and HPSG.

[12] The feature and tree structures reproduced here do not correspond exactly to the CAT2 notation. As a notational variant, I have chosen an HPSG like representation, since graphic support for these types of structures is available within LaTeX.

tree structure
constraints
lazy evaluation
constraint concatenation

Tree structures are represented by a mother node followed by square brackets indicating ordered sets of daughter nodes. Every node of a tree is described by a feature structure as shown in the following example:

(8)

$$\{\text{cat=s}\}\left[\begin{array}{l}\left\{\begin{array}{l}\text{cat=n,}\\ \text{lex=fritz}\end{array}\right\},\\ \{\text{cat=vp}\}\left[\begin{array}{l}\left\{\begin{array}{l}\text{cat=v,}\\ \text{lex=lieben}\end{array}\right\},\\ \left\{\begin{array}{l}\text{cat=n,}\\ \text{lex=maria}\end{array}\right\}\end{array}\right]\end{array}\right] \quad ==$$

```
            s
      _____|____
      |        vp
      |     ____|____
      n     v       n
   fritz  lieben  Maria
```

The values of the attributes may be either positive (as in the preceding examples {`attribute=value`}), or constrained by a negation ({`attribute~=value`}) or a disjunction ({`attribute=(value1;value2)`}). The advantages of these functionalities which are now standard in modern unification-based NLP systems are described in [Karttunen84]. Positive, negative and disjunctive constraints may be used in any logically meaningful combination, without any need to distribute positive and negative constraints over the disjunction (cf. [Eisele and Dörre90]). Disjunctive and negative constraints remain unresolved until unambiguous resolution is possible.
Within all rule types, the operator & serves for multiple description of the same value. As shown in the following example, all possible agreement values of the German article *'der'* are captured by the 'concatenation' of positive and disjunctive constraints[13].

(9)

$$\left\{\begin{array}{l}\text{string=der,}\\ \text{head=}\left\{\begin{array}{ll}\text{cat=d,} & \\ \text{ehead=} & \{\text{cat=n,infl=def}\}\\ & \&\ (\{\text{num=plu,case=gen}\}\\ & \quad ;\{\text{num=sing}\}\\ & \quad\ \&\ (\ \{\text{case=(gen;dat),gen=fem}\}\\ & \qquad ;\ \{\text{case=nom,gen=masc}\}))\end{array}\right\}\end{array}\right\}$$

The grammar and the lexicon contain rules in the form of tree structures. These rules interact in a defined way and store the current state of the processing in the form of objects, which equally have the form of tree structures. The feature and tree structures of an object and a rule **unify** when they do not contain any contradictory information, i.e. for none of the attributes in the object is there an attribute in the rule which has a value which is incompatible with that of the object. As a result of this **unification** any additional feature contained

[13] For a discussion of the features **head** (= head features) and **ehead** (= extended head features) see Chapter 6. `string` is the morphological realization of the lexeme of the entry (e.g. string=adoptions); its value must be atomic.

variable binding
b-rules

in the grammar rule is instantiated in the object, as shown in the following example[14].

(10)
$$\begin{array}{l}
\text{object1:}\left\{\begin{array}{l}\text{lex=mann,}\\ \text{head=}\left\{\text{cat=n}\right\}\end{array}\right\}\\
\sqcup\\
\text{rule:}\left\{\text{frame=}\left\{\text{arg1=}\left\{\text{role=nil}\right\}\right\}\right\}\\
\Longrightarrow\\
\text{object2:}\left\{\begin{array}{l}\text{lex=mann,}\\ \text{head=}\left\{\text{cat=n}\right\}\text{,}\\ \text{frame=}\left\{\text{arg1=}\left\{\text{role=nil}\right\}\right\}\end{array}\right\}
\end{array}$$

Following Prolog convention, logical variables beginning with upper-case letters can be used to bind two values within one object (cf. [W.F. and C.S.84]). Once a variable is instantiated by unification, all other variables bound to it become equally instantiated and share all the same constraints. Variables can refer to a feature structure within a node (11), just as to a complete node in a tree structure (12). In (12) the value of the attribute must unify with the adjacent node in the tree. Given such a construct, it is easy, for example, to state subcategorization requirements in one element that must be fulfilled by an adjacent element.

(11)
$$\{\ldots\}\left[\begin{array}{l}\left\{\text{attribute=VAR}\right\}\\ \left\{\text{attribute=VAR}\right\}\end{array}\right]$$

(12)
$$\{\ldots\}\left[\begin{array}{l}\left\{\text{attribute=VAR}\right\}\\ \text{VAR}\end{array}\right]$$

The described type of data structure, i.e. tree structures with annotated feature bundles, are used in five types of rules. These rules are called **b-rules**, **f-rules**, **l-rules**, **t-rules** and **tf-rules**.

b-rules (i.e. building rules) describe the valid tree structures for every level of representation. They correspond to the re-writing rules of generative grammars or the immediate dominance schemata of HPSG with the difference that **b-rules** specify not only the immediate dominance, but also the linear precedence, i.e. the linear order of elements in time or space. The rewriting rule **S -> NP, VP** could be expressed in CAT2 as:

[14] For purposes of presentation I use $\sqcup$ as the unification operator and $\Longrightarrow$ for designating the outcome of this operation.

(13) @rule(b).

f-rules

$$\left\{\text{cat=s}\right\}\left[\begin{array}{l}\left\{\text{cat=np}\right\},\\ \left\{\text{cat=vp}\right\}\end{array}\right]$$

The head-adjunct scheme of HPSG could be expressed in CAT2 as follows:

(14) @rule(b).

$$\{\}\left[\begin{array}{l}\left\{\text{synsem=S}\right\},\\ \left\{\text{synsem=}\left\{\text{loc=}\left\{\text{cat=}\left\{\text{head=}\left\{\text{mod=S}\right\}\right\}\right\}\right\}\right\}\end{array}\right]$$

f-rules (i.e. feature rules) apply to objects constructed by b-rules. An f-rule may assign default values, percolate values from one node to another or filter out an object if it does not meet the well-formedness condition expressed by the rule. If applied to lexical entries, f-rules operate as lexical redundancy rules, which calculate unexpressed information from stated information. E.g. such a rule may be responsible for different case assignments in the active and passive forms of a verb[15]:

(15) @rule(f).

$$\left\{\begin{array}{ll}\text{head=} & \left\{\text{cat=v}\right\},\\ \text{frame=} & (\left\{\begin{array}{l}\text{dia=act,}\\ \text{arg1=}\left\{\text{head=}\left\{\text{ehead=}\left\{\text{case=nom}\right\}\right\}\right\},\\ \text{arg2=}\left\{\text{head=}\left\{\text{ehead=}\left\{\text{case}^{\sim}\text{=nom}\right\}\right\}\right\}\end{array}\right\}\\ & ;\left\{\begin{array}{l}\text{dia=pass,}\\ \text{arg1=}\left\{\text{head=}\left\{\text{ehead=}\left\{\text{pform=by}\right\}\right\}\right\},\\ \text{arg2=}\left\{\text{head=}\left\{\text{ehead=}\left\{\text{case=nom}\right\}\right\}\right\}\end{array}\right\})\end{array}\right\}$$

If the `dia` feature has already been specified as being active (`dia=act`) or passive (`dia=pass`), (15) assigns nominative case to the first or second argument respectively. If both active and passive interpretations are still possible, the

[15] `dia` indicates the language-specific morpho-syntactic diathesis marking. This value is not transferred to the target language and must therefore be recalculated in the target language according to the theme-rheme structure of the sentence, the presence/absence of arguments and the type of the arguments (pronouns vs. nouns). Possible values are `(act;pass;erg)`, i.e. active, passive or ergative.

l-rules
t-rules
tf-rules
implicational constraint

information expressed by this rule is retained as a constraint until the ambiguity is resolved. While **f-rules** apply during runtime, **l-rules** apply when the lexicon is compiled.

t-rules (i.e. transfer rules) are used to relate different levels of representation within a language and between languages. They are responsible for the second type of operations, i.e. transformations. **t-rules** map the tree structure described on one side of the arrow to the tree structure described on the other side of the arrow. **t-rules** may be bidirectional or unidirectional and apply to atoms as well as to complex structures:

(16) a. @rule(t).

$$\left\{\text{lex=car}\right\}\Leftrightarrow\left\{\text{lex=voiture}\right\}$$

b. @rule(t).

$$\left\{\text{lex=cheese}\right\}\Rightarrow\left\{\text{lex=fromage}\right\}$$

c. @rule(t).

$$\{\}\left[\left\{\text{role=funct}\right\},\text{a:}\{\}\right]\Rightarrow \text{a:}\{\}.$$

(16a) translates the lexeme *car* into the lexeme *voiture* and *voiture* into *car*. (16b) is unidirectional and as a consequence every occurrence of the lexeme *cheese* is translated as *fromage* but not the other way round. A possible reason for the existence of such a rule may be that not every occurrence of the lexeme *fromage* can be translated as cheese (e.g. *fromage frais*) (cf. [Jakobson63]). (16c) is a non-atomic t-rule which maps a tree structure consisting of a functional word and its argument onto a structure which no longer contains this functional word. Rules of this type are used to get rid of nodes in a tree structure which should not enter the translation process. The '**a:**' which precedes the empty feature bundle '{}' is called a 'marker'. Such markers are recursive calls to other t-rules by which the marked structure is to be translated. With the help of these markers structures of unlimited complexity can be top-down decomposed, transferred and bottom-up reconstructed.

tf-rules (transfer feature rules) apply simultaneously to two tree structures which are related via a t-rule (i.e. a tree which has been transferred by a t-rule and the result of this transfer). Their main function is to transfer values from the source structure onto the target structure.

(17) @rule(f).

$$\left\{\text{head=}\left\{\text{ehead=}\left\{\text{sem=SEM}\right\}\right\}\right\}\Rightarrow\left\{\text{head=}\left\{\text{ehead=}\left\{\text{sem=SEM}\right\}\right\}\right\}$$

Within **f-rules**, **l-rules** and **tf-rules**, a feature bundle may be divided into two parts by the $\gg$ operator. If the feature bundle unifies with the left side of this operator, it also has to unify with the right side of the operator ≫. Otherwise the object is not well-formed and filtered out. If the content of sem=SEM in (17) does not unify with the right- and the left-hand side, this rule does not apply and, as a consequence, has no effect on the translation. Not so in (18):

monotonicity

(18) @rule(f).

$$\{\}\gg\left\{\text{head}=\left\{\text{ehead}=\left\{\text{sem}=\text{SEM}\right\}\right\}\right\}\Rightarrow \{\}\gg\left\{\text{head}=\left\{\text{ehead}=\left\{\text{sem}=\text{SEM}\right\}\right\}\right\}$$

As every object on the left- and right-hand side unifies with the empty feature bundle {}, they also have to unify with the description in sem=SEM, where the value of SEM on the right-hand side is instantiated with the values of the left-hand side[16].

2.2 Modularity

Within the CAT2 system different language components have been developed, among them German, English, French ([Haller91]), [Maas et al.95]), Dutch ([Streiter90]), Russian ([Iomdin94]), Spanish ([Streiter96]), Arabic ([Pease and Boushaba96]), Chinese ([Streiter96]) and Korean ([Choi95]). In order to maximize the functionality of these resources and to reduce at the same time the complexity of the system and the time needed for development and maintenance, the system is bidirectional and multilingual. This however can best be realized in a modular architecture.

Modularity has proven to be a useful concept in computer programming and the writing of large scale grammars (cf. [Erbach and Uszkoreit90]). A modular architecture offers the possibility of developing a system at different sites, as long as the input and output structures of every module are well defined. It allows for independent specification, testing and compilation of the modules so that the development of the system can be largely reduced to the development of the submodules, each of which is more easy to handle than the whole system. For the user of the system, a high degree of modularization may facilitate the adaptation of the system to personal needs, its integration in other tools and the isolation of modules for different applications.

[16] F-rules and tf-rules which do not use the ≫ operator are instances of the so-called default unification. The default argument (the new information) is only added to the strict argument (the given information) if this information is compatible with that already specified in the strict argument. The default unification is nonmonotonical since the order of application may influence the final outcome and it is non-symmetric as the outcome of the default unification is the strict argument if the default does not apply (cf. [Bouma and Nerbonne94]). F-rules and tf-rules which use the ≫ operator are monotonic and symmetrical if the condition part, i.e. the feature bundle which precedes the ≫ operator is empty.

MPRO
multi word units

Two different views on modularization can be maintained. The static view on modularization concerns the distribution of resources, e.g. which data are put together in one module according to the nature of the data. The dynamic view on modularization concerns the distribution of actions through time, how the treatment of the data can be separated into steps which can be realized by stages.

2.2.1 Dynamic Modularity

The most common approach for modularizing an MT system is stratification. Stratification refers to the fact that analysis and generation is effected along several hierarchically ordered stratal systems which stand in a relation of mutual interaction (cf. [Mel'čuk74]). This approach can be found in various forms and to various degrees in nearly all MT[17] and text generation systems[18]. In most cases the proposed stratification is based on the division into morphology, syntax and semantics, but other partitions are possible and are equally found. The stratificational approach of CAT2 can be illustrated by the following example:

`Der Mann hat Angst vor einem unaussprechbaren Satz`

1 Sentence⇒MS:Morphological analysis is done by an external morphological component called **MPRO*** which analyses words according to inflection, derivation and composition. In this light, note that the lexical value calculated for the adjective *unaussprechbar* has been reduced to its verbal root *sprechen*, plus the modal value `ability` and the value `negation` and the noun *Satz* has been derived from the verb *setzen*.

MS: `d mann haben angst p d sprechen setzen`

2 Between MS and CS the outcome of the morphological analysis may be reshaped in order to meet syntactic and semantic requirements. Multiword expressions like *Vitamin A* can be reduced to one node, parts of sentences can be reduced to a speech act (such as *wie wärs mit* (how about) to the speech act "proposal" and idioms may be reduced to one node so that no internal analysis need take place.

3 Syntactic and semantic analysis is done at level CS. A context free parser transforms the linear structure into a tree structure according to the specifications of the lexicons and the grammars.

[17] The stratificational approach is taken in ARIANE (cf. [Boitet et al.85]), E-STAR (cf. [Allegranza and Soma93]), Eurotra* (cf. [Malnati and Paggio90], [Cencioni91], [Mehrjerdian92], METAL (cf. [Thurmair90]), LOGOS (cf. [Scott89] [Scott92]), NTVECMT (cf. [Chen and Chen92]), CHARON*, ROSETTA (cf. [Landsbergen87], MIMO (cf. [Arnold and Sadler90], [van Noord et al.90], [van Noord et al.91]), ETAP*

[18] cf. [Danlos87], [Patten88].

function words
support verbs

(19)

```
  ________________|___
__|_     ____________|___
d  n     v     _________|___
|  |     |     n   ________|__
|  |     |     |   p ________|_
|  |     |     |   | d    ____|____
|  |     |     |   | |    a       n
d mann haben angst p d sprechen setzen
```

4 The CS representation has to be reshaped in order to facilitate the translation into another language. This transformation is achieved through the application of t-rules and tf-rules. The IS, the target of these transformations is derived by two intermediate levels called T1 and T2. These levels have no theoretical status, since their only purpose is to map more easily the CS structure onto the IS structure and vise versa.

a CS⇒T1
All function words (determiners, case marking prepositions, degree words, auxiliaries etc ...), are removed.

b T1⇒T2
Binary structures are transformed into flat structures. In addition, pronouns are introduced as an internal argument of modifier relations. These pronouns are coindexed with the externally realized argument (e.g. *Satz/setzen*).

c T2⇒IS
Support verbs (e.g. *haben*) and copula verbs are removed and the element bearing the argument structure (e.g. *Angst*) is moved into the position of the copula. This structure is necessary when support verb constructions or copula constructions are to be translated into simple verb constructions or if the TL has a zero copula.

(20)

```
 _____________|______
 n     n         ___|______
 |     |      ___|____    n
 |     |      a      d    |
mann angst sprechen pro setzen
```

5 IS⇒IS
The source IS is translated into the TL by replacing the lexical atoms of the SL by the corresponding lexical atoms of the TL. The choice of the lexical atom and its morphological derivation is constrained through the transfer of semantic information. Since the semantics, but not the part of speech is controlled, the predicative noun *Angst* can be translated into the adjective *afraid* and the deverbal adjective *unaussprechbar* into the

support constructions

verb *pronounce*. The values regarding modality and negation as they are expressed by the type of derivation of the German adjective *unaussprechbar* must then be expressed differently in English as this kind of derivation is not possible in English.

(21)

```
 _______________|_____
 n    a          ____|_______
 |    |        __|_____     n
 |    |        v      d     |
man afraid pronounce pro sentence
```

6a IS⇒T2
Support verbs, copulative verbs and other supporting lexical material such as the 'generic' support (cf. Chapter 11) are inserted into the structure according to the specifications of the lexical items involved. In our example the copula *be* is introduced.

b T2⇒T1
Word order is rearranged and left recursive structures are distinguished from right recursive structures. Elements are moved to wh-landing sides and topic positions.

c T1⇒CS
Flat structures are transformed into binary branching structures and functional categories are generated as necessary. The CS of the TL is the ultimate control instance for these operations, which only confirms only well-formed structures.

(22)

```
                        ___|__________________________
  ______________________|____                      text
__|_     ____________________|___                    |
d  n  ___|__    ________________|__                  |
|  |  v    a    p ________________|_                 |
|  |  |    |    | d    _____________|____            |
|  |  |    |    | |    n      _________|___          |
|  |  |    |    | |    |      d    _______|___       |
|  |  |    |    | |    |      |    v    _____|__     |
|  |  |    |    | |    |      |    |    v      v     |
d man be afraid p d sentence pro cannot be pronounce .
```

7 Between the CS and MS of the target structure, the word structure resulting from the syntax can be reshaped. Thus one node representing a concept may be transformed into a set of words, which is then treated by the morphological component (e.g. $because_of \Rightarrow because\ of$).

8 Morphological generation is done with the help of the external morphological module **mpro***. The surface string is generated from the basic lexeme and the information about compounding, derivation and inflection.

monotonicity

(23)

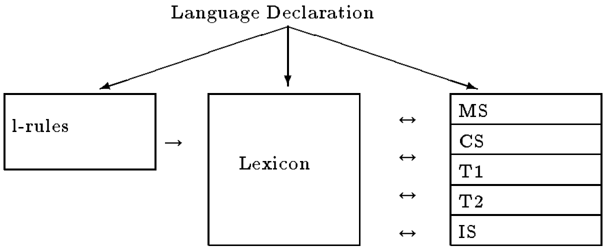
```
MSEN
The man is afraid of a sentence which cannot be pronounced.
```

The advantages of this stratificational approach are the reduced complexity of every submodule and the restriction of certain modules to specific functionalities. One risk with this approach, however, is that some information might not be accessible at a given level though necessary for unambiguous processing. In the face of ambiguities at a given level, stratificational systems produce a number of possible structures which have to be filtered out at higher levels of processing. This approach is not efficient if the points where an ambiguity is introduced and the point where it is resolved are distant (cf. [Mehrjerdian92]). A second problem with this approach is the possible incompatibility at different levels due to the nonmonotonicity introduced by different strata. In order to maintain the advantages of the stratificational approach, thereby reducing to a minimum the risks incurred, a static modularity is introduced in CAT2 which operates orthogonally on the stratification modules.

2.2.2 Static Modularity

In order to assure unambiguous processing at every stratificational level, every type of information should in principle be accessible throughout all levels. As the main data come from the lexicon, the lexicon should be present at every level, supplying the levels even with those types of information which are not typically required at that level. Ambiguities at one level (e.g. different syntactic and semantic properties of the verb *to be*) should be represented so as not to create an overgeneration at a level where these ambiguities are not relevant (e.g. the morphological level). As CAT2 uses one lexicon to which there is access at every level for each type of information, semantic information can already be used at level CS in order to exclude spurious objects, or morpho-syntactic information can be used in generation at an early stage in order to speed up generation. In order to ensure the consistency of the lexicon and the grammar modules, a language declaration system, including a feature declaration system and a macro system are employed.

l-rules
macro defintions



(24)

The feature declaration system, implemented in a program external to the CAT2 formalism, defines possible feature structures, i.e. the possible values of an attribute.
The second component of the Language Declaration is the macro definition which allows the expression of frequent cooccurences of independent features. These macros thus capture linguistic concepts which include more than one feature. The most important macros that will be referred to here are:

(25)

macro	feature definition
ABSTR	abstract~=nil
INFO	abstract={'T'=info}
ACTION	abstract={'T'=action}
EMOT	abstract={'T'=emot}
RELAT	abstract={'T'=relat}
ENTITY	abstract=(nil;{temp=nil})
A-ENTITY	abstract={temp=nil},concr=nil
TIME	abstract={'T'=time,temp=time},concr=nil
LANGUAGE	abstract={'T'=language,temp=nil},concr=nil
EVENT	abstract={temp={aspect~=nil}},concr=nil
CONCR	concr~=nil
INSTR	abstract=nil,concr={'T'=instr,sex=nil}
BODY	abstract=nil,concr={'T'=body,sex=nil}
VEHICLE	abstract=nil,concr={'T'=vehicle,sex=nil}
BUILDING	abstract=nil,concr={'T'=building,sex=nil}
MATERIAL	abstract=nil,concr={'T'=material,sex=nil}
PLANT	abstract=nil,concr={'T'=(plant;material),sex=nil}

SOLID	abstract=nil
	concr={'T'=material,sex=nil,state=solid}

LIQUID	abstract=nil
	concr={'T'=material,sex=nil,state=liquid}

GAS	abstract=nil
	concr={'T'=material,sex=nil,state=gas}

HUMAN	abstract={'T'=info,temp=nil}
	concr={'T'=hum,sex=(male;female)}

COLL	abstract={'T'=info,temp=nil}
	concr={'T'=hum,sex=nil}

S-ENTITY	abstract={'T'=info,temp=nil}
	concr={'T'=instr,sex=nil}

INSTITUTION	abstract={'T'=info,temp=nil}
	concr={'T'=buidinging,sex=nil}

POSSIBLE	modal={'T'=epist,epist=poss}

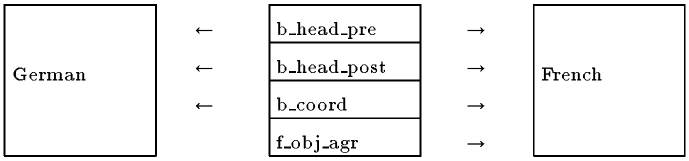
macro	feature definition
OBLIGATION	`modal={'T'=deont,deont=oblig}`
CONTROLLER	`ehead={sem=S,cat=C,num=N,gen=G,per=P,refindex=I}`
CONTROLLEE	`subj={head={CONTROLLER}}`

parametric grammar
common modules
customized grammar

Another type of static modularization is the modularization across languages. Thus, grammar rules may be used by more than one language component. In order to guide the application of such common grammar rules, languages may employ parameters which block or direct these grammar rules. The practical advantages of this approach are of three types: (i) New grammar modules (not the lexicons) are created easily through the combination of 'common' modules. (ii) Rules must be written and modified only once, which augments the consistency of the grammatical information. Improvements within a common module are immediately present in all language modules. (iii) The size of the grammar and the required memory space is considerably reduced, since the common module is loaded into memory only once. This approach allows for a parametrized grammar, where the language-specific grammar is composed of universal rules and language parameters following [Chomsky65], [Chomsky80], [Chomsky81]. Parameters may be clustered together according to typological classifications of languages and the implications one parameter has for other parameters [Greenberg63]. Such a parametrized approach to MT has already been argumented for and implemented by [Dorr90], [Dorr94].

I isolated from the existing grammars a set of about 50 rules which are loaded with the system so that every language has access to every subroutine. They are applied only in those modules in which they are called. Thus, the rule responsible for object agreement is activated only in the French language component, while other rules are activated for more language components, e.g. French and German (26).

(26)



A third type of modularization concerns the adaptation of the system to special user requirements. Such requirements may concern two different aspects: lexical variants and grammatical variants. Different users may prefer different translations for the same expression, e.g. the German word *Kupplung* has to be translated in the context of the automobile industry as English:*clutch* and French:*embrayage*, but in the context of mechanical engineering as English:*coupling* and French:*couplement*.
Grammatical user requirements may refer to the style of speech (e.g. telegram

style) and to special types of syntactic structures which are not part of standard text style or which can be excluded within a given text type in order to increase the speed of processing. Different user requirements are then defined in a specific set of rules (e.g. MUELLER) which have to be activated in order to supplement or replace standard language modules.

(27)

MUELLER		STANDARD GERMAN
terminology	→	lexicon
telegram	→	no-telegram
no-svc	→	svc
		b_head_pre
		b_head_post
		...

function words

Chapter 3

Divergences across Languages

In this chapter I give an overview of some of the problems an MT system may be confronted with. The reasons for the divergences between SL and TL to be listed in this chapter are different in nature and will be discussed only partially in this chapter. The aim of this chapter is (a) to have a representative collection of examples I can refer to in later parts of the thesis and (b) to show that all phenomena underlying these translation problems may interact in a way which cannot be predicted from the SL.

3.1 The Use of Function Words

Functions, as defined in the introduction, may be linguistically realized by function words. Function words do not refer to a concept but mark structures with values. The word *not*, for example, does not refer to a concept comparable to the word *house*, but marks the expression it applies to as 'negated'. The article *a*, to give another example, marks the referential status of a noun, and the auxiliary *is* in *He is called Tim* is a marker for a passive structure. While some languages use similar types of markers, other languages use different means for marking. English and German, for example, use articles to mark the referential status of nouns, while Russian and Chinese have no articles and express the referential status by other means or leave it unexpressed (cf. [Krušel'nickaja61], [Adamec66], [Giusti81]). But even if two languages use the same kinds of markers, they use them differently, i.e. they relate markers with comparable morpho-syntactic properties to different **functions** (cf. [Wandruszka69]). This is exemplified in (28) through the article of different languages, examples

(28f-g) are taken from [Vinay and Darbelnet58], pg.114[1].

(28) a. *German:*
Er ist Lehrer.
he is teacher.
'He is a teacher'

b. Die Industrie ergreift Maßnahmen.
the industry is taking measures
'Industry is taking measures.'

c. *Italian:*
Mangio delle mele.
eat$_1$ of the apples
'I eat apples.'

d. In Italia si mangia la pasta.
in Italy one eats the pasta
'In Italy people eat pasta.'

e. *French*
L'usage du WC est interdit!
the use of the lavatory is forbidden
'Lavatory should not be used!'

f. Il a les yeux bleus.
he has the eyes red
He has red eyes.

g. Aux Etats-Unis l'essence coûte 30 cents le gallon.
in United States the gasoline costs 30 cents the gallon
In the United States gasoline costs 30 cents a gallon.

3.2 Function Words vs. Affixes

As an alternative to the use of function words, i.e. the use of syntactic markers for functions, languages may use morphological markers: Negation may be marked by prefixes (29), determination may be marked in Bulgarian by a determiner added to the first inflected element of the nominal projection (data from [Guentcheva90] and [Walter and Kirjakova90]) and passive may be marked by suffixes as in Russian:

(29) happy - unhappy

[1] For the usage of the article in Arabic cf. [Harder and Schimmel89] pg.23.

analytic marking
synthetic marking
grammaticalization
lexicalization
obligatory opposition

(30) *Bulgarian:*

a. брашно - брашното
brašno - brašnoto
flour - flour$_{DEF}$
'flour - the flour'

b. **новият молив**
novijat moliv
new$_{DEF}$ pencil
'the new pencil'

c. **моят нов молив**
mojat nov moliv
my$_{DEF}$ new pencil
'my new pencil'

(31) *Russian:*
Решение обсуждается парламентом.

Rešenie obsuždaetsja parlamentom.
decision$_{NOM}$ taken$_{PASS}$ parliament$_{INSTR}$

'The decision is taken by the parliament.'

The distinction of these two equivalent types of functional markings has been proposed as far back as 1818 by [Schlegel1818], who called the syntactic marking **analytic** and the morphological marking **synthetic**. Cf. equally [Comrie81] Chapter 3.

3.3 Grammaticalization vs. Lexicalization

Another way to mark semantic oppositions is lexical choice. Thus, while some semantic oppositions are marked by function words or functional affixes, the same effect can be achieved by the choice of a different lexeme. Thus, while Turkish may use an infix to express 'causativization', English may express 'causativization' through a special lexeme (Turkish data and transliteration in (3.3a) from [Simone90], in (3.3b) from [Comrie81], pg.160).

a. *Turkish:*
anla-mak - anla-t-mak
listen - listen$_{CAUS}$

'listen - tell'

transposition

b. öl - öl-dim
die - dieCAUS
'die - kill'

Other examples of lexicalizations come from the degree forming. While languages often use function words or affixes to mark the degree of a word (e.g. *big, bigger, very big, too big, the biggest*), the degree of a word can be expressed as well lexically. According to [Vinay and Darbelnet58], the English verb *sprawl* expresses the high degree of *spread*. In the same way the German adjective *weltberühmt* (world-famous) expresses the high degree of *berühmt* (famous).

While lexicalization represents an optional choice (e.g. between *listen* and *tell*), grammaticalization represent an obligatory choice (cf. [Simone90]), where 'obligatory' means that you must mark the semantic opposition and you cannot leave it unresolved. In the case of the omission of a marker, the default semantic value is assigned, i.e. it is impossible not to express the grammaticalized semantic opposition.

3.4 Transposition of the Part of Speech

The **transposition** of the part of speech, as it is called in translation theory (cf. [Vinay and Darbelnet58], [Podeur93]), is the translational operation which changes the part of speech of a word in order to avoid an otherwise ill-formed word-for-word translation. In (4) on pg.5, I gave an example of syntactic constraints on the part of speech. In this example, a verbal argument must be rephrased as a noun group in the TL if the target matrix verb does not subcategorize for verbal phrases.

General syntactic constraints may equally cause a **transposition**. In Turkish, for example, relative clauses must be expressed with the help of nominalizations. The Turkish translation of the English sentence *I ate the potato Hasan gave to Sinan* can be glossed *I ate the potato of Hasan's giving to Sinan* (cf. [Comrie81], pg.135).

Stylistic considerations can also motivate a paraphrase. Where German frequently makes use of noun phrases to express events in argument position, French prefers verbal constructions. Where French uses prepositional phrases as nominal modifiers, Italian prefers strongly adjectives (examples (33) and (34) come from [Podeur93] pg.43).

(32) a. *German:*
Er schlägt einen Spaziergang vor.
he proposes a walk PREF

chain transposition

b. *French:*
Il propose de faire une promenade.
he proposes to make a walk

(33) a. entreprise de construction
firm of construction

b. impresa edile
firm constructional

(34) a. esprit de compétition
spirit of competition

b. spirito competitivo
spirit competitional

A further constraint may come from the possible usages a lexeme is allowed to acquire in a language. Thus, while the German verb *erpressen* can express the modal value 'possibility' with the help of a verb-to-adjective derivation, the equivalent derivation is not possible in English, so that the German adjective phrase must be translated as an English verb.

(35) *German:*
der erpressbare Mann
the blackmailable man

'the man who can be blackmailed '

der Mann ist erpressbar
the man is blackmailable

'the man can be blackmailed'

3.5 Chain Transposition

Besides the item in question the **transposition** of the part of speech may effect all the other constituents it has to combine with. In such cases we speak of a **chain transposition** (cf. [Podeur93]). If, for example, the concept underlying an adverb finds no adverbial but an adjectival realization in the TL, a nominal paraphrase of the verbal predicate saves the translation (cf. (36), taken from [Chuquet and Paillard89], pg.18). It goes without saying that in such cases prepositions and complementizers must be mutually translatable (37) and articles must be generated without an overt equivalent in the verbal construction (cf. (28e) pg.26).

syntactic functions
argument switching

(36) *French:*

d'une blancheur frappante
of a whiteness striking

'remarkably white'

(37) *French:*

Ne pas ouvrir avant l'arrêt du train!
NEG open before the stop of the train

'Do not open before train stops!'

3.6 Syntactic Functions

A further difference between languages comes from the (different) syntactic functions languages assign to the arguments of a predicate. Syntactic functions, also known as syntactic relations, i.e. subject, direct object, indirect object, agentive, entail syntactic and morphological properties of the arguments and, depending on the language in question, the predicate, so that different syntactic functions entail differences in case assignment, word order, agreement patterns or the choice of functional prepositions. This is exemplified with the 'experiencer' argument of the predicate *cold*, which is subject in French (nominative case, verb agreement) and indirect object in Russian (dative case, no verb agreement).

(38) a. *French:*
J'ai eu froid.
I$_{NOM}$ have$_{1SG}$ had cold

b. *Russian:*
Мне было холодно .
mne bylo xolodno
I$_{DAT}$ cold was$_3$

In cases of so-called argument-switching, the syntactic functions in the SL and TL are exchanged. In (39) the subject of the French sentence becomes the direct object in the English translation and the à-object of the French sentence becomes the subject of the English translation.

(39) *French:*

Jean plaît à Marie.
John pleases to Mary

'Mary likes John.'

pronouns

Equally, different diathesis markings on the predicate may be the cause of different assignments of syntactic functions to the arguments. According to [Comrie81] pg. 75, the English sentence (40a) translates more naturally into (40b) then into (40c), so that the subject of the English sentence becomes the agentive in Russian and the English object becomes the Russian subject.

(40) a. Masha was killed by Tanja.

b. *Russian:*
Машу убила Таня.
Mašu ubila Tanja.
Masha$_{ACC}$ kill$_{ACT}$ TANJA$_{NOM}$

c. **Маша была убита Таней.**
Maša byla ubita Tanej.
Masha$_{NOM}$ was killed$_{PASS}$ Tanja$_{INSTR}$

This different diathesis marking is due to the different functions the diathesis has in both languages. In English diathesis helps to identify the topic of a sentence (cf. [Creider79], [Comrie81]). In Russian it is word order which identifies the topic and the choice of the passive voice is related to written speech and the associated higher speech style (cf. [Xrakovskij72], [Comrie81]).

3.7 Dropping Pronouns

Some languages allow the suppression of pronouns in positions where other languages must realize them. In (41) the subject pronoun is not realized in Italian. If a sentence with a dropped object pronoun undergoes argument switching in transfer, the dropped object pronoun becomes the subject in the TL. This is shown in the French translation of the Chinese example.

(41) a. *Italian:*
Ti amo.
You love$_{1SG}$
'I love you.'

b. *Chinese:*
nǐ xǐ huān zhè dòng fáng zi mà? bù xǐ huān.
you like this piece house QU? NEG like
'Do you like this house? No I don't like it.'

c. *French:*
Est-ce que cette maison te plaît? No, elle ne me plaît pas.
QU this house you like? No, she NEG me pleases NEG

conflation

'Do you like this house? No I don't like it.'

Further examples of suppressed pronouns are English modifier structures. (??) brings examples of two semantically equivalent modifier structures, where the left but not the right variant uses pronouns.

(42) a. the running boy - the boy who is running
b. the boy running home - the boy who is running home

3.8 Conflation

Conflation, as described by [Dorr93], [Dorr94], is characterized by the incorporation of a word which is necessary to reconstruct the **meaning** of the expression. In the example (43) taken from [Dorr93], *puñalada* and *mudarse* have no direct English equivalent[2]:

(43) *Spanish:*
a. Yo le di puñaladas a Juan.
I him gave knife-wounds to John
'I stabbed John.'
b. La barca se mudaba flotando en el agua.
the boat moved floatingly in the water
'The boat floated on the water.'

My suggestion to cope with such divergences is different from that proposed in [Dorr94], as I would not classify these examples as 'conflation'. While [Dorr94] reconstructs the 'missing' elements in English, I propose to merge the two Spanish words *dar puñaladas* into one unit of transfer. The mechanism of this merging is that of **support verb constructions** discussed in Chapter 11. By the same token, *mudar flotando* is reduced to one unit of transfer, i.e. (*flotar*) by the mechanism of **generic support** discussed in the same chapter.
Another pattern of conflation, not mentioned in [Dorr94], is that of many adjectives, e.g. *wooden* or *golden*, which contain a 'nominal' and a 'predicative' part (e.g. the 'nominal' part *bois, wood* and the 'predicative' part *de* (preposition) and *-en* (affix)). These two parts may be conflated in SS and must be independently accessible for translation, in order to allow for the nominal translations.

[2] Note that although [Dorr93] glosses *puñaladas* as 'knife-wounds', *puñaladas* has beside the result reading she needs for her analysis the event reading which, I think, is dominant in the given example.

(44) *French:*

la table de bois/d'or
the table of wood/of gold

'the wooden/golden table'

inflation
default argument

The same pattern of conflation is quite common in languages like Finnish, where prepositions are conflated with the noun. The data in (45) is taken from [Wendt87]).

(45) *Finnish - English:*

a. talotta – without house

b. talona – as house

c. talossa – in the house

d. talosta – out of the house

e. taloon – into the house

f. talolla – on the house

g. talolta – away from the house

3.9 Inflation

By **inflation** I refer to the case where a constituent to which no **meaning** can be assigned appears in SS. In Chinese, for example, there exists a number of verbs which, if realized without an object pronoun, are interpreted as having a covert pronominal object. In order for these verbs to be interpreted as having no covert pronominal object, a default argument has to be inserted, so that the pronominal reading is blocked. In (46c) the default noun *shū* (book) is inserted to block the pronominal reading, creating however an ambiguity between an intransitive and transitive reading.

(46) *Chinese:*

a. Tā kàn zá zhi.
she read journal
'She reads a journal.'

b. Tā kàn.
she read
'She reads it.'

c. Tā kàn shū.
she read book
'She reads a book.' or 'She reads.'

3.10 Interaction of Phenomena

All the phenomena listed above may interact in many ways. The change in the part of speech, for example, may change the type of functional marking between syntactic, morphological or lexical marking. The influence may work the other way round as well: The availability of functional markers may influence the choice of the part of speech, as shown in the example *erpressbar – can be blackmailed* (35), pg.29. By the same token, the presence or absence of arguments may cause a change in the part of speech, as the examples in (47) show.

(47) *French:*

a. Le professeur m'a reproché d'avoir trop mangé.
the teacher me has reproached to have to much eaten
'The teacher reproached me for having eaten too much.'

b. * Le professeur m'a reproché.
the teacher me has reproached

c. Le professeur m'a fait des reproches.
the teacher me has made of the reproaches
'The teacher reproached me.'

The presence of an argument may equally require a 'support structure', as for example a copula verb (cf. (48)). These 'support structures' may in the same way be required by some types of functional markings, as shown in (cf. (49))

(48) a. the proud man
b. * the proud of his son man
c. the man who is proud of his son

(49) a. the proud man
b. * the too proud man
c. the man who is too proud

As the examples of **chain transpositions** show, the properties of one word may influence the structure of the whole sentence. Changes of the diathesis changes the syntactic functions of the arguments and with them their morphological and syntactic properties, while the choice of the diathesis depends on language specific constraints. The interaction between argument switching and the dropping of the pronoun has already been mentioned in (41b-c).

As I shall show in the following chapter, the mechanisms which have been developed to handle divergences between languages by reference to the **form** or the **functions** instead of the **meaning** may handle isolated phenomena of divergences but cannot account for the full range of interactions of these phenomena.

direct approach

Chapter 4

Structures for Translation

In order to handle divergences across languages, most MT systems resort to a mechanism that allows for a change in the structure of the sentence during the process of translation. Throughout the history of MT three approaches have been followed: (i) within the direct approach translation operates on (partial) syntactic structures, (ii) within the interface approach a standardized structure other than the syntactic structure is used for the purpose of translation and (iii) in the lexicalist approach translation operates only on the words appearing in the syntactic structure. My claim is that none of them work as long as the translation operates on **forms** and **functions**.

4.1 The Direct Approach

The direct approach to translation is historically the first MT approach, developed in the so-called computer-phase of MT (cf. [Bátori86]). Such systems, represented by the Georgetown System (cf. [Tucker87]) and the commercial systems SYSTRAN (cf. [Schwanke91]) And LOGOS (cf. [Scott89] [Scott92]), are designed for a specific SL and TL and achieve the translation through successive processing at a number of intermediate levels. Since these systems often do not realize a complete syntactic analysis of the SL sentences, the recognition of **functions** and **meaning** is impossible and therefore not considered for translation. The process of translation is based on similarities between the SSs of the SL and the TL and standard discrepancies, i.e. differences between SL and TL which are regular or highly probable. But even if a complete syntactic analysis were realized, the resulting syntactic structure cannot be regarded as a suitable structure for translation. The syntactic structure contains information which belongs to the **form** of that language, which (i) does not contribute

interface approach
interface structure
complex transfer

to the translation-relevant information and (ii) complicates translation if tree structures have to be transformed in order to match the requirements of the TL. In effect, translation based on syntactic structures proved to be difficult to manage, and researchers turned in a second, so-called linguistic period of MT, to the interface approach.

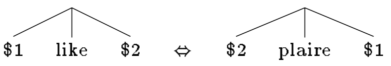
4.2 The Interface Approach

This approach, inspired by linguistic theories such as TG and Fillmore's Case Grammar, transforms the syntactic structure into a more suitable representation which is known as the **interface structure** (IS). This is a normalized structure in which properties of the language related to **form** and **function** are abstracted away. The ISs no longer contain, for example, semantically empty words (e.g. expletive pronouns) nor do they represent the word order of the SL. Conflated structures are reconstructed and pronouns may be inserted at IS if not present at CS.
The final shape of the IS is determined by the linguistic theory employed in the system. Since, with the exception of the Meaning⇔Text framework, no linguistic theory has been developed for the purpose of translation, incompatible structures may be assigned to translationally equivalent sentences. Therefore, most systems have at their disposal a mechanism for so-called complex translation where a structure larger than the basic units of translation is transformed in order to match the requirements of the TL. Thus, while regular structures are translated by lexical translation rules (cf. (50)) and rules which match structures onto isomorphic structures of the TL, divergence in structure is handled by complex translation rules which split the structure to be translated into a regular (convergent) and an irregular (divergent) part (cf. [Arnold and Sadler87]). The irregular part is handled in the complex translation rule while the regular parts (marked by '$') are delegated by a recursive call to other rules, the outcome of which is integrated on the target side of the complex transfer rule. In the following example, (50) provides the translations of *Mary* and *John*, which are integrated into the target structure of (51)[1]. Such rules could be used to treat the phenomenon of argument switching (cf. (39))

(50) Mary ⇔ Marie
John ⇔ Jean

[1] Transfer mechanisms operating on feature structures function similarly to the transfer mechanism operation on tree structures, so I use tree structures to illustrate both approaches. In Alep* for example, the operator '==' is used to trigger the recursive call for the top down decomposition of structures and the bottom-up reconstruction of the target structure (cf. [Theofilidis93]), marked here as '$'.

rule interaction
relaxed compositionality

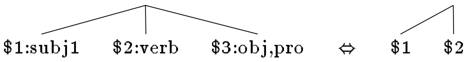
(51)


If, however, two phenomena requiring a complex translation appear within one sentence, two complex rules have to interact. If, for example, in addition to argument switching, the dropping of object pronouns is handled by complex translation rules, (51) and (53) have to interact for the translation of (41c) into (41b), repeated here as (52a-b).

(52) a. *Chinese:*
nǐ xǐ huān zhè dòng fáng zi mà? bù xǐ huān.
you like this piece house QU? NEG like
'Do you like this house? No I don't like it.'

b. *French:*
Est-ce que cette maison te plait? Non, elle ne me plait pas.
QU this house you like? No, she NEG me pleases NEG
'Do you like this house? No I don't like it.'

The rule (51), however, cooperates with other rules only at the level $1 and $2, so that $1 would be instantiated with *elle* and $2 of rule (51) would be instantiated with *me*. As a consequence, rule (53) can apply neither to $1 nor $2, i.e. the two rules cannot interact.

(53)


Attempts have been made to modify translation algorithms such that an interaction of two rules becomes possible (cf. [van Noord et al.91]). Although such 'relaxed' compositional translation algorithms may allow for some interaction of complex translation rules, cases where one word is involved in two complex transfer phenomena still cannot be handled: in such cases only one complex translation rule can apply to this word, thereby blocking the access of the second rule. As example might serve (54), taken from [Trujillo95].

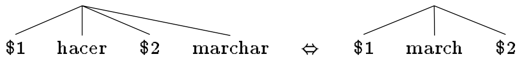
(54) *Spanish:*
Juan le hizo cruzar el valle a los soldados marchando.
John him made cross the valley to the soldiers marching

lexical approach
TAGs
shake-and-bake

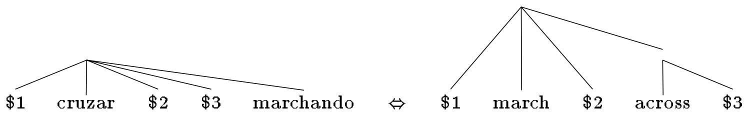
'John marched the soldiers across the valley'

The translation cannot be realized by two rules of the type (55) and (56). If (55) applies first to the English sentence, (56) cannot apply, since *march* is not present in $1 or $2.

(55)



(56)



4.3 The Lexical Approach

In order to overcome such translation problems a translation strategy has been developed which works without any structure. Translation rules operate only on the set of words of the analyzed source sentence. The structure of the TL is determined through the new combinations of the translated lexical items. Instances of the lexical approach are Shake and Bake (cf. [Beaven92], [Brew92], [Whitelock92]) and MT systems based on Tree Adjoining Grammars (TAGs) (cf. [Srinivas et al.94]). In Shake and Bake, the words of the SL are translated into an unstructured set, called 'the bag' of the TL language. The translated words are permuted in the bag of the TL in all possible combinations and parsed with the target grammar. A successful parse in the TL confirms the final translation.

In order to guide the translation and the reconstruction process in the absence of structural information, lexical approaches resort to a system of indices with the help of which the functional or semantic relations between words is expressed. The translation rule in (57), for example, could be used to assign the arguments correctly to the predicate. In those cases where complex relations between words cannot be expressed by indices, the lexical approaches have to use transfer rules which mention more than one lexical item.

Minimal Recursion Semantics

(57) $$\left[\begin{array}{l}\text{pred=like}\\ \text{subj=}\left[\text{index=I1}\right]\\ \text{obj=}\left[\text{index=I2}\right]\end{array}\right] \Rightarrow \left[\begin{array}{l}\text{pred=plaire}\\ \text{subj=}\left[\text{index=I2}\right]\\ \text{obj=}\left[\text{index=I1}\right]\end{array}\right]$$

This system of indices has been refined by [Copestake et al.95], [Copestake95] in the framework of Minimal Recursion Semantics (MRS) in order to reduce the number of cases where complex transfer rules have to be employed. But even with this elaborated system of indices which represent argument relations and quantifier scope, complex transfer rules are necessary whenever structures have to be generated without reference to **meaning**. The interaction of transfer rules which then becomes necessary is not possible. As a consequence, sentences like *John marched the soldiers across the valley* still cannot be treated: Through the application of the second rule of (58) the verb *march* is already eaten up when the third rule of (58) should apply. As a consequence, the translation in (59) fails[2].

(58) $\left[\text{john}\right] \Rightarrow \left[\text{juan}\right]$
$\left[\text{march}\right] \Rightarrow \left[\text{hacer marchar}\right]$
$\left[\text{march across}\right] \Rightarrow \left[\text{cruzar marchando}\right]$
$\left[\text{the}\right] \Rightarrow \left[\text{los}\right]$
$\left[\text{the}\right] \Rightarrow \left[\text{el}\right]$
$\left[\text{the}\right] \Rightarrow \left[\text{la}\right]$
$\left[\text{the}\right] \Rightarrow \left[\text{las}\right]$
$\left[\text{soldiers}\right] \Rightarrow \left[\text{soldados}\right]$
$\left[\text{valley}\right] \Rightarrow \left[\text{valle}\right]$

(59) $$\left[\begin{array}{l}\text{john}\\ \text{march}\\ \text{the}\\ \text{soldiers}\\ \text{across}\\ \text{the}\\ \text{valley}\end{array}\right] \Rightarrow \left[\begin{array}{l}\text{juan,}\\ \text{hacer marchar}\\ \text{los}\\ \text{soldados}\\ \text{?}\end{array}\right]$$

[2] In a personal communication Ralf Steinberger and Pete Whitelock suggested the following solution to this problem. The English sentence representation `Hannibal march AFF-past his soldiers across the Alps` can be transformed into the corresponding French sentence representation `Hannibal avoir AFF-pres faire AFF-pastp traverser AFF-inf les Alpes 'a les soldats en march AFF-presp` with the help of the following rules:
`1) march --> faire + march + AFF-inf + 'a`
`2) AFF-past --> avoir + AFF-pres + AFF-pastp`
`3) across --> traverser + en + AFF-presp`
Besides the undesirable presence of functional affixes in the representation used for translation, the question remains, how `AFF-inf` of rule (1) can be attached to `traverser` of rule (3) and how `AFF-presp` of rule (3) can be attached to `march` which figures in rule (1).

As the examples show, the expressivity of the lexical approach and the IS approach is comparable, with small advantages for the lexical approach. Within the lexical approach, the correct translation of a word can often only be triggered by reference to the context if the **meaning** of a word or its contribution to **meaning** has not been identified. As every time the context is referred to in complex translation rules the words of the context are eaten up, the problems illustrated above persist (cf. [Copestake95]). The suggestions made by [Trujillo95] to use bi-lexical rules may reduce the amount of complex transfer rules which must be written, but does not facilitate the interaction of these rules.

Chapter 5

Units for Translation

For the task of translation, all MT systems have to segment a text and to operate on the resulting units. These units may be paragraphs, sentences, words and morphemes. MT systems, however, differ wrt the question what the units for translation are. Thus, *mudarse flotando*, *dar puñaladas* and *wooden* may each be treated as one or two units for translation [1]. As can be seen from the preceding section however, translation has to start from equally structured **meaning** representations, so that the units of transfer can be mapped directly onto the units of the TL. This can be achieved only if the chosen units of translation are freed from all grammaticalized oppositions. But even if grammaticalized oppositions are analyzed wrt their semantic contribution and removed from the unit of translation, the units chosen may still be too large (e.g. treat *red car* as one unit of transfer), so that systems do not profit from the compositional aspects of language. If the units are too small, the compositional aspect of language is lost and an **over-translation** may result form it[2]. [Dorr93], for example, considers *dar puñaladas* to be two units of transfer and thus has to face two problems: to make reappear the covert *knife-wound* (cf. 43) page 32) and, secondly match *stab* onto *dar* (and not *golpear, herir* (hit, injure) or others).
A second example for an **over-translation** comes from [Kay et al.91] and [Tsujii95]. These authors decompose the German verb *entwerten* into a root *werten* (validate) plus a negation prefix *ent-* (in-) and observe that German

[1] [Vinay and Darbelnet58] pg.16 define the unit of translation as follows: "Le plus petit segment de l'énoncé dont la cohésion des signes est telle qu'ils ne doivent pas être traduits séparement (The smallest part of discourse which has such a cohesion of signs that these signs may not be translated separately).

[2] [Vinay and Darbelnet58] pg.14 define over-translation as follows: "SURTRADUCTION. Vice de traduction qui consiste à voir deux unités de traduction là où il n'y en a qu'une." (Defective translation which sees two units of translation, where in fact, there is only one). As an example of an over-translation these authors give the French expression *aller chercher* which is not *to go and look for*, but *to fetch.*

concepts
content words

uses a negative expression, while English uses a positive expression in order to refer to the same event (60a). As can be seen from (60b-c), however, there is no context in which *werten* and *entwerten* can be exchanged in order to achieve a negative polarity, so that the *ent-* prefix cannot be associated with a negative value. Thus, *entwerten* is one unit of transfer, disregarding the fact that it is morphologically composed.

(60) *German:*
a. ein Ticket entwerten
a ticket in-validate
'punch a ticket'
b. Ein Ticket *werten/entwerten.
a ticket *evaluate/in-validate

c. Den Sprung werten/*entwerten.
the jump allow/*in-validate

5.1 Concepts

Most MT systems use words as the basic unit for transfer. The rationale behind this is that words are prototypical identifiers of **concepts**. Thus, when MT systems use words in transfer, they intend to refer to concepts. **Concepts** are defined as units of knowledge and thought for which linguistic labels are available (cf. [Wüster985], [Picht and Draskau85], [Lethbridge94]). A word (e.g. *house*) is thus directly associated with **meaning**, i.e. the concept denoted by this word. Words that **function** in this way are called **content words**. Under the assumption that cultures employ **concepts** that are sufficiently close, the linguistic realization of a **concept** is generally assumed to be a good unit for transfer, even if the concrete realizations of that **concept** (e.g. house) differ between cultures (cf. [Larson84] pg.99)[3].

Concepts, however, are narrowly linked to the linguistic realizations of the **concept** and these linguistic realizations may include grammaticalized oppositions. Therefore, I claim, that **concepts** are not language independent meaning representations, but language-specific **functions** which realize **meaning**

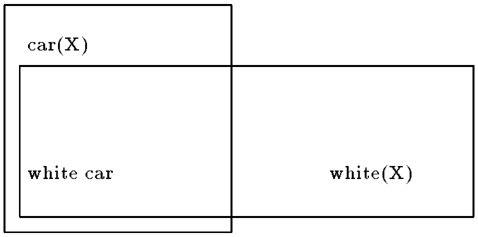
[3] Words and **concepts**, however, cannot be equated, although they often stand in a one-to-one relation. While a **concept** is a unit of **meaning**, word boundaries separate two disjoint sets of **functionally** equivalent operations: morphological operations inside a word and syntactic operations between words. Therefore, words may refer to more than one **concept** if the composition of **concepts** is realized in morphology (e.g. German:*Stahlnagel* (steel-nail)). By the same token, one **concept** can be expressed by more than one word, if the parts necessary for the composition of the **concept** are words instead of morphemes (e.g. *dar puñaladas*, *Japanese people*).

restrictive modification
prototype
notional domain
derivation

according to the possibilities and requirements of a language. In order to understand the **function** of **concepts** we have to look at the relations between **concepts** within a language and between languages.

There exist two major types of relations between **concepts** (cf. [Reed82]). The first type is the hierarchical ordering of **concepts** with the two basic relations of subordination and superordination. A subordinated **concept** is derived from a **concept** by the addition of information (the intension) which restricts the reference (the extension) of the **concept**. This is typically expressed by a restrictive modification (e.g. *the white car*)[4]. Such subordinated **concepts** can be conceived of as a cognitive subspace, inheriting all features from the spaces it is embedded in.

(61)



Ontogenetically prior to the hierarchical forming of **concepts** is the associative **concept** forming, where **concepts** are created through the identification of an entity (before it becomes a **concept**) with a known **concept**, the **prototype** or **center of attraction** (cf. [Wygotski86], [Luria82], [Loon-Vervoon84], [Loon-Vervoon86]). Two **concepts** may be associated due to their similarity in shape, use or their spatial or temporal coincidence. It is this common ground of a set of **concepts** which is called the **notional domain** (cf. [Culioli81], [Culioli90], [Streiter and Schmidt-Wigger95c]). At a later stage of the cognitive development, the new entity is separated and distinguished from the prototype through distinctive features (e.g. *bike* → *motorbike*)[5]. In this way the associated **concepts** have a common semantic part, i.e. the **notional domain** *bike*, and distinctive parts (motor vs. no motor), i.e. the **seme** (cf. [Fages67].
The most productive way for the association of **concepts** is the morphological derivation. This operation starts from a prototypical situation (the **notional**

[4] Restrictive modifications are generally opposed to appositive modifications. The restrictive modifications restrict the reference of the **concept** they apply to. *Cars are dangerous* is a statement about cars. *Cars which drive too fast are dangerous* is a statement about a subset of cars, i.e. all cars which drive too fast, i.e. restrictive modification. *Cars, which drive to fast, are dangerous* incorporates two statement about cars, one stating that *cars drive too fast* and the other that *cars are dangerous*, i.e. appositive modification.

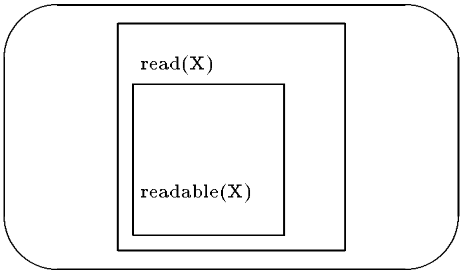
[5] Further psychological and philosophical background to this way of conceptualization can be found in [Wittgenstein84], [Rosch and Mervis75], [Rosch et al.75], [van Parreren and Carpey80], [Mervis and Rosch81].

base
conceptual inclusion
conceptual overlap

domain) and its linguistic label, which is called the **base**. Aspects or parts of the **notional domain**, the situation as such (e.g. *to paint, painting*), to its complement (e.g. *happiness* vs. *unhappiness*) or participants (e.g. *painter, a painting*), results, tools and locations, receive linguistic labels which are derived from the base. This operation of derivation is an economy operation where a high number of **concepts** related via a common **notional domain** is expressed with a **base** plus a limited set derivational morpheme, reducing the amount of **bases** a language needs. **Concepts** related through the operation of derivation may stand in the following relations:

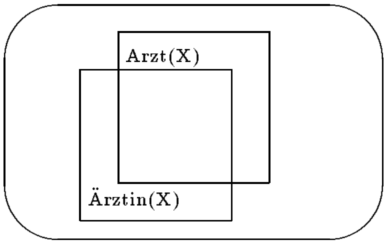
(1) One concept is included in the referential space of the other, representing a lexicalized or grammaticalized opposition. *unhappy*, for example, bears in addition to the reference to the state of happiness a negative value. *happy*, on the other side is not yet marked negatively or positively as it may appear in *He is happy* and *He is not happy*. The same goes for *to read* and *readable*: Both refer to the event of reading. *Readable* bears in addition the modal value 'possibility' by which the verb *to read* can be extended syntactically as in *can be read*. *To paint* and *the painting (of pictures)* refer both to the event of painting. The latter derivation however denotes only a possible subset compared to *to paint* since the temporal actualization* is not possible. The **notional domain** is indicated by the oval.

(62)



(2) Concepts may partially overlap wrt the semantic space they occupy. This can be exemplified with the German *Arzt* (doctor) and *Ärztin* (woman doctor), which differ only wrt gender.

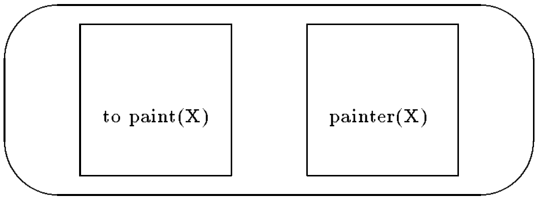
(63)



(3) Two concepts, e.g. *painter* and *to paint*, may be disjunct. Since *painter* can only be paraphrased as *the person who paints*, where a new head and with

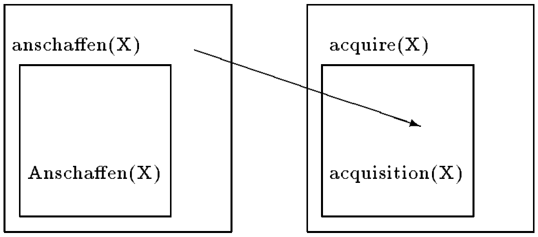
it new semantic properties are introduced, *painter* and *to paint* refer to disjunct **concepts**. These concepts are however still linked through the **notional domain**.

(64)

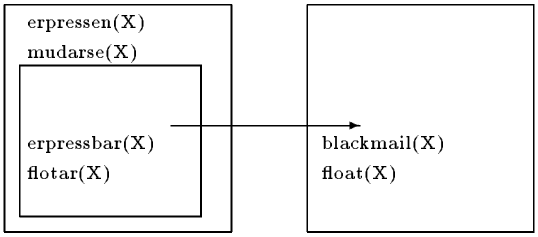


During the process of translation **concepts** of the SL may not be mapped directly onto the **concepts** of the TL. In example (4) page 5 *anschaffen* is mapped onto *acquisition* (cf. 65)), in example (35) page 29 *erpressbar* is translated into *blackmail*, or *flotar* into *float* (cf. 66))[6]. In (67) *Ärztin* and *Arzt* is translated into Russian:*vrač*.

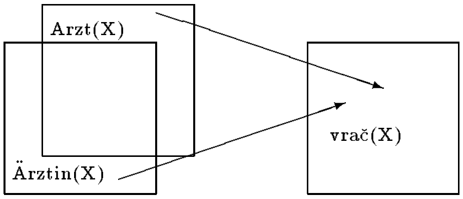
(65)



(66)



(67)



[6] Although the opposition *erpressen - erpressbar* is grammaticalized and the opposition *mudarse - flotar* is lexicalized, both represent relations of inclusion.

term
non-terms
transformation

5.2 Composed Concepts vs. Terms

What distinguishes different linguistic realizations of **concepts** is their degree of conventionalization, i.e. how far the bases of these expressions are composed in a way which is shared by the language community or how far the bases are composed in an ad-hoc manner by the language user. As for an MT system, conventionalized units should be treated as one unit, since the way **meaning** is achieved may be no longer comprehensible within a language (e.g. Italian: *pomodoro* "apple of gold" = tomato) and cannot be reproduced in other languages. These units of language referring to a **concept** in a conventionalized manner are called **terms**. There are no inherent limits wrt the size of **terms**: words, phrases, speech-acts and idioms can be classified as **terms** (e.g. *house, take a decision, how do you do?* and *to bite the dust*).
Concepts which are created creatively in an ad-hoc manner are **non-terms**. They have to be analyzed by reference to the morphological and syntactic means of composition. My claim is that only those composed **concepts** which correspond to the scheme (61) may be **non-terms**. All other composed **concepts** are necessarily **terms**. The inclusion relation of *flotar* and *mudarse*, for example, also known as 'relations of fusion' (cf.[Gross75a]) reflects its status as a *term* through its linguistic behaviour (cf.[Gross75b], [Danlos87]), not allowing for transformations which are meaning preserving. Thus although a *hot stone* may be transformed into *a stone which is hot*, *a hot dog* cannot be transformed into *a dog which is hot*. In the same way some meaning preserving transformations do not apply to *mudarse flotando*.

(68) *Spanish:*

a. El chico andaba cantando = El chico andaba y cantaba
The boy walked singing The boy walked and sang

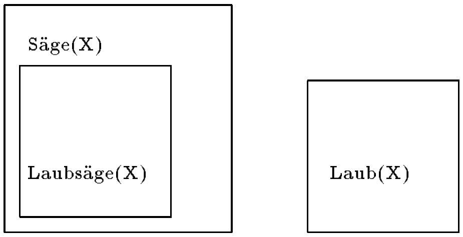
b. La barca se mudaba flotando ≠ La barca se mudaba y
The boat moved floating The boat moved and
flotaba
floated

A commonly accepted criterion for a **non-term** represented in scheme (61) is that the meaning of the composed structure must be more specific than the meaning of the head taken alone. Every composed **non-term** (e.g. *white car*) can be replaced by the head. The only effect which is achieved by this substitution is a loss of information: *He came in his white car* vs. *He came in his car.* If however the content of the information changes (e.g. *Er betrat den Steinbruch/*Bruch* (He entered the "stone-break"=quarry/*break)) the composed structure is a **term**. In this case *Steinbruch* and *Bruch* are disjoint **concepts**.

An additional criterion for a **non-term** is that even if the subordinate **concept** inherits its properties from the superordinate **concept** (e.g. *Laubsäge* 'leaves-saw' = fretsaw), the meaning of the whole must be derived from the meaning of the parts. The meaning of the whole can only be derived from the meaning of the parts if every part functions as a predication of the resulting **non-term**. If not every part can function as a predicate, this predication must be achieved via realized or implied language markers (mostly prepositions) which can denote such predicative relations. In our example, no implied preposition can be found which could link *Laub* and *Säge*. Therefore, the expression *Laubsäge* has to be considered as a **term**.

transfer approach

(69)



5.3 Representing Terms

5.3.1 The Transfer Approach

In transfer-based systems the units for transfer are represented by their monolingual properties, and bilingual rules relate SL and TL. In most transfer-based MT systems (e.g. Eurotra*), the relations between the unit of the SL and the unit of the TL are represented by lexical transfer rules (cf. 70)).

Since bilingual dictionaries are easily available, either in machine readable form or as paper dictionary, transfer dictionaries can be constructed within a reasonable amount of time. Problems with this approach arise only with multilingual MT systems since all language pairs have to be connected by a transfer module. With two language components one (bidirectional) transfer module is necessary, with 3 language components 3 transfer modules are necessary, with 4 languages 6 transfer modules and with 5 languages 10 transfer modules. Thus, the relative ease with which transfer modules are compiled is overshadowed by the fact that this work has to be multiplied.

Due to the erroneous assumption that different parts of speech represent different semantic types, most MT systems employ different lexical transfer rules for the translation of, for example, verbs and nouns (cf. [Schmidt88], [Apresjan et al.89], [Arnold and Sadler90], [Whitelock92]).

interlingua

As for the conditions for the transposition of the part of speech, such information cannot be coded at the level of the lexical transfer rule, since information about the context this word has to be integrated in (e.g. as argument, modifier or predicate) is necessary (cf. examples (4) to (37)). The delegation of the change of part of speech to structural transfer rules, as suggested in [Somers et al.88], is not possible since the governors, arguments and modifiers the item in question has to combine with are subject to the same degree of uncertainty, i.e. it is possible that they themselves must undergo a transposition, depending on their context. As a consequence, the possible change of the part of speech must be foreseen within the traditional transfer approach in all possible combinations by adding the corresponding translation rules to the transfer component.

(70) {lex=acquire}⇔{lex=anschaffen}
{lex=acquisition}⇔{lex=anschaffung}
{lex=acquire}⇔{lex=anschaffung}
{lex=acquisition}⇔{lex=anschaffen}

It goes without saying that such a redundant representation is not desirable for reasons of memory size and man power necessary for the construction of such lexicons. In reality, most systems foresee a limited rank of transfer pairs to cope with the encountered problematic cases, running the risk of never completing the description of all possible paraphrases[7]. In addition, assuming different semantic types for the different parts of speech signifies a renouncement of any semantic control when the part of speech shifts in transfer.

5.3.2 The Interlingual Approach

As an alternative to the transfer based approach, some MT systems adhere to the Interlingual Approach. In such systems the units for translation are freed from all language-specific properties, so that no lexical transfer rules are

[7] This incompleteness can be found in all traditional transfer dictionaries. In the Russian to German transfer lexicon of SUSY, for example, we find lexical transfer rules which translate relational adjectives as adjectives (e.g. 1,3) or as nouns (e.g. 2,5), but in many cases necessary variations are missing (e.g. 4,6).

1. *kvadratnyi* ⇒ *quadratisch*
2. *kvadratnyi* ⇒ *Quadrat-*
3. *gosudarstvennyi* ⇒ *staatlich*
4. MISSING:*gosudarstvennyj* ⇒ *Staats-(schulden)*
5. *policejskij* ⇒ *Polizei-*
6. MISSING: *policejskij* ⇒ *polizeilich(-e Ermittlungen)*

needed[8]. Instead, words are assigned codes in analysis and these codes are associated with words in generation. Thus, no lexical translation rules are needed.[9] Support for the Interlingual approach comes from the integration of knowledge bases in MT systems to support the process of analysis and translation (cf. [Cullingford and Onyshkevych85] [Bátori86]). Since such knowledge systems cannot be created anew for every language, the languages have to provide a uniform access key to the knowledge base. As a consequence, such KBMT systems are realized as interlingual MT systems. Instances of this paradigm are TRANSLATOR (cf. [Nirenburg et al.87]) and ANTHEM*.
Despite the advantages of the interlingual approach, there are a number of drawbacks, due to which this approach in its pure form is used very rarely (cf. [Boitet88] [Arnold and Sadler92]). The most serious problem relates to the difficulties of constructing a conceptual lexicon. As a consequence, interlingual systems are restricted to a fixed and highly specialized domain for which the conceptual lexicon can be constructed within a reasonable length of time (e.g. TRANSLATOR), or for which such a conceptual lexicon has already been built up by experts of that thematic field (e.g. ANTHEM*).

transposition

As for the treatment of transpositions in the context of an interlingual system, there are two possibilities. Either different parts of speech are assigned different codes, in the case of which no transposition is possible, or related words are assigned the same code. In ANTHEM*, for example, French:*os* (bone) and *osseux* (bony) are associated with the same code 'T-11000'. Even morphologically unrelated words may be associated with the same code such as French:*casser* (break) and *fracture* (fracture), i.e. 'M-12000'.

5.3.3 The CAT2 Approach

In CAT2 the interlingual and the transfer based approaches are combined, i.e. lexical units are translated either by bilingual transfer rules or interlingual codes. Bilingual transfer rules are used mainly with verbs, adjectives and adverbs. Interlingual codes are used for terminological nouns, if classifications are available. A description of the interlingual code systems currently used can be found in [Streiter96][10]. For **terms** which are inherently ambiguous (e.g. *das Unternehmen* (= the business/undertaking) lexical transfer rules are preferred since they can help to maintain the ambiguity, translating these words into

[8] Throughout the literature, the term 'interlingua' is used for the description of a number of different aspects of MT systems, e.g. the structure for translation, the units for translation, the extra-linguistic description of subject domains and the domain independent linguistic abstractions, cf. [Landsbergen87] [Tsujii93], [Hwee-Boon93]. Within this thesis I use the term interlingua only with respect to the status of the units of translations.

[9] Attempts to use natural languages as interlingua are described in [Schubert88]) [Slocum89] and [Guzmán de Rojas88].

[10] Lexical transfer in CAT2 is based on the feature `lex`, while interlingual units are transferred using only one general rule, based on the feature `slex` (secondary lexeme).

concept generation

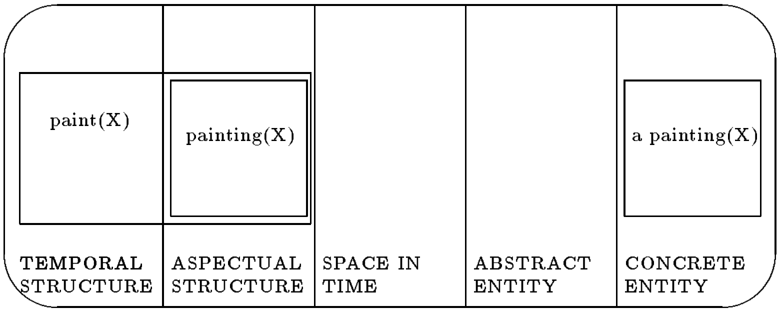
words with the same ambiguities (e.g. *enterprise*, fr:*entreprise*) (cf. [Prahl94]). Within these two approaches however, **concepts** are not mapped onto **concepts**. As the concept is a function which selects within a **notional domain** a subspace, the **concept** has to be reconstructed in the TL from the application of a language independent semantic classification on the **notional domain** of the TL. The **notional domain** is represented in the lexicon as the set of **concepts** associated through the base, each identified by its morpho-syntactic and semantic values (cf. (124) page 82).

The CAT2 transfer rules thus mention only the **notional domain**, represented by the **base**, e.g. *acquire* and the semantic content is transferred via the tf-rule (18) page 17, which then selects the possible **concepts** in the **notional domain** of the TL. With transfer rules only morphologically related words may be grouped in one domain, while with interlingual identifiers of **notional domains**, morphologically unrelated words (e.g. *to drive* and *chauffeur*) may be grouped in one **notional domain**.

(71) {lex=acquire}⇔{lex=anschaffen}
{lex=paint}⇔{lex=malen}

The semantic classification that accounts for the differences between the **concepts** of one **notional domain**, refers to all grammaticalized semantic values, e.g. modal values, negation, gender, the type of entity and the degree of actualization*. One dimension of this semantic classification and its discrimination of **concepts** in one **notional domain** is illustrated in (72).

(72)



The basic distinction in this semantic classification is that between events and entities (cf. [Zelinsky-Wibbelt86], [Nirenburg87]). Events (including states) are characterized by their aspectual structure, i.e. the internal structuring of the denoted time by reference to different phases of the event or state (cf. [Grimshaw90]), even if they are almost limited to a particular point in time (cf. [Comrie85]). The importance of the opposition between entities and events can be seen from the following examples. In (73a) the German verb can be translated into a noun if it is of the type event. In (73b), however, the translation is not possible, since the target noun refers to an entity.

event
entity
appositive modification

(73) *German:*

a. Er schlug vor, spazieren zu gehen.
he proposed PREF, walk to go
'He proposed a walk'

b. Er schlug vor, die Regierung zu beeinflussen.
he proposed PREF, the government to influence.
'He proposed to influence the government.'

* 'He proposed the influence of the government.'

The aspectual structure is further specified by the values for modality, Aktionsart, aspect and the tense value which distinguishes actualized* from non-actualized* events, and if actualized in what relation the events stand to the speech act (cf. [Reichenbach47]). Therefore, the translations in (74) are well formed and excluded only as translations in an actualized* context. These translations would be perfect in a title or caption under a cartoon.

(74) *German:*

Ein neues Auto wurde angeschafft.
a new car was acquired

' + The acquisition of a new car.'

'A new car has been acquired.'

In subordinate clauses, morpho-syntactic tense values on verbs do not have the same **function** as morpho-syntactic tense values in the main clause. The tense values do not relate the denoted event to the speech act but to the event in the main clause. As in this context no actualization* takes place, a nominal can replace the verb of the TL without any harm to the translation (cf. example (4) page 5). For modifiers we use the distinction of actualized* vs. non-actualized* events for the distinction of restrictive modifiers and appositive modifiers. Restrictive modifiers are non-actualized* and appositive modifiers are actualized*. This fact is expressed in TG by the derivation of an appositive relative clause from a coordinated phrase at the level of the matrix clause (cf. [Lujan80]). As example might serve the appositive relative clause in (75a) which is derived from (75b).

(75) a. All men, who are gentle, work hard.

b. All men are gentle and work hard

The different subtypes of entities will be discussed in Chapter 8.

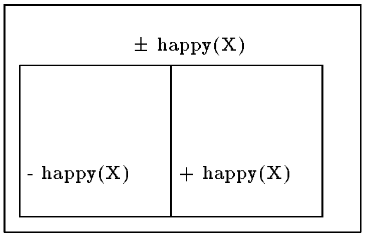
obligatory oppositions

Chapter 6

F–Structures

Function words are prototypical realizations of **functions**. They assign functional value to content words in the same way as functional affixes do, realizing obligatory grammatical oppositions, i.e. the semantic space denoted by the term is obligatorily segmented into one or another subspace:

(76)



In this chapter I shall concentrate on how function words and content words combine in syntax and semantics and on how the functional values of the unmarked variants are calculated. The matching of particular function words to semantic values will be exemplified in Chapter 7 dedicated to articles.

6.1 Functional Head–Structure

Function words and the structures they apply to form more complex syntactic structures of the type 'article + noun phrase', 'auxiliary + verb phrase' and 'complementizer + sentence'. These structures, called functional head-structures (F-Structure) are built up at level CS by a common **b-rule** reproduced in (77). The basic properties of this structure are listed from (a) to

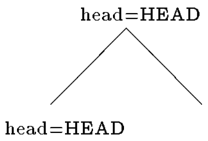
head features
complement
function words

(e).

(77) @rule(b).

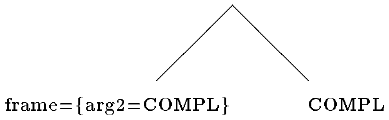
$$\left\{\begin{array}{l}\text{head=HEAD,}\\ \text{role=ROLE,}\\ \text{frame=FRAME}\end{array}\right\}. \left[\begin{array}{l} \left\{\begin{array}{l}\text{hpos=left,}\\ \text{role=funct,}\\ \text{frame=}\left\{\text{arg2=}\left\{\text{head=}\left\{\text{ehead=EH}\right\}\right\}\text{\&COMPL}\right\},\\ \text{head=}\left\{\text{ehead=EH}\right\}\text{\&HEAD}\end{array}\right\}, \\ \left\{\begin{array}{l}\text{role=ROLE,}\\ \text{frame=FRAME,}\\ \text{head=}\left\{\text{max=no}\right\}\end{array}\right\}\text{\&COMPL}\end{array}\right].$$

(a) The function word (**role=funct**) is assumed to be the head of the structure (cf. [Löbel90]). The head of a structure is the daughter which determines the syntactic properties of the mother node. These syntactic properties are described by the head features. Percolation of the head features is then realized through the variable binding **head=HEAD**. This is illustrated in (78). As the function word is the head of this structure, I will refer to the function word within this structure as the functional head.



(78)

(b) Functional heads select structures as their complements through the **COMPL** variable in the **frame** feature[1]. Functional heads are determiners and semantically empty prepositions and complementizers, prefixes, punctuation marks, degree words and coordinators. With the exception of coordinators which select two complements, all other function words select one complement. Coordination is therefore treated as a separate structure (cf. [Streiter96]).

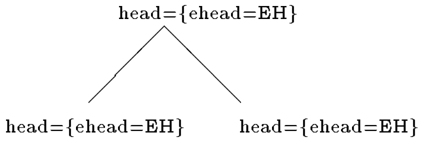


(79)

[1] For the sake of clarity I use the term **complement** for any structure selected in SS. The term **argument** refers to a structure which is assigned a thematic role.

extended head features
complement selection

(c) The selection of the complement of the functional head is restricted by the extended head feature principle. According to this principle formulated by [Grimshaw91], a subset of the head features of the function word and a subset of the head features of the complement have to unify. The subset of the head features which is shared by the function word and its complement is called the extended head, since the extended head is related to the 'extended' projection, i.e. the projection of the lexical head extended by the projection of its function word. Through the unification of the extended head features, syntactic and semantic information is shared between the functional head and its complement. The complement selection may be equally guided by functional values as by semantic values since both values have to unify in the extended head.



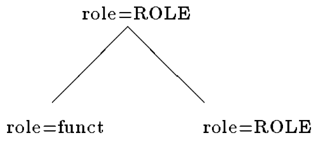
(80)

In addition to this principle, every function word may further specify syntactic restrictions on its complement, e.g. a determiner selects a nominal head and a complementizer selects a verbal head. This is illustrated in a simplified lexical entry where a preposition lists the possible structures it may select.

(81) @rule(b).

$$\left\{\begin{array}{l} \text{cat=p,} \\ \text{frame=}\left\{\text{arg2=}\left\{\begin{array}{ll}\text{head=} & (\{\text{cat=n}\} \\ & ;\{\text{cat=d}\})\end{array}\right\}\right\}\end{array}\right\}$$

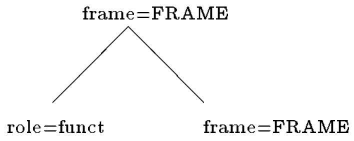
(d) The thematic role (cf. Chapter 8) is projected from the complement to the mother node. Thus, the projection of the determiner has the same thematic role as the projection of the noun the determiner subcategorizes for.



(82)

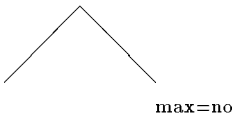
(e) The subcategorization frame of the complement (cf. Chapter 8) is percolated onto the mother node. This implies, for example, that auxiliary functional completeness

verbs do not have to be analysed in terms of raising structures, but once the auxiliaries have selected the main verb, the resulting structure inherits the subcategorization properties from the main verb and can bind the complements of the main verb.



(83)

(f) The subcategorized structure may or may not be a maximal projection. Maximal projections are marked max=yes, not maximal projections are marked max=no.



(84)

In the following two sections I give some more details on (i) how functional completeness can be controlled and (ii) how function words are treated after level CS.

6.2 Functional Completeness

Functional Completeness refers to the completeness of functional projections[2]. In some cases functional markers are obligatory for semantic or syntactic reasons. In (85a) the functional marker *on* has not been realized, so that the resulting structure is functionally incomplete. In the (85b) a determiner is missing in order for this structure to be complete.

(85) a. *It depends you.

b. *I saw animal.

As the realization of the functional markers is marked and controlled via the functional values percolated in the extended head, the control of functional completeness comes down to the puzzle of how a structure can be compatible

[2] Note, the term 'Functional Completeness' is used with a different meaning in LFG (cf. [Bresnan82b]).

default alternation

with a functional value, (e.g. how can *you* be compatible with the functional value pform=on), but on the other hand not have access to a slot for which this functional value is required, as long as no special marker has been realized.

In order to solve this problem, I associate with every part of speech a so-called default alternation (cf. [Streiter94]) which describes those functional values that have to be assigned to the projection if this structure becomes a phrasal (i.e. syntactic) non-head, i.e. a maximal projection without a functional head. This default alternation is centered around the disjunct ({max=yes};{max=no}): The max value of any projection must be max=yes[3] in order to allow this projection to be a non-head of a phrasal structure, i.e. to become an argument or modifier of a content word. The completeness requirement can be found in the **b-rules** for A-Structures (134) page 88 and M-Structures (156) page 105 in the following chapters. The {max=yes} part of the disjunction is combined with the default values as illustrated in (86). The first feature bundle of (86) describes the default alternation for nouns, saying that a bare noun is either not a maximal projection, or if it is a maximal projection it has to receive the functional values pform=nil, type=abs, wh=no and neg=no. If a noun becomes the non-head of a phrasal structure, in which case max=yes is instantiated, the values for determination, the prepositional form, the wh feature and negation are set to their default values.

(86)

$$\left\{\begin{array}{ll} \text{head}= & \{\text{cat=n}\}\& \\ & (\,\{\text{max=no}\} \\ & ;\left\{\begin{array}{l}\text{max=yes}\\ \text{ehead}=\left\{\begin{array}{l}\text{pform=nil,}\\ \text{type=abs,}\\ \text{wh=no,}\\ \text{neg=no}\end{array}\right\}\end{array}\right\}) \end{array}\right\} \sqcup \left\{\text{head}=\{\text{max=yes}\}\right\} \Longrightarrow$$

$$\left\{\text{head}=\left\{\begin{array}{l}\text{cat=n,}\\ \text{max=yes}\\ \text{ehead}=\left\{\begin{array}{l}\text{pform=nil,}\\ \text{type=abs,}\\ \text{wh=no,}\\ \text{neg=no}\end{array}\right\}\end{array}\right\}\right\}$$

If the noun is selected by a determiner, the extended head features of the function word and its complement unify. Only the disjunct max=no on the noun is retained, since the max=yes disjunct contains extended head features which are not compatible with the functional head. This is illustrated in (87).

[3] [Netter94] in a similar approach uses the feature FCOMPL +/- in order to indicate the functional completeness of the projection.

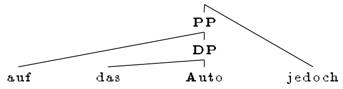
(87)
$$\left\{\begin{array}{ll} \text{head}= & \{\text{cat=n}\}\& \\ & (\{\text{max=no}\} \\ & ;\left\{\begin{array}{l}\text{max=yes}\\ \text{ehead}=\left\{\begin{array}{l}\text{pform=nil,}\\ \text{type=abs,}\\ \text{wh=no,}\\ \text{neg=no}\end{array}\right\}\end{array}\right\}) \end{array}\right\} \sqcup \left\{\text{head}=\left\{\text{ehead}=\left\{\text{type=def}\right\}\right\}\right\} \Longrightarrow$$
$$\left\{\text{head}=\left\{\begin{array}{l}\text{cat=n,}\\ \text{max=no,}\\ \text{ehead}=\{\text{type=def}\}\end{array}\right\}\right\}$$

After the determiner has selected the noun, the determiner is responsible for the control of the following functional projections. This is realized by the default alternation on the determiner, requiring `pform=nil` to be instantiated if this structure becomes a phrasal non-head (88)[4].

(88)
$$\{\text{head=HEAD}\}\left[\begin{array}{l} \left\{\begin{array}{ll}\text{lex=der,} & \\ \text{head}= & \left\{\begin{array}{l}\text{cat=d,}\\ \text{ehead}=\{\text{type=def}\}\end{array}\right\}\& \\ & (\{\text{max=no}\} \\ & ;\left\{\begin{array}{l}\text{max=yes}\\ \text{ehead}=\{\text{pform=nil}\}\end{array}\right\})\& \\ & \text{HEAD}\end{array}\right\} \\ \left\{\begin{array}{l}\text{lex=mann,}\\ \text{head}=\left\{\begin{array}{l}\text{cat=n,}\\ \text{max=no,}\\ \text{ehead}=\{\text{type=def}\}\end{array}\right\}\end{array}\right\}\end{array}\right]$$

The mechanism for the assignment of default values described above has to be completed by a second mechanism. Since every functional head solves the default alternation of its complement with the `max=no` disjuncts, this may have undesired results if one intermediate function word has not been realized, as in the case where a preposition directly selects a noun. In this case, the noun loses its default alternation, but no value for 'determination' is instantiated as was the case in (87). This structure would thus be compatible with any value for 'determination' (`type`).

[4] [Netter94] in this case simply assumes that the DP is functionally complete, which of course is not true if semantically empty prepositions or topic makers have to be added (e.g. *Auf das Auto jedoch war er nicht stolz* vs. **Das Auto jedoch war er nicht stolz*).



(89)
$$\left\{\begin{array}{ll}\text{head}= & \left\{\text{cat=n}\right\}\& \\ & (\left\{\text{max=no}\right\} \\ & ;\left\{\begin{array}{l}\text{max=yes} \\ \text{ehead}=\left\{\begin{array}{l}\text{pform=nil,} \\ \text{type=abs,} \\ \text{wh=no,} \\ \text{neg=no}\end{array}\right\}\end{array}\right\})\end{array}\right\} \sqcup \left\{\text{head}=\left\{\text{ehead}=\left\{\text{pform=von}\right\}\right\}\right\} \Longrightarrow$$
$$\left\{\text{head}=\left\{\begin{array}{l}\text{cat=n,} \\ \text{max=no,} \\ \text{ehead}=\left\{\text{pform=von}\right\}\end{array}\right\}\right\}$$

In order not to have such undesirable results, we add functional specifications to the selectional restrictions in (81). Thus, not only the part of speech of the possible complements are listed, but together with them the functional values which have to be assigned. Thus the German preposition *an* (*an das Auto, an ihn, *an Auto*) is specified as in (90), while the German preposition *am* (** am das Auto, *am ihn, am Auto*) is specified as in (91).

(90) @rule(b).
$$\left\{\begin{array}{l}\text{string=an,cat=p,} \\ \text{frame}=\left\{\text{arg2}=\left\{\begin{array}{ll}\text{head}= & (\left\{\begin{array}{l}\text{cat=n,} \\ \text{ehead}=\left\{\text{type=abs}\right\}\end{array}\right\} \\ & ;\left\{\text{cat=d}\right\})\end{array}\right\}\right\}\end{array}\right\}$$

(91) @rule(b).
$$\left\{\begin{array}{l}\text{string=am,cat=p,} \\ \text{frame}=\left\{\text{arg2}=\left\{\text{head}=\left\{\begin{array}{l}\text{cat=n,} \\ \text{ehead}=\left\{\text{type=def}\right\}\end{array}\right\}\right\}\right\}\end{array}\right\}$$

The presence of certain functional markers may be required by the content word. Some nouns, for example, (e.g. French:*Le Japon* (Japan), English: *The Netherlands*) must have an article; this information is entered in the lexical entry with the corresponding functional value `type=def`, which is compatible only with the `max=no` disjunct of the default alternation. This is illustrated in (92) where the unification of the lexical entry with the default alternation of nouns results in a violation of the completeness requirement, due to which this structure, i.e. *Japon*, cannot become a phrasal non-head.

(92)
$$\left\{\begin{array}{l}\text{lex=japon,}\\ \text{head=}\left\{\begin{array}{l}\text{cat=n,}\\ \text{ehead=}\left\{\text{type=def}\right\}\end{array}\right\}\end{array}\right\} \sqcup \left\{\begin{array}{ll}\text{head=} & (\left\{\text{max=no}\right\}\\ & ;\left\{\begin{array}{l}\text{max=yes}\\ \text{ehead=}\left\{\begin{array}{l}\text{pform=nil}\\ \text{type=abs}\end{array}\right\}\end{array}\right\})\end{array}\right\} \Longrightarrow$$

$$\left\{\begin{array}{l}\text{lex=japon}\\ \text{head=}\left\{\begin{array}{l}\text{cat=n,max=no}\\ \text{ehead=}\left\{\text{cat=n,type=def}\right\}\end{array}\right\}\end{array}\right\}$$

The requirement of having an article is also valid for singular count nouns except when they appear in a telegraphic speech style (`style=tele`) (e.g. *Problem with motor.*) or if the count noun functions as predicate (`role=pred`) or quasi predicate (`role=class`) (e.g. *We nominated him president*). In (93) the default alternation for German nouns is reproduced in a complete way.

(93) @rule(l).
$$\left\{\begin{array}{ll}\text{head=} & \left\{\text{cat=n}\right\}\&\\
& (\left\{\text{max=no}\right\}\\
& ;\left\{\begin{array}{l}\text{max=yes}\\ \text{ehead=}\left\{\begin{array}{l}\text{pform=nil,}\\ \text{type=abs,}\\ \text{wh=no,}\\ \text{neg=no}\end{array}\right\}\end{array}\right\}\&\\
& (\left\{\text{ehead=}\left\{\text{type}^{\sim}\text{abs}\right\}\right\}\\
& ;\left\{\text{ehead=}\left\{\text{type=abs}\right\}\right\}\&\\
& \qquad\qquad (\left\{\text{bound=mass}\right\}\\
& \qquad\qquad ;\left\{\text{bound=count}\right\}\&\\
& \qquad\qquad\quad (\left\{\text{ehead=}\left\{\text{num=plu}\right\}\right\}\\
& \qquad\qquad\quad ;\left\{\text{ehead=}\left\{\text{num=sing}\right\}\right\}\&\\
& \qquad\qquad\qquad (\left\{\text{style=tele}\right\}\\
& \qquad\qquad\qquad ;\left\{\begin{array}{l}\text{style}^{\sim}\text{=tele}\\ \text{role=(class;pred)}\end{array}\right\})))))
\end{array}\right\}$$

6.3 Transfer of Functional Categories

Function words are one possible surface realization of **functions**. As with all **functions**, they are analyzed in SL with respect to their contribution to **meaning** and regenerated as necessary in the TL. The **function** and the functional values are however not transferred, the functional marking may be realized in other contexts or languages by morphological, syntactic or lexical means.

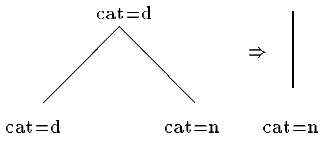
Function words are therefore not presented structurally at IS. Between CS and T1, they are elided by a common rule (94). The functional and semantic values, however, which were added by these elements, (e.g. pform, type, sem, ref etc ...) are not lost, since they are shared with the lexical head in the extended head features.

(94) @rule(t).

$$\{\}\begin{bmatrix} & \left\{\text{role=funct}\right\} \\ \text{arg:} & \{\} \end{bmatrix} \Rightarrow \text{arg:}\{\}$$

This transfers rule effects the transformations represented in (95).

(95)



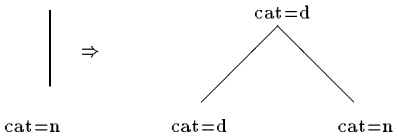
This strategy, that is to refer to the functional content accumulated in the extended head and not to the functional head itself, allows for a simple IS⇔IS component. Lexical heads are translated and the language and category independent semantic values derived from the functional values are transferred to the lexical head they belong to. The lexical head of the target language can then decide how its semantic value has to be realized. In generation, function words may again be introduced into the structure according to the extended head feature specifications by a set of generation rules, each one responsible for the generation of one specific functional head. I reproduce here the rule which generates determiners (lex=d) according to the specifications found in the extended head (ehead=EH). The right side unifies with the T1-structure and generates the structure on the left side at level CS.

(96) @rule(t).

$$\text{arg:}\left\{\text{head=}\left\{\begin{array}{l}\text{cat=n,}\\ \text{ehead=EH}\end{array}\right\}\right\} \Rightarrow \{\}\begin{bmatrix} & \left\{\begin{array}{l}\text{role=funct,}\\ \text{lex=d,}\\ \text{head=}\left\{\text{ehead=EH}\right\}\end{array}\right\} \\ \text{arg:} & \{\} \end{bmatrix}$$

This transfers rule effects the transformations represented in (97).

(97)



In the following chapter we will have a closer look at one class of function words.

neutralization

Chapter 7

Determination

Determination is a **function** associated typically with nominal projections. In this chapter I have a closer look at this **function** in order to illustrate some points of my argumentation. (1) Different languages use different types of marking systems in order to mark the determination and if they use similar systems, they may use the system in a different way: The determination of a noun can be marked by a noun prefix as in Bemba (cf. [Givón78]), a noun suffix as in Bulgarian and Danish, a noun prefix as in Arabic (cf. [Harder and Schimmel89]), a determiner as in Dutch, English, German, Italian, French and Spanish, an adjective suffix as in Bulgarian, by contraction of prepositions and determiners as in French, Italian, Spanish, German, the pre- or post-position of nouns with respect to the main verb for languages with a relatively free word order (e.g. Latin, Russian cf. [Birkenmaier79]) and the association of the subject function with 'definiteness' in languages with fewer possibilities of changing the word order (e.g. Bemba cf. [Givón78] and Chinese cf. [Van den Berg89]). Examples of different usages of determiners have already been given in Chapter (3).

(2) Neutralizations of the functions of determiners are easy to observe with proper nouns and idiomatic expressions. Thus, examples similar to the following can be discarded from the analysis of the **meaning** of determination (examples (98) from [Schwarze88] and (99) from [Booth and Gerritzen89]).

(98) *Italian:*

Il Cairo -	L'Aia -	La Mecca -	L'Avana
Cairo	The Hague	Mecca	Havanna

(99) a. to see **the** elephant

b. to drive pigs to market

quantification
referentiality

c. to raise a dust

(3) The **meaning** associated with determiners is applicable to parts of speech other than to nouns. The cross-categorical usage of the **meaning** allows to describe phenomena of translation which otherwise remain untractable. First, however, I give a short description of how determiners contribute to **meaning**.

7.1 Multiple Dimensions of Meaning

In those cases where the determiner fulfills a function, this function does not necessarily aim at only dimension of meaning. Languages tend to express by one marker semantic values that normally cooccur, running the risk that unlikely combinations of semantic values cannot be expressed (cf. [Givón78]).
The German indefinite article, for example, cannot be reduced to one semantic value. When opposed to the zero-article, the indefinite article introduces a singular quantification. While German: *Eis* refers to a not quantized (cumulative) amount of ice, *ein Eis* refers to one unit of ice. The historical origin of the indefinite article from the cardinal 'one' and its morpho-syntactic similarities underline the semantic relatedness of both operators (e.g. Dutch: *één* - *een*, Italian: *un* - *un*, German: *ein* - *ein* cf. [Bosco Coletsos88]). Languages without articles such as Chinese and Russian may resort to their cardinal to express indefiniteness when it has to be marked explicitly (e.g. [Birkenmaier79], [Van den Berg89]).

When opposed to the definite article, however, the indefinite article expresses the unfamiliarity of the hearer with the denoted entity (cf. [Zelinsky-Wibbelt91], [Francis et al.95]). As a consequence, two semantic classifications have to be associated with the indefinite article, that of quantification (cumululative vs. quantized) and that of "knownness" (unfam vs. fam). According to this model it is impossible to express a cumulative but known concept. The necessity to do this, however, is seldom felt since something not quantized can hardly be known to someone, except that the reference type changes and the concept itself is referred to instead of instances of the concept. In this case the determination is non-referential.

(100)

a/the child/children	*children*	*a child*	*the child*	*the children*
water	*water*	*a water*	*the water*	
non-referential	referential			
	cumulative	quantized		
		X=1		X>1
		unfam	fam	

7.2 Reference Type

reference type

Determination is used to express the reference type of the noun, that is the relation between discourse entities (e.g. the word *elephant*) and the conceptual classification system which is supposed to be identical for the sender and the receiver of a message: Either the discourse entity refers to instances of the concept (e.g. *There is an elephant in your bath*) or to the concept itself (e.g. *The elephant lives in Africa and Asia*) or to other discourse entities in order to build predications which assign discourse entities to concepts (e.g. *My favorite toy is an elephant*). This principled opposition is described by the referential vs. non-referential opposition in (101) where *'T'* represents the 'type' of the feature bundle which is going to be further specified.

(101)

There is an elephant ...	`ref={'T'=ref,...}`
The elephant lives in ...	`ref={'T'=nonref,...}`

If the sender presents a discourse entity only as an instance of a concept (e.g. *There is an elephant in your bath*), all the receiver knows about the discourse entity is its affiliation to the concept 'elephant'. The discourse entity is said to be unknown to the hearer. This first extraction of the discourse entity from the concept is prototypically marked by the indefinite article.
If, however, the extracted discourse entity is identified with another discourse entity which has previously been mentioned in the text (102), through a repetition of habitual action (103), or by an inaleniable relation to the subject (104) the discourse entity is assumed to be identifiable to the receiver of the message (cf. [Fauconnier84], [Chuquet and Paillard89]).

(102) I saw a woman. The **woman** drove a taxi. The **taxi** was blue.

(103) *French:*

Elle met **la table**.
she put the table

'She lays the table.'

(104) a. Il ferme les yeux.
he closes the eyes
'He closes his eyes.'

b. Il s'est cassé la jambe.
he himself has broken the leg
'He has broken his leg.'

deixis
generic reference
predicative reference

c. Il s'est casssé la jambe.
he himself has broken the leg
'He has broken his leg.'

d. Il a perdu la mémoire.
he has lost the consciousness
'He lost consciousness.'

Definite articles, demonstrative pronouns and possessive pronouns may be employed in order to mark the identification of the extracted discourse entity with another discourse entity related to discourse or the hearer's knowledge system. In our feature system, this opposition is marked as shown in (105), where the way familiarity is achieved (e.g. previous mentioning/repetitive action/ inalenable relation) is in the current state of the system not yet completely calculated. Only if familiarity is achieved via a pointing operation (deixis) in text or situation is the kind of deictic operation specified as being proximal or distal. The latter may be further specified as near or far if the language reflects this distinction[1]. If no deictic operation is involved the value of `deix` is `nil`.

(105)

An elephant	`ref={'T'=ref,ref=unfam,deix=nil}`
The elephant	`ref={'T'=ref,ref=fam,deix=nil}`
This elephant	`ref={'T'=ref,ref=fam,deix=prox}`
That elephant	`ref={'T'=ref,ref=fam,deix={dist=_}}`
Der Elefant da	`ref={'T'=ref,ref=fam,deix={dist=near}}`
Der Elefant dort	`ref={'T'=ref,ref=fam,deix={dist=far}}`

If the discourse entity (e.g. *elephant*) does not refer to an instance of the denoted concept, it may refer to the concept (or its prototypical representative), or it refers to a second discourse entity in order to form a predicative relation of the 'is a' type. The first case is called generic reference, while the second is called predicative reference, represented by the following feature structures:

(106)

The elephant lives in ...	`ref={'T'=nonref,nonref=generic}`
He is an elephant.	`ref={'T'=nonref,nonref=pred}`

Generic propositions denote what is normal or typical for members of a class. In most languages, these propositions are restricted to special verbal tense and aspect. Progressive aspect and aorist on one hand and generic interpretations

[1] The suggested structuring of features allows for a matching of the bipartition (cf. German: *diese/jene*, it:*questo/quella* hy:*zhè ge/nà ge*, English: *here/there*) onto a tripartition which is equally employed as for example in Spanish and German *aquí/allí/allà*, *hier/da/dort* (cf. [Ehrich82], [Hottenroth82]).

on the other hand are mutually exclusive. Perfective aspects and generic interpretations are unlikely to cooccur. Such contextual information may be used to calculate the reference value of a discourse entity. Predicative nouns appear only in argument positions of copula verbs, or, if the language doesn't use copula constructions, without copula. Specific and generic referents can be referred to anaphorically, predicative referents cannot, neither as antecedent (107b) nor as anaphora (107c, taken from [Lujan80], pg.19).

cumulative reference
quantized reference

(107) a. He$_i$ is a teacher$_j$. She likes him$_i$.

b. * He$_i$ is a teacher$_j$. She likes him$_j$.

c. *Spanish:*
* hablaré con el médico que es tu hermano.
I shall talk to the doctor who is your brother

A fourth type of reference is the unique reference of proper nouns. These discourse entities do not refer to a concept or an instance of a concept but to an individual occurrence which is associated with an individual label. For these entities we use `ref=unique`, which automatically blocks deictic operation on these proper nouns.

(108)

London	`ref=unique`

7.3 Conceptual Boundedness

Determiners specify not only the referential type of nouns, but function at the same time as quantifiers, as a primary conceptual bounding, comparable to the measure words of Chinese and English (e.g. English: *the sheet of paper, a cup of milk* versus German: *das Papier, eine Milch*). I follow the definition of conceptual boundedness given by [Quine60], according to whom an instance of a concept refers **cumulatively** if any sum of the instance is the instance itself, e.g *any sum of parts which are water is water* (pg.91). Otherwise concepts refer **quantized**. Traditionally better known than the cumul/quant distinction is the mass/count distinction which, however, is only of secondary importance for translation. This mass/count distinction refers to a lexical default value according to which the singular term refers quantized (e.g. *child*) or cumulative (e.g. *water*). This default however is frequently overwritten when mass nouns are pluralized or occur with an article or numeral (e.g. as serving unit *a beer, two coffees*, as a type *a beer I had in Germany*, or an instantiation *a war, two wars* (cf. [Bunt85])). The count/mass distinction as a lexical classification is

atelicity
telicity

irrelevant for translation as mass and count nouns may become mutual translation (cf. examples in (3, page 4 and in [Mufwene84]). The cumul/quant distinction on the other hand must necessarily be controlled in translation in order to ensure the translational equivalence. Cumulative concepts must be translated by cumulative concepts, and quantized concepts must be translated by quantized concepts, disregarding the mass/count classification, the number marking or article use in both languages[2]. A further advantage of the cumul/quant distinction is that it can be meaningfully applied to other parts of speech such as adjectives and verbs as explained in the following section.

7.4 Crossing the Part of Speech

7.4.1 Conceptual Boundedness

I assume conceptual boundedness to be applicable to all parts of speech (cf. Krifka's theory of homomorphism of objects and events [Krifka91]). The cumul/quant distinction, extended to verbal semantics, captures the well-known distinction between bounded and unbounded processes. Just as *water plus water* is *water*, *to run and to run* means *to run* and not *to run two times.* [Vendler67] illustrates this with the following example (pg.101): "*If it is true that someone has been running for half an hour, then it must be true that he has been running for every period within that half hour*". Traditionally these verbs are referred to as **atelic**. Bounded verbal concepts, that is verbal concepts which are inherently lexically bounded (e.g. *to reach, to attain*), or which become bounded by the presence of an object (e.g. *to run a mile, to drink a cup of tea*) refer quantized (cf.[Verkuyl72]): *to obtain* plus *to obtain* means to obtain two times. These verbs are traditionally referred to as **telic** verbs (cf. [Dahl81], [Dahl85]).

Controlling the conceptual boundedness in translation has repercussions when the part of speech changes during translation. If, for example, the event expressed by a verb is bounded, due to its inherent lexical semantics or contextual influences, this boundedness is transferred to the target language, where, according to the monolingual specifications, it may trigger the generation of a definite or indefinite article on a noun (cf. (109a)). If the verb refers cumulative, the translation with a definite nominalization becomes impossible (cf. (109b)).

[2] The relation between number marking and the cumul/quant distinction has already been discussed in Chapter 1 page 6.

(109) *English - German:*
a. He suggested to me to knit a pullover.
Er schlug mir vor, einen Pullover zu stricken.
Er schlug mir **das** Stricken eines Pullovers vor.

b. He suggested to me to knit.
Er schlug mir vor zu stricken.
* Er schlug mir das Stricken vor.
Er schlug mir Stricken vor.

As the following examples show, conceptual boundedness is a property of adjectives as well. Conceptual unbounded adjectives can be paraphrased only by cumulative PPs (110a-c) while conceptually bounded adjectives need a quantized paraphrase (110d).

(110) *German:*
a. Eine anspruchsvolle Waise
a demanding orphan
‘eine Waise mit Anspruch’
‘eine Waise mit Ansprüchen’
* ‘eine Waise mit **den** Ansprüchen’
* ‘eine Waise mit **dem/einem** Anspruch’
b. Eine bedeutsame Weise
a meaningful tune
‘eine Weise mit Bedeutung’
* ‘eine Weise mit **der/einer** Bedeutung‘
c. Die picklige Weise
the pimply wise woman
‘eine Weise mit Pickeln.’
* ‘eine Weise mit **dem/einem** Pickel’

The conceptual boundedness is equally important for the choice of the derivation type. In example (111a) the conceptually unbounded verb *wandern* is translated into the gerund nominalization *walking* and not into the zero-derivation *walk*, due to the boundedness implied by this last derivation type, while in the case of the conceptual bounded predicate *wandert nach Bonn* in (111b), the zero-derivation is a legal translation.

(111) *German:*
a. Er wandert gerne.
he walks likingly
‘He likes walking.’
*‘He likes the walk.’

b. Er wandert gerne nach Bonn.
he walks liking to Bonn
'He likes walking to Bonn.'
'He likes the walk to Bonn.'

7.4.2 Reference Type

In the same way as all features representing **meaning**, the `ref` feature is transferred regardless of the part of speech of the target expressions. Verb phrases do not refer **unique** but refer generically or referentially. Verb phrases which are neither related to the speech act nor to a second event are likely to refer generically (cf. [Dahl75]), especially if they have non-referential arguments (112).

(112) Dogs bark.

If a verb refers to a second event (including the speech act), it refers referentially (e.g. *suggested* in (113)). The knownness to the hearer, however, is seldomly marked. As a consequence, the subordinate verb in (113) may refer to a known or unknown event.

(113) a. He suggested to occupy the city.
b. He suggested the/an occupation of the city.

Examples for the markings of the knownness of events expressed by verbs are the French *puisque- parce que* difference for the complementizer *since*: sentences introduced by *puisque* refer to a known event (cf. (114a)) and sentences introduced by *parce que* refer to an unknown event (cf. (114b)). In German the marker *ja* may be used to mark the familiarity of the event (cf. (114c)). Word order may equally be used to distinguish known from unknown events, although the difference between (114d) and (114e) is not as clear, the first is more likely to refer to a known event than the latter.

(114) a. *French:*
Puisqu'il avait déjà mangé, il n'avait pas faim.
since he had already eaten, he had not hunger
b. Parce qu'il avait déjà mangé, il n'avait pas faim.
because he had already eaten, he had not hunger
c. *German:*
Er hatte keinen Hunger, da er ja schon gegessen hatte.
he had no hunger, since he well already eaten had

d. Da er schon gegessen hatte, hatter er keinen Hunger.
since he already eaten had, had he no hunger

e. Er hatte keinen Hunger, da er schon gegessen hatte.
he had no hunger, because he already eaten had

thematic role
syntactic functions

Chapter 8

A-Structure

Functions may relate **function** and **meaning** in a one-to-one relation. Thus, one function is mapped onto one dimension of **meaning** and vice versa. Such one-to-one relations can be found with degree words where the morphological or syntactic 'comparative' corresponds to the **meaning** 'comparative'. In many cases, however, one-to-many or many-to-many relations can be found. In Chapter 7, I described a one-to-many relation, where a system of articles maps onto two dimensions of **meaning**, i.e. the reference type and the conceptual boundedness. In this chapter I discuss a many-to-one relation which relates a set of functions to one dimension of **meaning**. This **function**, known as Argument Structure (henceforth A-Structure), relates word order, case marking, prepositional marking and the part of speech to a system of thematic roles. Among others, this **function** accounts for the divergences (4) page 5, (39) page 30 and (41c) page 31. The A-Structure is a set of more specific **functions** known as syntactic functions (e.g. subject, direct object, indirect object) which relate these properties for one constituent to a thematic role, where the thematic role represents the relevant dimension of **meaning**. In (115), the verb *phone* and its various translations assign two thematic roles, `theme` for the person who phones and `goal` for the person who is phoned. The `goal` must be marked for dative in Russian, an accusative in German, marked for the prepositional form `a` in Spanish and the prepositional form `gei` in Chinese.

(115) a. Boris phones Bill.

b. *Russian:*
Борис звонит Ивану.

Boris zvonit Ivanu.
Boris$_{NOM}$ phones Ivan$_{DAT}$

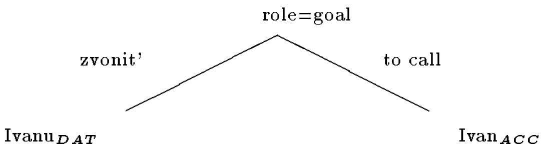
predicate
argument
thematic role
argument switching

c. *German:*
Boris ruft Peter an.
Boris$_{NOM}$ phones Peter$_{ACC}$ PREF.

d. *Spanish:*
Jaime llama a Ignacio.
Jaime phones to Ignacio

e. *Chinese:*
Jié gěi Xiǎohóng dǎ diàn huà.
Jie give Xiaohong hits telephone

The word which assigns thematic roles is necessarily a content word and is called the **predicate**. Constituents which are assigned thematic roles by the predicate are called **arguments**. The morpho-syntactic properties of the arguments are listed in the lexical entries of the predicate, together with the thematic role which functions as unique identifier of the participant. In analysis, the morpho-syntactic properties allow the constituent to be identified for thematic role assignment. In generation, the thematic role transferred from the SL to the TL associates the constituent unambiguously with its morpho-syntactic properties. This is illustrated in the following diagram, where the meaning `role=goal` is mapped onto different morpho-syntactic properties, according to the specifications of the verbal entries.



(116)

Further motivation for the use of thematic roles comes from the cases of so-called argument-switching, mentioned in Chapter (3), example (39). In order to treat such phenomena, I assign the role `goal` to the subject of the English verb *like* and the role `theme` to the object. Within the French lexicon, the à-object is assigned to the `goal` and the subject to the `theme`. By keeping the thematic roles constant between languages, the change of the syntactic functions becomes a side effect of the monolingual coupling of syntactic functions and thematic roles. We thus follow the suggestions of [Steiner87], [Steiner et al.88], and do not match functions onto functions, as suggested by [Kaplan et al.89], since this would prevent the TL from choosing the appropriate functions according its internal constraints (cf. [Butt94]). The alternative approach, i.e. the attempt to assign the arguments to their descriptions within the transfer rules as illustrated in rule (57) page 40 makes the system incompatible with the interlingual matching of translation units and does not work with grammaticalized causativizations, since the 'causer' cannot be assigned within the transfer rule which triggers the basic non-causative lexeme.

thematic roles

In addition to the influence of the predicate on the morpho-syntactic properties of the arguments, arguments may determine the morpho-syntactic properties of the head: In the Section 5.3.3 I showed how **concepts** of the TL are reconstructed out of the **notional domain** through a semantic classification. The A-Structure of a **concept** equally influences the choice of the appropriate **concept**. This is illustrated through (35). If we add the **agent** to the English sentence and inverse the direction of translation, the adjectival realization of the German predicate is no longer possible, as the adjective expresses the **agent** only with great difficulties:

(117) a. the man who can be blackmailed by his wife

b. *German:*
?? der von/durch seiner Frau erpressbare Mann
the by his wife blackmailable man

c. der Mann, der von seiner Frau erpreßt werden kan
the man who by his wife blackmailed be can

8.1 Thematic Roles

Although our approach is similar to that suggested in [Steiner et al.88], our set of thematic roles is much more limited. [Steiner et al.88] use the thematic roles for two different purposes, (i) the reading distinction of, for example, mental and action verbs through the use of different sets of thematic roles and (ii) the correct linking of argument constituents to the argument slot in the TL, independent of the linking in the SL (cf. the examples of argument switching). For various reasons I use the thematic roles only for the latter purpose, so that the system of thematic roles has to serve only this aim[1]. In our system `role=agent`, `role=theme` and `role=goal` are the most important thematic roles to which are added `role=location`, `role=provenance` and `role=direction` for movement verbs and `role=class` for copulative-like constructions[2]. The set of thematic roles of the predicate represents the coupling of the participants to the participants of a prototypical situation, where the `theme` is the entity the event is

[1] The reasons for working with a limited set of thematic roles are the following. (i) with a larger number of possible roles, the consistency between coders is very low. (ii) A large number of thematic roles may introduce ambiguities which cannot be motivated on monolingual grounds. (iii) The control of the readings in transfer (e.g. action verbs are translated as action verbs) does not work if the arguments are not realized, so that the readings are distinguished instead by the semantic coding of the predicate itself as 'action', 'mental process' or other.

[2] `role` is used also to specify the syntactic role of nodes (non-argument roles) as there are: `role=funct` for every functional category, `role=mod` for every modifier and `role=pred` for predicative nouns in support verb constructions.

argument hierarchy
syntactic functions

centered on, which is moved, transformed, effected, transmitted or considered. The `goal` is the entity to which the action or the theme is directed, which is the receiver or beneficiary of the action. The **agent** finally is the origin of energy, and movement (cf. [Ruus and Spang-Hanssen86]). This prototypical situation can be schematized as in (118)

```
(118)              ------------------------------>
          agent               theme                    goal
                   ------------------------------>
```

8.2 The Hierarchy of Arguments

The task of the A-Structure is to map the thematic roles (as part of the **meaning**) onto properties of the SS such as case, word order and prepositional marking, clustered under the header of 'syntactic functions'. As one thematic role can appear in various syntactic functions, depending on the diathesis marking of the verb and the derivation of the lexeme (cf. (123)), the thematic roles cannot be mapped directly onto the syntactic functions, but must be mapped onto an intermediate representation, which is the hierarchy of arguments. The hierarchy of arguments is described in the `frame` feature by the attributes `arg1` to `arg4`, each containing the description of one argument. Thus, in generation, a thematic role is not directly matched onto the SS but assigned to one argument slot. The SS is then determined according to the argument slot and the realizations of the argument and the predicate. This hierarchy of arguments is the **predicative structure** of the predicate. The **predicative structure** is illustrated in (119a) and (119b) for *to like* and *plaire*. It is this different assignment of thematic roles onto the hierarchy of arguments which accounts for the translational divergence in (39).

(119) a. @rule(b).

$$\left\{\begin{array}{l}\text{lex=like}\\ \text{frame=}\left\{\begin{array}{l}\text{arg1=}\left\{\text{role=goal,}\dots\right\}\\ \text{arg2=}\left\{\text{role=theme,}\dots\right\}\end{array}\right\}\end{array}\right\}$$

b. @rule(b).

$$\left\{\begin{array}{l}\text{lex=plaire}\\ \text{frame=}\left\{\begin{array}{l}\text{arg1=}\left\{\text{role=theme,}\dots\right\}\\ \text{arg2=}\left\{\text{role=goal,}\dots\right\}\end{array}\right\}\end{array}\right\}$$

The distributions of thematic roles for some English verbs is given in the following table:

diathesis

(120)

verb	arg1	arg2	arg3	arg4
abuse	agent	theme		
abut		theme	goal	
arrive		theme	location	
bring	agent	theme	provenance	direction
call	agent	theme	class	
define	agent	theme	goal	
give	agent	theme	goal	
groan		theme		
have	goal	theme		
hear	goal	theme	agent	
receive	goal	theme	agent	
take	goal	theme	agent	
tell	agent	theme	goal	

The hierarchical ordering of arguments correspond to the order of arguments in the initial stratum of relational grammar (cf. [Blake90]). This order of arguments is that of the hierarchy of syntactic functions (cf. [Comrie81], [Radford88], [Blake90] [Pollard and Sag87]) for the unmarked mapping of the argument hierarchy to syntactic functions. As I follow [Xrakovskij72] and [Duranti and Elinor79] and assume the active voice to be the unmarked variant wrt to the passive voice, arg1 contains the description of the subject, arg2 the description of the direct object, arg3 the description of the indirect object of an active sentence and arg4 the description of directional or instrumental arguments[3].

The mapping of the hierarchy arg1 to arg4 to the syntactic functions depends on the promotion or demotion of a constituent (e.g. passivization or intransitivization) (cf. [Siewierska84]). In CAT2 arg1 corresponds to the subject of an active sentence. In a passive sentence arg2 is the subject and arg1 the by-object. (121a) and (121b) represent the active and passive variant of *like*, where the subject is marked by case=nom and the by-object by pform=by respectively. These variants are obtained from an otherwise unmarked lexical entry by the application of a language-specific diathesis rule as illustrated in (122).

[3] In addition, I specify a so-called referential object arg0 which, following [Higginbotham85], denotes the referent, either an entity or a situation. With nouns, verbs, prepositions and complementizers the value of sem in arg0 unifies with the value sem in ehead, i.e. the selected semantic values. Only with a group of adjectives and adverbs are the values in arg0 and ehead different. This will be illustrated in Chapter 10.

(121) a.

$$\left\{\begin{array}{l}\text{lex=like}\\ \text{frame=}\left\{\begin{array}{l}\text{dia=act}\\ \text{arg1=}\left\{\text{role=goal,head=}\left\{\text{ehead=}\left\{\text{cat=n,case=nom}\right\}\right\}\right\}\\ \text{arg2=}\left\{\text{role=theme,head=}\left\{\text{ehead=}\left\{\text{cat=n,case=acc,pform=nil}\right\}\right\}\right\}\end{array}\right\}\end{array}\right\}$$

b.

$$\left\{\begin{array}{l}\text{lex=like}\\ \text{frame=}\left\{\begin{array}{l}\text{dia=pass}\\ \text{arg1=}\left\{\text{role=goal,head=}\left\{\text{ehead=}\left\{\text{cat=n,pform=by}\right\}\right\}\right\}\\ \text{arg2=}\left\{\text{role=theme,head=}\left\{\text{ehead=}\left\{\text{cat=n,case=nom}\right\}\right\}\right\}\end{array}\right\}\end{array}\right\}$$

(122) @rule(f).

$$\left\{\begin{array}{ll}\text{head=} & \left\{\text{cat=v}\right\},\\ \text{frame=} & (\left\{\begin{array}{l}\text{dia=act,}\\ \text{arg1=}\left\{\begin{array}{l}\text{role}^{\sim}\text{=nil,}\\ \text{head=}\left\{\text{ehead=}\left\{\text{case=nom}\right\}\right\}\end{array}\right\},\\ \text{arg2=}\left\{\text{head=}\left\{\text{ehead=}\left\{\text{case}^{\sim}\text{=nom}\right\}\right\}\right\}\end{array}\right\}\\ & ;\left\{\begin{array}{l}\text{dia=act,}\\ \text{arg1=}\left\{\text{role=nil}\right\},\\ \text{arg2=}\left\{\text{head=}\left\{\text{ehead=}\left\{\text{case=nom}\right\}\right\}\right\}\end{array}\right\}\\ & ;\left\{\begin{array}{l}\text{dia=erg,}\\ \text{arg1=}\left\{\begin{array}{l}\text{role}^{\sim}\text{=nil,}\\ \text{pos=nil,}\\ \text{head=}\left\{\text{ehead=}\left\{\text{index='IMPERS'}\right\}\right\}\end{array}\right\}\\ \text{arg2=}\left\{\text{head=}\left\{\text{ehead=}\left\{\text{case=nom}\right\}\right\}\right\}\end{array}\right\}\\ & ;\left\{\begin{array}{l}\text{dia=pass,}\\ \text{arg1=}\left\{\text{head=}\left\{\text{ehead=}\left\{\text{pform=by}\right\}\right\}\right\},\\ \text{arg2=}\left\{\text{head=}\left\{\text{ehead=}\left\{\text{case=nom}\right\}\right\}\right\}\end{array}\right\})\end{array}\right\}$$

Besides the standard active diathesis (top) and the standard passive diathesis (bottom) this rules provides for the so-called 'ergative' diathesis *The door opens* and the possibility of having only a second argument and no first argument in the case of which the second argument (the highest in the hierarchy) becomes the subject of passive sentences. The main motivation for having a second argument and no first argument is to obtain a correct pattern of argument inheritance, where certain derivations require a second argument which otherwise is not available.

argument inheritance

8.3 Argument Inheritance

According to the principle of Argument-Inheritance (cf. [Olsen92]), derivationally related words share the **predicative structure**. During the process of derivation the **predicative structure** of the **base** of the derivation is percolated to the new head, in the same way as shown in the F-Structure in the form of the percolation of the `frame` feature (cf. 58). The fact that the **predicative structure** is completely inherited is sometimes not directly visible since not every derivation can find a syntactic realization of every argument. The subject, for example, may be realized with the verb, the action nominalization, but not with the -able derivation. That the argument slot is blocked for syntactic reasons can be seen from the copulative variant, where the subject can again be realized. This is illustrated in (123).

(123) *German:*

a. der Mann adaptiert das Gerät
the man adapts the device

b. das Gerät wird durch den Mann adaptiert
the device becomes by the man adapted

c. der Mann adaptiert Geräte
the man adapts devices

d. die Adaptation des Gerätes
the adaptation of the device

e. die Adaption durch den Mann
the adaptation by the man

f. die Adaptation des Geräts durch den Mann
the adaptation of the device by the man

g. das adaptierbare Gerät
the adaptable device

h. ? das durch den Mann adaptierbare Gerät
the by the man adaptable device

i. das Gerät ist durch den Mann adaptierbar
the device is by the man adaptable

The principle of argument inheritance offers the possibility of specifying the argument slot only once with the **base**, the identifier of the **notional domain**. Together with the different derivations, the semantic values are coded. Action derivations may be coded with the macro THING, A-ENTITY or EVENT, the adjectival *able*-derivation with the macro POSSIBLE or OBLIGATION.

The following lexical entry of *adaptieren* encompasses the lemmata *adaptieren, die Adaptation*, *das Adaptieren*, *adaptierbar*, *unadaptierbar*, *unadaptiert* and *der Adapter*). With the help of the alternating head, the coupling of the morphological derivation to its syntactic and semantic values is made. As the `frame` is coded only once in this lexical entry, all derived forms have all arguments listed in the `frame`, accounting for the data in (123). The access of these arguments in SS is however severely limited by the syntactic and semantic properties of the derived words as will be described below.

(124) @rule(b).

$$\left\{\begin{array}{ll}
\text{lex=adaptieren} & \\
\text{head=} & (\left\{\text{cat=v}\right\} \\
 & ;\left\{\begin{array}{l}\text{cat=n,deriv=}\left\{\text{suff=ation}\right\},\\ \text{ehead=}\left\{\text{sem=}\left\{\text{concr=nil}\right\}\right\}\end{array}\right\} \\
 & ;\left\{\begin{array}{l}\text{cat=n,deriv=}\left\{\text{suff=inf}\right\},\\ \text{ehead=}\left\{\text{sem=}\left\{\text{EVENT}\right\}\right\}\end{array}\right\} \\
 & ;\left\{\begin{array}{l}\text{cat=a,deriv=}\left\{\text{suff=bar}\right\},\\ \text{ehead=}\left\{\text{POSSIBLE,sem=}\left\{\text{EVENT}\right\}\right\}\end{array}\right\} \\
 & ;\left\{\begin{array}{l}\text{cat=a,deriv=}\left\{\text{suff=un,deriv=}\left\{\text{aff=bar}\right\}\right\},\\ \text{ehead=}\left\{\text{mneg=yes,POSSIBLE,sem=}\left\{\text{EVENT}\right\}\right\}\end{array}\right\} \\
 & ;\left\{\begin{array}{l}\text{cat=n,deriv=}\left\{\text{suff=er}\right\},\\ \text{ehead=}\left\{\text{sem=}\left\{\text{INSTR}\right\}\right\}\end{array}\right\}) \\
\text{frame=} & \left\{\begin{array}{l}\text{arg1=}\left\{\text{role=agent,head=}\left\{\text{ehead=}\left\{\text{cat=n}\right\}\right\}\right\},\\ \text{arg2=}\left\{\text{role=theme,head=}\left\{\text{ehead=}\left\{\text{cat=n}\right\}\right\}\right\},\\ \text{arg3=}\left\{\text{role=goal,head=}\left\{\text{ehead=}\left\{\text{cat=n,pform=an}\right\}\right\}\right\}\end{array}\right\}
\end{array}\right\}$$

8.4 Argument Slots

argument selection

The argument slots, i.e. the value of the attributes arg1 ..arg4 consists of three parts. These are beside the thematic role of the argument, (i) the morphological and syntactic properties of the syntactic role associated with the thematic role and (ii) the semantic selectional restrictions.

8.4.1 Morpho-syntactic constraints

The morpho-syntactic constraints the thematic roles are most frequently linked to are the part of speech of the lexical head (ehead={cat=_}), the prepositional form (ehead={pform=_})[4], the case marking of nominal projections (ehead={case=_})[5], the tense of verbs ehead={tense=_}[6] which is used to distinguish untensed verbal arguments (tense=nil) from tensed verbal arguments (tense~=nil) and the mood of verbs ehead={mood=_} which is used to distinguish declarative from interrogative subordinated clauses (*that* = mood=declarative vs. *whether* = mood=interrogative).
Most selectional restrictions are extended head features, since the use of extended head features allows the part of speech of the lexical head to be referred to, although it may no longer be the head of the structure if embedded in F-Structures: if, for example, the arguments of a verb and its nominalizations may be a nominal phrase (123c), a determiner phrase (123a) or a prepositional phrase (123e,f), only that piece of information that is common to all these realizations can be entered in the argument slots. Thus, not the part of speech of the head (cat=p, cat=d, cat=n) but the part of speech of the extended head (cat=n) is coded in the argument slot.

Due to the implementation of argument inheritance and the lexical treatment of passive and reflexivizations, I distinguish case and prepositional forms assigned in the lexicon (valid for all realizations of the given lexeme), and those which are assigned by default rules (varying according to different realizations of

[4] pform indicates the preposition of a prepositional phrase, abstracting away from the morphological realization of the preposition. Thus the German preposition *an, am* and *ans* have the same pform=an. The same for French: *à, au, aux* have the same pform=a. The pform may equally specify the complementizer if the preposition of the nominal argument and the complementizer of the verbal argument are identical as may be the case in French and Spanish (e.g. *l'alternative d'acheter une voiture - l'alternative de l'acquisition*). pform is used as well for the specification of the prepositional form of German correlates (e.g. de:*Er denkt* **an** *sie. Er denkt dar***an***, daß sie kommt*).

[5] case indicates the morpho-syntactic case marking of nouns, determiners and prepositions. Contrary to classical linguistic conceptions, the extended feature convention assigns case to prepositions, determiners and the nominal head.

[6] tense indicates the morpho-syntactic tense marking for a specific language. This value is not transferred into the TL.

the lexeme). If the second argument of a verb requires `pform=nil` and the nominalization of the verb `pform=of`, the corresponding values should not be entered into the lexicon but calculated by general principles (e.g. structural case assignment, *of*-insertion, adjacency restrictions) after the part of speech and the derivation type have been identified. As the examples in (123) show, the first argument of *adaptieren* must be assigned `pform` and `case` depending on the part of speech of the lexeme by grammatical default rules. The second argument can be assigned `pform=nil` in the lexicon, since this value is identical for all realizations of this lexeme. The `case` of the second argument is again calculated according to the part of speech by grammatical default rules. Below we reproduce rules responsible for the default assignment of `case` and `pform` for German nouns. The first rule restricts noun arguments without prepositions to the genitive case (cf. (123e)) or the preposition *von* or *durch* (cf. (123f)). A similar rule applies to the second argument of nouns. If the second argument is assigned `pform=nil`, only the `case=gen` variant is possible. Similar rules apply to verbs and assign `pform=nil` as default to the second argument.

(125) a. @rule(f).

$$\left\{\begin{array}{llll}
\text{head=}\left\{\text{cat=n}\right\} & & & \\
\text{frame=}\,|\,\text{arg1=}\,|\,\text{head=}\,|\,\text{ehead=} & (\left\{\text{cat=n}\right\} & \& & (\left\{\text{pform=(von;durch)}\right\} \\
 & & & \left\{\begin{array}{l}\text{pform=nil}\\ \text{case=gen}\end{array}\right\}) \\
 & ;\left\{\text{cat=v}\right\}) & &
\end{array}\right\}$$

b.

$$\left\{\begin{array}{llll}
\text{head=}\left\{\text{cat=n}\right\} & & & \\
\text{frame=}\,|\,\text{arg2=}\,|\,\text{head=}\,|\,\text{ehead=} & (\left\{\text{cat=n}\right\} & \& & (\left\{\text{pform}^{\sim}\text{=nil}\right\} \\
 & & & \left\{\begin{array}{l}\text{pform=nil}\\ \text{case=gen}\end{array}\right\}) \\
 & ;\left\{\text{cat=v}\right\}) & &
\end{array}\right\}$$

In cases where the default assignment does not apply for verbs and nouns equally (e.g. `pform=nil` for the verb and `pform=against` instead of `pform=of` for the noun), as shown in the examples in (126) taken from [Danlos et al.90] (pg.11), the corresponding alternation has to be entered in the frame (e.g. `pform=(nil;against)`). The default assignment with verbs `pform=nil` is still possible and resolves the alternation to `pform=nil`. Since English nouns cannot structurally bind nominal arguments without a preposition the `pform=nil` alternative is excluded when a nominal argument is encountered and `pform=against` is the only possible alternative.

(126) a. John attacked Mary

b. John's attack against Mary

8.4.2 Semantic Constraints

Beside the syntactic constraints, semantic constraints figure in the argument slot of every lexeme. They are used to restrict the argument slots to structures of a specific semantic type. The semantic constraints not only help to exclude impossible structures in analysis, but disambiguate possible readings of arguments and to find the right translation of them. As the German verb *bemalen* selects a concrete entity as its argument (`concr~=nil`), the desambiguated argument can only be translated as *body* (`concr~=nil`) and not as *erection* (`concr=nil`).

(127) *German - English:*
a. Aufbau - body/erection

b. Er bemalte den Aufbau. - He painted the body.

By the same token, the translation of the predicate may be restricted by the semantic constraints on the arguments (cf. [Sakamato et al.86]). As the examples from [Hawkins83] pg.30 show, the ambiguity of the English predicate wrt its German translation is resolved through the constraint that *streichen* takes a concrete (`concr~=nil`) argument.

(128) *English - German:*
a. to paint - malen/streichen

b. to paint a landscape - eine Landschaft malen

The attribute `sem` is used to describe the semantic constraints. The system of semantic constraints is based on the system developed by [Zelinsky-Wibbelt88], [Zelinsky-Wibbelt89], aiming at a semantic classification of nouns. Its actual implementation represents only a subset of the feature structure proposed by Zelinsky-Wibbelt, arranged in such a way that it can be applied to all parts of speech. The basic dimensions of this classification are (i) the conceptual boundedness discussed in Chapter 1 and 7, (ii) the material nature of a concept and (iii) the abstract nature of a concept.

(129)
$$\left\{ \text{head=} \mid \text{ehead=} \mid \text{sem=} \left\{ \begin{array}{l} \text{bound=}\{\text{'T'=}\ldots\} \\ \text{concr=}\{\text{'T'=}\ldots\} \\ \text{abstract=}\{\text{'T'=}\ldots\} \end{array} \right\} \right\}$$

Concrete are entities which can be accessed by the tactile sense, or which are parts of collections of such entities. Besides through tactile perception, entities can be perceived either visually, olfactorily or auditively, but none of the latter senses on its own allows an entity to be described as concrete, i.e. you can see *children paly*, the *children* are 'concrete' but not the *playing*. Words referring to a class of entities must share the value of `concr` with that of the elements of that class, since by the relation of hyponymy, every superordinate word can enter into a slot of a subordinate word (e.g. *Sie jagen Löwen, Sie jagen Säugetiere*). The `concr` feature is specified by a general 'type' `'T'=TYPE` which gives an encyclopaedic characterization of the entity in question. The values of `'T'=TYPE` are:

(130)

`'T'=animal`	animal	e.g. *lion*
`'T'=body`	body part	e.g. *leg*
`'T'=building`	building	e.g. *post office*
`'T'=clothes`	clothes	e.g. *trousers*
`'T'=hum`	human being	e.g. *fisherman*
`'T'=instr`	instrument	e.g. *chair*
`'T'=plant`	plant	e.g. *tree*
`'T'=region`	region, area	e.g. *beach*
`'T'=stuff`	stuff, material	e.g. *wood*
`'T'=vehicle`	vehicle	e.g. *car*

Features referred to more frequently within the `concr` feature, as they are independent of the encyclopaedic classification, are:

(131)

`dim=1`	one-dimensional	e.g. *point*
`dim=2`	two-dimensional	e.g. *along the road*
`dim=3`	three-dimensional	e.g. *in the car*
`sex=female`	female	e.g. *woman*
`sex=male`	male	e.g. *man*
`sex=nil`	no sex	e.g. *car*
`state=solid`	solid	e.g. *gold*
`state=liquid`	liquid	e.g. *water*
`state=gas`	gaseous	e.g. *oxigen*

Abstract are all types of events and those entities the meaning of which cannot be restricted to its material realization. This may be the case for entities which have no material realization at all, i.e. temporal indications, or for semiotic entities, which beside their material realization have an separate ideational nature.

The `abstract` feature is further specified by a general 'type' `'T'=TYPE`. The values are:

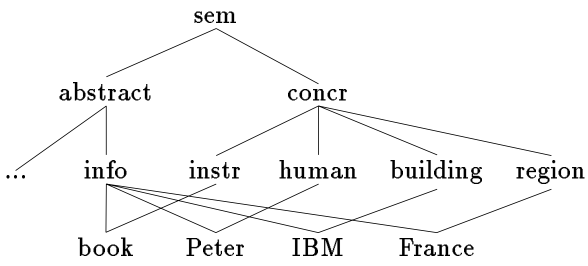
(132)

`'T'=action`	physical action	e.g. *wind, rain*
`'T'=info`	sources of information	e.g. *book, Peter*
`'T'=time`	units of time	e.g. *hour*
`'T'=emot`	type of emotion	e.g. *love, hatred*
`'T'=relat`	relations between entities	e.g. *fatherhood*
`'T'=language`	languages	e.g. *Spanish*
`'T'=field`	cognitive region	e.g. *mathematics agriculture*
`'T'=measure`	non-temporal units of measure	e.g. *meter kilogram*

This double description of the abstract and the concrete nature of an entity is unusual as most linguistic descriptions characterize the entity exclusively as abstract or as concrete. But for those entities which have a concrete and an ideational nature, an exclusive coding as abstract or concrete may lead to conflicting values. If, however, both dimensions, the concreteness and the abstractness, are coded independently, the semantic description can accommodate different demands. An example for such a double nature are semiotic entities (S-ENTITY), e.g. *book, letter, picture*, which denote on the one hand the abstract informational content and on the other hand the concrete instrumental support of this content. A disjunctive coding as 'abstract' or "concrete' is not satisfying as both properties may be required at the same time as in *the book I wrote is on the table.*

Semiotic entities are thus marked as being abstract and concrete, so that they can enter argument slots restricted to `abstract={'T'=info,temp=nil}` (e.g. *the book argues against*) or `concr={'T'=instr}` (e.g. *the book corrodes*). This multiple inheritance of semantic values from the abstract and concrete dimension is illustrated for semiotic entities (*book*), humans (*Peter*), institutions (*IBM*) and countries (*France*).

(133)



8.5 A-Structure in Syntax

The **predicative structures** project differently into syntax. If the predicate is the head of the syntactic structure, I speak of an **A-Structure**. If the predicate is the non-head of the structure and an argument of the head, I speak of a modifier structure (M-Structure). An example of a **A-Structure** is *The man is singing*, where *sing* is the head and *the man* the argument. An example of M-Structures is *singing men* in the case of which *men* is the head and *singing* the modifier. As M-Structures will be treated in later chapters, I shall discuss here only the **A-Structure**, the most direct projection of the **predicative structure** into syntax.

The **A-Structure** is implemented in the form of eight binary branching **b-rules**. These rules are responsible for the detection of the first, second, third and fourth arguments in right and in left position to the head. Within these rules an argument is detected when it unifies directly with the argument description of the head. The headedness of this structure is indicated by the percolation of the extended head features. A simplified head-argument scheme illustrates the line of projection (**ehead=EHEAD**) and the selection of the argument by the predicate (**arg1=ARG1**) for the selection of the first argument in left position.

(134) @rule(b).

$$\left\{\text{head}=\left\{\text{ehead=EHEAD}\right\}\right\}\begin{bmatrix}\text{ARG1\&}\left\{\text{head}=\left\{\text{max=yes}\right\}\right\} \\ \left\{\begin{array}{l}\text{head}=\left\{\text{ehead=EHEAD}\right\} \\ \text{frame}=\left\{\text{arg1=ARG1}\right\}\end{array}\right\}\end{bmatrix}$$

8.6 Accessibility of the Head Position

In addition to the restrictions on the arguments listed above there are a number of restrictions on the head of the **A-Structure**. That is, although the argument frame is inherited from the **base**, the realization of the argument in syntax is restricted and argument slots remain empty through the restrictions on the head position.

The first case of an unaccessible argument slot occurs when the argument is realized externally to the projection line of the head, resulting in a M-Structure as discussed in the following chapter, or a support structure to be discussed in Chapter 11. An inflected German adjective, for example, can only function as a modifier, so that the second argument slot adjectives usually have cannot be filled in the A-Structure. Instead, the argument slot has to remain empty until it is coindexed with the nominal to be modified. If the adjective is not inflected, it may function as a sentencial adverb, in which case the second argument becomes coindexed with the main verb of the sentence or the subject of the sentence, or as a predicate, in which case the the realization of the second argument is delegated to the copulative verb.

In this and similar cases, **pos=nil** is assigned to the argument description of **arg2**, thereby blocking the unification of the adjective with the **b-rule** responsible for the binding of the second argument (135). As the second argument will be filled at the IS level, **pos=nil** is only active at the syntactic level.

event reading

A second type of constraints on the head position of **A-Structures** comes from the semantic properties of the head. The first, third and fourth argument can only be realized if the head is an event (cf. [Grimshaw90]). This becomes relevant for the reduction of lexical ambiguity. While action nominalizations are often ambiguous with respect to event versus entity reading, the entity reading is automatically excluded when slots other than the second argument slot are filled. In order to realize a second argument within a phrasal structure, animateness or abstractness of the concept in question suffices, while in sub-phrasal **A-Structures** (e.g. in compounds) even these constraints may not apply (e.g. English:*can opener*, Dutch:*blikopener*, German:*Büchsenöffner*, Spanish:*abrelatas* (cf. [Olsen92], [Carulla94])). The constraints on the properties of the head are entered in the **b-rule** for **A-Structures**, which I reproduce in (135). It can be seen that the head of a **A-Structure** must be an event and secondly that **pos=nil** can block the access to this structure.

(135) @rule(b).

$$\left\{\text{head}=\left\{\text{ehead}=\text{EHEAD}\right\}\right\}\left[\begin{array}{l}\text{ARG1},\\ \left\{\begin{array}{l}\text{head}=\mid\text{ehead}=\left\{\text{sem}=\left\{\text{EVENT}\right\}\right\}\&\text{EHEAD},\\ \text{frame}=\left\{\text{arg1}=\left\{\text{pos}=\text{pre}\right\}\&\text{ARG1}\right\}\end{array}\right\}\end{array}\right]$$

Chapter 9

Pronouns

It is not only functions words which cannot be directly equated with a concept. Pronouns have an uncertain status as concepts, which is reflected linguistically by the different and inconsistent forms pronouns are associated with: Pronouns are expressed either as words, as inflectional suffixes, as pronominal affixes, or they are not expressed at all. In the following Spanish examples the pronouns are expressed by an inflectional markers in *quier***o**, as word **lo** or attached to the main verb *hacer***lo**.

(136) *Spanish:*

a. Lo quiero hacer
   it want$_{1P}$ do
   'I want to do it'

b. Quiero hacerlo
   want$_{1P}$ do-it
   'I want to do it'

9.1 Contextual Pronouns

Classically, one distinguishes two types of pronouns, discourse related pronouns (contextual pronouns) and text related pronouns (textual pronouns) (cf. [Halliday85]). Examples for contextual pronouns are *I, you, this* and *that.* Through the use of discourse pronouns, the speech actualized, i.e. anchored in the direct experience of the interacting partners and provides the partners with common points of reference. In the following context consisting of three cars (`<0=0>`) and two persons (+) and the speech act "B:*I want this*", *I* is identified with B through the turn taking. *This* refers to an inanimate entity which is not involved in the speech situation and which is close to the speaker. Thus *this* refers to A or C, but not to E.

```
<o=o>  +   <o=o>  +  <o=o>

  A    B     C    D    E
```

Since B refers to A or C, the partner D can easily react to B's statement. It would have been more difficult if B would have said *Wittgenstein liked to eat bread and cheese*, where neither *Wittgenstein* nor *bread* and *cheese* are part of the context.

For a correct translation of contextual pronouns, the entity of the context referred to has to be identified so that the TL can use an appropriate pronoun to refer to this contextual entity. The translation of *this* should be *dieses* in German, not **diese* or **diesen*, since the German *Auto* (car) is neuter. In Italian the translation should be *questa*, not **questo*, since the Italian *macchina* (car) is feminine. Thus at least the contextual pronouns like *this* and *that* are not concepts.

Concerning the pronouns like *I* and *you*, they seem to refer to concepts as 'speaker' and 'hearer'. In these cases they are easily translatable, at least in Western languages (English:*I*, French:*je*, Russian:я Italian:*io* etc...). In Asian languages, the choice of the adequate discourse pronoun may depend on the social relation the discourse partners maintain, and the translation of these pronouns becomes difficult without knowing to which persons they refer.

The identification of the entity referred to by the pronoun cannot be afforded by the linguistic component of an MT system alone. All the linguistic component can do is to restrict the referential scope of the entity referred to in the context. The choice of the morpho-syntactic properties of the pronoun and its linguistic context may characterize a pronoun as being an entity or an event, animate or inanimate, far or near, as being high or low in social hierarchy. This information is transferred onto the contextual pronoun of the TL, where appropriate forms and markers have to be found. If information of the SL is not sufficient for the correct choice of the target pronoun, a default pronoun should be generated. It would be necessary to identify the referent of the pronouns with the help of another system and to replace the pronoun at level IS by this concept. As in the current state of CAT2 this identification of the contextual entity is not possible, the pronoun is not replaced at level IS.

9.2 Textual Pronouns

textual pronouns
cohesion

Text related pronouns like *it* or *he* replace a concept of the text (e.g. *it* replaces *the car*) and, maintaining the reference of the original concept (*it* refers to the same entity as *the car* did), equate the replaced concept with a concept that has previously occurred in the text, which may be *a car*. The function pronouns have in a text can be illustrated as follows: Even if neither *Wittgenstein* nor *bread* and *cheese* are present in the actual context of the speech act, the sentence *Wittgenstein liked to eat bread and cheese* can lead to or be part of a communication if either *Wittgenstein*, *bread* and *cheese*, other philosophers or other eating habits have been mentioned previously. This would allow this sentence to be linked to prior parts of the communication. This linking of parts of texts through common semantic domains, spans or common referents is known as **cohesion**. Textual pronouns mark the cohesion of a text (cf. [Halliday85]), or as expressed by [Larson84] pg.397, they are a "*cohesive device of discourse*". They replace a concept which is coreferential with a concept previously mentioned in the text (the antecedent) and declare both concepts to be coreferential. While **coreference** is a property of concepts, **cohesion** is a property of text, i.e. the degree to which the linking of parts of the text is marked. Coreference is just one possibility to mark the cohesion of the text and pronouns are just one means to mark coreference.

If two concepts are coreferential (e.g. *Pete … this man*), this is a piece of information which is part of the **meaning**. The choice whether to mark this coreference with anaphora or by other means is a language-specific function. Thus different languages, different contexts and different speech styles may employ different structures to mark cohesion (cf. [Larson84] [Schwarze88]). The following structures, including the omission of the concept, are all possible structures for coreference marking.

(137) a. Peter … he … he … he
b. Peter … Peter … Peter … Peter
c. Peter … he … Peter … he
d. Peter … ⊘ … Peter … ⊘[1]

The translation of a pronoun depends on a number of factors. (i) If a pronoun is translated by a pronoun, the semantic and morpho-syntactic properties of the concept the pronouns refers to select the form of the pronoun ('car' is female in Russian and neuter in German (138a) vs. (138b)). (ii) Pronouns may be expressed or characterized either as pronouns (138b), as inflection markers (first part of the Italien example (138c)), or indirectly as agreement markers as in

[1] I use ⊘ to mark covert. i.e. unrealized pronouns.

pronominalization

the second part of (138c). (iii) If two concepts are coreferential there may be the choice which of the concepts has to be realized as a pronoun and which not. This is illustrated in (138d) where in Chinese *Hans* is the subject of two clauses without an overt pronoun. The English translation has the choice of pronominalizing the subject of the subordinate or the matrix clause, according to the theme-rheme structure of the sentence.

(138) a. *Russian:*
У меня естьмашина. Она красная.
U menja est' mashina. Ona krasnaja.
at me is car. She red.
'I have a car. It is red.'

b. *German:*
Ich habe ein Auto. Es ist rot.
I have a car. It is red.
'I have a car. It is red.'

c. *Italian:*
Ho una maccina. E rossa.
have$_{1P}$ a car. Is red$_{FEM,SING}$.
'I have a car. It is red.'

d. *Chinese:*
qù dà xué yĭ hòu hàn sī măi lė shū
go university after Hans buy PAST book
'After he went to the university, Hans bought some books.'

'After Hans went to the university, he bought some books.'

'Hans bought some books, after he went to the university.'

* 'He bought some books, after Hans went to the university.'

As these examples show, pronouns cannot be translated without knowing what they refer to: They are substitutes for concepts and not concepts themselves. This supports the general claim of this book, that functions cannot be translated, and, accordingly the IS structure should not reflect the **functions** employed, but the resulting **meaning**. Therefore, textual pronouns should be represented at IS by the concepts they replace plus a referential index that shows their coreferentiality with the preceding concept. The question as to whether the coreferential concepts should be expressed at CS by a pronoun, a concept, or by an omission, is solved with the cohesion operation the language usually employs and the concrete syntactic constraints.

This complex behaviour of pronouns in translation has lead to a large number of publications in this field. Most of this literature is concerned pro-drop

with the question of how the antecedent can be identified (cf. [Sidner86], [Mitkov95], [Nakaiwa and Ikehara95]), others mention the difficulties of translating pronouns due to missing textual and contextual parameters (cf. [Choi95], [Mitkov et al.95]). Within this chapter I shall discuss only covert pronouns in the context of argument structures, as modifier structures and the function of pronouns within them will be the content of Chapter 10. The themes discussed are the pro-drop for subjects, the pro-drop in imperative clauses and in control structures.

9.3 Subject Pro-Drop

The attempts of GB to identify parameters which describe languages and account at the same time for phenomena of language acquisition have led to the stipulation of the so-called pro-drop parameter (cf. [Cook88]). This parameter, if set to "on", allows a subject pronoun to be dropped at the level of SS. In order to have equivalent structures at level IS the pronoun (e.g. the concept 'speaker') must be reconstructed for Italian and Spanish in analysis.

(139)

pro-drop on	pro-drop off
Ti amo	*I love you*
Te quiero	*Ja ljublju tebja*

In order to reconstruct the pronoun, the information concerning the subject is stored at level CS in the subject feature. This feature, which has properties of the `AGRsubj` feature of GB (cf. [Chomsky93]) as of the `subj` attribute of the split argument list in HPSG (cf. [Sag95], [Riehemann95]), is unified with the argument slot of the subject (the first, second or third argument of the `frame` feature according to the diathesis marking). (140) is a modified version of rule (15) discussed above, showing how the `subj` feature is instantiated[2]. The pronoun will be generated from this feature.

[2] For the defintion of the macro COINDEX cf. page 22.

(140) @rule(f).

$$\left\{\begin{array}{ll} \text{role}\tilde{}= & \text{funct}, \\ \text{head}= & \left\{\begin{array}{l}\text{cat=v},\\ \text{ehead}=\left\{\text{subj}=\left\{\begin{array}{l}\text{role=ROLE},\\ \text{head}=\left\{\text{ehead}=\left\{\text{COINDEX}\right\}\right\}\end{array}\right\}\right\}\end{array}\right\}, \\ \text{frame}= & (\left\{\begin{array}{l}\text{dia=act},\\ \text{arg1}=\left\{\begin{array}{l}\text{role=ROLE},\\ \text{head}=\left\{\text{ehead}=\left\{\text{COINDEX}\right\}\right\}\end{array}\right\}\end{array}\right\} \\ & ;\left\{\begin{array}{l}\text{dia=pass},\\ \text{arg2}=\left\{\begin{array}{l}\text{role=ROLE},\\ \text{head}=\left\{\text{ehead}=\left\{\text{COINDEX}\right\}\right\}\end{array}\right\}\end{array}\right\}) \end{array}\right\}$$

Between CS and IS, a pronoun is introduced according to the `subj` feature of every content word, if no subject is already realized. The rule responsible for the generation of the missing pronoun is reproduced in (141). A content word marked as `head` is translated into an identical structure, plus a pronoun `lex=pro` which unifies with the specifications in `subj=SUBJ`.

(141) @rule(t).

$$\text{head:}\left\{\text{head}=\left\{\text{ehead}=\left\{\text{subj=SUBJ}\right\}\right\}\right\} \Rightarrow \{\}\left[\begin{array}{ll}\text{head:} & \{\}\\ & \left\{\text{lex=pro}\right\}\&\text{SUBJ}\end{array}\right]$$

At level IS the contextual pronouns are currently still preserved, i.e. not replaced by the original concept. If the IS structure is translated into a language which does not allow for pro-drop the pronoun is translated until the CS structure of the TL. If the TL allows for pro-drop, the pronoun has to be preserved until the level T2 in order to guarantee the completeness of the structure. Between T2 and T1 the subject pronoun is removed by the inversed rule of (141). The unification of pronoun and `subj` feature does not only control the validity of the pro-drop, but transmits the information coming from the subject pronoun to the `subj` feature, so that this information can be used in syntax for various phenomena as agreement phenomena with the covert subject (eg it:*Sono amalato* vs *Sono amalata* (I am ill), spoken by a man vs. a woman).

9.4 Imperative Clauses

imperative clauses
control structures

The subject of a direct imperative clause can be equated with the second person *you* (142a)[3]. Indirect imperative clauses (142b) will be treated as a control structure, the functioning of which is described below.

(142) a. Make noise!
b. He ordered him to make noise.

Since the marking of the subject of the direct imperative clause as second person may be redundant, many languages do not employ a pronoun for the subject. Only if the choice of the pronoun may relate to the speech style or the attitude of the speaker wrt the receiver of the message may the pronoun be employed.

Languages use markers in order to indicate the imperative nature of the clause. Such markers may be sentential markers such as *!*, special word order, or imperative markers such as the Chinese *bȧ*. These markers may be used to unify the index 'hearer' into the subj feature. This is illustrated for the imperative marker *!* which requires the verb to have a 'hearer' subject via the extended head feature principle (cf. 57).

(143) @rule(b).

$$\left\{\begin{array}{l}\text{role=funct}\\ \text{lex='!',}\\ \text{head=}\left\{\text{ehead=}\left\{\text{subj=}\left\{\text{head=}\left\{\text{ehead=}\left\{\text{index=hearer}\right\}\right\}\right\}\right\}\right\}\end{array}\right\}$$

Between CS and IS the pronoun is introduced as illustrated above and translated into the TL. In generation the pronoun may be dropped between the IS and CS when specific syntactic structures are generated. Again, the dropped pronoun is unified with the subj feature of the verb in order to have the appropriate agreement values on the verb (or elsewhere).

9.5 Control Structures

Another instance of an anaphoric relation is found in so-called control structures. In these structures there is an anaphoric relation between an unexpressed

[3] At the time being, imperative clauses with a 'speaker' subject, e.g. English: *Let's go*, Chinese *zǒu bȧ* are not treated.

subject and an expressed or unexpressed antecedent. The unexpressed subject is referred to as the controlled element and the antecedent as the controller[4] (cf. [Bresnan82a])[5].

Control structures can be subdivided into syntactic (obligatory, functional) control and pragmatic (optional, anaphoric) control (cf. [Bresnan82a], [van Riemskijk and Williams86]). Syntactic control structures appear typically in argument position (lexical control), (e.g. *John$_i$ wants PRO$_i$ to leave*) and in non-fintite modifiers (*PRO$_i$ Sure of winning, Mary$_i$ entered the room*). Since modifier relations will be treated in chapter 10 and the treatment of pragmatic control is out of the scope of this book, I shall restrict the discussion in this chapter to the phenomenon of lexical control[6].

The cohesion marked by the covert pronouns in syntactic control structures must be expressed differently in different contexts and languages. (144a) shows an instance where the covert pronoun in SL becomes an overt pronoun in the TL due to the fact that the matrix verb does not have a non-finite complement and a finite subjectless sentence would be ungrammatical. In (144b-c) the covert English pronoun must be overt in German due to different assignments of syntactic functions. In these examples, different diathesis markings or argument switching make it impossible to control the pronoun (example from [Hawkins83], pg.53-54 and pg.57). (144d-h) shows that in the Balkan languages Greek, Bulgarian, Macedonian and Serbo-Croatian (cf. [Ruge86], [Hill91], [Rehder91a], [Rehder91b]) or in Portuguese, the subordinate verb is marked for agreement values with the unexpressed subject, expressing by this marking the cohesion. The examples are taken from [Norbert and Faensen76] pg.17 for Serbo-Croatian and [Walter and Kirjakova90] pg.60 for Bulgarian respectively:

(144) a. *French:*
Le directeur m'a dit d'être satisfait.
the director me has told to be satisfied
'The director told me that he was satisfied.'

b. *German:*
Er hoffte, daß ihm geholfen würde.
he hoped, that he helped was

[4] In the current implementation two macros are used to identify the controller (#CONTROLLER) and the controlled element (#CONTROLLEE). The definition of these macros is given in (25).

[5] In a control structure only referential properties are controlled. This is what makes control structures different from raising constructions. Control structures are not the result of NP movement, since in a control structure both verbs, the matrix verb and the subordinate verb, assign thematic roles, so that the Θ-Criterion would be violated if the NP was moved out of the subordinate clause into another Θ-marked position (see [van Riemskijk and Williams86]).

[6] Pragmatic control structures appear in infinitival subject positions (e.g. *PRO to leave would be nice*) and in indirect questions (e.g.*It is unknown to us$_i$, what PRO$_i$ to do*).

'He hoped to be helped.'

subject control
object control

c. Ich hoffe, daß mir dieses Buch gefällt.
I hope, that me this book pleases
'I hope to like this book.'

d. *Serbo-Croatian:*
Milica ide u kino da gleda film.
Milica go$_{3SG}$ to cinema to see$_{3SG}$ film
'Milica goes to the cinema to see a film.'

e. Milica i Ahmet idu u kino da gledaju film.
Milica and Achmet go$_{3PL}$ to cinema to see$_{3PL}$ film
'Milica and Achmet go to the cinema to see a film.'

f. *Bulgarian:*
Мария, искаш ли да напьлнееш?
Maria, iskaš li da napâlneeš?
Maria, want$_{2SG}$ QU that gain$_{2SG}$ weight
'Maria, do you want to gain weight?'

g. **Мария, Милева, искате ли да напьлнеете?**
Maria, Mileva, iskate li da napâlneete?
Maria, Mileava, want$_{2PL}$ QU that gain$_{2PL}$ weight
'Maria, Mileva, do you want to gain weight?'

In these cases of lexical control, the anaphoric relation between the unexpressed subject and the subject or the object of the matix verb is directly entered into the dictionary entry of the matrix verb (noun, adjective) in the form of a coindexation of the controller and the subject of the infinitival sentence.

(145) @rule(b).

$$\left\{\begin{array}{l} \text{lex=absicht} \\ \text{head=}\left\{\text{cat=n}\right\}, \\ \text{frame=}\left\{\begin{array}{l} \text{arg1=}\left\{\begin{array}{l}\text{role=agent,}\\ \text{head=}\left\{\text{ehead=}\left\{\text{COINDEX}\right\}\right\}\end{array}\right\}, \\ \text{arg2=}\left\{\begin{array}{l}\text{role=theme,}\\ \text{head=}\mid\text{ehead}\mid\text{subj}\mid\text{head}\mid\text{ehead}\mid\text{COINDEX}\end{array}\right\}\end{array}\right\}\end{array}\right\}$$

The control behavior of the German noun *Absicht* (intention) is an example of subject control, where the subject of the subordinate infinite clause is coindexed with the subject of the matrix clause (146a). The German noun *Befehl* (order) is an example of object control with the indirect object of *Befehl* as the controller of the subordinated clause (146b). The corresponding lexical entry is specified in (147).

(146) *German:*

a. Peters$_i$ Absicht, PRO$_i$ Lieder zu singen.
Peters intention, PRO songs to sing
'Peters intention to sing songs'

b. Peters Befehl an Paul$_i$, PRO$_i$ Lieder zu singen.
Peters order to Paul, PRO songs to sing
'Peters order to Paul to sing songs'

(147) @rule(b).

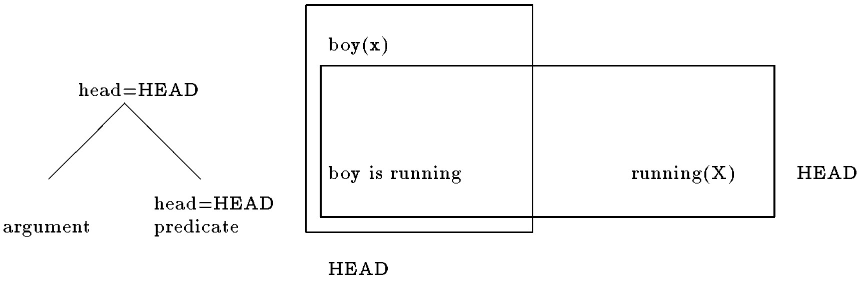
$$\left\{\begin{array}{l}
\text{lex=befehlen}\\
\text{head=}\left\{\text{cat=n}\right\}\\
\text{frame=}\left\{\begin{array}{l}
\text{arg1=}\left\{\text{role=agent}\right\},\\
\text{arg2=}\left\{\begin{array}{l}\text{role=theme,}\\ \text{head=}\mid\text{ehead}\mid\text{subj}\mid\text{head}\mid\text{ehead}\mid\text{COINDEX}\end{array}\right\},\\
\text{arg3=}\left\{\begin{array}{l}\text{role=goal,}\\ \text{head=}\left\{\text{ehead=}\left\{\text{COINDEX}\right\}\right\}\end{array}\right\}
\end{array}\right\}
\end{array}\right\}$$

Chapter 10

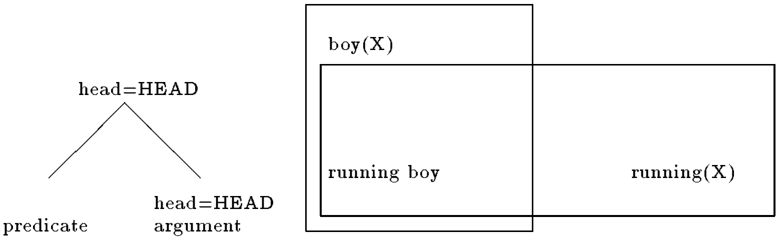
M-Structures

From a semantic point of view, Modifier Structures (**M-Structures**) are **predicative structures**, in the same way as Argument Structures (**A-Strucutres**). The difference between **M-Structures** and **A-Structures** is purely syntactic. While in **A-Structures** the predicate is the syntactic head (e.g. *the boys are* **running** (148)), in **M-Structures**, not the predicate, but the complement is the syntactic head (e.g. *running* **boys** (149)) (cf. [Langacker81], [Langacker87]). In other words, arguments which are realized internally to the (extended) projection of the predicate realize an **A-Structure**, arguments which are realized externally to the projection of the predicate realize an **M-Structure**.

(148)



(149)



relativization
cohesion

10.1 From Internal to External Arguments

In order to indicate which of the arguments is realized externally to the syntactic projection, the external argument is coindexed with the internal argument position. In cases where different internal argument positions are possibly realized externally, overt pronouns are used to mark the coindexation explicitely. This is the case in relative clauses of European languages where the relative pronoun specifies whether the subject, the direct object or the indirect object has been realized externally (cf. [Hawkins83]).

(150) The girl **whom** I showed the baby to
The girl **who** I showed to the baby.

Therefore, **M-Structures** necessarily involve pronominal structures and are thus part of the cohesion operation of a language. Instead of saying *The girl entered. The girl was happy*, we say:

(151) The girl entered happily.
The girl, **who** was happy, entered.
The girl, **who** entered, was happy.

In those SSs where only one argument slot of the modifier can be realized externally, no overt pronouns are required (e.g. with adjectival modifier). Prepositions and complementizers have two arguments, the first is realized externally while the second is realized internally. The German preposition *wegen* (because of), for example, assigns two roles to these arguments, `theme` (caused) to the external argument and `agent` (causer) to the internal argument. The role assignment is thus the same as for the verb *to cause*. The external argument is coindexed with a covert pronoun.
The external argument is linked to the **predicative structure** through the feature `restr`[1]. This feature has to unify with an argument that is bound externally in a modifier relation. The content of `restr` is determined by the head of the **M-Structure**. With prepositions and complementizers, for example, `restr` is coindexed with the first argument and inherits the semantic constraints from this argument. `pos=nil` makes sure that no internal realization of the first argument is possible. The second argument is realized internally[2].

[1] The feature `restr` is thus equivalent to the feature `mod` used in HPSG.

[2] A similar analysis of modifiers in the framework of HPSG can be found in [Abb and Maienborn94].

(152) @rule(b).

$$\left\{\begin{array}{l}
\text{lex=wegen,}\\
\text{role=mod,}\\
\text{head=}\left\{\text{cat=p,restr=}\left\{\text{head=}\left\{\text{COINDEX}\right\}\right\}\right\},\\
\text{frame=}\left\{\begin{array}{l}\text{arg1=}\left\{\begin{array}{l}\text{role=theme,pos=nil,}\\ \text{head=}\left\{\text{COINDEX}\right\}\end{array}\right\},\\ \text{arg2=}\left\{\text{role=agent}\right\}\end{array}\right\}
\end{array}\right\}$$

With adjectives, the internal argument position and the external argument **restr** may be linked in two possible ways. With "active" adjectives, i.e. deverbal adjectives whose first argument is being predicated on, the external argument is copied from the first argument and with "passive" adjectives the external argument is coindexed with the second argument. In underived adjectives it is the second argument which is coindexed with the external argument position, but with derived adjectives, the base of the derivation decides whether the first or second argument becomes the external argument. An example of an "active" adjective is the French adjective *méditatif* which means *qui médite* (who meditates) (153). A passive adjective is the French adjective *applicable*, which is derived from the French verb *appliquer* and means *qui peut être appliqué* (which can be applied) (154). The lexical entries of these adjectives are coded together with the verbs they are derived from.

(153) @rule(b).

$$\left\{\begin{array}{ll}
\text{lex=} & \text{me1diter,}\\
\text{head=} & (\left\{\text{cat=v}\right\}\\
 & ;\left\{\text{cat=a,deriv=}\left\{\text{aff=if}\right\}\text{,ehead=}\left\{\text{voice=act}\right\}\right\}\\
 & ;\left\{\text{cat=n,deriv=}\left\{\text{aff=ation}\right\}\right\})\\
\text{frame=} & \left\{\begin{array}{l}\text{arg1=}\left\{\text{role=agent}\right\},\\ \text{arg2=}\left\{\text{role=theme,head=}\mid\text{ehead=}\left\{\text{cat=n,pform=(nil;sur)}\right\}\right\}\end{array}\right\}
\end{array}\right\}$$

(154) @rule(b).

$$\left\{\begin{array}{ll}
\text{lex=} & \text{appliquer,} \\
\text{head=} & (\left\{\text{cat=v}\right\} \\
 & ;\left\{\text{cat=n,deriv=}\left\{\text{aff=eur}\right\}\text{,COINDEX}\right\} \\
 & ;\left\{\text{cat=a,deriv=}\left\{\text{aff=able}\right\}\text{,ehead=}\left\{\text{voice=pass}\right\}\right\} \\
 & ;\left\{\text{cat=n,deriv=}\left\{\text{aff=ation}\right\}\right\}) \\
\text{frame=} & \left\{\begin{array}{l}
\text{arg1=}\left\{\text{role=agent,head=}\left\{\text{COINDEX}\right\}\right\}\text{,} \\
\text{arg2=}\left\{\text{role=theme}\right\}\text{,} \\
\text{arg3=}\left\{\text{role=goal,head=}\left\{\text{ehead=}\left\{\text{pform=(a;sur)}\right\}\right\}\right\}
\end{array}\right\}
\end{array}\right\}$$

The mapping from the **predicative structure** (`frame`) of the adjective to the external argument (`restr`) is done by a lexical rule (155) which coindexes the internal and external argument slot, depending on the voice marking of the adjective.

(155) @rule(f).

$$\left\{\begin{array}{ll}
\text{head=} & \left\{\begin{array}{l}
\text{cat=a} \\
\text{ehead=}\left\{\text{voice=V}\right\}\text{,} \\
\text{restr=} \mid \text{head=}\left\{\text{COINDEX}\right\}
\end{array}\right\}\text{,} \\
\text{frame=} & (\left\{\text{dia=act\&V,arg1=}\left\{\text{head=}\left\{\text{COINDEX}\right\}\right\}\right\} \\
 & ;\left\{\text{dia=pass\&V,arg2=}\left\{\text{head=}\left\{\text{COINDEX}\right\}\right\}\right\})
\end{array}\right\}$$

In languages with covert (relative) pronouns, e.g. Chinese, the possibilities of realizing arguments externally are severely limited (cf. [Keenan72], [Keenan75], [Comrie81]), as the language cannot mark which argument position has be relativized.

10.2 M-Structures at CS

At the level CS, a constituent is analyzed as a modifier of another constituent if the external argument description (`restr`) unifies with the description of the constituent to be modified. The modified element is the head of the SS. The principles of the **M-Structure** are implemented in the form of two binary branching b-rules responsible for the detection of modifiers in right and in left position to the head (156). Within these rules the modifier selects the head when the head unifies with the external argument description of the modifier (`restr=ARG`). The headedness of this structure is indicated by the percolation of the extended head features (`ehead=EHEAD`). The semantic role is percolated with the line of the head projection (`role=ROLE`). As the modifier is the non-head of this structure, it must be a maximal projection, in order to control the functional completeness (cf. pg.58). Furthermore we require every modifier to have the properties of an event, i.e. to have aspectual values. This requirement forces us to analyze *Peter's dog* as *the dog owned by Peter*, where the relative clause has these aspectual properties. Restrictive modifiers are non-actualized*, while appositive modifiers are actualized* (cf. example (75) pg.54).

(156) @rule(b).

$$\left\{\begin{array}{l}\text{role=ROLE,}\\ \text{head=}\left\{\text{ehead=EHEAD}\right\}\end{array}\right\} \left[\begin{array}{l} \left\{\begin{array}{l}\text{role=ROLE,}\\ \text{pos=pre,}\\ \text{head=}\left\{\text{ehead=EHEAD}\right\}\end{array}\right\}\&\text{ARG,}\\ \left\{\begin{array}{l}\text{pos=post,}\\ \text{role=mod,}\\ \text{head=}\left\{\begin{array}{l}\text{max=yes,}\\ \text{restr=ARG,}\\ \text{ehead=}\left\{\text{sem=}\left\{\text{EVENT}\right\}\right\}\end{array}\right\}\end{array}\right\}\end{array}\right]$$

10.3 M-Structures for Translation

As we have already mentioned, modifier relations involve necessarily anaphoric structures, which can be seen from the relative clauses that may serve as paraphrases and translations (cf. (157)).

(157) a. The red car
The car **which** is red

b. The running girl
The girl **who** is running

c. The recognized error
The error **which** is recognized

d. The car on the table
The car **which** is on the table

e. He could find the way due to his knowledge of Korean.
He could find the way **which** was due to his knowledge of Korean.

The necessity of having such paraphrases is illustrated in example (35) pg.29, example (44) pg.33 and example (158). For a similar Japanese-English example cf. [Nagao and Tsujii86].

(158) *German:*

a. der auf seine Frau überaus stolze Mann
the on his wife extremely proud man

'the man who is extremely proud of his wife'

b. *Chinese:*
mù tóu zuò de zhuō zi
wood make ATT table
'the wooden table'
'the table made of wood'

At the deepest level of analysis the antecedent and the anaphor should be expressed by two identical and coindexed structures (cf. [Nagao and Tsujii86]) which in the current implementation of CAT2 is not yet achieved. This structure can be found in the SS of, for example, Hindi, where two coreferential nouns are used to form relative clauses (example and transliteration from [Comrie81] pg. 139).

(159) *Hindi:*

ādmī ne jis cākū se murgī ko mārā thā, us cākū ko
man$_{ERG}$ which knife with chicken$_{ACC}$ killed that knife$_{ACC}$
Rām ne dekhā.
Ram$_{ERG}$ saw

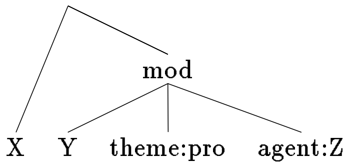
'Ram saw the knife with which the man killed the chicken.'

As a first step of implementation, however, a covert internal argument is reconstructed between CS and IS, in order to allow paraphrases with an overt internal argument (e.g. relative clauses) or paraphrases where at level IS an internal argument is required, but dropped later on (e.g. *the running boys*).

In order to specify the structure aimed at level IS, let us consider first prepositional modifiers. These are represented at level IS with two internal arguments. The first internal argument is a pronoun which is coreferential with the external argument (X). The second internal argument is that which has already been internal at the syntactic level (Z). The structure is represented in (160), where X may refer to *the inundation of the city*, Y to *because of* and Z to *breach of the dyke*.

incorporation

(160)



This structure can be mapped onto a relative clause with an overt or covert internal argument where Y refers to *caused* and Z refers to *by the breach of the dyke.* This structure can equally be matched onto a structure where Y refers to *since* or *because* and *Z the dyke broke.* Thus, this structure is appropriate for the following paraphrases.

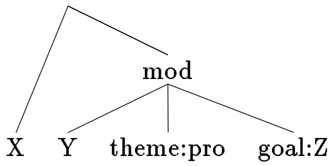
(161) a. the inundation of the city because of the breach of the dyke

b. the inundation of the city caused by the breach of the dyke

c. the city was inundated because the dyke broke

The same structure (160), where X refers to *Flöte* (recorder), Y to *aus* (from) and Z to *Holz* (wood) can be used for the translation into an adjective headed phrase (*a wooden recorder*) if we assume that the adjective is split into two components, a referential part *wood* and a part which is responsible for the predicative relation 'material' with the aspectual properties required for the modifier relation (cf. example (44) pg.33). X refers to *recorder*, Y to the abstract relation 'material' incorporated at CS into the adjective and made overt by the German preposition *aus* or the Chinese verb *yòng* (158c-d) and Z refers to *wood.*

(162)



This structure however implies that copulative structures of these adjectives are represented such that the copula *be* or German: *sein* is replaced by the predicative relation coming from the adjective. In other words, the copula is used at SS not as a support for the adjective, but as a syntactic support for the predicative component of the adjective, which is separated from its referential argument

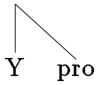
referential part. These adjectives are described lexically as specified in (163), where **psup** (predicative support) specifies the head of the predicative relation and **arg0** specifies the referential properties. Thus the adjective *silvery* will be represented at level IS as *silver* headed by the marker for MAT, e.g. the German preposition *aus*.

(163) @rule(b).

$$\left\{\begin{array}{l} \text{lex=silver,} \\ \text{head=}\left\{\begin{array}{l}\text{psup=}\left\{\text{slex='MAT'}\right\}, \\ \text{ehead=}\left\{\text{sem=}\left\{\text{EVENT}\right\}\right\}\end{array}\right\}, \\ \text{frame=}\left\{\text{arg0=}\left\{\text{head=}\left\{\text{ehead=}\left\{\text{sem=}\left\{\text{MATERIAL}\right\}\right\}\right\}\right\}\right\} \end{array}\right\}$$

With other adjectives (e.g. *afraid*), such a representation is not motivated, since such adjectives, when placed into the head position of a copulative sentence, can be mapped directly onto a verb or the predicative noun of a support verb construction. The sentences *He is afraid*, de:*Er fürchtet sich* and fr:*Il a peur*, can best be represented by a structure where the head position is occupied by Y, the adjective *afraid*, the verb *fürchen* or the noun *peur*.

(164)



We therefore have to separate the adjectives into two groups. The first group creates the predicative relation on its own (e.g. *afraid*), while the second is conflated with a predicative relation which has to be reconstructed at level IS. According to our distinction of *terms* and *non-terms* we separate an adjective into the two components (the referential part and the predicative part) if there are special markers for the expressed relation in the languages of the system. Adjectives like *silvery* are thus classified as non-terms and decomposed for translation. Possible relations and their markers are listed below in (165). Due to the existence of no appropriate functional marker we have to classify adjectives like *afraid* as **term**, i.e. one unit for transfer.

(165)

relation	English	French	German
CAUSE	*because of* *due to* *own to*	*á cause de*	*wegen*

relation	English	French	German
	because	*parce que*	weil
	since	*puisque*	da
	as		
	to cause	*causer*	*verursachen*
CONCESSION	*although*	*bien que*	*obwohl*
	though		
	inspite of	*malgé*	*trotz*
	despite		
FUNCT	*to function*	*fonctionner*	*funktionieren*
	as	*comme*	*als*
		en tant que	
MAT	*of*	*en*	*aus*
	out of		
	to make	*faire*	*machen*
EQUAL	*as*	*comme*	*wie*
	like		
	equal	*egal*	*gleich*
	to equal		*gleichen*
TRANSPORT	*by*	*en*	*mit*
CONDITION	if	si	falls
			sofern
			wenn
PURPOSE	*for*	*pour*	*um*
	for	*àfin*	*zu*
			zwecks
LOC	*be*	*être*	*sein*
	in, on, ...	*dans, sur, ...*	*in, auf, ...*
	to situate	*se trouver*	*sich befinden*
	to locate		

Chapter 11

Lexical Functions

In the same way as MT systems have to provide a special treatment for function words, a special treatment has to be developed for the cases where content words do not have their habitual meaning but behave in a similar fashion to function words. This is the case for the so-called **lexical functions** (LF)[1]. The essence of the notion of the Lexical Function is that a word A selects a second word B in order to realize a special syntactic or semantic structure related to A, which is called the LF. Take A to be *smoker.* In order to realize the high degree of this concept, which is not possible by morphological operations for English nouns, A selects B = *heavy* as its modifier in order to realize through the expression *heavy smoker* the high degree of *smoker.*

As functions cannot be translated into functions, A_{source} but not B_{source} can be translated literally into another language. A literal translation of B_{source} could possibly result in a mismatch with the specification of B_{target} found in the A_{target}. Thus, the Dutch *harde storm* (hard storm) is not correctly translated as *harter Sturm* in German, although *harde steen* (hard stone) is a *harter Stein* in German. Since the B for the LF Magnifier for *Sturm* is *schwer*, *harde storm* has to be translated as *schwerer Sturm.* Thus no translation rule of the type `hard=>schwer` is involved in the translation of this example. In the same way as in F-Structures, the total semantic value (e.g. the high degree) has to be calculated in analysis, removing the marker (e.g. *hard*). The TL then has to decide which function has to operate in order to express that meaning, with

[1] The concept of **Lexical Functions** has been developed by Mel'čuk in the framework of his Meaning⇔Text Model (cf. [Mel'čuk74], [Mel'čuk84], [Mel'čuk and Pertsov87], [Mel'čuk88], [Apresjan91]). Since then the importance of this concept for purposes of multilingual NLP has been more and more recognized (cf. [Magnúsdóttir88], [Bloksma and van der Kraan91], [Heylen92], [Heylen et al.92]), which finally resulted in the integration of this concept in a number of MT systems (cf. [Apresjan et al.89], [Apresjan et al.92a], [Heylen et al.93], [Streiter94], [Streiter95]).

magnifier
degree

the help of function words, lexical functions or lexicalizations. If a LF has to be employed in the TL as well, A of the target language has to choose its B according to its lexical specifications.

11.1 Magnifier

Magnifier is the name of a LF, where the degree of a concept is not expressed by a degree word, inflection or derivation, but by a content word, which in this restricted context is interpreted as functor. [Vinay and Darbelnet58] call these constructions 'locutions d'intensité', i.e. expression of intensity. German adjectives may form the high degree (elative) with the function word *sehr* or a specific noun (the magnifier) with which they form a compound (166). The difference between the analytic degree forming and the synthetic degree forming is one of speech style. For the treatment of style in CAT2 cf. [Streiter96]. Similar structures are found in English, examples of [Vinay and Darbelnet58] pg. 40. Nouns generally take adjectives as their magnifier (168) and verbs take adverbs (169).

(166) a. *German:*

sehr gemein = hundsgemein
very mean = 'dog-mean'

b. sehr kalt = arschkalt
very cold = 'arse-cold'

c. sehr trocken = furztrocken
very dry = 'fart-dry'

d. sehr groß = riesengroß
very big = 'giant-big'

e. sehr blöd = saublöd
very silly = 'sow-silly'

(167) a. stone deaf

b. stark mad

c. stark naked

d. dead tired

e. dripping wet

(168) a. *German:*

magnifier

herbe Kritik
sharp criticism
'harsh criticism'

b. saftige Rechnung
juicy bill
'enormous bill'

(169) a. innig lieben
heartfelt love
'to love deeply'

b. heftig schimpfen
heavily scold
'to scold severely'

The lexeme a word combines with to mark the high degree is marked in the lexical entry with the feature `magn=`{`lex=`_} . If no magnifier is specified in the lexicon, `magn=nil` is assigned per default.

(170)
$$\left\{\begin{array}{l}\text{lex=kritik,}\\ \text{head=}\left\{\text{magn=}\left\{\text{lex=herb}\right\}\right\}\end{array}\right\}$$

The semantic values aimed at by the operation of degree forming are the same as those triggered by analytic or synthetic operations of degree forming, expressed by the extended head feature `dvalue`.

(171)

`dvalue=pos`	positive	*bad*
`dvalue=comp`	comparative	*worse*
`dvalue=super`	superlative	*worst*
`dvalue=exz`	excessive	*too bad*
`dvalue=equal`	equality	*as bad*
`dvalue=acc`	acceptability	*bad enough*
`dvalue=elat`	elative	*very bad*
`dvalue='QU'`	degree questioned	*how bad*

Between CS and IS, the magnifier is removed and the `dvalue` of the head is calculated according to the `dvalue` of the magnifier. If the magnifier itself has the positive degree (e.g. *heavy smoker*, the magnified structure receives the value `elative`. In all other cases, the `dvalue` is copied from the magnifier onto the magnified.

(172) a. *German:*

ausführlich
detailed$_{POS}$
'in detail'

b. ausführlicher
detailed$_{COMP}$
'in greater detail'

c. ausführlichst
detailed$_{SUPER}$
'in the greatest detail'

d. zu ausführlich
detailed$_{EXZ}$
'in too great a detail'

e. so ausführlich
detailed$_{EQUAL}$
'in as great a detail'

f. ausführlich genug
detailed$_{ACC}$
'in great enough detail'

g. sehr ausführlich
detailed$_{ELAT}$
'in great detail'

h. wie ausführlich
detailed$_{QU}$
'in how great a detail'

11.2 Minifier

In the same way as the magnifier, the minifier represents a content word which is used to express the low degree of another content word. The semantic value triggered by this construction is `dvalue=low` (e.g. *a little bad*) Examples for this construction are:

minifier
generic support

(173) a. *German:*

leichte Rötung
light reddening

'slight reddening'

b. kleine Hitze
little heat
'low heat'

The lexical items that are employed as minifier are stored in the feature `mini={lex=_}`. As pointed out in [Streiter95], the value of the minifier cannot be derived from the value of the magnifier, so that two separate specifications are justified.

(174) @rule(b).

a.
$$\left\{\begin{array}{l}\text{lex=roeten,}\\ \text{head=}\left\{\begin{array}{l}\text{mini=}\left\{\text{lex=leicht}\right\},\\ \text{magn=}\left\{\text{lex=stark}\right\}\end{array}\right\}\end{array}\right\}$$

b.
$$\left\{\begin{array}{l}\text{lex=hitze,}\\ \text{head=}\left\{\begin{array}{l}\text{mini=}\left\{\text{lex=klein}\right\},\\ \text{magn=}\left\{\text{lex=gross}\right\}\end{array}\right\}\end{array}\right\}$$

c.
$$\left\{\begin{array}{l}\text{lex=sturm,}\\ \text{head=}\left\{\begin{array}{l}\text{mini=}\left\{\text{lex=leicht}\right\},\\ \text{magn=}\left\{\text{lex=schwer}\right\}\end{array}\right\}\end{array}\right\}$$

The treatment of these items corresponds to that of the magnifiers. They are removed between CS and IS and regenerated if syntactic forms of degree forming becomes necessary.

11.3 Generic Support

Generic Support as described in [Mel'čuk74] is a third type of LFs. Adjectives may be defective with respect to the possibility of deriving nouns from them by morphological means. In these cases, as for stylistic purposes, adjectives may build a nominal structure with the help of their hyperonyms (e.g. red→colour). In English for example, support nouns are required for some singular nationality adjectives. A 'support noun' such as *man* and *woman* must be inserted in order to establish reference. Further examples come from Russian and Serbo-Croatian for the concept of 'Russian' and 'Serbian'. Spanish frequently uses support nouns for colours:

(175) a. *German:*
ein Japaner - eine Japanerin
a Japan$_{MASC}$ - a Japan$_{FEM}$
'a Japanese **man** - a Japanese **woman**'

b. ein Weiser - eine Weise
a wise$_{MASC}$ - a wise$_{FEM}$
'a wise **man** - a wise **woman**'

c. *Russian:*
на русском языке

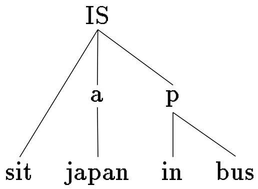
na russkom **jazyke**
on Russian language
'in Russian'

d. *Serbo-Croatian:*
na srpskom **jeziku**
on Serbian language
'in Serbian'

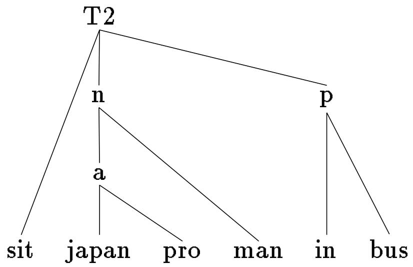
e. *Spanish:*
de color **naranja**
of colour orange
'orange'

In any of these cases the concept is expressed by an adjective plus a supporting noun, where the semantic space denoted by the adjective is included in the space denoted by the noun (cf. (62), pg.46). Comparing the inclusion relation of *Japanese* and *man* with the concept relation of *white car* in (61) pg.45, it becomes clear that we have to use the adjective as the base for the translation. Accordingly, the German noun *Japaner* is mapped onto the English adjective *Japanese* (cf. (176a)). The support noun *man* has to be regenerated according to monolingual lexical specifications between IS and CS (cf. (176b)).

(176) a.

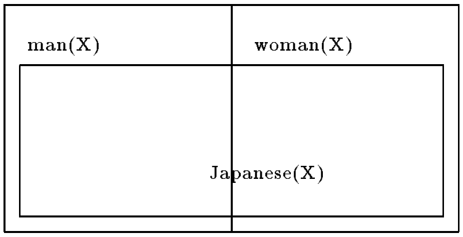


b.



If more than one support is possible for a given modifier (*Japanese* → *Japanese man, Japanese Woman, Japanese people*), the selection of the support may add a distinctive feature which must be compatible with the referential argument of the modifier. The inclusion relation is thus as illustrated below:

(177)



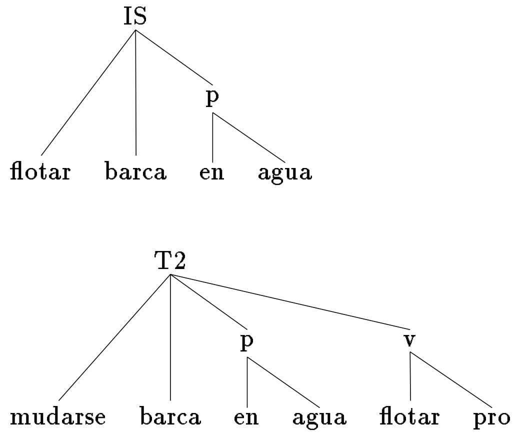
Thus restrictions on the referential argument of the adjective (`{'T'=hum,sex=male}` or `{animate=hum,sex=female}`) trigger the corresponding support noun, *man* or *woman*. Thus if in the following entry `sem=SEM` is instantiated by `sem=`{`concr=`{`sex=female`}}, only the lexical entry of *woman* can unify with the specification found in `nsup`.

$$\left\{\begin{array}{l} \text{string=japanese,}\\ \text{lex=japan,}\\ \text{frame=}\left\{\text{arg0} \mid \text{head=} \mid \text{ehead=}\left\{\text{sem=SEM}\right\}\right\},\\ \text{head=}\left\{\text{nsup=}\left\{\begin{array}{l}\text{frame=} \mid \text{head=} \mid \text{ehead=}\left\{\text{sem=SEM}\right\},\\ \text{head=}\left\{\text{lex=(man;woman)}\right\}\end{array}\right\}\right\}\end{array}\right\}$$

Generic support is not restricted to specific parts of speech. Verbs may form a generic support for sentencial modifiers by the same mechanism. We thus can account for the divergence described in (43b) pg.32 without a complex transfer rule, deriving (178b) from (178a).

support verb
constructions

(178) a. b.



Contrary to the LF Magnifier or Minifier, the generic support inserts a new head. This lexical function can thus be conceived of as a syntactic pattern of derivation (cf. [Streiter and Schmidt-Wigger95c]), which shows again the functional equivalence of syntactic and morphological operations.

11.4 Support Verb Construction

Another LF I want to present is the Support Verb Construction (SVC). This LF accounts for conflation phenomena as exemplified in (43a) pg.32. As the generic support, SVCs represent a syntactic style of derivation where (in most cases) a noun selects a verb (or its derivations) as its head. The support verb (SV) inherits the **predicative structure** from the noun, leaving open whether the arguments are expressed under the nominal or verbal head. As for the generic support I assume that the referential argument of the noun is shared with the SV. The verb then has the function to express the meaning which the noun bears but cannot express, i.e. tense and modality. As with all functions, SVCs cannot be translated compositionally. This is illustrated through the following examples taken from [Mesli91] (pg.4).

(179) *French-German:*

a. prendre une décision - einen Beschluß fassen
take a decision a decision grap
'to take a decision'

b. prendre l'initiative - die Initiative ergreifen
take the initiative the initiative grap
'to take the initiative'

c. prendre peur - Angst bekommen
take fear fear get
'to become afraid'

conversion

d. prendre des risques - Risiken eingehen
take of the risks risks enter
'to run a risk'

e. prendre de l'importance - an Bedeutung gewinnen
take of the importance at importance win
'to become important'

f. prendre la peine de - sich die Mühe machen
take a la effort himself the effort make
'to make an effort'

The fact that the verb has lost the **predicative structure** and serves mainly the purpose of marking tense, aspect, mood and modality, allows the verb to be replaced by other verbs, in order to receive new variants of the SVC with this predicative noun. In the example (180) the ambiguous verb *ankern*, which may be either transitive or intransitive, loses its ambiguity when realized by a SVC.

(180) *German:*

a. vor Anker gehen ⇔ ankern
before anchor go anchor
'to anchor'

b. vor Anker liegen ⇔ ankern
before anchor ly anchor
'to be anchored'

c. den Anker lichten
the anchor lift
'to weigh'

The basic assumption underlying our treatment of SVCs is that the choice of the SV reflects one semantic variant of the underlying predicative relation. *Einen Befehl bekommen* (to receive an order) for example, expresses the converse (change of direction) of the situation expressed by the expression *Einen Befehl geben* (to give an order), just as `to buy` expresses the converse relation of `to sell` or a passive sentence to an active sentence, where subject and object are exchanged in SS, in order to change the relation between the reference point and the located point (cf. [Lyons73], [Herrmann-Dresel87] [Langacker87]). The correct assingment of the actants to the argument slots is accounted for by the semantic role as explained in Chapter 8.

aktionsart
argument transfer
support verb

Other variants of SVCs differ wrt the Aktionsart they assign to the whole process. Since the field of Aktionsarten is very large and the number of Aktionsarten apparently unlimited (cf. [Švedova80], [Schwenk91]), our treatment concentrates on Aktionsarten which may be paraphrased with the help of verbs like *to start, to continue, to stop* etc..., without excluding other Aktionsarten a priori.[2]

(181)

`akt=neut`	neutral	*Angst haben* 'to be afraid'
`akt=incho`	inchoative (ingressive)	*Angst bekommen* 'to become afraid'
`akt=term`	terminative (egressive)	*Angst verlieren* 'to lose one's fear'
`akt=contin`	continuitive	*bei der Behauptung bleiben* 'to stick to one's claim'

As explained above, the **predicative structure** of SVCs are determined by the non-verbal, predicative part of the SVC. This element assigns semantic roles and defines the selectional restrictions on the arguments. These are copied onto the frame specification of the SV the moment the SV subcategorizes for the predicative noun. This operation is known as the **argument transfer** (cf. [Grimshaw and Mester88]), a syntactic variant of the morphological argument inheritance. Contrary to argument inheritance, the arguments are not necessarily realized as complements to the verb but may be realized as complements to the predicative noun if the syntactic completeness requirements of the verb do not require a special argument to be filled (cf. [Kuhn94])

The verb and the predicative noun assign default `case` and `pform` to the arguments and to the predicative element. These values are not percolated from the noun to the verb since they represent syntactic restrictions which are different depending on whether the arguments of the SVC are realized syntactically under the predicative noun or under the SV.

The lexical entry of a SV and its morphological derivatives is exemplified in (182). The entry already contains all variable linking necessary to copy the arguments of the predicative noun (`role=pred`) onto the argument slots of the SV.

[2] Note my notion of continuitive Aktionsart subsumes the so-called **iterative** Aktionsart e.g. *John multiplied his attacks against Mary.* ([Danlos et al.90] pg.20). Iterativity and continuity refer to the same kind of operators. What continuity is for an unbounded process iterativity is for a bounded process. We therefore can limit the number of Aktionsarten to 4.

(182) @rule(b).

$$\left\{\begin{array}{ll}
\text{lex=} & \text{ergreifen,}\\
\text{head=} & \text{HEAD\&}\\
 & (\left\{\text{cat=v}\right\};\left\{\begin{array}{l}\text{cat=n,}\\ \text{deriv=}\left\{\text{aff=ung}\right\}\end{array}\right\};\left\{\begin{array}{l}\text{cat=n,}\\ \text{deriv=}\left\{\text{aff=inf}\right\}\end{array}\right\};\left\{\begin{array}{l}\text{cat=a,}\\ \text{deriv=}\left\{\text{aff=bar}\right\}\end{array}\right\})\\
\text{frame=} & \left\{\begin{array}{l}
\text{arg1=}\left\{\begin{array}{l}\text{role=ROLE1,}\\ \text{head=}\mid\text{ehead=}\left\{\begin{array}{l}\text{sem=SEM1,}\\ \text{tense=TENSE1,}\\ \text{mood=MOOD1}\end{array}\right\}\end{array}\right\},\\
\text{arg2=}\left\{\begin{array}{ll}\text{role=} & \text{pred,}\\
\text{frame=} & \left\{\begin{array}{l}
\text{arg1=}\left\{\begin{array}{l}\text{role=ROLE1,}\\ \text{head=}\mid\text{ehead=}\left\{\begin{array}{l}\text{sem=SEM1,}\\ \text{tense=TENSE1,}\\ \text{mood=MOOD1}\end{array}\right\}\end{array}\right\},\\
\text{arg2=}\left\{\begin{array}{l}\text{role=ROLE3,}\\ \text{head=}\mid\text{ehead=}\left\{\begin{array}{l}\text{sem=SEM3,}\\ \text{tense=TENSE3,}\\ \text{mood=MOOD3,}\\ \text{case=CASE3,}\\ \text{pform=PF3}\end{array}\right\}\end{array}\right\},\\
\text{arg3=}\left\{\begin{array}{l}\text{role=ROLE4,}\\ \text{head=}\mid\text{ehead=}\left\{\begin{array}{l}\text{sem=SEM4,}\\ \text{tense=TENSE4,}\\ \text{mood=MOOD4,}\\ \text{case=CASE4,}\\ \text{pform=PF4}\end{array}\right\}\end{array}\right\},\\
\text{pred=}\mid\text{vsup=HEAD}
\end{array}\right\}
\end{array}\right\},\\
\text{arg3=}\left\{\begin{array}{l}\text{role=ROLE3,}\\ \text{head=}\mid\text{ehead=}\left\{\begin{array}{l}\text{sem=SEM3,}\\ \text{tense=TENSE3,}\\ \text{mood=MOOD3,}\\ \text{case=CASE3,}\\ \text{pform=PF3}\end{array}\right\}\end{array}\right\},\\
\text{arg4=}\left\{\begin{array}{l}\text{role=ROLE4,}\\ \text{head=}\mid\text{ehead=}\left\{\begin{array}{l}\text{sem=SEM4,}\\ \text{tense=TENSE4,}\\ \text{mood=MOOD4,}\\ \text{case=CASE4,}\\ \text{pform=PF4}\end{array}\right\}\end{array}\right\}
\end{array}\right\}
\end{array}\right\}$$

In this example, the SV subcategorizes the predicative noun as its second argument, by means of which the noun is assigned accusative case in active and nominative case in passive sentences. In other SVCs, other argument slots are opened for the predicative noun.

The entry of the predicative noun (`lex=macht`), predicts the SV necessary for its predicative use in a sentence (`vsup=...`) and the aktionsart this verb is related to, as well as its own form when appearing together with the predicative

predicative noun negation

verb (`xpred=...`).[3]. This information (e.g. `pform, type` and `num`) cannot be coded directly as an extended head feature of the noun since the predicative properties can be delegated to a dependent relative pronoun, in which case these constraints apply to the relative pronoun and not to the predicative noun itself (e.g. *He was proud* of *the decision which his daughter had made*). The value of `vsup` are head features and the value of `xpred` are extended head features. The aspectual structure of the noun is unified with that of the SV, by way of which the semantic tense and aspect are shared between the predicative noun and the SV. In addition, the SV and the predicative noun share information about the scope of negation.

(183) @rule(b).

$$\left\{\begin{array}{l}
\text{head}=\left\{\begin{array}{l}\text{lex=macht cat=n,}\\ \text{ehead}=\left\{\begin{array}{l}\text{sem=SEM\&}\left\{\text{abstract= | temp= | aspect=incho}\right\}\text{,}\\ \text{sNEG=SNEG}\end{array}\right\}\end{array}\right\}\text{,}\\
\text{frame}=\left\{\begin{array}{l}
\text{arg1}=\left\{\begin{array}{l}\text{role=agent,}\\ \text{head}=\left\{\text{ehead}=\left\{\text{cat=n}\right\}\right\}\end{array}\right\}\text{,}\\
\text{arg2}=\left\{\text{role=nil}\right\}\text{,}\\
\text{arg3}=\left\{\text{role=nil}\right\}\text{,}\\
\text{pred}=\left\{\begin{array}{l}\text{vsup}=\left\{\begin{array}{l}\text{lex=ergreifen,}\\ \text{ehead}=\left\{\begin{array}{l}\text{sNEG=SNEG,}\\ \text{sem=SEM}\end{array}\right\}\end{array}\right\}\text{,}\\ \text{xpred}=\left\{\text{pform=nil,type=def,num=sing}\right\}\end{array}\right\}
\end{array}\right\}
\end{array}\right\}$$

The result of the subcategorization of the predicative noun by the SV is represented in the following structure, where the argument slot for the subject is still accessible.

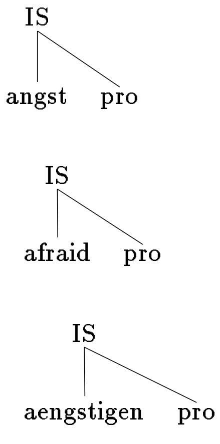
[3] In the treatment implemented in CAT2 the predicative noun selects the SV together with the aktionsart this verb is associated with. [Krenn and Erbach94](pg.389) suggest the SV select the predicative noun and have a fixed aktionsart. This approach, however, is not only intractable for generation where the predicative noun somehow has to generate its SV, but implies that every SV has a list of hundreds of lexemes of predicative nouns it may select and third, that multiple entries become necessary if the same SV is used for different aktionsarten (cf. (179)).

(184)

$$\left\{\begin{array}{ll}
\text{head}= & \text{HEAD\&} \\
& \left\{\begin{array}{l}\text{lex=ergreifen,cat=v} \\ \text{ehead} \mid \text{sem} \mid \text{abstract} \mid \text{temp} \mid \text{aspect} \mid \text{akt=incho}\end{array}\right\}, \\
\text{frame}= & \left\{\begin{array}{l}
\text{arg1}=\left\{\begin{array}{l}\text{role=agent,} \\ \text{head=} \mid \text{ehead=} \mid \text{cat=n}\end{array}\right\}, \\
\text{arg2}=\left\{\begin{array}{l}
\text{role=pred,} \\
\text{head}=\left\{\text{lex=macht}\right\}, \\
\text{frame}=\left\{\begin{array}{l}
\text{arg1}=\left\{\begin{array}{l}\text{role=agent,} \\ \text{head} \mid \text{ehead} \mid \text{cat=n}\end{array}\right\}, \\
\text{arg2}=\left\{\text{role=nil}\right\}, \\
\text{arg3}=\left\{\text{role=nil}\right\}, \\
\text{pred} \mid \text{vsup=HEAD,} \\
\text{xpred}=\left\{\begin{array}{l}\text{type=def,} \\ \text{pform=nil,} \\ \text{num=sing,} \\ \text{cat=n}\end{array}\right\}
\end{array}\right\}
\end{array}\right\}, \\
\text{arg3}=\left\{\text{role=nil}\right\}, \\
\text{arg4}=\left\{\text{role=nil}\right\}
\end{array}\right\}
\end{array}\right\}
\left[\begin{array}{l}\left\{\text{lex=macht}\right\}, \\ \left\{\text{lex=ergreifen}\right\}\end{array}\right]$$

At level IS, the SV is no longer represented. The semantic predicate (the predicative noun) enters the position of the governor (cf. Chapter 1). By these means we obtain compatible structures for German:*Er hat Angst* (He has fear), English:*He is afraid* and German:*Er ängstigt sich* (He fears REFL).

(185)



Modifiers and arguments of the predicative noun can (or must) be translated as sentence modifiers. This is illustrated in (186a) where the modifier is realized at the level of the SV or in (186b) at the level of the predicative noun (examples default argument inflation

from [Danlos et al.90] pg.31). Therefore, all arguments and modifiers that were originally adjuncted to the noun are raised to the sentence level.

(186) a. John frequently gives advice to Mary
b. John gives frequent advice to Mary

11.5 Default Arguments

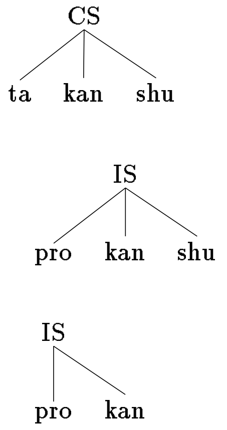
A further type of LFs I want to describe is the so-called default argument. This lexical function serves the purpose of filling the argument position which is otherwise interpreted as pronominal. In Chapter 3.9, I characterized such structures as **inflation**, the appearance of a syntactic constituent to which no **meaning** can be assigned. Examples are given in (46) pg.33. Default arguments are stored in the `default` feature in the `frame` description.

(187) @rule(b).

$$\left\{\begin{array}{l}\text{lex=kàn,}\\ \text{frame=}\left\{\text{default=}\left\{\text{lex=shū}\right\}\right\}\end{array}\right\}$$

In analysis the indirect object is optionally removed from the structure if it unifies with the default description, so that two IS structures will result from it, representing the ambiguity of that expression.

(188)



In generation the default argument is generated in postverbal position, if no other indirect object is present.

11.6 Copulative Structures

copulative structures
copulative verb

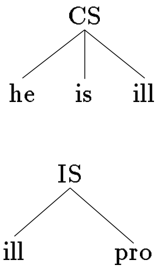
Copulative Structures have already been mentioned in Chapter 10. Copulative structures placing an adjective in a predicative position are treated in our implementation at level CS along the same lines as SVC, implying the adjective to be the semantic predicate of the sentence. The lexical entries of the copula is equivalent to those of the SV, describing an argument shift from the external argument position of the adjective to the first argument position of the verb. Supplementary arguments can be realized following the description given by the adjectival frame.

(189) @rule(b).

$$\left\{\begin{array}{l}
\text{lex=be,}\\
\text{head=}\left\{\text{cat=v}\right\}\text{,}\\
\text{frame=}\left\{\begin{array}{l}
\text{arg1=}\left\{\begin{array}{l}
\text{role=ROLE1,}\\
\text{head=}\mid\text{ehead=}\left\{\begin{array}{l}\text{sem=SEM1,}\\ \text{tense=TENSE1,}\\ \text{mood=MOOD1}\end{array}\right\}
\end{array}\right\}\text{,}\\
\text{arg2=}\left\{\begin{array}{l}
\text{role=pred,}\\
\text{head=}\left\{\text{ehead=}\left\{\text{cat=a,voice=pass}\right\}\right\}\text{,}\\
\text{frame=}\left\{\begin{array}{l}
\text{arg2=}\left\{\begin{array}{l}
\text{role=ROLE1,}\\
\text{head=}\mid\text{ehead=}\left\{\begin{array}{l}\text{sem=SEM1,}\\ \text{tense=TENSE1,}\\ \text{mood=MOOD1}\end{array}\right\}
\end{array}\right\}\text{,}\\
\text{arg3=}\left\{\begin{array}{l}
\text{role=ROLE3,}\\
\text{head=}\mid\text{ehead=}\left\{\begin{array}{l}\text{sem=SEM3,}\\ \text{tense=TENSE3,}\\ \text{mood=MOOD3,}\\ \text{case=CASE3,}\\ \text{pform=PF3}\end{array}\right\}
\end{array}\right\}\text{,}\\
\text{pred=}\mid\text{vsup=HEAD}
\end{array}\right\}
\end{array}\right\}\text{,}\\
\text{arg3=}\left\{\begin{array}{l}
\text{role=ROLE3,}\\
\text{head=}\mid\text{ehead=}\left\{\begin{array}{l}\text{sem=SEM3,}\\ \text{tense=TENSE3,}\\ \text{mood=MOOD3,}\\ \text{case=CASE3,}\\ \text{pform=PF3}\end{array}\right\}
\end{array}\right\}
\end{array}\right\}
\end{array}\right\}$$

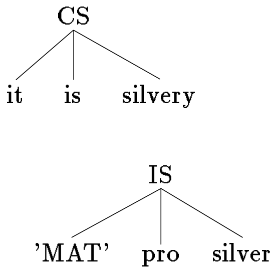
At level IS the copula is dropped in cases where the adjective has an autonomous predicative structure (i.e. *afraid* and the adjective enters the position of the verb). This treatment is identical for the treatment of SVC at level IS (cf. (185)).

(190)



For the class of adjectives which are composed of a referential part (e.g. *silver* in *silvery*) and a predicative part (e.g. 'MAT'), the copula is replaced at level IS by a representative of the abstract predicative relations.

(191)



Summary and Conclusions

Summary

In this thesis I assume that translation is meaning-preserving in the contexts where MT is currently used, i.e. technical and scientific types of texts. I show that any attempt to handle divergences between languages by reference to form or functions, instead of meaning, must necessarily fail, as the transfer rules employed for this purpose are local rules which cannot take into account the totality of the constraints of the target language. Transfer rules which refer to the structure of a sentence instead of the meaning, may handle isolated phenomena of divergences between languages but cannot account for the full range of constraints and interactions of these phenomena. It is therefore only natural to use meaning for the modeling of translation, as the mapping of various surface structures to and from one type of meaning representation is tractable.

The meaning expressed in the surface structure of a language is encoded in the form of possible choices of that language. Whenever a choice between different surface structures is possible a change in meaning results from it. The choices a language offers and the associated meaning are represented in the form of functions. In other words, functions map choices in surface structures onto meaning (in analysis) and meaning onto surface structures (in synthesis).

Most MT systems, however, even if they abstract away from the form of a language, use functions for the purpose of translation. This is not a promising approach for a multilingual system, as the choice of the appropriate function in the target language depends on the interaction of the form and all constraints on the functions of the target language. These constraints cannot be predicted from the function used in the source language and cannot be correctly described by local translation rules.

Following this I analyze different types of functions, classified according to their linguistic realizations. As a first example of a function which is commonly referred to in translation rules I mention the 'concept' which, I argue, is not a meaning representation but a language-specific function which foregrounds a region in a notional domain. Therefore, the concept cannot be directly mapped between languages but must be reconstructed in the target language.

Functions realized via function words are analysed in the Functional Head–Structure with the help of the extended head feature convention, allowing the semantic contributions coming from the function words to be added to the lexical head. A possible meaning representation triggered by function words is developed for articles, where special attention is given to the fact that the meaning representation must be applicable to all parts of speech.

In Argument Structures one dimension of meaning, i.e. the thematic roles of arguments, is mapped onto different surface structures, depending on the different realizations of the predicate. As the realization of the arguments depends on the predicate in the same way that the realization of the predicate depends on the predicates for which argument positions have to be supplied, and both depend on the possibilities of realizing the necessary functional marking, no algorithm can be realized which matches the syntactic functions of the source language directly onto the syntactic functions of the target language.

Pronouns are functions which connect a concept to a textual or contextual entity, in order to achieve coherence and actualization. As pronouns may be covert, I restricted myself to the reconstruction of covert pronouns in argument positions.

Modifier structures are analyzed as predicative structures where the predicate is the syntactic non-head, resulting in an anaphoric structure with an overt or covert pronoun. In order to translate freely all types of modifiers the covert pronoun has to be reconstructed and conflated concepts have to be decomposed into their basic terms.

Lexical Functions finally are functions which use content words as markers instead of functional markers. These content words are coded in the lexical entries which realize these functions. As with all functions, a direct translation is not possible. Magnifiers, Minifiers, Support Verbs and Generic Supports are introduced and discussed with respect to their implementation and relevance for translation, showing how the mechanism of lexical functions can cope with divergences between languages which are otherwise not tractable.

Conclusions

In this thesis I have shown that the translation process cannot be modeled by reference to functions, which is what most MT systems do. Instead the modeling has to refer to meaning. The meaning must necessarily be designed such that it is applicable to all parts of speech and all functions the different parts of speech may employ. At the same time I show through the examples of implementation that this approach, ambitious as it might seem, is not only feasible, but finally the most straightforward one, as the process of translation follows regular and understandable lines of analysis and generation.

Further Outlook

The implementation of the CAT2 MT system as described in this thesis is continued in the context of different national and international projects, a description of which can be found at URL:http://www.iai.uni-sb.de/cat2/home.html. Within these projects, the ongoing activities are (i) the integration of new language modules (e.g. Arabic cf. [Pease and Boushaba96]), (ii) the enlargement of existing lexicons, (iii) the coverage of until now untreated semantic and syntactic phenomena, (iv) the adaptation of the system to special user requirements for the medical, banking and automobile sector and (v) as a consequence, the integration of knowledge, i.e. domain-specific meaning distinctions, into the general meaning description presented in this thesis.

actualization
ALEP
ANTHEM

Glossary

Actualization

By actualization I understand the passage from the virtual system (la langue) to the actual process (la parole) by the embedding of the propositional content in a complex system of relations that are based on the speech act (cf. [Greimas and Joseph89], [Mainguenau81], [Bouscaren and Chuquet87], [Danon-Boileau87], [Culioli90]). While the verbal concept actualizes the proposition with the help of tense values, nominal concepts cannot actualize on their own the denoted event, i.e. they need a support verb. Tense, aspect and mood belong to the actualizing function of the verb (cf. [Mesli and Bresson92]).

ALEP

The Advanced Language Engineering Platform (ALEP), an initiative of the EC, provides the natural language research and engineering community in Europe with a general purpose research and development environment. The environment is intended to ease and speed up the transitions from research to laboratory prototype, and from prototype to marketable product. It is designed as a basic, low-cost, non-proprietary platform for a broad range of research and technology development activities related to NLP (cf. URL: http://www.iai.uni-sb.de/alep). For more information about ALEP cf. [Alshawi et al.91], [ALE93], [Theofilidis93], [Schütz95] and [Simpkins95].

ANTHEM

ANTHEM develops a prototype of a natural language interface that allows users of Healthcare Information Systems to enter medical diagnostic expressions in a multilingual environment. Within ANTHEM, CAT2 is used to analyze Dutch and French diagnostic expressions. The system (i) translates these into German, Dutch and French and (ii) produces a semantic representation which is then passed to the ANTHEM Expert System for automatic coding in ICD*.

interlingual approach
sublanguage approach
CHARON
ETAP
Eurotra

For both purposes the interlingual approach has been assumed, which consists of a set of basic word-concepts identified by their SNOMED* code and a limited set of semantic relations which link word-concepts to form complex statements. By the integration of a semantic model representing the medical knowledge necessary to interpret the diagnostic expressions the number of spurious objects can be considerably reduced. A feature structure is used as an interface to an expert system, which checks the diagnosis for internal consistency and allows automatic encoding of the diagnostic statements. For an overview of this project cf. [Ceusters et al.94b], [Ceusters et al.94c], [Ceusters et al.94a], [Streiter and Schmidt-Wigger95a], [Streiter and Schmidt-Wigger95b]).

CHARON

Charon is an experimental MT system developed between 1985 and 1992 at the University of Stuttgart in order to test the possible applications of LFG to MT. The developed language pair is German ⇒ French. The basic mechanism used within this approach is that of codescription, which maps representations of one level to the representation of another level. The codescriptions are stated in the lexicon and in the c-structure rules of the source languages. For an overview of the system cf. [Kaplan et al.89]. A critical appraisal of this approach can be found in [Sadler93] and [Butt94].

ETAP

ETAP is an MT system developed at the Akademija Nauk Russii. The system integrates some aspects of the Meaning⇔Text Model, as developed by Mel'čuk, Pertsov and Apresjan ([Mel'čuk84], [Mel'čuk and Pertsov87], [Mel'čuk88]) and of its variant called *integral'nij slovar'* (Integral Dictionary) (cf. [Apresjan91]. Developed language pairs are French⇒Russian (ETAP1), English⇒Russian (ETAP2) and English⇔Russian (ETAP3). For more information about the system cf. [Apresjan et al.89], [Apresjan et al.92a], [Apresjan et al.92b], [Apresjan et al.93].

EUROTRA

In order to cope with the enormous translation load of European industry and trade as well as of the EC institutions themselves, the Commission launched in 1983 a multilingual machine translation project called Eurotra. The technical objective of Eurotra was the creation of a machine translation prototype capable of dealing with all nine official Community languages (Danish, Dutch, English, French, German, Greek, Italian, Portuguese and Spanish) in all possible directions, thus forming 72 language pairs. The project finished in 1992. Although this project could not develop a functional prototype, the linguistic staff of the Eurotra groups did very valuable research into morphology, syntax and semantics, resulting in one of the best documentations of NLP systems, the Eurotra Reference Manual [ERM90].

ICD
WHO
MPRO
SNOMED

ICD

ICD, the International Classification of Diseases is a system developed by the World Health Organization, and is designed for the classification of morbidity and mortality information for statistical purposes. For further information about ICD cf. [WHO93].

MPRO

MPRO is the morphological analyzer developed at IAI by Heinz-Dieter Maas. The system is implemented in Sicstus Prolog and, with minor modifications, runs under SWI-Prolog and YAP as well. The morphological dictionary of mpro contains morphemes (and their allomorphs and writing variants), all with extensive descriptions of their behaviour in word formation processes (derivation, compounding, and inflection). A more detailed description of the system can be found in [Seewald93], [Maas94b], [Maas94a].

SNOMED

SNOMED, the Systematized Nomenclature of Medicine is one of the most exhaustive systems used to codify elementary concepts in medicine. It joins seven types of medical elements (topography, morphology, etiology, function, disease, procedure and occupation) to form combinatorial expressions. It is used world-wide and represents the great majority of all symptoms and phenomena needed for healthcare data handling. For further information about SNOMED cf. [Côté et al.93], [Rothwell et al.93].

Bibliographical Abbreviations

COLING-84 = Coling-84: 10th International Conference on Computational Linguistics, 22nd Annual Meeting of the Association for Computational Linguistics, Proceedings of Coling 84, 2-6 July 1984, Stanford University, California.

COLING-86 = Coling-86: 11th International Conference on Computational Linguistics, Proceedings of Coling 86, 25-29th August 1986, Bonn.

COLING-90 = Karlgren, H. (ed.) Coling-90: papers presented to the 13th International Conference on Computational Linguistics, August 1990, Helsinki, Finland.

COLING-92 = Boitet, C. (ed.) Coling-92: Proceedings of the 14th International Conference on Computational Linguistics, August 1992, Nantes, France.

COLING-94 = COLING-94, The 15th International Conference on Computational Linguistics. Proceedings, August 5-9 1994, Kyoto, Japan.

ERM = Eurotra Reference Manual, edition 7.0, 1990

IAI WP = Working Paper, IAI, Institut der Gesellschaft zur Förderung der angewandten Informationsforschung e.V. an der Universität des Saarlandes, Martin-Luther-Straße 14, 66111 Saarbrücken, BRD.

KONVENS '94 = Harald Trost (ed.) KONVENS '94,Tagungsband, 2. Konferenz ”Verarbeitung natürlicher Sprache” Wien, 28-30 September 1994 Springer-Verlag, Berlin.

Meta-92 = Meta 37(4), December 1992.

MT TMI = Sergei Nirenburg (ed.), Machine Translation. Theoretical and Methodological Issues, Studies in Natural Language Processing, Cambridge

University Press, Cambridge, 1987.

TMI-90 = The Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages, 11-13 June, University of Texas, Austin.

TMI-95 = Proceedings of the Sixth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages, 5-7 July, Leuven, Belgium.

Bibliography

[Abb and Maienborn94] Bernd Abb and Claudia Maienborn. 1994. Adjuncts in HPSG. In *KONVENS, '94*.

[Adamec66] Přmysl Adamec. 1966. *Porjadok slov v sovremennom russkom jazyke*. Československé akademie věd, Prag.

[ALE93] 1993. *ALEP Document Package, Vol. I and II*. Luxembourg.

[Allegranza and Soma93] Valerio Allegranza and Enzo Soma. 1993. The E-Star grammatical formalism. Working paper, Gruppo DIMA, Torino.

[Alshawi et al.91] H. Alshawi, Doug J. Arnold, Ralf Backofen, D.M. Carter, J. Lindop, Klaus Netter, Steve G. Pulman, J. Tsujii, and Hans Uszkoreit. 1991. ET6/1 formalism study - final report. DG XIII, CEC, Luxembourg.

[Apresjan et al.89] Jurij D. Apresjan, Igor M. Boguslavskij, Leonid L. Iomdin, Alexandre V. Lazurskij, Vladimir Z. Sannikov, and Leonid L. Tsinman. 1989. *Lingvističeskoe obespečenie sistemy ETAP-2*. Izdatel'stvo "Nauka", Moskva.

[Apresjan et al.92a] Jurij D. Apresjan, Igor M. Boguslavskij, Leonid L. Iomdin, Alexandre V. Lazurskij, Vladimir Z. Sannikov, and Leonid L. Tsinman. 1992a. ETAP-2: The linguistics of a Machine Translation system. *Meta*, 37(1):97–112.

[Apresjan et al.92b] Jurij D. Apresjan, Igor M. Boguslavskij, Leonid L. Iomdin, Alexandre V. Lazurskij, Vladimir Z. Sannikov, and Leonid L. Tsinman. 1992b. *Lingvističeskij processor dlja složnyx informacionnyx sistem*. Izdatel'stvo "Nauka", Moskva.

[Apresjan et al.93] Jurij D. Apresjan, Igor M. Boguslavskij, Leonid L. Iomdin, Alexandre V. Lazurskij, Vladimir Z. Sannikov, and Leonid L. Tsinman. 1993. Système de traduction automatique ETAP. In P.Buillon and A.Clas, editors, *La Traductique*. Les Presses de l'Université de Montréal, Montréal.

[Apresjan91] Jurij D. Apresjan. 1991. Ob integral'nom slovare russkogo jazyka. *Semiotika i informatika*, 32:3–15.

[Arnold and Sadler87] Doug Arnold and Louisa Sadler. 1987. Non-Compositionality and translation. Working paper in language processing no.1, University of Essex, Department of Language and Linguistics, Essex.

[Arnold and Sadler90] Doug Arnold and Louisa Sadler. 1990. The theoretical basis of MiMo. *Machine Translation*, 5(3):195–222.

[Arnold and Sadler91] Doug Arnold and Louisa Sadler. 1991. Transfer formalisms. *Machine Translation*, 6(3):201–214, September.

[Arnold and Sadler92] Doug Arnold and Louisa Sadler. 1992. Unification and Machine Translation. *Meta-92*, pages 657–680.

[Arnold et al.86] Doug Arnold, S. Krauwer, Rosner M., Louis des Tombe, and G.B. Varile. 1986. The <C,A>,T framework in EUROTRA: A theoretically committed notation for MT. In *COLING-86*, pages 297–303.

[Balari et al.90] S. Balari, L. Damas, N. Moreira, and G.B. Varile. 1990. CLG(n): Constraint logic grammars. In *COLING-90*.

[Bateman89] John A. Bateman. 1989. Upper modelling for machine translation: a level of abstraction for preserving meaning. unpublished draft, September.

[Bátori86] István Bátori. 1986. Die Paradigmen der Maschinellen Sprachübersetzung. In István Bátori and Heinz J. Weber, editors, *Neue Ansätze in maschineller Sprachübersetzung: Wissensrepräsentation und Textbezug*, Sprache und Information, Band 13. Niemeyer, Tübingen.

[Beaven92] John L. Beaven. 1992. Shake-and-bake Machine Translation. In *COLING-92*, pages 603–609.

[Bech and Nygaard88] Anneliese Bech and A. Nygaard. 1988. The E-framework: A formalism for natural language processing. In *Proceedings of COLING'88*, pages 36–39. Budapest.

[Bech90] Anneliese Bech. 1990. The virtual machine. In *ERM*.

[Bech91] Anneliese Bech. 1991. Description of the Eurotra framework. In Steven Krauwer Charles Copeland, Jacques Durand and Bente Maegaard, editors, *The Eurotra Formal Specifications*, Studies in Machine Translation and Natural Language Processing, Volume 2. Commission of the European Communities, Brussels & Luxembourg.

[Bennett95] Paul Bennett. 1995. *A course in Genral Phrase Structure Grammar*. Computational Linguistics. UCL Press, London.

[Birkenmaier79] Willy Birkenmaier. 1979. Artikelfunktionen in einer artikellosen Sprache. Studien zur nominalen Determination im Russischen. In Dimitij Tschiževskij, editor, *Forum Slavicum*. Wilhelm Fink Verlag, München.

[Blake90] Barry J. Blake. 1990. *Relational Grammar*. Croom Helm Linguistic Theory Guides. Routledge, London and New York.

[Bloksma and van der Kraan91] Laura Bloksma and Mark van der Kraan. 1991. Collocations and categorial grammar. In Jan van Eijck and Wilfried Meyer Viol, editors, *Computational Linguistics in the Netherlands*, pages 12–25. Utrecht.

[Boitet et al.85] Christian Boitet, P. Guillaume, and M. Quezel-Ambrunaz. 1985. A case study in software evolution: From Ariane-78.4 to Ariane-85. In *Proceedings of the Conference on Theoretical and Methodological Issues of Natural Languages, Colgate University*, pages 1–14. Hamilton, NY.

[Boitet88] Christian Boitet. 1988. Pros and cons of the pivot and transfer approaches in multilingual Machine Translation. In Dan Maxwell, Klaus Schubert, and Toon Witkam, editors, *New Directions in Machine Translation*. Foris Publication, Dordrecht - Holland.

[Booth and Gerritzen89] Cheri Booth and Christian Gerritzen, editors. 1989. *Lexikon der englischen Umgangssprache*. Bechtermünz Verlag GmbH, Eltville am Rhein.

[Bosco Coletsos88] Sandra Bosco Coletsos. 1988. *Storia della lingua Tedesca*. Garzanti, Milano.

[Bouma and Nerbonne94] Gosse Bouma and John Nerbonne. 1994. Lexicons for feature-based systems. In *KONVENS, '94*.

[Bouscaren and Chuquet87] Janine Bouscaren and Jean Chuquet. 1987. *Grammaire et Textes Anglais, Guide pour l'analyse linguistique*. OPHYS, Paris.

[Bouvier77] Irmgard Bouvier, editor. 1977. *RAK, Regeln für die alphabetische Kategorisierung*. Dr. Ludwig Reichert Verlag, Wiesbaden.

[Bresnan (ed.)82] Joan Bresnan (ed.). 1982. *The Mental Representation of Grammatical Relations*. The MIT Press, Cambridge, Massachusetts.

[Bresnan82a] Joan Bresnan. 1982a. Control and complementation. In Joan Bresnan, editor, *The Mental Representation of Grammatical Relations*. The MIT Press, Cambridge, Massachusetts.

[Bresnan82b] Joan Bresnan. 1982b. The passive in lexical theory. In Joan Bresnan, editor, *The Mental Representation of Grammatical Relations.* The MIT Press, Cambridge, Massachusetts.

[Brew92] Chris Brew. 1992. Letting the cat out of the bag: generation for shake-and-bake MT. In *COLING-92*, pages 610–616.

[Brown and Frederking95] Ralf Brown and Robert Frederking. 1995. Applying statistical English language modelling to symbolic Machine Translation. In *TMI-95.*

[Brown et al.90] Peter F. Brown, J. Cocke, Stephen A. Della Pietra, Vincent J. Della Pietra, F. Jelinek, Robert L. Mercer, and Roossin P.S. 1990. A statistical approach to language translation. *Computational Linguistics*, 16:79–85.

[Brown et al.93] Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. 1993. The mathematics of statistical Machine Translation: Parameter estimation. *Computational Linguistics*, 32:263–311.

[Bunt85] Harry C. Bunt. 1985. *Mass Terms and Model-theoretic Semantics.* Cambridge Studies in Linguistics. Cambridge University Press, Cambridge.

[Butt94] Miriam Butt. 1994. Machine Translation and complex predicates. In *KONVENS, '94.*

[Carpenter et al.91] Bob Carpenter, Carl Pollard, and Alex Franz. 1991. The specification and implementation of constraint-based unification grammars. In *Proceedings of the Second International Workshop on Parsing Technology*, pages 143–153, Cancun, Mexico.

[Carpenter92] Bob Carpenter. 1992. *The Logic of Typed Feature Structures with Applications to Unification-based Grammars, Logic Programming and Constraint Resolution*, volume 32 of *Cambridge Tracts in Theoretical Computer Science.* Cambridge University Press, New York.

[Carulla94] Marta Carulla. 1994. Relational adjectives in the translation from Germanic nominal compounds into Romance languages. In Pierrette Bouillon and Dominique Estival, editors, *Proceedings of the Workshop on compound Nouns: Multilingual Aspects of Nominal Composition*, pages 103–107, Geneva, 2-3 December. ISSCO.

[Cencioni91] Roberto Cencioni. 1991. The Eurotra software environment - A broad overview. In Steven Krauwer Charles Copeland, Jacques Durand and Bente Maegaard, editors, *The Eurotra Formal Specifications*, Studies in Machine Translation and Natural Language Processing, Volume 2. Commission of the European Communities, Brussels & Luxembourg.

[Ceusters et al.94a] Werner Ceusters, Guy Deville, Emmanuel Herbigniaux, Pierre Mousel, Oliver Streiter, and Geert Thienpont. 1994a. The ANTHEM Prototype. IAI WP 31.

[Ceusters et al.94b] Werner Ceusters, Guy Deville, Oliver Streiter, Emmanuel Herbigniaux, and Jos Devlies. 1994b. A computational linguistic approach to semantic modeling in medicine. In *Belgo-Dutch Congress on Medical Informatics '94*, pages 311–319, Veldhoven.

[Ceusters et al.94c] Werner Ceusters, Guy Deville, Oliver Streiter, Emmanuel Herbigniaux, and Jos Devlies. 1994c. When Machine Translation meets automatic encoding. In *Proceedings of the 1st Language Engineering Convention, held at Paris, 6-7 July 1994*, Paris.

[Chen and Chen92] Kuang-Hua Chen and Hsin-Hsi Chen. 1992. Machine Translation via unification. In *Advances on the Research of Machine Translation, ji qi fan yi yan jiu jin zhan*, pages 290–313, Beijing, August. dian zi gong yie chu bang she.

[Choi95] Sung-Kwon Choi. 1995. Unifikationsbasierte Maschinelle Übersetzung mit Koreansisch als Quellsprache. IAI WP 34.

[Chomsky65] Noam Chomsky. 1965. *Aspects of the Theory of Syntax.* The M.I.T. Press, Cambridge, Massachusetts.

[Chomsky80] Noam Chomsky. 1980. *Rules and Representations.* Columbia University Press, New York.

[Chomsky81] Noam Chomsky. 1981. *Lectures on Government and Binding. The Pisa Lectures.* Studies in Generative Grammar 9. Foris Publication, Doordrecht Holland & Cinnaminson U.S.A.

[Chomsky93] Noam Chomsky. 1993. A minimalist program for linguistic theory. In Kenneth Hale and Jay Keyser, editors, *The View from Building 20:Essays in Linguistics in Honour of Sylvain Bromberger*, Current studies in linguistics series; 20. The MIT Press, Cambridge, Massachusettes.

[Chuquet and Paillard89] Hélène Chuquet and Michel Paillard. 1989. *Approche linguistique des problèmes de traduction anglais ↔ français.* OPHRYS, Paris.

[Collins96] Bróna Collins. 1996. Similarity and adaptability guided retrieval in EBMT. In *ms.*

[Comrie81] Bernard Comrie. 1981. *Language Universals and Linguistic Typology.* Basil Blackwell, Oxford.

[Comrie85] Bernhard Comrie. 1985. *Aspect, An Introduction to the Study of Verbal Aspect and related Problems.* Cambridge University Press, Cambridge.

[Cook88] V.J. Cook. 1988. *Chomsky's Universal Grammar. An Introduction.* Applied Languages Studies. Blackwell, Oxford UK & Cambridge USA.

[Copestake et al.95] Ann Copestake, Flickinger Dan, Rob Malouf, Susanne Riehemann, and Ivan Sag. 1995. Translation using minimal recursion semantics. In *TMI-95.*

[Copestake95] Ann Copestake. 1995. Semantic transfer in Verbmobil. Verbmobil report 93, Universität Stuttgart and CSLI.

[Côté et al.93] Roger A. Côté, David J. Rothwell, Ronald S. Beckett, and James L. Palotay, editors. 1993. *Developing a standard data structure for the systematized Nomenclature of Human and Veterinary Medicine. SNOMED International. Introduction.* College of American Pathologists & American Veterinary Medical Association.

[Creider79] Chet A. Creider. 1979. On the explanation of transformations. In Talmy Givón, editor, *Syntax and Semantics. Volume 12 Discourse and Syntax.* Academic Press, New York.

[Crystal87] David Crystal. 1987. *The Cambrige Encyclopedia of Language.* Cambridge University Press, Cambridge.

[Culioli81] Antoine Culioli. 1981. Sur le concept de notion. *BULAG*, 8:62–79.

[Culioli90] Antoine Culioli. 1990. *Pour une linguistique de l'énonciation, tom 1.* Collection L'Homme dans la Langue. OPHYS, Paris.

[Cullingford and Onyshkevych85] R.E. Cullingford and B.A. Onyshkevych. 1985. Lexicon-driven Machine Translation. In *Proceedings of the Conference on Theoretical and Methodological Issues of Natural Languages, Colgate University*, pages 75–111. Hamilton, NY.

[Dahl75] Östen Dahl. 1975. On generics. In E.L. Keenan, editor, *Formal Semantics of Natural Language.* Cambridge University Press, Cambrdige.

[Dahl81] Östen Dahl. 1981. On the definition of the telic -atelic (bounded-nonbounded) distinction. In Philip Tedeschi and Annie Zaenen, editors, *Syntax and Semantics. Volume 14 Tense and Aspect.* Academic Press, New York.

[Dahl85] Östen Dahl. 1985. *Tense and Aspect Systems.* Basil Blackwell, Oxford.

[Danlos et al.90] Laurence Danlos, Anneliese Bech, Folker Caroli, B. Daille, Nadia Mesli, F. Namer, and S. Nohr. 1990. Support verbs and predicative nouns. In *ERM.*

[Danlos87] Laurence Danlos. 1987. *The linguistic basis of text generation.* Studies in Natural Language Processing. Cambridge University Press, Cambridge.

[Danon-Boileau87] Laurent Danon-Boileau. 1987. *Enonciation et Référence.* Collection L'Homme dans la Langue. OPHYS, Paris.

[des Tombe et al.85] Louis des Tombe, Doug Arnold, Lieven Jaspaert, Rod Johnson, Steven Krauwer, MMike Rosner, G.B. Varile, and Susan Warwick. 1985. A preliminary linguistic framework for EUROTRA, june 1985. In *Proceedings of the Conference on Theoretical and Methodological Issues of Natural Languages, Colgate University*, pages 283–288. Hamilton, NY.

[Dorr90] Bonnie Jean Dorr. 1990. A cross-linguistic approach to translation. In *TMI-90.*

[Dorr93] Bonnie Jean Dorr. 1993. *Machine Translation: A View form the Lexicon.* MIT Press, Cambridge, Massachusetts. London, England.

[Dorr94] Bonnie Jean Dorr. 1994. Machine Translation divergences. *Computational Linguistics*, 20(4):597–633.

[Duranti and Elinor79] Alessandro Duranti and Ochs Elinor. 1979. Left-dislocation in Italian conversation. In Talmy Givón, editor, *Syntax and Semantics. Volume 12 Discourse and Syntax.* Academic Press, New York.

[Ehrich82] Veronika Ehrich. 1982. *Da* and the system of spatial deixis in German. In Jürgen Weissenborn and Wolfgang Klein, editors, *Here and There. Cross-linguistic Studies on Deixis and Demonstration*, Pragmatics & Beyond III 2/3. John Benjamins Publishing Company, Amsterdam& Philadelphia.

[Eisele and Dörre90] Andreas Eisele and Jochen Dörre. 1990. Disjunctive unification. IWBS report 124, IBM, Wissenschaftliches Zentrum, Institut für Wissensbasierte Systeme.

[Erbach and Uszkoreit90] Gregor Erbach and Hans Uszkoreit. 1990. Grammar engineering: Problems and prospects. Report on the Saarbrücken grammar engineering workshop. CLAUS report nr. 1, Computerlinguistik an der Universität des Saarlandes, Universität des Saarlandes, 66125 Saarbrücken, Germany.

[ERM90] 1990. *ERM.*

[Fages67] J.-B. Fages. 1967. *Comprendre le structuralisme.* Édouard Privat, Toulouse.

[Fakhour93] Sabah Fakhour. 1993. Morphologie du français et grammaire binaire du français dans le formalisme CAT2. Rapport de stage, Jussieu, Université de Paris VII, Paris.

[Fauconnier84] Gilles Fauconnier. 1984. *Espaces Mentaux. Aspect de la construction du sens dans les languages naturelles.* Les Editions de Minuit, Paris.

[Francis et al.95] Bond Francis, Ogura Kentaro, and Kawaoka Tsukasa. 1995. Noun phrase reference in Japanese-to-English Machine Translation. In *TMI-95.*

[Fung and Wu95] Pascale Fung and Dekai Wu. 1995. Coerced markov models for cross-lingual lexial-tag relatons. In *TMI-95.*

[Furuse and Iida92] O. Furuse and H. Iida. 1992. An example-based method for transfer-driven Machine Translation. In *The Third International Conference on Theoretical and Methodological Issues, Empiristic vs. Rationalist Methods in MT*, Montréal, 25-27 June.

[Gazdar et al.85] Gerald Gazdar, Ewan Klein, Geoffrey Pullum, and Ivan Sag. 1985. *Generalized Phrase Structure Grammar.* Basil Blackwell, Oxford, UK.

[Giusti81] Francesca Fici Giusti. 1981. La referenza nominale in una lingua senza articolo. Analisi comparative del russo e dell'italiano. *Studi di grammatica italiana*, X:109–214.

[Givón78] Talmy Givón. 1978. Universal grammar, lexical structure and translatability. In F. Guenthner and M. Guenthner Reutter, editors, *Meaning and Translation. Philosophical and Linguistic Approaches.* Duckworth, London.

[Goddman and Nirenburg (eds.)92] Kenneth Goddman and Sergei Nirenburg (eds.). 1992. *The KBMT Project: A case study in Knowledge Based Machine Translation.* Morgan Kaufamn Publishers, San Mateo, California.

[Greenberg63] Joseph H. Greenberg. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg, editor, *Universals of Language*. The M.I.T. Press, Cambridge, Massachusetts, 2 edition.

[Greenberg65] Josef H. Greenberg. 1965. Some generalizations concerning initial and final consonant sequences. *Linguistics*, 18:73–81.

[Greimas and Joseph89] Algirdas Julien Greimas and Courtés Joseph. 1989. *Sémiotique, Dictionnaire raisonné de la théorie du langage*. Langue Linguistique Communication. Hachette, Paris.

[Grimshaw and Mester88] Jane Grimshaw and Armin Mester. 1988. Light verbs and θ-marking. *Linguistic Inquiry*, 19(2):205–232.

[Grimshaw90] Jane Grimshaw. 1990. *Argument Structure*. Linguistic Inquiry Monographs. The MIT Press, Cambridge, Massachusettes.

[Grimshaw91] Jane Grimshaw. 1991. Extended Projection. Brandeis University, Waltham MA 02254, ms, July.

[Gross75a] Maurice Gross. 1975a. *Méthodes en Syntaxe*. Hermann, Paris.

[Gross75b] Maurice Gross. 1975b. On the relation between syntax and semantics. In E.L. Keenan, editor, *Formal Semantics of Natural Language*. Cambridge University Press, Cambrdige.

[Guentcheva90] Zlatka Guentcheva. 1990. *Temps et Aspect, L'exemple du bulgare contemporain*. Édition du Centre National de la Recherche Scientifique, Paris.

[Guzmán de Rojas88] Iván Guzmán de Rojas. 1988. ATAMIRI - interlingual MT using the Aymara language. In Dan Maxwell, Klaus Schubert, and Toon Witkam, editors, *New Directions in Machine Translation*. Foris Publication, Dordrecht - Holland.

[Haller91] Johann Haller. 1991. EUROTRA – Das Forschungs- und Entwicklungsprojekt der EG zur Maschinellen Übersetzung: Französisch- Deutsche Übersetzung mit der Seitenlinie CAT2. In Rolshoven and Seelbach, editors, *Romanistische Computerlinguistik, Linguistische Arbeiten*. Max Niemeyer Verlag, Tübingen.

[Haller93] Johann Haller. 1993. CAT2, Vom Forschungssystem zum präindustriellen Prototyp. In Horst P. Pütz and Johann Haller, editors, *Sprachtechnologie: Methoden, Werkzeuge, Perspektiven*, pages 282–303, Hildesheim. GLDV, Georg Olms AG. URL: http://www.iai.uni-sb.de/cat2/docs.html.

[Halliday85] M.A.K. Halliday. 1985. *An Introduction to Functional Grammar*. Edward Arnold, London.

[Harder and Schimmel89] Ernst Harder and Annemarie Schimmel. 1989. *Arabische Sprachenlehre*. Julius Groos Verlag, Heidelberg, 16 edition.

[Hauenschild and Busemann88] Christa Hauenschild and Stephan Busemann. 1988. A constructive version of GPSG for Machine Translation. In Erich Steiner, Paul Schmidt, and Cornelia Zelinsky-Wibbelt, editors, *From Syntax to Semantics, Insight from Machine Translation*. Pinter Publishers Ltd., London.

[Hawkins83] John A. Hawkins. 1983. *A comparative Typology of English and German*. Croom Helm, London & Sydney.

[Herrmann-Dresel87] Eva Herrmann-Dresel. 1987. *Die Funktionsverbgefüge des Russischen und des Tschechischen*. Heidelberger Publikationen zur Slavistik. Peter Lang, Frankfurt am Main.

[Heylen et al.92] Dirk Heylen, Lee Humphreys, Susan Warwick-Armstrong, Nicolette Calzolari, and Murison-Bowie Simon. 1992. Lexical functions for multilingual lexicons. In *International Workshop on the Meaning-Text Theory*, pages 173–185, Darmstadt. Arbeitspapiere der GMD Band 671.

[Heylen et al.93] Dirk Heylen, Kerry G. Maxwell, and Susan Warwick-Armstrong. 1993. Collocations, dictionaries, and MT. In *Building Lexicons for Machine Translation*, Standford. AAAI Spring Symposium Series.

[Heylen92] Dirk Heylen. 1992. Lexical functions and knowledge representation. In *Proceedings of the Second International Workshop on Computational Lexical Semantics*, Toulouse.

[Higginbotham85] J. Higginbotham. 1985. On semanitcs. *Linguistic Inquiry*, (16):547–593.

[Hill91] Peter Hill. 1991. Das Bulgarische. In Peter Rehder, editor, *Einführung in die slavischen Sprachen.* Wissenschaftliche Buchgesellschaft, Darmstadt, 2 edition.

[Hottenroth82] Priska-Monika Hottenroth. 1982. The system of local deixis in Spanish. In Jürgen Weissenborn and Wolfgang Klein, editors, *Here and There. Cross-linguistic Studies on Deixis and Demonstration*, Pragmatics & Beyond III 2/3. John Benjamins Publishing Company, Amsterdam& Philadelphia.

[Huáng and Liào83] Huáng and Liào. 1983. *Xiàn dái hàn yǔ.* Gān Sù rén míng dà xīn shè. Lán Zhōu.

[Hutchins86] W.J Hutchins. 1986. *Machine Translation. past, present, future.* Computers and Their Applications. Ellis Horwood Limited, New York.

[Hutchins93] John Hutchins. 1993. Latest developments in Machine Translation technology: Beginning a new era in MT research. In *Proceedings of the Fourth Machine Translation Summit*, pages 11–34, Kyoto, Japan.

[Hwee-Boon93] Low Hwee-Boon. 1993. Some lessons learnt by a new comer. In *Proceedings of the Fourth Machine Translation Summit*, pages 181–187, Kyoto, Japan.

[Iomdin94] Leonid Iomdin. 1994. Automatic syntactic analysis of Russian in the CAT2 MT system. IAI WP 32.

[Isabelle93] Pierre Isabelle. 1993. Machine-aided human translation and the paradigm shift. In *Proceedings of the Fourth Machine Translation Summit*, pages 177–178, Kyoto, Japan.

[Jakobson63] Roman Jakobson. 1963. Aspects linguistiques de la traduction (on translation, 1959). In *Essais de linguistique générale*, Collection double. Editions de Minuit, Paris.

[Jakobson86] Roman Jakobson. 1986. *Kindersprache, Aphasie und allgemeine Lautgesetze (1941).* Edition Suhrkamp 330. Suhrkamp Verlag, Frankfurt a.M.

[Jones95] Daniel Jones. 1995. *Analogical natural language processing.* Computational Linguistics. UCL Press, London.

[Kaplan et al.89] Ronald M. Kaplan, Klaus Netter, Jürgen Wedekind, and Annie Zaenen. 1989. Translation by structural correspondences. In *Fourth Conference of the European Chapter of the Association for Computational Linguistics*, pages 272–281, Institute of Science and Technology Manchester, England, April. Association for Computational Linguistics.

[Karttunen84] Lauri Karttunen. 1984. Features and values. In *COLING-84.*

[Kay et al.91] M. Kay, J.M. Gawron, and P. Norvig. 1991. *Verbmobil: A Translation System for Face-to-Face Dialog. BMFT Study.* CSLI, Stanford University, Stanford.

[Keenan72] E.L Keenan. 1972. On semantically based grammar. *Linguistic Inquiry*, 3(4):413–461.

[Keenan75] E.L Keenan. 1975. Logical expressive power and syntactic variation in natural languages. In E.L. Keenan, editor, *Formal Semantics of Natural Language.* Cambridge University Press, Cambrdige.

[Krenn and Erbach94] Brigitte Krenn and Gregor Erbach. 1994. Idioms and support verb constructions. In Pollard Nerbonne, Netter, editor, *German Head-Driven Phrase Structure Grammar.* CSLI Lecture Notes Number 46, Standford.

[Krifka91] Manfred Krifka. 1991. Thematic relations as links between nominal reference and temporal constitution. In Ivan Sag and Anna Sabolesi, editors, *Lexical Matters.* Chicago Univeristy Press, Chicago.

[Krušel'nickaja61] K.G. Krušel'nickaja. 1961. *Očerki po sopostavitel'noj grammatike nemeckogo i russkogo jazykov.* Biblioteka filologa. Izdatel'stvo literatury na inostrannyx jazykax, Moskva.

[Kudo and Nomura86] Ikuo Kudo and Hirosato Nomura. 1986. Lexical-functional transfer: A transfer framework in a Machine Translation system based on LFG. In *COLING-86*, pages 112–114.

[Kuhn87] Thomas S. Kuhn. 1987. *De structuur van wetenschappelijke revoluties.* Boon Meppel, Amsterdam.

[Kuhn94] Jonas Kuhn. 1994. Die Behandlung von Funktionsverbgefügen in einem HPSG-basierten Übersetzungsansatz. Verbmobil Bericht 66, Universität Stuttgart.

[Kulagina89] O. S. Kulagina. 1989. Mašinnyj perevod: sovremennoe sostojanie. *Semiotika i informatika*, 29:5–33.

[Kuz'min75] Ju.G. Kuz'min. 1975. Perevod kak myslitel'no-rečevaja dejatel'nost'. In L.S. Barchudarov, editor, *Tetradi perevodčika.* Izdatel'stvo "Meždunarodnye otnošenija", Moskva.

[Laffling91] John Laffling. 1991. *Towards high-precision Machine Translation. Based on Contrastive Textology.* Distributed Language Translation 7. Foris Publications, Berlin & New York.

[Landsbergen87] Jan Landsbergen. 1987. Montague grammar and Machine Translation. In Peter Whitelock, M.M. Wood, Harold L. Somers, R.Johnson, and Paul Bennet, editors, *Linguistic Theory and Computer Application.* Academic Press.

[Langacker81] Ronald W. Langacker. 1981. The nature of grammatical valence. *Linguistic Notes from La Jolla*, (10):33–59.

[Langacker87] Ronald W. Langacker. 1987. *Foundations of Cognitive Grammar. Volume 1 Theoretical Prerequisites.* Stanford University Press, Stanford, California.

[Larson84] Mildred L. Larson. 1984. *Meaning-based Translation: A Guide to Cross-language Equivalence.* University Press of America, Lanham, New York, London.

[Lehrberger and Bourbeau88] John Lehrberger and Laurent Bourbeau. 1988. *Machine Translation. Linguistic characteristics of MT systems and general methology of evaluation.* Studies in French & General Linguistics. Études en Linguistique Française et Générale. John Benjamins Publishing Company, Amsterdam & Philadelphia.

[Lethbridge94] Timothy Christian Lethbridge. 1994. *Practical Techniques for Organizing and Measuring Knowledge.* Ph.D. thesis, School of Graduate Studies and Research at the University of Ottawa, Ottawa.

[Löbel90] Elisabeth Löbel. 1990. D und Q als funktionale Kategorien in der Nominalphrase. *Linguistische Berichte*, 127:233–264.

[Loon-Vervoon84] W.A. Loon-Vervoon. 1984. Voorstelbaarheid en prototypicaliteit von woorden. In *Leezing gehouden op het vakgroepsymposium funktieleer*, Utrecht, 9 november.

[Loon-Vervoon86] W.A. Loon-Vervoon. 1986. Sensorimotor versus linguistically based word imagibility: The importance of age of word acquisition. In *European Workshop on Imagery and Cognition*, Paris, 24-26 september.

[Luckhardt and Maas83] Heinz-Dirk Luckhardt and Heinz-Dieter Maas. 1983. SUSY Handbuch für Transfer und Synthese. Linguistische Arbeiten, Neue Folge, Heft 7, Sonderforschungsbereich 100. Elektronische Sprachforschung. Universität des Saarlandes, Saarbrücken.

[Luckhardt and Zimmermann91] Heinz-Dirk Luckhardt and Harald H. Zimmermann. 1991. *Computergestützte und Maschinelle Übersetzung.* Sprachwissenschaft-Computerlinguistik Band 14. AQ-Verlag, Saarbrücken.

[Luckhardt87] Heinz-Dirk Luckhardt. 1987. *Der Transfer in der Maschinellen Übersetzung.* Sprache und Information. Max Niemeyer Verlag, Tübingen.

[Lujan80] Marta Lujan. 1980. *Sintaxis y semantica del adjetivo.* GGT, Grammatica generativa transformacional del español. Ediciones Cátedtra, Madrid.

[Luria82] Alexander Romanowitsch Luria. 1982. *Sprache und Bewußtsein.* Studien zur Kritischen Psychologie. Pahl-Rugenstein Verlag, Köln.

[Lyons73] John Lyons. 1973. *Einführung in die moderne Linguistik.* C.H.Beck, München, 3 edition.

[Maas et al.95] Heinz-Dieter Maas, Antje Schmidt-Wigger, and Oliver Streiter. 1995. Report on WP6-3 - Integration of French. CAT2-EDS Deliverable D6-3, Commission of the European Communities, Luxembourg.

[Maas84] Heinz-Dieter Maas. 1984. SUSY-II-Handbuch. Linguistische Arbeiten, Neue Folge, Heft 14, Sonderforschungsbereich 100. Elektronische Sprachforschung. Universität des Saarlandes, Saarbrücken.

[Maas94a] Heinz-Dieter Maas. 1994a. Analysis and translation of compound nouns in Mpro. In Pierrette Bouillon and Dominique Estival, editors, *Proceedings of the Workshop on compound Nouns: Multilingual Aspects of Nominal Composition*, pages 162–172, Geneva, 2-3 December. ISSCO.

[Maas94b] Heinz-Dieter Maas. 1994b. Report on WP1 - Extensive German morphology. CAT2-EDS Deliverable D1, Commission of the European Communities, Luxembourg.

[Magnúsdóttir88] Gudrún Magnúsdóttir. 1988. Problems of lexical access in Machine Translation. In Martin Gellerstam, editor, *Studies in Computer-Aided Lexicology*, Data linguistica 18. Almqvist & Wiksell International, Stockholm.

[Mainguenau81] Dominique Mainguenau. 1981. *Approche de l'Enonciation en linguistique Française, Embrayeurs, "Temps", Discours rapporté.* Langue Linguistique Communication. Hachette, Paris.

[Malnati and Paggio90] Giovanni Malnati and Patrizia Paggio. 1990. The Eurotra User Language. In *Eurotra Reference Manual.*

[Mehrjerdian92] Hooshang Mehrjerdian. 1992. *ATSTEP1: Automatische Übersetzung englischer Fachtexte ins Persische.* Ph.D. thesis, Universität Bonn, Bonn.

[Melby89] Alan Melby. 1989. Machine Translation: General development. In István S. Bátori, Winfried Lenders, and Wolfgang Putschke, editors, *Computational Linguistics. Computerlinguistik. An International Handbook on Computer Oriented Language Research and Application. Ein internationales Handbuch zur computergestützten Sprachforschung und ihrer Anwendungen.* Walter de Gryter, Berlin & New York.

[Mel'čuk and Pertsov87] Igor Aleksandrovič Mel'čuk and Nikolaj V. Pertsov. 1987. *Surface Syntax of English, a formal model within the meaning-text framework.* Linguistic & Literary, Studies in Eastern Europe. John Benjamins Publishing Company, Amsterdam, Philadelphia.

[Mel'čuk74] Igor Aleksandrovič Mel'čuk. 1974. *Opyt teorii lingusticeskix modelej Smysl⇔Tekst. Semantika, sintaksis.* Izdatel'stvo "Nauka", Moskva.

[Mel'čuk84] Igor Aleksandrovič Mel'čuk. 1984. *Dictionnaire explicatif et combinatoire du français contemporain, Rechereches lexico-sémantiques I.* Les Presses de l'Université de Montréal, Canada.

[Mel'čuk88] Igor Aleksandrovič Mel'čuk. 1988. *Dictionnaire explicatif et combinatoire du français contemporain, Rechereches lexico-sémantiques II.* Les Presses de l'Université de Montréal, Canada.

[Mervis and Rosch81] C.B Mervis and E. Rosch. 1981. Categorization of natural objects. *Annual Review of Psychology*, 32:89–115.

[Mesli and Bresson92] Nadia Mesli and Daniel Bresson. 1992. Nominale Prädikate und Funktionsverben bei Martin Luther am Beispiel des lexikalischen Bereichs der Gefühle und der zwischenmenschlichen Beziehungen. *Cahier d'Etudes Germaniques, Beiträge zur Lexikologie und Lexikographie des Deutschen, Revue Semestrielle*, (23):75–99.

[Mesli91] Nadia Mesli. 1991. Analyse et traduction automatique de constructions à verbe support dans le formalisme CAT2. IAI WP 19b.

[Mitkov et al.95] Ruslan Mitkov, Sung-Kwon Choi, and Randall Sharp. 1995. Anaphora resolution in Machine Translation. In *TMI-95*.

[Mitkov95] Ruslan Mitkov. 1995. A new approach for trecking center. In *Proceedings of the International Conference "New Methods in Language Processing"*, Mancheser, UK. UMIST.

[Mufwene84] Salikok S. Mufwene. 1984. The count/mass distinction and the English lexicon. In David Testen, Yeena Mishra, and Joseph Drogo, editors, *Papers from the Parasession on Lexical Semantics*, pages 200–221, Chicago, 27-28 April. Chicago Linguistic Society.

[Nagao and Tsujii86] Maloto Nagao and Jun-ichi Tsujii. 1986. The transfer phase of the Mu Machine Translation system. In *COLING-86*, pages 97–103.

[Nakaiwa and Ikehara95] Hiromi Nakaiwa and Saturo Ikehara. 1995. Intrasentential resolution of japanes zero pronouns in a Machine Translation system using semantic and pragmatic constraints. In *TMI-95*.

[Netter94] Klaus Netter. 1994. Towards a theory of functional heads: German nominal phrases. In Pollard Nerbonne, Netter, editor, *German Head-Driven Phrase Structure Grammar*. CSLI Lecture Notes Number 46, Standford.

[Nirenburg et al.87] Sergei Nirenburg, Victor Raskin, and Allen B. Tucker. 1987. The structure of interlingua in TRANSLATOR. In *MT TMI*, pages 1–15.

[Nirenburg (ed.)87] Sergei Nirenburg (ed.). 1987. *Machine Translation, Theoretical and Methodological Issues*. Studies in Natural Language Processing. Cambridge University Press, Cambridge.

[Nirenburg87] Sergei Nirenburg. 1987. Knowlege and choices in Machine Translation. In *MT TMI*, pages 1–15.

[Nirenburg93] Sergei Nirenburg. 1993. A direction of MT development. In *Proceedings of the Fourth Machine Translation Summit*, pages 189–193, Kyoto, Japan.

[Nitta86] Yoshihiko Nitta. 1986. Idiosyntractic gap: A tough problem to structure-bound Machine Translation. In *COLING-86*, pages 107–111.

[Norbert and Faensen76] Reiter Norbert and Johannes Faensen. 1976. *30 Stunden Serbokroatisch*. Langenscheidt, Berlin.

[Olsen92] Susan Olsen. 1992. Zur Grammatik des Wortes. Argumente zur Argumentvererbung. *Linguistische Berichte*, (137).

[Patten88] Terry Patten. 1988. *Systemic Text Generation as Problem Solving*. Studies in Natural Language Processing. Cambridge University Press, Cambridge.

[Pease and Boushaba96] Catherin Pease and Abd Al-Aziz Boushaba. 1996. ARAMED. Extension and integration of Arabic lingware components in a unification-based MT system for the field of medical terminology and classification. In *First KFUPM Workshop on Information & Computer Science (WICS)*, Dhahran, June 9.

[Pease93] Catherine Pease. 1993. The Analysis of German Compounds and their Translation into English. M.sc. thesis, UMIST, Manchester.

[Pereira and M.87] Fernando C.N. Pereira and Shieber Stuart M. 1987. *PROLOG and Natural-Language Analysis.* CSLI Lecture Notes Number 10. Center for the Study of Language and Information, Standford.

[Picht and Draskau85] Heribert Picht and Jennifer Draskau. 1985. *Terminology: An introduction.* The University of Surrey, Guildford, England.

[Podeur93] Josiane Podeur. 1993. *La Pratica della Traduzione. Dal francese in italiano e dell'italiano in francese.* Liguori editore, Napoli.

[Pollard and Sag87] Carl Pollard and Ivan Sag. 1987. *Information-Based Syntax and Semantics.* CSLI Lecture Notes 13, Stanford.

[Pollard and Sag94] Carl Pollard and Ivan Sag. 1994. *Head-Driven Phrase Structure Grammar.* Studies in Contemporary Linguistics. The University of Chicago Press, Chicago & London.

[Prahl94] Birte Prahl. 1994. Menschliche Desambiguierungsstrategien für die Maschinelle Übersetzung? Ein Beispiel für ambiguitätserhaltende Übersetzungen. Verbmobil memo 54, Universität Hildesheim.

[Quine60] Willard Van Orman Quine. 1960. *Word and Object.* The M.I.T. Press, Cambridge, Massachusetts, 14 edition.

[Radford88] Andrew Radford. 1988. *Transformational Grammar, a first course.* Cambridge Textbooks in Linguistics. Cambridge University Press, Cambridge.

[Raffy92] Caroline Raffy. 1992. L'emploi des structures de traits dans les dictionnaires français du système de traduction automatique CAT2. Mémoire du DEA d'informatique fondamentale, Jussieu, Université de Paris VII, Paris.

[Reed82] Stephen K. Reed. 1982. *Cognition. Theory and Applications.* Brookes/Cole Publishing Company, Monterey, Califonia.

[Rehder91a] Peter Rehder. 1991a. Das Makedonische. In Peter Rehder, editor, *Einführung in die slavischen Sprachen.* Wissenschaftliche Buchgesellschaft, Darmstadt, 2 edition.

[Rehder91b] Peter Rehder. 1991b. Das Serbokroatische. In Peter Rehder, editor, *Einführung in die slavischen Sprachen.* Wissenschaftliche Buchgesellschaft, Darmstadt, 2 edition.

[Reichenbach47] H. Reichenbach. 1947. *Elements of Symbolic Logic.* The Free Press, New York.

[Riehemann95] Susanne Riehemann. 1995. The HPSG formalism. unpublished, May.

[Rohrer86] Christian Rohrer. 1986. Maschinelle Übersetzung mit Unifikationsgrammatiken. In István Bátori and Heinz J. Weber, editors, *Neue Ansätze in maschineller Sprachübersetzung: Wissensrepräsentation und Textbezug*, Sprache und Information Band 13. Niemeyer, Tübingen.

[Rombouts19] Fr. S. Rombouts. 1919. *De psychologie der kleutertaal - verklaard voor taalleeraren, pedologen, opvoeders en kindervrienden.* Uitgever van den Apostolischen Stoel, Malmberg-Nijmegen.

[Rosch and Mervis75] E. Rosch and C.B. Mervis. 1975. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology,* 7:573–605.

[Rosch et al.75] E. Rosch, C.B. Mervis, W.D. Gray, D.M. Johnson, and P. Boyes-Braem. 1975. Basic objects in natural catgories. *Cognitive Psychology,* 8:382–440.

[Rothwell et al.93] David J. Rothwell, Roger A. Côté, J.P. Cordeau, and M.A. Boisvert. 1993. Developing a standard data structure for medical language, the SNOMED proposal. In *Proceedigs of the Annual Symposim on Computer Applications in Medical Care'93,* pages 695–699.

[Ruge86] Hans Ruge. 1986. *Grammatik des Neugriechischen. Lautlehre, Formlehre, Syntax.* Romiosini, Köln.

[Ruus and Spang-Hanssen86] Hanne Ruus and Ebbe Spang-Hanssen. 1986. A theory of semantic relations for large-scale natural language processing. In *COLING-86*, pages 20–22.

[Sadler93] Louisa Sadler. 1993. Codescription and translation. In Frandk von Eynde, editor, *Linguistic Issues in Machine Translation*, Artificial Intelligence Series. Pinter Publlishers.

[Sag95] Ivan A. Sag. 1995. HPSG: Background and Basis. unpublished.

[Sakamato et al.86] Yoshiyuki Sakamato, Tetsuya Ishikawa, and Masayuki Satoh. 1986. Concept and structure of semantic markers for Machine Translation in Mu-project. In *COLING-86*, pages 13–19.

[Schlegel1818] A.W. Schlegel. 1818. *Observatrions sur la langue et littérature provençales.*

[Schmidt-Wigger91] Antje Schmidt-Wigger. 1991. Traitement du grec moderne. Rapport de stage, Jussieu, Université de Paris VII, Paris.

[Schmidt88] Paul Schmidt. 1988. Transfer strategies in EUROTRA. In Erich Steiner, Paul Schmidt, and Cornelia Zelinsky-Wibbelt, editors, *From Syntax to Semantics, Insight from Machine Translation.* Pinter Publishers Ltd., London.

[Schubert88] Klaus Schubert. 1988. The architecture of DLT - interlingua or double direct? In Dan Maxwell, Klaus Schubert, and Toon Witkam, editors, *New Directions in Machine Translation.* Foris Publication, Dordrecht - Holland.

[Schütz94] Jörg Schütz. 1994. *Terminological Knowledge in Multilingual Language Processing.* Studies in Machine Translation and Natural Language Processing, Volume 5. Commission of the European Communities, Brussels, Luxembourg.

[Schütz95] Jörg Schütz. 1995. The ALEP formalism in a nutshell. URL: http://www.iai.uni-sb.de/alep/alep-docs.html.

[Schwanke91] Martina Schwanke. 1991. *Maschinelle Übersetzung. Ein Überblick über Thorie und Praxis.* Springer Verlag, Berlin.

[Schwarze88] Christoph Schwarze. 1988. *Grammatik der italienischen Sprache.* Max Niemeyer Verlag, Tübingen.

[Schwenk91] Hans-Jörg Schwenk. 1991. *Studien zur Semantik des Verbalaspekts im Russischen.* Verlag Otto Sagner, München.

[Scott89] Bernard E. Scott. 1989. The logos system. In *Paper delivered at the MT SUMMIT II Conference in Munich, on August 1989*, Munich.

[Scott92] Bernard E. Scott. 1992. Computability in search of a paradigm. Logos Corporation Mt. Arlington, NJ USA, ms, June.

[Seewald93] Uta Seewald. 1993. Automatische Wortformerkennung im Deutschen - Analyseverfahren und Systeme. Ein Bericht über die auf dem 1.GLDV-Workshop zur automatischen Wortformerkennung in Erlangen präsentierten Entwicklungen. *LDV-Forum*, 10(2):5–16.

[Sharp and Streiter92] Randall Sharp and Oliver Streiter. 1992. Simplifying the Complexity of Machine Translation. *Meta-92*, pages 681–692. URL: http://www.iai.uni-sb.de/cat2/docs.html.

[Sharp and Streiter95] Randall Sharp and Oliver Streiter. 1995. Applications in Multilingual Machine Translation. In *Proceedings of The Third International Conference and Exhibition on Practical Applications of Prolog, Paris, 4th-7th April.* URL: http://www.iai.uni-sb.de/cat2/docs.html.

[Sharp86] Randall Sharp. 1986. A parametric NL translator. In *COLING-86*, pages 124–126.

[Sharp88] Randall Sharp. 1988. CAT2 - implementing a formalism for multi-lingual MT. In *Proceedings of the 2nd International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Language.* Carnegie Mellon University, Pittsburgh, PA.

[Sharp91] Randall Sharp. 1991. CAT2: An experimental Eurotra alternative. *Machine Translation*, 6(3):215–228.

[Sharp94] Randall Sharp. 1994. CAT2 Reference Manual Version 3.6. IAI WP n.27.

[Shieber86] S. Shieber, editor. 1986. *An Introduction to Unification-Based Approaches to Grammar.* SLI Lecture Notes 4, CSLI, Stanford.

[Sidner86] C. Sidner. 1986. Focusing in the comprehesion of definite anaphora. In B. Webber B. Grosz, K.Jones, editor, *Readings in Natural Language Processing.* Morgan Kaufmann Publishers.

[Siewierska84] Anna Siewierska. 1984. *The Passive. A Comparative Linguistic Analysis.* Linguistics Series. Croom Helm, London, Sydney, Wolfeboro, New Hanpshire.

[Simone90] Raffaele Simone. 1990. *Fondamenti di linguistica.* Editioni Laterza, Roma-Bari.

[Simpkins95] N.K. Simpkins. 1995. ALEP (advanced language engineering platform). an open architecture for language engineering. URL: http://www.iai.uni-sb.de/alep/alep-docs.html.

[Slocum89] Jonathan Slocum. 1989. Machine Translation: A survey of active systems. In István S. Bátori, Winfried Lenders, and Wolfgang Putschke, editors, *Computational Linguistics. Computerlinguistik. An International Handbook on Computer Oriented Language Research and Application. Ein internationales Handbuch zur computergestützten Sprachforschung und ihrer Anwendungen.* Walter de Gryter, Berlin & New York.

[Somers et al.88] Harold Somers, Hideki Hirakawa, Seiji Miike, and Shinya Amano. 1988. The treatment of complex English nominalizations in Machine Translation. *Computers and Translation*, 3(1):3–21, March.

[Somers90] Harold L. Somers. 1990. Current research in Machine Translation. In *TMI-90.*

[Srinivas et al.94] Bangalore Srinivas, Dania Egedi, Christy Doran, and Tilman Becker. 1994. Lexicalization and grammar development. In *KONVENS, '94.*

[Steinberger92a] Ralf Steinberger. 1992a. Beschreibung der Adverbstellung im deutschen und englischen Satz im Hinblick auf Maschinelle Übersetzung. IAI WP 22.

[Steinberger92b] Ralf Steinberger. 1992b. Der Skopus von Gradpartikeln: Seine Übersetzung und seine Implementierung im Maschinellen Übersetzungssystem CAT2. IAI WP 23.

[Steinberger94] Ralf Steinberger. 1994. Lexikoneinträge für deutsche Adverbien. In *KONVENS '94.*

[Steiner et al.88] Erich Steiner, Ursula Eckert, Birgit Roth, and Jutta Winter-Thielen. 1988. The development of the EUROTRA-D system of semantic relations. In Erich Steiner, Paul Schmidt, and Cornelia Zelinsky-Wibbelt, editors, *From Syntax to Semantics, Insight from Machine Translation.* Pinter Publishers Ltd., London.

[Steiner87] Erich Steiner. 1987. Semantic relations in LFG and in EUROTRA-D - A comparison. in: IAI WP 3.

[Streiter and Schmidt-Wigger95a] Oliver Streiter and Antje Schmidt-Wigger. 1995a. The integration of linguistic and domain specific knowledge: CAT2 within ANTHEM. In *Proceedings of the Conference on Health Telematics95*, pages 387–392, Ischia, July 2-6. URL: http://www.iai.uni-sb.de/cat2/docs.html.

[Streiter and Schmidt-Wigger95b] Oliver Streiter and Antje Schmidt-Wigger. 1995b. The integration of linguistic and domain specific knowledge: CAT2 within ANTHEM. *MT News International*, 11:15–16. URL: http://www.iai.uni-sb.de/cat2/docs.html.

[Streiter and Schmidt-Wigger95c] Oliver Streiter and Antje Schmidt-Wigger. 1995c. Patterns of derivation. In *TMI-95.* URL: http://www.iai.uni-sb.de/cat2/docs.html.

[Streiter et al.94] Oliver Streiter, Randall Sharp, Johann Haller, Catherine Pease, and Antje Schmidt-Wigger. 1994. Aspects of a unification based multilingual system for computer-aided translation. In *Proceedings of Avignon '94, 14th International Conference*. URL: http://www.iai.uni-sb.de/cat2/docs.html.

[Streiter90] Oliver Streiter. 1990. Construction d'un système de traduction automatique néerlandais → allemand dans le formalisme CAT2. Rapport de stage, Jussieu, Université de Paris VII, Paris.

[Streiter94] Oliver Streiter. 1994. Komplexe Disjunktion und Erweiterter Kopf: Ein Kontrollmechanismus für die MÜ. In *KONVENS '94*. URL: http://www.iai.uni-sb.de/cat2/docs.html.

[Streiter95] Oliver Streiter. 1995. Lexical functions in ANTHEM. ANTHEM Document, 5.1.1995.

[Streiter96] Oliver Streiter, 1996. *Linguistic Reference Manual of the CAT2 Machine Translation System, Version 0.2.* Martin-Luther-Straße 14, 66111 Saarbrücken, BRD, January.

[Sumita and Iida95] Eiichiro Sumita and Hitoshi Iida. 1995. Heterogeneous computing for example-based translation of spoken lanugage. In *TMI-95.*

[Theofilidis93] Axel Theofilidis. 1993. ET-9/1 lingware report phase 2. IAI WP 26.

[Thurmair90] Gregor Thurmair. 1990. Complex lexical transfer in metal. In *TMI-90.*

[Trujillo95] Arturo Trujillo. 1995. Bi-lexical rules for multi-lexeme translation in lexicalist MT. In *TMI-95.*

[Tsujii93] Jun-Ichi Tsujii. 1993. After linguistic-based MT. In *Proceedings of the Fourth Machine Translation Summit*, pages 197–198, Kyoto, Japan.

[Tsujii95] Jun-Ichi Tsujii. 1995. MT research: Productivity and conventionality of language. In *Proceedings of the International Conference on "Recent Advances in Natural Language Processing"*, Tzigov Chark, Bulgaria, 14-16 September.

[Tucker87] Allen B. Tucker. 1987. Current strategies in Machine Translation research and development. In *MT TMI*, pages 22–42.

[Van den Berg89] M.E. Van den Berg. 1989. *Modern standaard chinees - een funktionele grammatika.* Coutinho, Muiderberg.

[van Noord et al.90] Gertjan van Noord, Joke Dorrepaal, Pim Van der Eijk, Maria Florenza, Herbert Ruessink, and Louis des Tombe. 1990. The MiMo2 research system. In *TMI-90.*

[van Noord et al.91] Gertjan van Noord, Joke Dorrepaal, Pim Van der Eijk, Maria Florenza, Herbert Ruessink, and Louis des Tombe. 1991. An overview of MiMo2. *Machine Translation*, 6(3):201–214, september.

[van Parreren and Carpey80] C.F. van Parreren and J.A.M Carpey. 1980. *Sovjetpsychologen over onderwijs en cognitive ontwikkeling.* Leerpsychologie en onderwijs 4. Wolters Noordhoff, Groningen.

[van Riemskijk and Williams86] Henk van Riemskijk and Edwin Williams. 1986. *Introduction to the Theory of Grammar.* Current Studies in Linguistics. The MIT Press, Cambridge, Massachusetts.

[Vendler67] Zeno Vendler. 1967. Verbs and times. In *Linguistics in Philosophy.* Cornell University Press, Ithaca and London.

[Verkuyl72] H.J. Verkuyl. 1972. *On the Compositional Nature of the Aspects.* Foundations of Language. Supplementary Series. Volume 15. D. Reidel Publishing Company, Dordrecht.

[Vinay and Darbelnet58] J.-P. Vinay and J. Darbelnet. 1958. *Stylstique comparée du français et de l'anglais.* Bibliothèque de Stylistique Comparée. Didier, Paris.

[Švedova80] N.Ju. Švedova, editor. 1980. *Russkaja grammatika Tom I.* Izdatel'stvo "Nauka", Moskva.

[Walter and Kirjakova90] H. Walter and E.G. Kirjakova. 1990. *Lehrbuch der bulgarischen Sprache.* VEB Verlag Enzyklopädie Leipzig, Leipzig.

[Wandruszka69] Mario Wandruszka. 1969. *Sprachen. Vergleichbar und unvergleichbar.* Piper & Co. Verlag, München.

[Weisweber94] Wilhelm Weisweber. 1994. *Termersetzung als Basis für eine einheitliche Architektur in der maschinellen Sprachübersetzung. Das experimentelle MÜ System des Berliner Projekts der Eurotra-D Begleitforschung KIT–Fast.* Sprache und Information. Max Niemeyer Verlag, Tübingen.

[Wendt87] H.F. Wendt. 1987. *Fischer Lexikon: Sprachen.* Fischer Taschenbuchverlag, Frankfurt am Main.

[W.F. and C.S.84] Clocksin W.F. and Mellish C.S. 1984. *Programming in Prolog.* Springer-Verlag, Berlin, Heidelberg, New York and Tokio.

[Whitelock and Kilby95] Peter Whitelock and Kieran Kilby. 1995. *Linguistic and computational techniques in Machine Translation system design.* Computational Linguistics. UCL Press, London.

[Whitelock92] Peter Whitelock. 1992. Shake-and-bake translation. In *COLING-92*, pages 784–789.

[WHO93] WHO. 1993. *ICD-10: International Statistical Classification of Diseases and Related Health Problems.* World Health Organization, Geneva.

[Wittgenstein84] Ludwig Wittgenstein. 1984. *Tractatus logico-philosophicus. Tagebücher 1914-1916. Philosophische Untersuchugen.* Suhrkamp Verlag, Frankfurt a.M.

[Wittgenstein89] Ludwig Wittgenstein. 1989. *Vorlesungen 1930-1935.* Suhrkamp Verlag, Frankfurt a.M.

[Wüster985] Eugen Wüster. 19985. *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie.* THe Copenhagen School of Economics, Copenhagen.

[Wygotski86] Lew Semjonowitsch Wygotski. 1986. *Denken und Sprechen.* Fischer Wissenschaft. Fischer Taschenbuch Verlag, Frankfurt am Main.

[Xrakovskij72] V.S. Xrakovskij. 1972. Aktivnye u passivnye konstrukcii v jazykax ėrgativnogo stroja. *Voprosy Jazykoznanja*, (5), 9-10.

[Zadoenko and Xuan93] T.P Zadoenko and Shuin Xuan. 1993. *Osnovy kitajskogo jazyka.* "Nauka", izdatel'skaja firma "Vostočnaja Literatura", Moskva.

[Zelinsky-Wibbelt86] Cornelia Zelinsky-Wibbelt. 1986. An empirically based approch towards a system of semantic features. In *COLING-86*, pages 7–12.

[Zelinsky-Wibbelt88] Cornelia Zelinsky-Wibbelt. 1988. From cognitive grammars to the generation of semantic interpretation in Machine Translation. In Erich Steiner, Paul Schmidt, and Cornelia Zelinsky-Wibbelt, editors, *From Syntax to Semantics. Insight from Machine Translation.* Pinter Publishers Ltd., 25 Floral Street, London.

[Zelinsky-Wibbelt89] Cornelia Zelinsky-Wibbelt. 1989. Machine Translation based on cognitive linguistics: What lexical semantics contributes to the semantic unity of a sentence. IAI WP 16.

[Zelinsky-Wibbelt91] Cornelia Zelinsky-Wibbelt. 1991. Reference as a universal cognitive process: A contrastive study of article use. IAI WP 21.

Subject Index