# Lexical Ambiguity in Machine Translation: Expressing Regularities in the Polysemy of Danish Motion Verbs

Bolette Sandford Pedersen

Center for Language Technology
University of Copenhagen

October 1997

# Contents

# Acknowledgements

This research has been funded by The Danish Research Council of the Humanities within the research project UDOG (Research into Danish Vocabulary and Grammar) carried out in the years 1993-1997 under the management of Professor Bente Maegaard, Center for Sprogteknologi. The PhD study has been supervised by Professor Ebbe Spang-Hanssen at Institut for Almen og Anvendt Sprogvidenskab, University of Copenhagen.

First of all, I would like to thank my supervisor for his valuable support and advice. I would also like to thank all my colleagues at Center for Sprogteknologi for their help; in particular, thanks should go to Bente Maegaard who has encouraged me throughout this process and who has commented on several parts of this thesis, to Patrizia Paggio, Sussi Olsen, Anne Neville, and Daniel Hart who have also provided valuable comments and input to parts of the thesis, to Costanza Navarretta and Bradley Music for introducing me to the ALEP formalism and to the LSGRAM-DK implementation, to Anna Braasch for her willingness to let me benefit from her lexicographical expertise, and to Bjarne Ørsnes for the fruitful talks I have had with him on HPSG-related issues.

Also thanks to all the participants of the UDOG project for the inspiring discussions at the UDOG seminars; in particular I am grateful for the interest that Lene Schøsler and Sabine Kirchmeier-Andersen have shown in my work and for the collaboration we have had in the project.

I would also like to thank Sanni Nimb who helped me work with the Corpus of Den Danske Ordbog, to Karen Margrethe Pedersen who opened my eyes to the variety of motion expressions in Danish dialects, and to Hanne Ruus and Lone Schack Rasmussen for the fruitful discussions I had with each of them at an early stage of this work.

Finally, I am indebted to my two kids, Kaspar and Ida who have inspired me continuously in my work, and to my husband, Hedin, without whose full support it would have been very difficult to carry out this work.

# Dansk resumé

Formålet med denne afhandling er at undersøge nye metoder til behandling af leksikalsk flertydighed i maskinoversættelse, således at man kan opnå en bedre håndtering af polysemi i det datamatiske leksikon. Specielt ønskes alternative metoder til behandling af den *systematiske* polysemi, hvor flere ord følger det samme betydningsændrende mønster, undersøgt.

Et mere empirisk delformål består i at foretage en korpusbaseret undersøgelse af danske bevægelsesverber, som ikke tidligere har været studeret ud fra denne problematik, på trods af at de udviser en høj grad af flertydighed, som gør dem særdeles vanskelige at kategorisere og behandle i datamatiske sammenhænge. Formålet med undersøgelsen er at kortlægge, hvilken systematik der gør sig gældende i bevægelsesverbernes evne til at ændre betydning i forskellige kontekster.

I afhandlingen diskuteres det, hvordan leksikalsk flertydighed traditionelt håndteres i maskinoversættelse, og det vises, hvorfor en simpel læsningsopregnende strategi, hvor relaterede betydninger (polysemi) og ikke-relaterede betydninger (homonomi) alle udtrykkes som ikke-relaterede i det datamatiske leksikon, er uhensigtmæssig i flere oversættelsesmæssige sammenhænge.

Det undersøges derfor, hvorledes man inden for nyere leksikalsk-semantisk forskning håndterer flertydighedproblematikken. Specielt anses to tilgange til verbernes polysemi for at være særligt interessante i forbindelse med det forestående analysearbejde, nemlig på den ene side Cognitive Frame Semantics (Atkins & Fillmore 1994) som inden for specifikke semantiske domæner forsøger at identificere de centrale semantiske parametre ('frame elements') og beskrive disses udmøntning i syntaksen, og på den anden side Alternationsmetoden (Levin 1993, Levin & Rappaport Hovav (under udg.)) hvor man ud fra distributionstests forsøger at kortlægge relationen mellem syntaktisk konstruktionspotentiale og *elasticitet* i betydning.

På basis af korpusanalysen etableres et lille sæt 'frame elements' for bevægelsesdomænet, nemlig *Mover, Direction, Manner, Vehicle* og *Protagonist Control*, ligesom der etableres et sæt af mulige aspektuelle værdier inden for domænet: *activity, accomplishment* og *achievement*, samt et hierarki over typiske selektions-

restriktioner på omgivende led. Ud fra disse semantiske parametre etableres en taksonomi over de danske bevægelsesverbers prototypiske betydninger, og undergruppernes syntaktiske potentiale undersøges ved hjælp af en række distributionstests. Disse tests viser, at der er en klar relation mellem syntaks og semantik: ord der har et eller flere betydningskomponenter til fælles har også en kraftig tendens til at udvise det samme syntaktiske konstruktionspotentiale.

Afdækningen af dette forhold mellem syntaks og semantik udgør en væsentlig kilde til identifikationen af forskellige typer betydningsændringer. Således identificeres tre forskellige slags systematisk polysemi i forbindelse med danske bevægelsesverber, nemlig *betydningsudvidelse, betydningsindsnævring* og *co-composition.* Beskrevet i meget korte træk afdækkes der følgende relationer mellem taksonomien og den systematiske polysemi:

- Retningsverber (*ankomme, afgå, krydse*) er meget begrænsede både syntaktisk og semantisk og kan stort set ikke undergå nogen systematiske betydningsændringer.
- Kropsverber (*gå, løbe*) og transportmiddelverber (*køre, sejle*) er syntaktisk meget fleksible og undergår også hyppigt betydningsændringer af typen betydningsudvidelse og co-composition. Verber placeret meget lavt i taksonomien (og dvs. med en meget specifik betydning) er dog noget mere restriktive, således kan verber som f.eks. *liste, promenere* og *flakke* ikke ændre betydning i samme grad som mere generelle verber fra gruppen.
- Ikke-kropsverber (*daske, drive*) kan generelt undergå alle typer betydningsændringer: betydningsudvidelse, betydningsindsnævring og co-composition, omend man også i denne gruppe finder meget restriktive verber.
- Gruppeverber (*myldre, vrimle, sværme*) er relativt begrænsede og kan kun undergå betydningsudvidelse.

Dette mønster i polysemiske egenskaber forklares bl.a. ud fra Levin & Rappaport Hovavs generelle hypotese om, at verber, der som iboende egenskab angiver en sluttilstand (som *ankomme* gør), er mindre elastiske i betydning end verber, der ikke angiver en sådan sluttilstand. De sidstnævnte verber har nemlig muligheden for via polysemiske ændringer at tilføje en sluttilstand, en mulighed som retningsverberne af gode grunde ikke har.

I den sidste del af afhandlingen beskrives, hvorledes en datamatisk ordbog med danske bevægelsesverber kan udtrykkes i en HPSG-lignende formalisme med anvendelse af typedefinerede trækstrukturer, som muliggør en høj grad af *organisering* af de udvalgte syntaktiske og semantiske oplysninger. Det vises yderligere, hvorledes den systematiske polysemi kan udtrykkes ved hjælp af leksikalske regler, således at det kun er nødvendigt at gengive de *prototypiske* betydninger direkte i ordbogen. Endvidere etableres en blokeringsmekanisme, der hindrer overgenerering ved de verber, der af etymologiske eller andre årsager udviser en mindre grad af fleksibilitet, end man kunne forvente ud fra deres semantiske tilhørsforhold. Hermed håndteres to modsatrettede egenskaber i polysemi, som ikke synes at være til at komme udenom: nemlig det systematiske aspekt på den ene side og det idiosynkratiske aspekt på den anden side. Endelig beskrives en implementering i ALEP-formalismen af 100 danske bevægelsesverber som integreres i en kørende analysegrammatik for dansk. Et testkorpus på 42 danske sætninger giver lovende kørselsresultater: i 39 af de udvalgte eksempler entydiggøres bevægelsesverberne korrekt.

Det konkluderes derfor, at den udviklede metode synes at være både operationel og tilstrækkelig nuanceret til at fange de semantiske parametre som er centrale i entydiggørelsesprocessen, og det anses for sandsynligt - omend det må komme an på nærmere undersøgelser - at metoden kan anvendes med samme held på andre semantiske domæner.

# English Summary

The primary aim of this thesis is to investigate new methods for handling lexical ambiguity in machine translation in order to achieve a more adequate treatment of polysemy in the lexicon. More specifically, we are particularly interested in finding alternative ways of treating *systematic* polysemy which is characterised by several words following the same pattern of meaning change.

A more empirical subgoal consists in performing a corpus-based analysis of Danish motion verbs which have not previously been examined from this view point in spite of the fact that they are very ambiguous and therefore very difficult to treat in machine translation. The purpose of this investigation is to identify the regularities that Danish motion verbs exhibit in the ability to undergo meaning change.

In the thesis we discuss how lexical ambiguity is traditionally treated in machine translation and it is shown why simple sense enumerating techniques, where there is no distinction between unrelated meanings (homonomy) and related meanings (polysemy), are inadequate for several analysis and translation purposes.

We therefore turn to recent research within the field of lexical semantics in order to examine how the problem of verb polysemy is dealt with here. In particular, two approaches to the problem are considered to be adequate for our purpose, namely on the one hand Cognitive Frame Semantics (Atkins & Fillmore 1994) where a taxonomical approach to verbs is taken and where, for each semantic domain, the central semantic parameters (frame elements) are identified and examined with regards to their syntactic realisation, and on the other hand the alternation method (Levin 1993, Levin & Rappapport forthcoming) which aims at an identification of the relation between syntactic potential and *elasticity* in meaning.

On the basis of a corpus analysis, a small set of frame elements is identified for the motion domain, namely *Mover, Direction, Manner, Vehicle* and *Protagonist control*; likewise, a set of possible aspectual values for motion events are established: *activity, accomplishment* and *achievement*, and finally we present a partial sort hierarchy for nouns meant for encoding selectional restrictions on motion verbs.

On the basis of these semantic features, a taxonomy of the prototypical meanings of Danish motion verbs is established and, in order to clarify their syntactic potential, each subgroup of verbs from the taxonomy is tested on a set of distributional tests. The tests expose a more or less transparent relation between semantics and syntax: words which share certain semantic features also show a strong tendency of exhibiting a similar syntactic behaviour.

The exposure of this close relationship between syntax and semantics in Danish motion constructions constitutes a first step towards an identification of different types of meaning change. Thus, basically 3 kinds of systematic polysemy are identified for Danish motion verbs: *meaning extension, meaning reduction*, and *co-composition*. Described in very short terms, the following relation between the taxonomy and the systematic meaning changes is exposed:

- Direction verbs (*ankomme* (arrive), *afgå* (depart), *krydse* (cross)) are very restricted both syntactically and semantically and can hardly undergo any systematic meaning changes whatsoever.
- Body verbs (*gå* (walk), *løbe* (run)) and body-by-vehicle verbs (*køre* (drive), *sejle* (sail)) are syntactically very flexible and often undergo meaning extension and co-composition. However, verbs placed low in the hierarchy, i.e. with a very specific meaning, generally show a much more restricted behaviour; thus, verbs like for instance *liste* (tiptoe), *promenere* (promenade), and *steppe* (tap dance) do not change meaning to the same degree as more general verbs from the group.
- Non-body verbs (*daske* (flap), *drive* (ooze)) can generally undergo all kinds of systematic meaning change, even if we in this group find also very restricted verbs.
- Group verbs (*myldre* (teem), *sværme* (swarm)) are relatively restricted and can only undergo meaning extension.

These results are partially explained on the basis of Levin & Rappaport Hovav's analysis, where it is claimed that direction verbs are 'saturated' in meaning because they inherently denote a final state of the event and therefore can't be varied, whereas manner verbs (comprising body verbs, body-by-vehicle verbs, non-body verbs and group verbs) denote activities, which are simple non-finite events, and are therefore open to variation.

In the second part of the thesis we describe how the lexicon of Danish motion verbs can be expressed in an HPSG-like formalism which applies Typed Feature Structures and which facilitates a high degree of *organisation* of the selected syntactic and semantic information. Furthermore, it is demonstrated how systematic polysemy can be formalised by means of lexical rules making it necessary only to encode the *prototypical* senses directly in the lexicon. Also, a blocking mechanism is established which hinders overgeneration in those cases where a verb for etymological or other reasons exhibits a less flexible behaviour than one should expect considering its semantic affiliation. In this way, two in essence opposite directed characteristics of polysemy are treated: the systematicity on the one hand and the idiosyncracy on the other. Finally, we describe an implementation of 100 Danish motion verbs in the ALEP formalism which we integrate in a running analysis grammar for Danish. A test suite of 42 Danish sentences comprising motion events are tested on the system and the results are very promising: in 39 of the 42 examples the motion verbs are disambiguated correctly.

We therefore conclude that the elaborated method seems to be both operational and sufficiently nuanced as to capture the semantic parameters which are central in the disambiguation process and we consider it to be plausible - although further investigations are required - that the method can be successfully extended to comprise other semantic domains as well.

# Chapter 1

# Introduction

## 1.1 Preliminaries

### 1.1.1 Lexicography - a renaissance in linguistics

During the last decade, the lexicon has been studied with increasing interest both within theoretical and computational linguistics. Part of this interest can be seen as a result of the development within the NLP (natural language processing) community where the need for well-composed and rich lexica has proven to be indispensable when dealing with large-scale NLP applications.

In other words, the fact that the lexicon always proves to be 'the bottleneck' in a practical NLP application has forced computational linguists to examine more closely the theoretical aspects of the lexicon, the layers of information that it must contain and the ways that these layers of information can interact.

A new epoch in computational lexicography began with the development of *Typed Feature Structures* which enable the linguist to express lexical information in a highly structured fashion as opposed to what is feasible in 'flat' lexical representations where each chunk of information is in essence atomic and unstructured. Also grammatical theories have been influenced by the possibility of lexical structuring of information: so-called *lexicalist* grammatical theories which allow complex and highly structured lexical entries expressing a conflation of syntactic and semantic information have had great impact within computational linguistics in recent years.

### 1.1.2 Semantic information in the lexicon

A natural consequence of the demand for better lexica is the demand for better but still operational semantic description models to put in lexica. The growing need from the NLP community for rich semantic lexica that can be applied for advanced applications such as machine translation and natural language understanding has reinforced considerable methodological research into this area.

A question that arises in this context is whether the semantic information required is the same in all kinds of NLP applications. In computational linguistics in general, truth conditional semantics (cf. Montague 1974) and its successors such as dynamic semantics (cf. Kamp 1980 and Heim 1982) has up to now been the most dominant and successful approaches to meaning representation. This is probably because of their computational tractability and because the referential account of meaning within truth conditional semantics and related theories is unsurpassed in the tasks of identifying data in databases and the like.

In contrast, when dealing with translation, more linguistically as well as cognitively based accounts of meaning seem more relevant. Basically what is required in machine translation is not necessarily a full meaning representation but rather a model which enables the system to *disambiguate* the input properly so that a correct translation can be performed. And since the disambiguation task here is another and to some extent more complex one than the one foreseen in for instance natural language-based interfaces to databases, the semantic analysis in machine translation must take other aspects of meaning into account than the referential and truth-conditional aspects.

### 1.1.3 Machine translation and the task of lexical disambiguation

Machine translation has developed rapidly during the last decade especially within the area of commercial systems; however, the problem of lexical ambiguity has not been overcome. One way to get around or at least *reduce* the problem of lexical ambiguity in machine translation has been to focus on the development of systems tuned towards very specific subdomains. Subdomains are characterised by exposing a low degree of word ambiguity but even within these restricted domains lexical ambiguity proves to constitute one of the most error-triggering problems in machine translation.

It has been acknowledged for a long time that one of the ways to improve machine translation is to put more semantics or even world-knowledge into the lexicon and there have been several attempts towards integration of such knowledge; attempts which in smaller experimental systems have proven to improve the quality of translation (see for instance Goodman & Nirenburg 1991).

Nevertheless, the *dynamics* of word meaning and the extremely complex relations between different meanings of a word in different contexts are not successfully represented in machine translation systems today. The fact that words are polysemous in nature, and that their meaning can only by vaguely defined, or be defined by means of prototypes and polysemous regularities has not yet been given much attention within the field.

Rather, the technique traditionally applied in the MT-lexicon is a static enumeration of all relevant senses of a word with no distinction between related meanings (polysemy) and unrelated meanings (homonomy) (see for instance ten Hacken et al. 1991). This means that the automatic translation process cannot fully benefit from the productive mechanisms in language which prove to play a dominant role in almost all kinds of language production - even within sublanguages.

Recent research into lexical semantics has focused on lexical ambiguity and on the *elasticity* in word meaning in order to come up with more satisfactory ways of representing word senses in the dictionary than that of pure sense enumeration. The Generative Lexicon (Pustejovsky 1995) , Cognitive Frame Semantics (Atkins & Fillmore 1994) as well as The Hypothesis of Semantic Determination (Levin 1993, Levin & Rappaport Hovav forthcoming) are all examples of such attempts. One of the aims of this thesis is to apply this research in the field of machine translation.

### 1.1.4 The ambiguity of Danish motion verbs

We have chosen Danish as the source language for our empirical research and more specifically Danish motion verbs which constitute a group of highly polysemous words in Danish. This relates to a general aspect of Danish verb lexemes, namely that these are in broad terms semantically very ambiguous and difficult to classify (see Herslund 1993). This fact about Danish lexemes highlights the importance of capturing the systematic patterns that may exist between different senses of a verb in order to be able to treat them adequately in machine translation.

Motion verbs are interesting for several reasons:

- From a cognitive point of view, verbs that denote how we move around in time and space are considered fundamental in language. Motion verbs therefore constitute a convenient starting point for analysing the nature of verb polysemy, cf. Miller & Johnson-Laird (1976) p.527:

  > 'If one wished to identify the most characteristically verbal of all the verbs, therefore, one would turn to the verbs of motion, the

verbs that describe how people and things change their places and their orientation in space.'

- Verbs that denote these very fundamental aspects of human conduct are also common in *conventionalised metaphors* by which we mean metaphors which are so regular and 'frozen' in meaning that they have lost the dynamic and pragmatically related characteristics of the traditional metaphor. We thus see conventionalised metaphors as a particular and highly frequent instance of systematic polysemy; i.e. of cases of polysemy where a group of words follows the same pattern of meaning change[1].

- A systematic syntactic and semantic description of this group of verbs is lacking for Danish. Thus, from a purely linguistic point of view, some interesting empirical research is foreseen.

To give an example of some of the aspects of ambiguity that we shall be dealing with, consider the notion of vagueness in the Danish motion verb *gå* (walk) in the following contexts, extracted from the Corpus of Den Danske Ordbog (henceforth DDO Corpus, see Andersson 1994):

(1) *når julefrokosten har været så våd, at man ikke kan gå*
(when the Christmas dinner has been so wet, that one cannot walk)

(2) *han gik pludselig ned i bådens agterende*
(he suddenly went down to the stern of the boat)

(3) *når der er gået 7 dage efter at den sidste pille blev taget, begyndes på en ny pakke*
(when 7 days have passed by since the last pill was taken a new package is initiated)

(4) *en sandet vej gik gennem plantagen der lå mellem to gårde*
(a sandy road went through the plantation which was situated between two farms)

The prototypical sense of the word *gå* is expressed in example (1), where we focus on the ability to move around in space by putting one foot in front of the other on the ground as opposed to other ways of moving around. In (2), we are still dealing with the concrete sense of *gå* but the meaning components are weighted differently. The manner component - putting one foot in front of the other on the ground - is neutralised and another meaning component has come in focus, namely the directional component of going somewhere within a short distance. (3) illustrates the conventionalised metaphor which is frequently seen with these

[1]Nevertheless, it should be noted that we shall not be dealing with the more cognitive aspects of conventionalised metaphors in this thesis.

verbs where time acts like the moving entity in the motion frame; (4) on the other hand denotes a conventionalised metaphor where locations are conceptualised as moving entities.

One might argue that we are here dealing with subtle and peripheral sense differences that we would not necessarily want to capture in our lexical descriptions. If we take machine translation as our target application, some few basic cross-linguistic examinations illustrate, however, that exactly such subtle meaning shifts are the ones that determine which construction to use in a given target language, since the lexicalisation patterns for motion in some cases differ radically from language to language. Translations of *gå* into English and Spanish in the given examples illustrate the case:

(5) *Four translations of gå*

| Example | Danish | English | Spanish |
|---|---|---|---|
| (1) | *gå* | *walk* | *caminar* |
| (2) | *gå* | *go* | *ir* |
| (3) | *gå* | *pass by* | *pasar* |
| (4) | *gå* | *go/be* | *haber* |

These examples of lexical mismatches between languages show that we often need to identify exactly which semantic parameters are triggered in a given context if we want to translate correctly.

## 1.2 The aims of the thesis

The purpose of this thesis can now be summarised as twofold:

- The general purpose is to investigate new methods for the treatment of lexical ambiguity in machine translation. Our intention is, on the basis of Typed Feature Structures, to try and adapt recent developments in lexical semantics within the treatment of polysemy to machine translation where pure sense enumeration techniques have proven to be inadequate to handle several disambiguation and translation problems. In other words, it is our intention to try and reduce the gap between traditional MT-lexicons and lexical semantic research and still come up with a practically feasible and implementable model.

- Another more specific purpose is that of providing an analysis of some not very well-studied phenomena in Danish. Danish motion verbs constitute a group of very heterogeneous and very polysemous verbs which have not to our knowledge been studied from this viewpoint before. Our intention is to examine the syntactic and semantic potential of these verbs on the basis of corpus data and, from this work, to draw a conclusion concerning the regularities in meaning change that these verbs exhibit.

## 1.3 Composition of the thesis

The thesis consists of 7 chapters. In the next chapter, we look into lexical disambiguation in machine translation and we sketch out the strategies that are traditionally applied in machine translation in order to overcome the problem of lexical ambiguity. In Chapter 3, we look at lexical ambiguity from a more theoretical viewpoint: some influential approaches to the treatment of lexical ambiguity within lexical semantics are discussed. We begin by looking at some basic considerations on lexical ambiguity and proceed to examine some recent approaches to *verb* polysemy in particular. Finally, some tentative decisions wrt. the semantic approach to be used for Danish are made. Chapter 4 focuses on motion verbs and provides a survey of different accounts of the meaning of motion verbs spanning from Conceptual Semantics to Cognitive Frame Semantics over to aspectual theories on the concept of motion, just to mention some of the approaches described.

In Chapter 5 we proceed to the empirical research of the thesis; we start out by describing the corpus work which forms a starting point for our research and proceed to the descriptive work of Danish motion verbs. Having decided on a basic apparatus based primarily on Cognitive Frame Semantics combined with some minor aspects of aspectual theories, we establish a taxonomy of Danish motion verbs and we examine which syntactic alternations the verbs can undergo in order to see to which extent our material exhibits a transparent relation between semantic categorisation and syntactic potential. These examinations help us achieve an overview of the *elasticity* in meaning that Danish motion verbs exhibit. A primary goal is to identify different *categories* of elasticity and to define *which subgroups* of verbs are sensitive to which meaning changes.

Chapter 6 is devoted to the formalisation and implementation of the dictionary. Applying a lexical sign inspired by Head Driven Phrase Structure Grammar (HPSG), which is one of the most well-known Typed Feature Structure-based theories within computational linguistics today, we provide suggestions as to how the examples of systematic polysemy given in Chapter 5 can be formalised by means of productive devices in the lexicon. In addition, we show how a frame semantic analysis can be exploited in transfer in relation to divergences between language pairs, so that word specific complex transfer rules can be avoided, allowing a more general approach. Last but not least, we give an implementation of the dictionary containing 100 Danish motion verbs in the Typed Feature Structure-based formalism, ALEP (Advanced Linguistics Engineering Platform). It has furthermore been possible to test the dictionary in the analysis phase by integrating it in the Danish grammar LSGRAM-DK which was developed within a EU-project during 1996. By parsing a test suite of 42 sentences, a first impression of the performance of the dictionary in relation to lexical disambiguation *in textual contexts* is given.

Finally, Chapter 7 contains the conclusions of the research.

## 1.4 Clarifying notes

*Corpus material:* As has already been stated, the linguistic research of this thesis builds to a large extent on corpus examples (see Section 5.2 for further details). Therefore, nearly all examples are taken from the DDO Corpus comprising approx. 36 mill. running words. One exception is the examples provided in relation to distributional tests in Sections 5.3.4 and 5.3.6; these examples are all constructed by the author.

*Translation of examples:* All Danish examples are translated into English. In cases where the structure of the Danish example differs from the parallel construction in English and where this difference is relevant for the discussion, both a literal and a paraphrased translation is given (literal translations are marked with "). However, in some cases, where there is a relevant difference in structure or lexical choice, but where the correct meaning is easily deduced from the literal translation, only the literal one is provided.

*Previous publications*: Part of the research presented here has been published just before the finalisation of this thesis: in Pedersen 1997, an earlier version of the linguistic descriptive work is presented, in Pedersen 1997a the basic strategies for formalisation of regularities in polysemy are accounted for. It should be stressed, however, that the thesis contains the most extensive and updated versions of this research.

# Chapter 2

# Lexical Disambiguation and Machine Translation

## 2.1 Introduction

Along with more and more refined parsing strategies, more computational power, and more robust mechanisms to account for unforeseen grammatical and lexical constellations, one of the remaining problems in machine translation is the disambiguation of words, and - as a consequence of this: - the problem of selecting the correct translations in the target language. Examples (1) to (4) clearly demonstrate this problem, since a proper analysis of the Danish examples is indispensable in translating correctly into English or Spanish.

Although we are mainly concerned with the *monolingual disambiguation phase* and shall only rudimentarily describe how actual transfer into the target language(s) takes place (see Chapter 6), we find it relevant in this chapter to relate more directly to machine translation in order to establish a basic framework in which the present work is to be understood. For example, it is of crucial importance to stress that we are not necessarily interested in a *very* deep semantic interpretation in the monolingual phase; actually we strive towards a simple and practically applicable model, which however, should ideally provide us with exactly the amount of information necessary for performing the relevant disambiguation. We shall go deeper into these questions in Chapter 3 where we provide a survey of existing lexical-semantic approaches to this problem and relate them to our specific purpose; in this chapter we simply aim at setting up the background for which such a disambiguation is to take place.

In the first section we briefly describe some basic strategies within machine translation today and we argue that a somewhat complex system with a thorough analysis phase is necessary if one wants to perform a proper monolingual disambiguation. This is followed by a section which describes different approaches to the task of monolingual disambiguation within machine translation systems; and, finally, we look closer at some approaches to the contrastive problem cases, namely lexical mismatches and lexical divergences between languages. In relation to this we briefly look into the selected domain for this work, namely Danish motion verbs, in order to get a first impression of the contrastive problems that occur when translating these verbs into English and Spanish, respectively.

## 2.2 Basic strategies in machine translation

It is convenient to distinguish between three basic strategies in machine translation, namely the *direct* strategy, the *interlingua* strategy and the *transfer* strategy. The direct strategy results in the most simple model with a very shallow level of monolingual analysis, in most cases limited to a morphological analysis of the input words, as illustrated in the following figure (Hutchins & Somers 1992:p.72):

(6) *The direct strategy*

source language input ⇒
morphological analysis ⇒
bilingual dictionary look-up ⇒
local reordering ⇒
target language output

However, if one aims at performing lexical disambiguation on a semantic basis, a direct strategy is clearly insufficient. Being much more sophisticated, interlingua- and transfer-based systems constitute better platforms for such an analysis, although these systems can vary from very simple ones which perform only a superficial analysis to very complex systems with several levels of analysis and generation. The difference between the two is illustrated in the figures below, where (7) illustrates the interlingua model and (8) illustrates the transfer model:

(7) *The interlingua strategy*

English analysis \ / English generation
INTERLINGUA
French analysis / \ French generation

(8) *The transfer strategy*

French analysis — French-to-English TRANSFER – English generation

English analysis – English-to-French TRANSFER – French generation

In a multilingual translation system the interlingua model may seem the most attractive, since adding a new language to the model solely consists in performing the analysis and generation of that particular language. In a transfer model, on the contrary, transfer pairs have to be constructed for each language combination in addition to the analysis and generation modules. On the other hand, the difficulties of defining a *real* interlingua which interprets input on a fully language-neutral basis, speaks in favour of adopting the more practically oriented transfer-based approach. In particular when dealing with problems like polysemy, the difficulties of establishing neutral representations for ambiguous lexical units are overwhelming. In this context a transfer model seems more reliable since it opens for the possibility of combining a semantic analysis with bilingual transfer of lexical items. In other words, in a transfer-based model a much more controlled mapping from source language to target language can be performed. It should be noted however, that in practice, many interlingua-based systems avoid lexical mapping into the interlingua by simply applying lexical transfer.

## 2.3 Monolingual disambiguation in machine translation

### 2.3.1 Ontological approaches to the lexicon

Let us first look briefly into the methodology that 'real' interlingua-based systems apply to the lexicon and to the problem of ambiguity. One famous example of a knowledge-based interlingua system is the KBMT (Knowledge-Based Machine Translation) system developed at Carnegie-Mellon in the years 1988-89 translating PC manuals from English to Japanese and vice versa (see Goodman & Nirenburg 1991). In this system an ontological approach to the lexicon is taken, mapping lexemes from the monolingual dictionaries onto what should ideally be a language-independent concept dictionary, or *ontology*, as it is also called. An ontology is a set of hierarchically ordered concepts where ambiguity in essence does not exist. Possible ambiguities in source and target languages are in theory resolved during the mapping phase when they are mapped onto an unambiguous concept. In practice, however, establishing such ontologies is a difficult task, as is also exemplified in Goodmann & Nirenburg (1991:p.180) with the English verb *remove*. It is claimed that *remove* in English is ambiguous wrt. to meaning, since it can mean either REMOVE FROM INSIDE ANOTHER OBJECT or REMOVE FROM ANOTHER OBJECT'S SURFACE; an ambiguity which in Japanese is expressed by two different lexemes. Such ambiguity must be resolved in the ontology, so the solution suggested is to split the originally established REMOVE concept into two different concepts.

Melby (1995) describes a similar working process in relation to another interlingua-based machine translation system with no specific sublanguage domain. In this project it turns out that it is in practice impossible to extend and elaborate on the ontology as the project develops; whenever a new sense of a word is encountered in the source language the ontology has to be changed correspondingly in what seems to be a never ending process. Concludingly the author claims (p.48-49):

> My view of language had changed. I had become convinced that what we had been looking for: *the* universal, language-independent set of sememes did not exist.(...) I was beginning to sense that language is more dynamic and ambiguous in a deeper way than I had ever supposed and that there is not just one built-in, correct way to view the world.

However, as is also stated by Melby, the concepts of a narrow domain and the concepts of general language are of a fundamentally different nature. In other words, there exists another kind of language than general language: *domain-specific* language exhibits a much more restricted performance than general language, and therefore concept hierarchies over specific domains come up with fruitful results in machine translation; see for instance Braasch et al. (1994) for a presentation of a concept hierarchy on motor car mechanics used for improving machine translation between English and Danish within this domain.

### 2.3.2 Disambiguation by means of domain typing

In most operational translation systems, a completely different and much more practical view to lexical ambiguity is taken. Actually, lexical ambiguity is to some extent ignored in many systems in the sense that only the most frequent or for the context relevant sense of a word is encoded in the dictionary. This naturally leads to incorrect translations in several cases, which are then to be corrected by a human translator in the postediting phase.

However, since commercial machine translation systems are often tuned towards - or especially designed for - specific subdomains, a large number of the words that occur in the relevant input texts can be categorised as domain specific terms. A main characteristics of domain specific terms is that they are mostly unambiguous, at least within the same domain. Several systems, for instance SYSTRAN, LOGOS, METAL (Schneider 1993) and PaTrans (Bech 1992), therefore apply a kind of *domain typing* with domain specific term dictionaries which override the other senses that a word may have in other contexts.

To give an example of such domain specific senses, consider the motion verb *rulle* in the following example taken from a Danish word processing manual:

(9) *brug piletasterne for at rulle gennem brevet*
(apply the arrow keys to scroll through the letter)

The most likely translation of a word like *rulle* is *roll*, however, in a computational context, this translation is overridden by the domain-specific term *scroll*.

### 2.3.3 Disambiguation by means of semantic typing

When it comes to the general vocabulary (to which the majority of motion verbs belong) disambiguation by means of subdomains is insufficient. In more advanced machine translation systems such as EUROTRA (see Copeland et. al. 1991) and METAL (see Schneider 1993 and Christoffersen 1995), *semantic typing* is employed, so that some disambiguation can take place by means of selectional restrictions. A rudimentary example of this typing is given below with the verb *udvikle*:

(10) *kogende vand*(non-human)*udvikler damp*(non-human)
(boiling water generates vapour)

(11) *forskerne* (human) *udvikler til stadighed nye metoder*(abstract) *til kloning*
(the researchers constantly develop new methods for cloning)

On the basis of such a typing on nouns, selectional restrictions can be assigned to valency-bearing words and thereby the ambiguity of a verb like *udvikle* can be described formally in the following fashion:

(12) *udvikle*1 (generate): first_argument=non-human, second_argument=non-human
*udvikle*2 (develop): first_argument=human, second_argument=entity

The assignment of *argument structure* also eases disambiguation in cases where different senses of a word differ with respect to number and type of arguments, as in:

(13) *røgen driver stadig henover byen*
(the smoke still drifts over the town)

(14) *albanerne driver dem ud af landet*
(the Albanians are driving them out of the country)

(15) *han driver en blomstrende forretning*
(he runs a blossoming business)

In some systems semantic roles are assigned to arguments in order to further describe the semantic characteristics of each argument. Together with a semantic typing on nouns, this may lead to a more fine-grained disambiguation, as is seen below:

(16) *han gav plankeværket*(goal) *maling*(theme)
('he gave paint to the hoarding')
(he painted the hoarding)

(17) *han gav hende*(recipient) *bogen*(theme)
(he gave her the book)

### 2.3.4 Preference in disambiguation

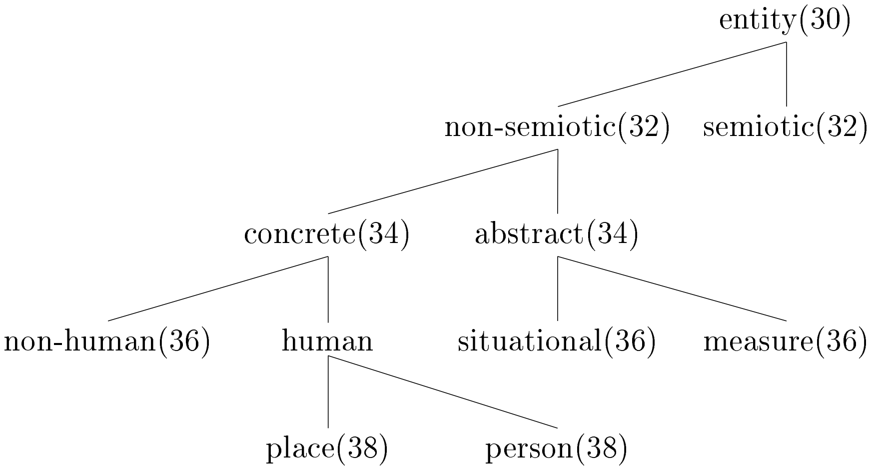
In the two previous sections the notion of selectional restrictions established on the basis of either domain typing or semantic typing was briefly presented. It is a well-known fact that selectional criteria should rather be described by means of *preferences* than by means of 'hard core' restrictions (see Jackendoff 1983) since vague or graded criteria better reflect the dynamics of words in different contexts. Therefore, some systems incorporate preference mechanisms in the lexical disambiguation phase in order to establish more robust systems that do not crash just because of a type error. An example where such soft criteria come into play is given below:

(18) *tiden gik*
('the time walked') (time passed by)

*Tiden* is 'abstract' and does not fulfil the default restriction given on the argument of *gå*, however, the sentence is fully acceptable and represents what Lakoff & Johnson (1980) denote a *conventionalised metaphor*.

In an experimental version of EUROTRA, a *graded* semantic type hierarchy is introduced in order to treat more fine-grained disambiguation tasks; see the illustration below (Pedersen 1993, p. 55)

(19) *Weighted sort hierarchy*




This system applies a notion of *semantic distance* with the basic assumption that the smaller the semantic distance between a noun sense and the slot restrictions imposed by a verb sense, the better the solution. This is implemented by means of a preference system assigning weights (referred to by the numbers in the hierarchy above) to the interpretations in question. Interpretations with high weights override interpretations with lower weights.

Such an approach has two advantages:

1. it prevents the failure of sentences where the complements do not match the requirements of the frame-bearing word, since there will always be one interpretation that 'survives' the preference module.
2. it can perform a disambiguation even in cases where two or more interpretations fulfil the requirements of the frame-bearing word, since one interpretation will most presumably have a smaller semantic distance between slot and slot filler than the other interpretations.

Point 2 can be illustrated by the following example (Pedersen 1993:p56):

(20) *industrien fremstiller antenner*
(the industry produces antennas)

where *fremstille* has got two readings in the dictionary:
*fremstille_1* (present): restriction on second argument: entity
*fremstille_2* (produce): restriction on second argument: concrete.

Since reading 2 (produce) has a more specific restriction and thus exhibits a smaller distance between slot and slot filler in the given example (which for *antenne* is 'concrete'), reading 2 wins with a preference of 4 (34 minus 30) and thus overrides the other alternative interpretation.

A similar strategy of calculation of semantic distance is used in several ontology-based approaches to machine translation (see Mahesh et al. 1997).

### 2.3.5 Different approaches to the identification of *senses* in machine translation

As we have seen, identifying the semantic *senses* that are to be established in a lexicon constitutes a major problem in machine translation, as well as in more general approaches to lexicography (see among others Cruse 1991). At present, the enumeration technique of semantic senses is normally applied in machine translation, as was illustrated in the previous section on semantic typing, with the verbs *udvikle* and *fremstille* given as examples. This means that there is no differentiation between unrelated meanings (homonomy) and related meanings (polysemy) in these lexica, since all senses are in essence treated as unrelated meanings. In opposition to this, we find generative approaches to the lexicon, as described in Pustejovsky (1995), Copestake & Briscoe (1995) and Atkins & Fillmore (1994), where the aim is to account for the productive mechanisms in polysemy and thereby minimalise the use of sense enumeration which is considered to be inadequate as a descriptive device for words that exhibit a pattern of systematic polysemy. These approaches have not yet to our knowledge been applied in machine translation at a larger scale.

Within pure enumeration techniques, however, different strategies can be applied; in bilingual systems for instance, it is tempting to look at the target language and define exactly the amount of senses in the source language that seem relevant when translating into the particular target language. However, such an approach is not satisfactory neither in a reuse perspective (if the system is extended to include a third language, for instance), nor from a theoretical point of view, since it leads to the identification of senses that is not necessarily justified in the given source language. Therefore, several works in the field (see among others Ten Hacken 1990 and Boje & Schøsler (eds.) 1992) suggest to apply a monolingual medium-size dictionary as a starting point for the monolingual codings in order to ensure that the splitting of meanings is actually founded in the given language. However, the sense distinctions established in a monolingual dictionary for human beings can hardly ever be applied directly in a machine translation context often due to unclear and/or inconsistent criteria for the splitting of meanings. Therefore, it is often convenient to let sense candidates from traditional dictionaries undergo tests like the following:

- Test 1: distributional examination, for example by means of the so-called zeugma-test[1].
- Test 2: formal examination, i.e. definition of formal readings: can the semantic distinctions found by the distributional tests be maintained as formal readings described unambiguously by means of the semantic typing system ?

Dictionary senses that do not pass the two tests must either be discarded or merged with another related sense in the computational lexicon.

We shall not here go deeper into the different kinds of semantic typing that can be applied for dictionary coding but just remark that a lexicon for machine translation purposes (and for NLP purposes in general) should not include a finer sense distinction than what is allowed by the given feature system since such a coding leads to severe overgeneration.

### 2.3.6 Statistical methods to disambiguation

Another completely different approach to word sense disambiguation which we shall at least mention here although we shall not go deeper into the different aspects of it, is the method of *statistics*. Statistical approaches typically make use of corpora by which the operator trains the disambiguation system by presenting to it the correct interpretation of words in different contexts. Such a training enables the system to disambiguate words in new contexts by using advanced pattern matching algorithms which assign felicity scores to each interpretation and selects the interpretation with the highest score (see Teo et al. 1997).

More refined systems also make use of a thesaurus or a concept hierarchy and encode the polysemous words with semantic categories from the hierarchy. The underlying idea of such an encoding is that different conceptual classes tend to appear in recognizable different contexts which can be identified on statistical grounds (see Yarowsky 1992).

[1] The zeugma-test is the following: A word has two senses, S1 and S2, if it is possible to construct two contexts which both contain the word and which both are correct in isolation but cannot be used in a coordinated structure without making the sentence ungrammatical or strange, e.g. making a zeugma (Schøsler & Boje (ed.) (1992) p.19). To illustrate this, take the examples with *udvikle* given above; we cannot say *vandet og forskerne udvikler nye metoder* (the water and the researchers develop new methods) or anything of the like; therefore we must be dealing with two senses of the verb. See also Cruse 1991 p.61-62

Several statistical methods come up with rather promising results as regards percentages of correct analyses (from 80 to 90). However, to our knowledge, a statistically based disambiguation tool has not yet been fully implemented in a running machine translation system.

## 2.4 Specific problem cases in translation: lexical mismatches and lexical divergences

Although monolingual disambiguation is the first step towards better translations, it is a well-known fact that other problems of correct lexical mapping can occur in the translation phase; namely problems of *lexical mismatches* and *lexical divergences* (see Dorr 1993 and Barnett et al. 1993, among others). Another closely related aim of the monolingual analysis is therefore to capture the types of information that are necessary to handle such problem cases.

*Lexical mismatches* occur when source and target language differ wrt. the information conveyed by the lexical item. Typically this discrepancy concerns vagueness or ambiguity in one language which in another language is expressed by different lexemes. An example of this is the English word *wall* which in Danish is either *mur* or *væg* depending on whether one refers to the vertical sides of a room or to the long narrow vertical structure made of bricks. If we translate from Danish to English we generalise into the English term *wall* or actually *throw away* information from the noun; if we translate the other way around we *add* information to the noun about which part of the wall we are talking about:

(21) *der hang et billede på væggen* → *a picture was hanging on the wall*

(22) *he climbed the wall* → *han klatrede over muren* (**han klatrede over væggen*)

Also in the case of motion verbs, lexical mismatches are frequent; consider the example below:

(23) *he went to the Baker's to buy bread* → *han tog/kørte/gik til bageren for at købe brød*

In the Danish translation we are forced to add information either of a long distance (the Danish verb *tage* indicates a distance longer than walking distance) or of a manner component which is not present in the English sentence, but which in Danish is expressed by the verbs *gå* (walk) or *køre* (drive).

*Lexical divergences*, on the other hand, occur when source and target language apply different structures for expressing the same information. An example of lexical divergence is seen below:

(24) *me gusta comer* → *Ich esse gern*
(I like to eat)

where the content of the verb *gustar* is expressed by an adverbial phrase in German (*gern*). Also motion verbs present many cases of divergences, especially when translating between Germanic and Romance languages. Consider the example below:

(25) *han svømmede ind i grotten* → *entró en la cueva nadando*
(he swam into the cave)

The example is inspired by Talmy (1985) who is concerned with the different lexicalisation patterns in motion events within different language families. This example shows that Danish and Spanish lexicalise different aspects of motion in the matrix verb and leave different aspects of motion to be expressed by means of modifiers. We shall come back to this discrepancy in the following chapters.

In transfer-based systems, such restructuring can be performed by means of structural transfer-rules which map the structure of the source language into the appropriate structure of the target language. However, if such complex transfer-rules are established on the basis of a shallow semantic analysis they cannot be defined on a general basis but must be word-specific. Since the potential of word combinations is very large, word specific transfer-rules are in many cases inconvenient because they often lack the ability of covering the linguistic phenomenon as a whole, but only examples of it.

Therefore, if information is to be added or thrown away *automatically* (as is the case with lexical mismatches), or if a phrase is to be completely restructured in order to convey the same information in a target language (as in the case with lexical divergence), a semantic analysis deeper than that of semantic typing is called for. This fact has been acknowledged within the field of machine translation for a long time, and several research-oriented approaches have been developed through the last decades with the aim of overcoming problems of this nature.

One of these is the KBMT system which as we have already seen strives towards representing the 'full' meaning of the input text, comprising a relatively large spectrum of semantic and pragmatic aspects such as speech-act assignment, discourse structure, modality, temporal calculations, assignment of thematic roles, semantic typing of objects on the basis of semantic networks comprising both subworlds and the 'everyday-world', etc. In this system, problems of lexical mismatches and divergences should ideally disappear, since the input sentence is fully neutralised wrt. syntax and lexical choices.

A more lexicalist approach to machine-translation which also tries to overcome the problem of lexical mismatches and especially lexical divergences is UNI-

TRAN (see Dorr 1993). This system is also interlingua-based and applies a lexical-semantic representation based on Jackendoff's *Lexical Conceptual Structure* (see Chapter 4). Again an *abstraction* of language-independent properties from structural idiosyncrasies is aimed at and by defining a small number of divergence types, a systematic mapping between interlingua and syntactic structure is achieved.

## 2.5 Summary

To sum up, we have in this chapter shown that lexical ambiguity constitutes a major problem in machine translation and we have roughly described two lines in the treatment of the topic[2]:

- An *operational* line which either more or less ignores the problem of ambiguity or handles the topic by means of domain typing and in some cases by use of simple selectional restrictions;
- a *research-oriented* line, which tries to give a 'deep' meaning representation and thereby overcome the problem of lexical ambiguity

Although it can be questioned whether the extraction of the 'full' meaning is necessary in order to achieve adequate machine translation, it is difficult - in theory - to present valid objections to knowledge-based interlingua systems providing such 'full' semantic representations. There are always cases - especially when translating between languages that are structurally far apart - where 'full' abstraction is required in order to get an acceptable translation and it is extremely difficult beforehand to predict the proper candidates for deep analyses. On the other hand, the practical difficulties of defining the framework of such 'full' analyses and the severe problems of extending research-oriented systems, which treat carefully selected domains, into systems that are actually practically useful, has resulted in a general scepticism towards the future applicability of such systems. Rather, it should be accepted that not all linguistic divergences can be treated appropriately in machine translation and that a subset of sentences will therefore always be left over to the postediting phase.

Another objection that can be made to interlingua systems is that they do not pay sufficient attention to the syntax in the source language and that this lack of attention may mean that important information is actually lost during analysis and therefore not represented in the target text. This objection has led to

[2]We exclude now from the discussion the purely statistical methods to word sense disambiguation.

the question of whether a truly 'universal' and language-independent interlingua *can* actually be defined; an interlingua which at the same time captures all the different nuances represented in syntax.

To conclude, we have demonstrated that there is a need for an *intermediate* solution to the treatment of lexical ambiguity in machine translation which on the one hand is operationally feasible and on the other goes beyond simple selectional coding and captures other basic aspects of lexical meaning relevant for disambiguation.

As a part of this discussion, it should also be questioned whether the sense enumeration technique traditionally used in machine translation is the most adequate for handling lexical ambiguity in this field; could it be so that techniques that actually capture the relatedness in meaning between different senses of a word would actually ease the disambiguation phase and/or improve the quality of translation ? This question constitutes a working hypothesis for the following chapters, where one of the aims is to reduce the gap between the somewhat 'conservative' lexical approach traditionally applied in machine translation and some of the more recent approaches to the lexicon described in computational linguistics and lexicography today.

Finally, it has become clear that our analysis should be textually anchored and pay sufficient attention to syntax, since these two levels of analysis are closely interrelated. Especially when dealing with a *satellite-framed* language like Danish[3] which is characterised by having a relatively small vocabulary but a complex syntax with a heavy semantic weight on particles and adverbs, much focus is required on syntax. As a consequence of this we opt for a transfer-based approach which allows us to refer more directly to the syntax of the source language.

[3]The term satellite-framed is taken from Talmy (1985) as opposed to verb-framed languages such as Spanish.

# Chapter 3

# Some Influential Approaches to the Treatment of Lexical Ambiguity

## 3.1 Introduction

In the previous chapter we have demonstrated that lexical ambiguity is a central problem case in machine translation and we have roughly sketched out some of the methods that are applied in machine translation as an attempt to overcome this problem. We also proposed that machine translation could benefit from a more flexible approach to the lexicon in order to capture the correspondences between related meanings and thereby improve translation.

In this chapter we shall therefore look into the problem of lexical ambiguity from a more theoretical point of view by referring to some of the more influential approaches to this topic within the field of lexical semantics.

The chapter falls into two parts. In the first part we refer to some standard works within lexical semantics which treat lexical ambiguity in general, but which focus primarily on the ambiguity of *nouns*. In the second part we shall look at some recent approaches to the treatment of *verb* ambiguity which obviously incorporate basic characteristics of ambiguity but which at the same time display another complexity than what is found in standard works.

## 3.2 Homonomy and polysemy - some basic considerations on lexical ambiguity

Lyons (1977:p.550-569 as well as 1995:p.46-60) refers to two kinds of lexical ambiguity: *homonomy* and *polysemy*. A standard criterion for the division into homonomy and polysemy is that of *etymology*; if two senses are historically derived from the same word then they can be considered to be polysemous, otherwise we must be dealing with homonomy. Another central criterion is *relatedness in meaning*: if two senses are related in meaning then they can be categorised as polysemous; otherwise they are homographs.

Lyons objections to these criteria are that they are not operational; we do not always know the etymology of words, and more important: we are not always aware of etymology when using words. As regards the second criterion, we cannot, according to Lyons, define relatedness in meaning on operational grounds. Relatedness in meaning constitutes a continuum and it is therefore in practice impossible to define a clear borderline between one sense and another, and therefore also between homonomy and polysemy[1].

Cruse (1991:p.49-83) adopts an intermediate position between those linguists that believe that a word form is associated with a number of discrete senses and those (like Lyons) who believe that such discreteness is fundamentally illusory. He does so by defining different kinds of polysemy, namely *sense modulation* and *sense selection*. Sense modulation can, according to Cruse, not be expressed in discrete 'jumps', since every context in essence modulates its own 'sense' or interpretation of a word. From a lexicological point of view it is therefore clearly inconvenient to try and describe such modulations as individual senses, partly for theoretical reasons, partly for practical reasons since the dictionary would explode with redundancies.

*Highlighting* and *back-grounding* is according to Cruse an important aspect of sense modulation; each context seems to highlight and background specific aspects of a sense, as can be seen in the examples below (Cruse 1991:p.53):

(26) *We can't afford a car*
(27) *Our car couldn't keep up with his*
(28) *The car crushed Arthur's foot*

In these examples, the price, the speed and the weight of the car are the aspects of meaning focused upon, respectively, but this does not mean that individual

[1]Note that this does not necessarily imply that word meaning cannot be defined on discrete grounds. Wierzbicka (1985) illustrates how the notion of vagueness *can* be defined in discrete terms.

senses of a car should be defined in the lexicon.

In opposition to sense modulation, sense selection can, according to Cruse, be described in terms of discrete senses. He puts forward a list of distributional tests to be applied in order to test sense selection, or *real* ambiguity. One of these tests is the zeugma test, which we mentioned in Chapter 2, another is the test of oppositeness which is illustrated below (Cruse 1991:p.55):

(29) *The room was painted in light (≠ dark) colours*

(30) *Arthur has rather a light (≠ heavy) teaching load*

In these examples, the antonyms to *light* differ, and this should be an indication of two individual senses.

Unfortunately, as is also stated by Cruse, distributional tests do not always give clear answers to the question of sense distinctions. Consider the following example (Cruse 1991:p.57):

(31) *a thin woman (≠ fat) woman*

(32) *a thin (≠ thick) tree-trunk*

In spite of different antonyms, it seems counter-intuitive to consider the two occurrences of *thin* as expressing a case of two individual senses.

In Chapter 5 on Danish motion verbs, we shall take Cruse's division into sense modulation and sense selection as our starting point when defining polysemes to the verbs in question. Also we shall consider more closely to which degree distributional tests can be applied for the testing of ambiguous verbs.

## 3.3 Some recent approaches to verb ambiguity

Since we have chosen Danish motion verbs as our specific case study, we are particularly interested in approaches to *verb* ambiguity. In opposition to nouns, which denote entities, verbs denote events and can therefore not be assumed to behave in completely the same way as nouns with respect to ambiguity. In the following we shall look closer at some recent approaches to verb ambiguity.

### 3.3.1 Polysemy and diathesis alternations

Levin (1993) focuses on the relation between polysemy and *syntagmatic effects*, such as diathesis alternations. One of her basic assumptions is that of the *Hypothesis of Semantic Determination*, which includes the following assumptions:

a. If the members of a set of verbs share some meaning component, then the members of this set can be expected to exhibit the same syntactic behaviour, and

b. If the members of a set of verbs exhibit the same syntactic behaviour, then the members of this set can be expected to share some meaning component.

Levin pursues the determination hypothesis by first examining alternate ways in which verbs express their arguments in English. From this a methodology emerges for classifying verbs from a distributional point of view and hereafter follows a classification of more than 3000 English verbs according to shared meaning and behaviour with samples of possible alternations for each subclass.

Following the determination hypothesis, Levin is able to uncover a wide range of syntactic alternations that reflect verb meaning, and thus brings us closer to the understanding of the underlying mechanisms in verb polysemy. Furthermore, although Danish differs from English in several respects, her categorisation work on English motion verbs has proven to be of considerable importance to the categorisation work that we describe in Chapter 5 on Danish motion verbs.

In more recent work, Levin & Rappaport Hovav (forthcoming) look into some general characteristics of verbs which delimit the ways in which verbs can extend their meanings. They look into two major groups of verbs, namely *manner verbs* and *result verbs* and it is demonstrated how result verbs are restricted with respect to syntactic potential and thereby also with respect to meaning change, whereas manner verbs show a wide range of syntactic and semantic potential, as is demonstrated below (Levin & Rappaport Hovav p.1 and p.7):

Manner verbs:

(33) *Terry swept*

(34) *Terry swept the floor*

(35) *Terry swept the crumbs into the corner*

(36) *Terry swept the leaves off the sidewalk*

(37) *Terry swept the floor clean*

Result verbs:

(38) *Kelly broke the dishes*

(39) **Kelly broke*

(40) **Kelly broke the dishes off the table*

(41) **Kelly broke the dishes into a pile*

In order to formally describe this difference in elasticity, Levin & Rappaport Hovav first identify two aspects to verb meaning: a *structural* and an *idiosyncratic* one[2] (p.10-11):

- STRUCTURAL: linguistically analysed information; has 'grammatical life'; determines verb behaviour (e.g. the expression of arguments), defines linguistically relevant classes;
- IDIOSYNCRATIC: cognitively analysed, but linguistically atomic, 'inert' information; distinguishes among the members of a semantic class.

To represent the structural aspects of meaning a set of basic event structure templates is established on the basis of the traditional Vendlerian verb classes (p.12):

(42) activity: [x $\mathrm{ACT}_{<MANNER>}$]
state: [x <*STATE*>]
achievement: [ BECOME [x <*STATE*>]]
accomplishment: [ [x $\mathrm{ACT}_{<MANNER>}$ ] CAUSE [ BECOME [y <*STATE*>]]]
accomplishment: [ CAUSE [ BECOME [x <*STATE*>]] ]

The idiosyncratic aspects of meaning on the contrary are expressed by means of an open-ended set of constants, like for *sweep*: 'with a broom', to give an example. Finally a set of realisation rules combines the two levels of meaning by associating constants of idiosyncratic meaning with templates of structural meaning.

In conclusion, difference in syntactic and semantic potential is described on the basis of the event structure analysis where result verbs are seen as complex event structures denoting *accomplishments* or *achievements* which are - so to speak - already saturated with respect to meaning potential since a final state of the event is inherent in the verb as can be seen in following entry for *break*:

(43) [ [x ACT ] CAUSE [ BECOME [y <*BROKEN*>]]]

Manner verbs on the contrary inherently denote a simple event structure, namely an activity:

(44) [x $\mathrm{ACT}_{<SWEEP>}$ y[3]]

[2]The distinction corresponds roughly to Grimshaw's in Grimshaw 1993 (apud Levin & Rappaport Hovav (forthcoming)) who labels the two *semantic structure* vs. *semantic content*, as well as to the distinction made by Tenny (1995).

[3]*y* refers to surface and is mapped to the object, as in *he swept the floor*.

The meaning can, however, be extended by adding a state and thereby turn the event into an accomplishment as is depicted below:

(45) $[\ [\text{x ACT}_{<SWEEP>}\ \text{y}\ ]\ \text{CAUSE}\ [\ \text{BECOME}\ [\text{y} <STATE>]]]$

Since we see the same division in motion verbs, namely a distinction between manner and result verbs (for instance *svømme* (swim) vs. *ankomme* (arrive)), we shall come back to this analysis in Chapter 5 on Danish motion verbs when we look more closely at the meaning changes that Danish motion verbs can undergo.

### 3.3.2 The Generative Lexicon

Another recent approach to verb polysemy is that of Pustejovsky (1995) generally referred to as *The Generative Lexicon*. Pustejovsky argues against the traditional sense enumeration technique which he claims is inadequate for describing the dynamics of word senses. He therefore aims at a composite lexical model which incorporates four central aspects of meaning and which takes into account the interaction of the semantics of different word classes: (p.61):

1. Argument Structure: Specification of number and type of logical arguments, and how they are realized syntactically.
2. Event Structure: Definition of the event type of a lexical item and a phrase. Sorts include STATE, PROCESS, and TRANSITION, and events may have subeventual structure.
3. Qualia Structure: Modes of explanation, composed of FORMAL, CONSTITUTIVE, TELIC, and AGENTIVE roles.
4. Inheritance Structure: Identification of how a lexical structure is related to other structures in the type lattice, and its contribution to the global organisation of the lexicon.

Although new aspects are added to argument structure and event structure, the most controversial part of Pustejovsky's theory is constituted by the Qualia Structure which in our opinion touches the borderline between linguistic knowledge and world-knowledge. The levels of Qualia Structure are further described below:

- CONSTITUTIVE: the relation between an object and its constituent parts (material, weight, parts and component elements);
- FORMAL: that which distinguishes it within a larger domain (orientation, magnitude, shape, dimensionality, colour, position);

- TELIC: its purpose and function (purpose that an agent has in performing an act, built-in function or aim that specifies certain activities);
- AGENTIVE: factors involved in its origin or 'bringing it about' (creator, artifact, natural fact, causal chain).

So where traditional lexical semantics can only account for the two interpretations of *bake* in examples like the following:

(46) *John baked the cake*

(47) *John baked the potato*

by assuming that we are dealing with two readings of *bake*, namely a create reading and a change-of-state reading; Pustejovsky claims that there is only one reading, namely a change-of-state reading as is seen in the following AVM (attribute-value matrix) representation :

(48)
$$\begin{bmatrix} \text{bake} & \\ \text{EVENTSTR} = \begin{bmatrix} \text{E}_1 = \text{e}_1\text{:process} \\ \text{HEAD} = \text{e}_1 \end{bmatrix} \\ \text{ARGSTR} = \begin{bmatrix} \text{ARG1} = \boxed{1}\begin{bmatrix} \text{animate\_ind} \\ \text{FORMAL} = \text{physobj} \end{bmatrix} \\ \text{ARG2} = \boxed{2}\begin{bmatrix} \text{mass} \\ \text{FORMAL} = \text{physobj} \end{bmatrix} \end{bmatrix} \\ \text{QUALIA} = \begin{bmatrix} \text{state\_change\_lcp} \\ \text{AGENTIVE} = \text{bake\_act}(\text{e}_1, \boxed{1}, \boxed{2}) \end{bmatrix} \end{bmatrix}$$

and that any other sense is derived through generative mechanisms in composition with its arguments. Thus, Pustejovsky puts more semantic 'weight' on the noun and sees creation as a central semantic feature of an artifact (*cake*) and change of state as a central semantic aspect of natural kinds *potato* (these are therefore assigned as features in the Qualia Structure of the given nouns). The example is considered to be a case of logical polysemy which can now be accounted for by means of a *generative operation*, in this case *co-composition*[4], and not by means of numerous reading distinctions in the lexicon. In an AVM representation, a

[4] Co-composition is defined on p.61 as cases where 'multiple elements within a phrase behave as functors, generating new non-lexicalised senses for the words in composition', in this case, therefore, both *cake* and *bake* behave as functors.

phrase like *bake a cake* would trigger a co-composition rule and be represented as follows (p.125):

(49)
$$\begin{bmatrix} \text{bake a cake} & \\ \text{EVENTSTR} = & \begin{bmatrix} \text{E}_1 = \text{e}_1\text{:process} \\ \text{E}_2 = \text{e}_2\text{:state} \\ \text{RESTR} = <_{\propto} \\ \text{HEAD} = \text{e}_1 \end{bmatrix} \\ \text{ARGSTR} = & \begin{bmatrix} \text{ARG1} = \boxed{1} \begin{bmatrix} \text{animate\_ind} \\ \text{FORMAL} = \text{physobj} \end{bmatrix} \\ \text{ARG2} = \boxed{2} \begin{bmatrix} \text{artifact} \\ \text{CONST} = \boxed{3} \\ \text{FORMAL} = \text{physobj} \end{bmatrix} \\ \text{D-ARG1} = \boxed{3} \begin{bmatrix} \text{material} \\ \text{FORMAL} = \text{mass} \end{bmatrix} \end{bmatrix} \\ \text{QUALIA} = & \begin{bmatrix} \text{create\_lcp} \\ \text{FORMAL} = \text{exist}(\text{e}_2,\boxed{2}) \\ \text{AGENTIVE} = \text{bake\_act}(\text{e}_1,\boxed{1},\boxed{3}) \end{bmatrix} \end{bmatrix}$$

What the representation claims, according to Pustejovsky, is that the creation sense of *bake* embeds the change-of-state reading within it by systematic rules of composition.

Seen from a machine translation perspective, many aspects of Pustejovsky's theory are attractive since he presents a formalised model in which sense enumeration can be avoided and overlapping meaning components in the different uses of a word can be represented. However, we believe that several parts of his theory are far from being as operational as one could wish for in a machine translation context. This relates in particular to the Qualia Structure where there seems to be some difficulties in defining the properties of each role (cf. among others Vikner & Jensen (1994)) and where one could expect a considerable amount of overgeneration as a consequence of the strong generative devices established. Having decided in Chapter 2 on a transfer model which provides some semantic analysis but not a 'full' analysis, a Qualia Structure analysis which includes several aspects of world-knowledge would therefore not be appropriate.

What we suggest to take over from Pustejovsky is therefore rather some very general aspects of his model: the combination of several levels of semantic structure, i.e. argument structure, event structure etc. together with the overall idea of generating new meanings from a set of core lexical representations. As regards the generative mechanisms of his lexicon, the notion of co-composition is of particular interest to us since it comes into action in many cases of verb polysemy, as will become clear in Chapter 5.

### 3.3.3 Corpus-based Frame Semantics

The last approach to verb polysemy that we shall take into account here is that of Fillmore & Atkins (1994) which relates to Frame Semantics, initially presented in Fillmore (1985) and further developed in Alonge et. al. (1993). In Frame Semantics, a taxonomic approach to verbs is adopted - as was the case with Levin - dividing verbs into semantic classes and subclasses so that words that can be considered semantic neighbours with similar sets of semantic parameters (henceforth called *frame elements*) are treated coherently. In Alonge et al. (1993) Frame Semantics places itself as follows:

> Frame Semantics is of all an approach to the understanding and description of the meaning of lexical items and grammatical constructions. It begins with the uncontroversial assumption that in order to understand the meanings of the words in language, we must first have access to the conceptual structures, - or as we call them, *semantic frames* - which provide the background and motivation for the existence of these words in the language and for their use in discourse. It is then assumed that an account of the meaning and function of a lexical item can proceed from a description of the underlying semantic frame to a characterisation of the manner in which the lexical item in question, and the phrase built around it, can convey information about aspects or instances of that frame.

In Fillmore & Atkins (1994), the particular case study concerns the verb *risk*. In this paper, the authors view the problem of verb ambiguity from the lexicographer's point of view and suggest to leave the traditional 'dictionary sense' aside in favour of what they define as *situation types* with sets of frame elements connected to each type.

On the basis of more than 3000 corpus citations of *risk* they establish the following set of frame elements for the risk scenario (p.16):

(50) Protagonist [Pr] the central person in the frame.
Bad [Ba] the possible bad outcome, or harm.
Decision [De] the decision that could trigger this.
Goal [Go] the desired outcome.
Setting [Se] the situation within which the risk exists.

Possession [Po] something or someone valued by the Protagonist and endangered in the situation.
Source [So] something or someone which could cause the harm.

Three basic situation types are now established:

- a type A, which includes primarily a Protagonist [Pro] and a possible bad outcome [Ba], as illustrated in *babies ran the risk of hypothermia there*,
- a type B, in which the protagonist [Pro] takes a decision [De] which makes him or her vulnerable to some sort of harm [Ba], as in *I had no idea I was risking my life*, and finally
- a type C, which furthermore includes the frame element Goal [Go], i.e. the desired outcome which the decision-maker has in mind, as expressed in: *she risked everything for a few dollars.*

An important ingredient of corpus-based Frame Semantics is to record the grammatical means by which frame elements are expressed in texts, as is exemplified in the database record below (Atkins & Fillmore 1994:p.17a):

(51)
```
EXAMPLE: I wouldn't risk letting you meet her
CATEGORY: verb
COMPLEMENT: ger-vp/letting you meet her
CLAUSE: main
CLAUSETYPE: negative
..
V_SUBJECT: I
V_GRCONTXT: [null]
V_TENSEETC: ind\wouldn't
PROTAG_LEX: I
BAD_GR: [null]
BAD_LEX: [null]
DECISION_GE ger-vp
DECISION_LEX: letting you meet her
..
```

In Atkins & Fillmore (1994) frame elements can either map to arguments or modifiers in the text or they can be implicit as inherent features to words or be deduced from a larger context. In other more restricted variants of Frame Semantics (see Alonge et al. 1993 and Heid & Krüger 1996), frame elements are directly comparable to semantic roles since they can only be assigned to arguments or modifiers in the surrounding context and thus not go beyond the lexicalisation border.

For our purpose, the basic notions of frame semantics seem very attractive, for the following reasons:

- frame elements constitute a good alternative to traditional semantic roles since they are defined more precisely on the basis of specific semantic domains[5] and not on a very general basis[6],
- being corpus-oriented, Frame Semantics pays much attention to syntax, i.e. the method is textually anchored to a degree such that also more complex syntactic configurations can be accounted for, and finally
- some variants of Frame Semantics allow for an assignment of frame elements which goes beyond the lexicalisation border, or in other words: which combines inherent and structural frame assignment; recall from Chapter 2 that this was an important aspect when dealing with contrastive divergences, as is the case in machine translation.

In spite of these attractive aspects of Frame Semantics, some reservations should be noted. In the case of Atkins & Fillmore (1994), for example, one can question to which degree frame element assignment can take place on an operational basis; in other words, for our purpose, the set of frame elements should be reduced to a minimum including only core frame elements that can be clearly defined within a sentence, preferably on a distributional basis.

Another reservation concerns the incompleteness of a frame semantic analysis alone: as has been mentioned in several recent lexical-semantic accounts (see for instance Levin & Rappaport 1991, Levin & Rappaport Hovav 1996, Tenny 1995 & 1995a, as well as Dini & Tomaso 1995), aspectual theories come up with classifications and descriptions that are in several cases more adequate than do localist theories; if we choose to apply a Frame Semantic approach we should therefore consider to combine it with aspectual information . We shall come back to this point in Chapters 5 and 6.

[5]Examples of semantic domains treated in Frame Semantics are: perception domain, speech-act domain, motion domain etc.

[6]It should be noted that there are of course other advantages of very generalised thematic role sets, see for instance Dowty (1991) for his work on proto-roles and argument selection.

## 3.4 Summary

In this chapter we have looked briefly into the theoretical problems of lexical ambiguity in general and we have looked more closely at three approaches to verb polysemy. If we compare with the methods described in Chapter 2 on machine translation, it has become clear that there is a considerable gap between approaches traditionally applied in machine translation and approaches launched recently within the field of lexical semantics. One of the aims of this work is to try and reduce this gap.

The hypothesis of semantic determination pursued by Levin (1993) and the work on restrictions in verb polysemy by Levin & Rappaport Hovav (forthcoming) contribute to our analysis as a methodology of testing which semantic parameters come into action when and what the consequences are for verb polysemy. Pustejovsky (1995) integrates three levels of semantic structure in his analysis and accounts for the interactions between these in word polysemy. We have argued that his model as a whole is too ambitious for machine translation purposes; several aspects of his theory are in our opinion not fully operational and sufficiently restricted. However, some basic assumptions of his model are taken over in this work; the general notion of new senses derived on the basis of core senses as well as the concept of co-composition.

Finally we have chosen to incorporate a Frame Semantic approach in the lexicon influenced by the work of primarily Fillmore (1985) and Atkins & Fillmore (1994). There are several reasons for this choice: frame elements constitute a good alternative to traditional semantic roles; they are defined within specific semantic domains and can be freely assigned across the lexicalisation border. Furthermore, Frame Semantics has a tradition of being highly corpus-oriented and thereby very textually anchored; an attractive aspect considering that we are dealing with a satellite-framed language, as has already been mentioned in Chapter 2. It was decided, however, that a frame semantic analysis should be supported by an aspectual analysis in the line of Levin & Rappaport Hovav (forthcoming).

# Chapter 4

# A Survey of Different Approaches to Motion Verbs

## 4.1 Introduction

In the previous chapters we have looked at verb ambiguity from a machine translation viewpoint (Chapter 2) as well as from a more theoretical viewpoint (Chapter 3). In this chapter we shall start looking into our specific semantic domain, motion verbs, and we shall do so by giving a survey of some previous approaches to motion verbs in other languages as well as in Danish. Although we have already in Chapter 3 made some very basic decisions as regards the apparatus to be applied in the treatment of Danish motion verbs, we see the need of getting a broad overview of how our specific semantic domain has been treated within other theories as well in order to ensure that all relevant aspects are taken into account.

Motion verbs have been the subject of numerous studies by linguists both within syntax and semantics. It can be argued that this major interest in motion verbs results from the fact that meaning is to some extent conceived as mental pictures. From this it follows that meaning primarily deals with spatial entities and relations and that everything else is simply secondary metaphors of this. Motion verbs therefore constitute one of the primary bases for an investigation of meaning and thereby for an investigation of lexical ambiguity.

### 4.1.1 Decompositional approaches to motion verbs

**Language and perception, Miller & Johnson Laird**

Miller & Johnson Laird are concerned with the relation between language and perception, and in this respect motion verbs constitute an interesting class since it has a strongly perceptual basis. Motion verbs are defined as verbs that describe 'how an object changes from a place p at time t to another place p' at a later time t+i'.

In order to establish some guidelines for the delimitation of motion verbs, the English verb *travel* is defined as a generic concept of change of location. This verb is used in a distributional test that should enable the language user to decide whether a verb belongs to the group of motion verbs. If the following sentences sound odd, then the tested verb (underlined) belongs to the motion verb group:

(52) ?*He <u>walked</u>, but he travelled*
?*He <u>walked</u>, but he didn't travel*

For the causative motion verbs a similar test is elaborated:

(53) ?*He <u>raised</u> it, but it changed location*
?*He <u>raised</u> it, but it didn't change location*

The area of interest is further limited to verbs that denote locomotion and not only motion, excluding therefore verbs like *revolve, spin, turn* etc.

The concept 'travel' is now schematized in the following notation, where AT characterises location (p. 532):

(54) *The 'travel' concept*

TRAVEL($x$): Something $x$ 'travels' if there is a time $t$ and a place $y$ such that:
Either: $R_{t-i}(notAT(x,y))$ & $R_t(AT(x,y))$
Or: $R_{t-i}(AT(x,y))$ & $R_t(notAT(x,y))$
where $i > 0$

Using a predicate decomposition approach, different manner components which are lexicalised as in *soar*, *drift* and *ride*, can now be formalised as follows (p.534):

(55) *soar* travel through the air (THROUGH(TRAVEL))(x,AIR)
*drift* travel on water (ON(TRAVEL))(x, WATER)
*ride* travel in conveyance (IN(TRAVEL))(x,CONVEYANCE)

Directional meaning components are expressed in a similar fashion:

(56) *ascend* (UPWARD(TRAVEL))(x)
*enter* (INWARD(TRAVEL))(x)

Finally, comments are made on other semantic components that can be part of the motion event, such as aspectual information which is expressed for example by the difference between *travel* and *reach*, deictic information as expressed by verbs like *come* and *go*, as well as iterative aspects of some motion verbs like *hop* and *bounce*.

### Motion verbs in Conceptual Semantics

Jackendoff (1983, 1990) adopts and further develops Gruber's localist approach within the theory of Conceptual Semantics. For motion verbs he thus follows Gruber's analysis, denoting the NP whose position is changed 'theme'. Elaborating what he calls a conceptual structure of the sentence, a typical motion event is represented as below applying a predicate-argument notation similar to that of Miller & Johnson-Laird:

(57) [Event GO ([Thing $x$ ], [Path $y$ ])]

where the variables $x$ and $y$ typically represent the information to be filled in from the subject and the prepositional phrase of the sentence, respectively, as in:

(58) *Bill flew around the pole*

Here GO is the primitive for an Event of the motion type, and Thing and Path are the semantic labels applied for the group of NP's or PP's that would typically fill in the given slots.

Causative motion verbs are formalised by means of an extra Event type, labelled CAUSE:

(59) [Event CAUSE ([Thing $x$],[Event GO ([Thing $y$],[Path $z$])])]

This representation would thus cover a sentence like:

(60) *Amy put the flowers in the vase*

In a later account (Jackendoff 1990), Jackendoff looks closer at a group of motion verbs which does not imply traversal of a path, and which therefore cannot be described with the original set of primitives provided in Jackendoff (1983). The examples given are:

(61) *Willy wiggled*

(62) *Debby danced*

(63) *The top spun*

(64) *The flag waved*

These can all be labelled manner of motion verbs or 'natural actions', and such are not, according to Jackendoff, the business of conceptual structure, since one can hardly express idiosyncratic manner of motion by decompositional features. Rather, he claims that:

> 'Conceptual structure has to encode primarily an appropriate argument structure, linked in the lexicon to a more detailed spatial structure encoding.' (p.88)

The argument structure is given in the following form, introducing now a new conceptual function or an additional semantic primitive, MOVE:

(65) [Event MOVE ([Thing $x$])]

As opposed to the functions GO and CAUSE, this function only takes a single argument.

Jackendoff concludes that future analysis of each individual verb within the context of the overall semantic class is necessary in order to fully determine the boundaries of explanation. Only then can it be judged whether a particular characteristics about the verb is principled or idiosyncratic. We shall come back to this point when examining Danish motion verbs in Chapter 5.

### 4.1.2 Semantic roles approaches to motion verbs

#### Cognitive Frame Semantics

Atkins (1994) has recently examined motion verbs from the perspective of Cognitive Frames Semantics (see Chapter 3) and she identifies the following set of frame elements for motion events:

(66) *The motion frame*

| Mover | |
|---|---|
| Path | Manner |
| Source | Sound |
| Goal | Speed |
| Distance | Vehicle |
| Area | Medium |

According to Atkins, these frame elements can be expressed in the following way in corpus:

(67) *the surveyor will crawl into the loft* (Mover)

(68) *Some bees were already crawling on the early clover* (Area)

(69) *They crawled round the stone* (Path)

(70) *The two hedgehogs crawled from the nest* (Source)

(71) *James had crawled into a cupboard* (Goal)

(72) *They crawled the last ten yards to the edge of a rise* (Distance)

(73) *He crawled on toes and elbows round the Land-Rover* (Manner)

(74) *I could picture them crawling silently through the mud* (Sound)

(75) *I crawled smartly after him* (Speed)

(76) *We crawled through the city in his car* (Vehicle)

(77) *You crawl along the ground looking for worms* (Medium)

As can be seen, frame elements in this work correspond to semantic roles in the sense that they include only constituents in the surrounding sentence; the exemplification of the frame element Manner fully illustrates this viewpoint. Manner is expressed by means of a modifier *on toes and elbows*, but from a purely semantic point of view, the verb *crawl* in itself expresses a relatively specific manner component (normally on knees and hands). This means that in a sentence like *they crawled along the sidewalk* it would be reasonable also to talk about a realised manner component, however, in this construction it is only expressed lexically.

### 4.1.3 Contrastive approaches: typological differences between languages in relation to motion verbs

#### Semantic structure of motion verbs in lexical forms as proposed by Talmy

If we relate again to translation, different lexicalisation patterns in different languages provide the evidence that a inter-lexical analysis is relevant if we want to translate correctly between languages. In this section on Talmy's work we shall look into an analysis that takes exactly such differences in lexicalisation patterns as one of its central issues.

Talmy provides an approach to the semantics of motion events in Talmy (1975) and Talmy (1985). In Talmy (1975), four semantic components are identified to be part of the motion event and to be expressed in any language. By means of transformations (in the sense of Chomsky) he shows how the two languages of interest, English and Atsugewi, realise these components in syntax.

The semantic components of motion events are as follows:

(78) Motion situation: Figure + Motion + Path + Ground

where Figure is the object that is considered as moving, Ground is the object with respect to which a first object is considered as moving, Path the respect in which one object is considered as moving, and Motion is the motion event in itself. It is now posited by Talmy that a motion situation is specified at the underlying level of all languages by a particular syntactic structure, namely the following:

(79) The figure-specifying constituent is a NOMINAL
The ground-specifying constituent is a NOMINAL
The path-specifying constituent is a preposition is a PREPOSITIONAL
The motion-specifying constituent is a VERB

A typical motion event in English is then realised as follows:

(80) *the bottle (Figure) floated (Motion) into (Path) the cove (Ground)*

In Talmy (1985) two more components are included in the basic motion schema, namely Manner and Cause. His main interest in this paper is to examine the differences in lexicalisation patterns between languages like English, Spanish, and Atsugewi and he comes up with some interesting characterisations. He claims that English is typically characterised by the fact that the Motion and Manner or Motion and Cause are both lexicalised in the motion verb, whereas Spanish typically lexicalises Motion and Path. This is illustrated by examples like the following:

(81) *the bottle floated into the cave* (verb denotes Motion and Manner)

(82) *la botella entró en la cueva (flotando)* (verb denotes Motion and Path)
(the bottle moved-in to the cave floating)

For English it is the case that the path component must be expressed outside the verb by means of a directional preposition, for Spanish, in contrast, the manner element is typically omitted or expressed by means of a gerund functioning as modifier.

For Cause the following examples are used for illustration:

(83) *I cut the wrapper off the package* (verb denotes Motion and Cause)

(84) *quité el papel del paquete cortándolo* (verb denotes Motion and Path)
(I moved off the paper form the package cutting it)

A general overview of lexicalisation patterns is given in Talmy 1985:p75:

(85)

| Language/language family | The particular components of a motion event characteristically represented in the verb root |
|---|---|
| Romance, Semitic, Polynesian Nez Perce, Caddo | Path + fact-of-Motion |
| Indo-European (except Romance) Chinese | Manner/Cause + fact-of-Motion |
| Atsugewi (and apparently most Nothern Hokan), Navajo | Figure + fact-of-Motion |

What is left unclear in Atkins approach to motion verbs, namely whether the motion frame elements can be realised as BOTH syntactic constituents in the surrounding construction AND be lexicalised in the motion verbs themselves, is being treated as a central issue in Talmy's work. In this respect Talmy's work provide a better overview of how differently semantic components are realised in syntax across languages and why - consequently - we need to get an overview of these patterns if we want to do translation. Nevertheless, it can be questioned whether Talmy's typology of language characteristics can pass the test of a corpus analysis. This issue is discussed in the following section on Slobin's work.

**Further developments of Talmy's contrastive approach**

Slobin (1996) takes Talmy's typology of motion events as his starting point for a comparison between verbs of motion in English and Spanish. He agrees with Talmy that certain languages tend to follow certain lexicalisation patterns, and he points out that this can also be seen by the amount of verbs of a specific kind in a particular language. For instance, English has a large set of manner-of-motion verbs, whereas Spanish has relatively few. However, Slobin calls into dispute the strictness of Talmy's typology and he gives counterexamples to the statement that a language like Spanish should not be able to express manner in the main verb expressing motion, in the same way as English:

(86) *la botella flotó hacia la cueva* (the bottle floated towards the cave)

Slobin claims that a critical factor between English and Spanish is rather whether the path crosses a boundary or not. In Spanish, the mover cannot cross the boundary (and that is: actually enter the cave) when using a manner-of-motion verb, since the crossing of a boundary requires a path-of-motion verb in Spanish. He concludes that the meanings of verbs of motion in a given language must, however, be considered in the light of both the typological frame of the language (verb-framed language or satellite-framed language) and the discourse in which such verbs occur. In Spanish, which is a typical verb-framed language, characteristically mapping the core schema into the verb and applying a very low use of satellites (particles etc.), explicit mention of both goal and source in the motion event is very rare, for instance. But special rhetorical conditions can overwrite this feature.

### 4.1.4 Motion verbs from the point of view of unaccusativity and event structure

**Levin & Rappaport Hovav**

Levin & Rappaport (1991) and Levin & Rappaport Hovav (1995) examine motion verbs from a completely different perspective than what we have seen in the sections above, namely from the point of view of unaccusativity.

The grouping of intransitive verbs into unaccusatives and unergatives was first suggested by Perlmutter (1978) as a means to express the difference in structure between these two groups of intransitive verbs. Seen from a Government and Binding approach, the difference in D-structure between the two types of verbs can be illustrated as follows:

(87) *Unergative Verb: NP [V]*
Example: *the man runs*
Comments: The subject is considered to be the external argument.

(88) *Unaccusative Verb: _ [V NP]*
Example: *the man arrives at the station*
Comments: the subject is considered to be the internal argument. Since it does not have an external argument it cannot assign accusative case to its direct object, therefore the name 'unaccusative'.

In other terms, one thing that characterises unaccusative verbs is that the subjects to these verbs share several properties with the objects of transitive verbs. Consequently, subjects to unaccusative verbs are considered to be internal arguments just like ordinary objects, whereas subjects to unergative verbs are considered to be external.

Levin & Rappaport agree with Perlmutter in the assumption that the distinction between the two classes of verbs seems to be syntactically represented but semantically determined. However, Levin & Rappaport point out that the relationship between the syntax and the semantics of unaccusatives is complicated and they take motion verbs as an example of this complexity.

It is claimed by the authors that most localist analyses of motion verbs simply lack the distinction between different subclasses of motion verbs, since only concepts of motion and location are considered to be central for the argument structure of these verbs. It is this lack of distinction between semantically different subclasses which makes the results of an unaccusative analysis confusing and apparently unsystematic from a semantic point of view.

So what is needed then to make motion verbs fit into the determination hypothesis is thus a more fine-grained semantic analysis which is concerned not only with the more general meaning components such as motion and location but also with the more nuanced notions of aspect and protagonist control.

Levin & Rappaport (1991) define three classes of intransitive motion verbs (p.252):

arrive class: *arrive, come, go, depart, fall, return, descend...*
roll class: *roll, slide, move, swing, spin, rotate...*
run class: *run, walk, gallop, jump, hop, skip, swim...*

They identify three components of meaning as to characterize these three groups, namely:

Direction
Manner
Direct External Cause

Arrive verbs are then characterised as having inherent Direction as the feature that distinguishes them from the other two classes. Roll verbs and run verbs, in contrast, have Manner as inherent component. In order to differentiate roll verbs from run verbs the feature Direct External Cause is introduced. This feature distinguishes verbs that denote actions that involve 'protagonist control' from verbs that do not, and is defined as follows (p. 253):

> 'If the meaning of a verb involves a direct external cause, it does not involve protagonist control, while if the meaning does not involve a direct external cause, then it typically involves protagonist control.'

This set of parameters enables the authors to put forward some syntactic generalisations about each verb class:

> *Arrive* verbs:
> Generalisation 1: Verbs whose meaning includes a specification of inherent direction are found in the unaccusative syntactic configuration
>
> *Roll* and *run* verbs:
> Generalisation 2: If the meaning of a verb implies a direct external cause, the verb is found in the unaccusative syntactic configuration; otherwise, it is found in the unergative syntactic configuration

It is further argued that aspectual information also has an influence on the unaccusativity hypothesis, since the attachment of a directional phrase to verbs of the run group changes partly the aspect of the phrase from an activity to an accomplishment and partly the syntactic behaviour whatsoever. Evidence for this is given in Italian where manner of motion verbs that can appear with a goal phrase select the auxiliary *essere* (to be) in the presence of a goal phrase (p.259):

(89) **Ne hanno corso due*
(Two of them ran)
(auxiliary essere (be) not selected)

(90) *Ne sono corsi due a casa*
(Two of them ran home) (auxiliary essere (be) selected)

Two lexical semantic representations for *run* are given:

(91) *run* (manner of motion): [x MOVE in-a-running-manner]
*run* (directional): [x GO TO y BY [x MOVE in-a-running-manner]

The effects of such alternations make Levin & Rappaport claim that aspectual theories of motion verbs seem to come up with more adequate classifications and descriptions of motion verbs than do localist theories applying traditional semantic roles such as agent, theme, and goal. They find semantic role assignment inadequate and instead they introduce an alternative set of features, specifically identified for motion verbs. As we saw in Chapter 3, recent work by Levin & Rappaport Hovav (forthcoming) suggest a more general way of treating this double nature of some verbs.

In Levin & Rappaport Hovav (1995), the double membership of agentive manner verbs (corresponding to the *run* group) is stated directly wrt. unaccusativity, p.187:

> These facts indicate that agentive verbs of manner of motion can indeed be associated with two related, though distinct, meanings, one consistent with an unaccusative and the other with an unergative classification of these verbs.

**Other aspectual approaches to motion verbs**

With a similar viewpoint as Levin and Rappaport Hovav (forthcoming), Tenny (1995 and partially (1995a)) looks into the aspectual properties of motion verbs and draws a parallel distinction between structural/templatic meaning in verbs - comprising particularly aspectual information - and 'gapfilling' idiosyncratic information which fills in the templates. The meaning aspect focussed upon is that of 'measuring-out', an element of meaning which is aspectual and which relates directly to what is traditionally labelled *temporal boundedness*. Also here, manner-of-motion verbs receive specific attention because of their elasticity with regards temporal boundedness, in contrast to verbs of inherently directed motion.

Also Dini & Tomaso (1995) apply an aspectual analysis to motion verbs in order to be able to treat the ambiguity of Italian locative expressions in an appropriate way. With an event structure analysis closely indebted to Pustejovsky's (1995), motion verbs are divided into

- *pure achievements* (*partire* (to leave))
- *pure accomplishments* (*andare* (to go)), and finally
- *pure processes* (*galleggiare* (to float)).

### 4.1.5 Deictic descriptions of motion verbs

Apart from more general semantic descriptions of motion verbs like the ones given in the sections above, several linguists have examined a small group of motion verbs from a deictic point of view (see Miller & Johnson-Laird 1976: p. 539-40 for an overview).

One account of these verbs in English is given in Fillmore (1966). He describes deixis as 'the name given to those aspects of language whose interpretation is relative to the occasion of utterance: to the time of utterance, and to times before and after the time of utterance; to the location of the speaker at the time of utterance; and to the identity of the speaker and the intended audience'.

In English, typical deictic motion verbs are *come* and *go*, and Fillmore puts forward a set of deictic rules which accounts for these verbs. The most substantial part of the rules concern the supposition of the speaker and hearer, as is seen in supposition rule for *go*:

(92) [[+Speaker][not][Present][[be] Location]]

The rule indicates that whatever the subject of the verb *go* may be, the place to which one goes is a place where I am not, making a sentence like the following impossible:

(93) ** I will go here again tonight*

*Come* is the antonym to *go* and the supposition rule for *come* is positive (excluding the [not]) but less restricted since it can refer to either your or my present location. 'Speaker' is therefore replaced by 'participant' (meaning either speaker or hearer or both) allowing for sentences like:

(94) *When did you come to the shop ?*

with the supposition that either you or I or both of us are in the shop at the time of speech.

What is not treated in Fillmore's account are the cross-linguistic differences in the restrictions on deixis. The ability of the English *come* for taking the listener's point of view, does not work for Spanish, to give an example. The English sentence *I'm coming* cannot translate into *ya vengo* [1]. Instead the equivalent to *go* is applied, as in: *ya voy* ('already I go'). In Danish, on the contrary, we find even less restrictions on these verbs than in English. For instance, the 'speaker

[1]Instead, *ya vengo* means *I'll be back (in a short while).*

not present' restriction doesn't count for the Danish equivalent to *go*, which can be seen from the perfectly acceptable sentence below:

(95) *Jeg tager herud igen i aften*
('I will go here-out again tonight')

It is therefore obvious that in order to produce good translations in contexts where these verbs are involved, some calculation of deixis and some descriptions of how deictic information is expressed in a given language is necessary. In the present work, however, we have excluded this aspect since deixis relies to a large extent on the discourse structure of the whole text; an aspect which is outside the scope of our work.

### 4.1.6 Descriptions of Danish motion verbs

Apart from what is represented in Danish lexica, Danish motion verbs are more or less untreated from a general semantic point of view.

One exception to this is some remarkable work done by Skautrup (1927) on a Danish dialect called Hardsysselmål. In order to give an overview of how different aspects of motion are expressed in this dialect, he established a taxonomy of Danish motion verbs. Below is given the pure taxonomy with exclusion of the specific dialectic expressions:

1. bevæge (move)
2. rask og langsom (quickly and slowly)
3. gang (walking)
4. krøble, stumle, styldre (ways of walking if one is crippled in some way)
5. sjakke, s(j)okke, traske o.lign. (mooch, trudge, plod etc.)
6. drive, løje, o.lign. (drift, go down (wind) etc.)
7. trine, træde (tread, step)
8. rende, trave, springe, hoppe (run, plod, jump, hop)
9. kravle, krybe (crawl, creep)
10. liste, smutte (tiptoe, slip)
11. komme, gå (come, go)
12. færdes, rejse (move (in good society, for instance) travel)
13. uordentlig, uregelmæssig bevægelse (improper, irregular movement)
14. flyve, sejle m.m. (fly, sail etc.)
15. vædske i bevægelse (liquids in movement)
16. bevægelse i eller med vædske, tyktflydende masse o.lign. (movements in or with liquids)

17. bevægelse op eller ned (movement up or down)
18. bevægelse ud og ind, frem og tilbage (movement in and out, back and forth)
19. skride, trille o.lign. (slip, roll etc.)
20. dreje, vende o.lign. (rotate, turn etc.)
21. ujævn , uventet enkeltbevægelse (irregular, unexpected movement)
22. regelmæssig gentaget bevægelse (regular, iterative movement)
23. trække, skyde o.lign (pull, push)
24. jage, drive, genne (chase, drive, shoo)
25. kyle, smide o.lign. (fling, throw etc.)
26. slå, banke, give hug o.lign. (hit, beat etc.)
27. bevægelse med legemsdele (movements with body parts)
28. retningsbestemt bevægelse (directional movement)
29. hente, fly, sende (fetch, hand, send)

After this work of categorisation, Skautrup concludes on the composition of the vocabulary in Hardsysselmål and Danish:

> (...) det er jo højst påfaldende, at mens der begrebsmæssigt ikke er noget til hinder for en særdeles indgående findeling af retningsbestemte bevægelser (nemlig ad matematisk vej), har sproget egentlig meget få selvstændige ord herfor, hvorimod det vrimler - og da vistnok især i dialekterne - med varierende og nuancerede, men særdeles vanskeligt bestemmelige udtryk for vaklende gang, slæbende gang, drivende gang, uordentlig gang (...)[2]

In accordance with what has been described above on different lexicalisation patterns in different languages (especially Talmy and Slobin), Skautrup acknowledges that this is not necessarily the case for all other languages, and he mentions French as an example of a language where the vocabulary of directional motion verbs is larger than the vocabulary for manner motion verbs. He further concludes that this could have to do with a lower faculty of abstraction among Danes, and especially among Danish peasants (!):

> (...) den primitivere mentalitet foretrækker det mere anskuelige - om end i nok så mange varierede former - frem for det mere abstrakte[3].

[2](...) it is thus very striking, that although there is nothing conceptual that hinders a fine-grainedness of directional verbs (in mathematical terms, that is), the language proves to have very few individual words for this, whereas there is a very large group of varied and nuanced words - especially in the dialects - for denoting reeling, staggering, doddling, and improper ways of walking (...)

[3](...) the more primitive mentality prefers the more intuitive - even if very varied forms - to the more abstract forms.

Still within the frames of the more philosophical considerations, Hjortø (1936) gives an account of some of the locational uses of motion verbs in Danish, where items that are physically non-moving are conceptualised as moving. Examples:

(96) *mit hus kommer først bag den tre etagers bygning du ser derhenne*
('my house comes not-until the three story building you see over-there')

The concept of something moving is thus transferred from the moving entity to the item that the moving entity is moving towards. For roads and paths, the notion of movement is also involved. Here it seems, according to the author, that the sight of a road or path evokes the notion of movement in itself, as in:

(97) *hvor går den vej hen ?*
('where walks that road over ?')

The width of the path seems to determine which motion verb to apply, so for instance you would say:

(98) *smøgen løber herfra ned til havnen*
('the passage runs from-here down to the harbour')

These observations lead to some very philosophical considerations on why man sees himself as the passive part, whereas roads, time, and life are considered to be the moving part, when actually it could just as well be defended to conceptualise it the other way around.

Where Hjortø describes cases of non-moving objects combined with motion verbs, E. Hansen (1972) describes the opposite case, namely cases of motion that are expressed without a motion verb, but by means of modal verbs. The phenomenon is quite normal and frequent in Danish, as can be seen from the following corpus examples:

(99) *DDR bør ind i FN, siger Sovjet*
('DDR ought in to UN, says Sovjet')
(DDR ought to enter the UN)

(100) *ustandselig hører man vræl, så man må jo ud og se, hvem af ungerne ...*
('constantly one hears cries, so one must therefore out and see, who of the kids...')

(101) *han skulle afsted til sanatoriet*
('he should off to the sanatory')
(he was going to the sanatory)

(102) *og saa vilde hun aldrig mere derhjem*
('and then she would never more there-home')
(and she would never again want to go there)

It is further discussed whether this phenomenon should be considered an ellipsis or whether we are simply dealing with some not very well-described aspects of meaning of Danish modal verbs. The latter hypothesis seems to be the right one, according to the author. Aa. Hansen (1967, vol.III:pp78) also discusses the phenomenon, and he too rejects the assumption that it should be a case of ellipsis.

Recent work on transitivity in Danish exposes some interesting aspects of Danish motion verbs, as is seen in Herslund (1993). He argues that Danish motion verbs - as well as several other verbs - undergo extensive derivations at clause level; a fact which enforces us to distinguish strictly between verb *stems*, which have certain inherent properties, and verbs at *clause level* which can have other derived properties[4]. In the case of motion verbs, therefore, we see that many verbs can denote either an atelic activity as in (p.44):

(103) *hun løber i parken* (she runs in the park)

or a telic action, as in:

(104) *han svømmede ind i fjorden* (he swam into the fjord)

A similar behaviour is noticed wrt. unaccusativity: most Danish motion verbs can be realised as either unaccusatives or unergatives depending on the context (p.53):

(105) *Boris har løbet langt* ('Boris has run long')

(106) *Boris er løbet hjem* ('Boris is run home')

These facts relate nicely to Levin & Rappaport Hovav's work on English verbs referred to in the previous chapter.

Other minor aspects of motion verbs can be found in Durst-Andersen & Herslund (1992), who deal with the verbs in Danish that exhibit a systematic alternation between direct objects and prepositional objects. Among these are quite a lot of motion verbs such as *dreje* (turn) and *ryste*:

(107) *han drejede knappen/han drejede på knappen*
(he turned the button/he turned on/at the button)

(108) *han ryster hånden/han ryster på hånden*
(he shakes his hands/he shakes on his hands)

[4]In our terminology this refers to *prototypical senses* vs. *derived senses*.

Herslund & Durst-Andersen distinguish two cases of prepositional objects: those denoting incomplete actions and those denoting repeated actions. These subclasses are then again seen as special cases of the more fundamental distinction between activity and action verbs. According to the authors, some verbs (of which motion verbs are part, as far as we can see) are, however, ambiguous to activity and action:

> 'by choosing the direct object construction one gets an action reading of the verb, and by choosing the prepositional construction one gets an activity reading.'

Finally, some ongoing work on a Danish valency dictionary, namely the Odense Valency Dictionary Project (see Schøsler 1994), includes a detailed description of valency patterns for some of the most prototypical Danish motion verbs. The Valency Dictionary is based on The Pronominal Approach, which is presented as a method capable of establishing language immanent criteria for the determination of valency and sense distinction by means of the pronominal paradigm (see examples of pronominal paradigms in the examples below). Thus, the project is highly concerned with the establishment of valid test sets to ensure intercoder consistency and explicit procedures for sense distinctions based on distributional patterns alone.

To give an example, the Danish motion verb *dreje* (turn, rotate) is assigned the following three senses, all based on distributional criteria (see Kirchmeier-Andersen, Pedersen & Schøsler 1994):

(109)

| *dreje_1* | argument slot 1 |
|---|---|
| pronominal paradigm: | hvad, denne her, *hvem, *det[5] |
| semantic features: | concrete |
| syntactic features: | subject |
| type of passive: | no passive |
| Aktionsart: | imperfective |
| type of auxiliary: | *have* |
| example: | kloden drejer (the earth rotates) |

(110)

| *dreje_2* | argument slot 1 | argument slot 2 |
|---|---|---|
| pronominal paradigm: | hvem, hvad, denne her, *det, | hvorhen |
| semantic features: | human, concrete | directional |
| syntactic features: | subject | adverbial complement |
| type of passive: | no passive | |
| Aktionsart: | perfective | |
| type of auxiliary: | *være* | |
| example: | *hun/bilen drejer (til højre*<br>(she/the car turns (left) | |

(111)

| *dreje_3* | argument slot 1 | argument slot 2 |
|---|---|---|
| pronominal paradigm: | hvem, *hvad | hvem, hvad, denne her, *det |
| semantic features: | human | human, concrete |
| syntactic features: | subject | object |
| type of passive: | inflectional and analytical | |
| Aktionsart: | perfective | |
| type of auxiliary: | *have* | |
| example: | *hun drejer barnet/stolen*<br>(she turns the child/the chair) | |

In Kirchmeier-Andersen, Pedersen & Schøsler (1994) it is argued that such sense distinctions based on distributional criteria correspond to a large degree to the senses that can be established on semantic grounds.

## 4.2 Summary

In this chapter we have given a survey of different approaches to motion verbs; from decompositional analyses in the works of Miller & Johnson-laird and Jackendoff to analyses more oriented towards semantic roles, as presented by Atkins. Others, such as Talmy and Slobin have focussed on typological differences between languages, where others again have looked at motion verbs from the point of view unaccusativity and aspectual theories (Levin & Rappaport 1991, Levin & Rappaport Hovav 1996, forthcoming, Tenny 1995, Dini & Tomaso 1995 and others).

If we relate again to our main problem area - lexical ambiguity - most of the aspects touched upon in these accounts are of relevance to us, although some, for instance a 'complete' decompositional analysis into semantic primitives, would not be feasible in a transfer-based machine translation system. Frame element assignments as well as the acknowledgement of the importance of unaccusativity, exemplified by the assignment of additional features of protagonist control and telicity as proposed by Levin & Rappaport are on the other hand aspects of meaning which are feasible to incorporate in our system and which proves to have importance in the disambiguation process as will be illustrated in the next chapter.

Talmy's and Slobin's works are oriented towards cross-linguistic research and they identify the different lexicalisation patterns that are found in different kinds of languages. Therefore, their work is highly relevant for all contrastive considerations in our work, and they will provide a basis for determining what kind of a language Danish is, seen from the perspective of motion verbs.

Finally, we have looked into the scarce accounts of Danish motion verbs and we can conclude that there is a need for a systematic semantic and syntactic description of these verbs in Danish if we want to get a picture of the differences in elasticity of meaning that occur with motion verbs.

Because the topic is so scarcely treated, the information hidden in lexica and corpora provides the most important basis for our work. As regards lexica, however, a large-scale dictionary on modern Danish is only just now under development, leaving for practical use only a medium-sized dictionary of modern Danish and a large-scale but highly antiquated dictionary of Danish. This means that much attention will be paid on corpora, which is therefore the topic treated in the first sections of the next chapter.

# Chapter 5

# Danish Motion Verbs

## 5.1 Introduction

In Chapter 3 we described some recent approaches to verb polysemy and in Chapter 4 we looked more closely at different approaches to motion verbs, in particular. Motion verbs proved to be much studied in linguistics in general but little studied from the point of view of Danish. It was concluded that Danish corpora therefore constitute one of the main sources for the empirical work that is foreseen in this chapter.

This chapter falls into two parts: In the first Section (5.2) we present the corpora used for this work and we demonstrate how corpus analysis has been undertaken as a starting point for the linguistic investigations.

In the second part (Section 5.3) we proceed with the descriptive work of Danish motion verbs: we start out by a delimitation of the verbs of interest and continue with a definition of the basic apparatus to be used in the analysis part. Three notions need to be specified more precisely here, namely the exact set of features relating to frame elements, event structure and selectional restrictions, respectively. On the basis of this apparatus we establish a taxonomy of Danish motion verbs and we examine which syntactic alternations these verbs can undergo in order to see to which extent our material exposes a transparent relation between semantic categorisation and syntactic potential.

Finally, these examinations enable us to proceed to one of the central aims of this chapter: to get an overview of the *elasticity* in meaning that Danish motion verbs exhibit. A primary goal is to identify different *categories* of elasticity and to define *which subgroups* of verbs are sensitive to which meaning changes.

## 5.2 Corpus analysis - crucial for lexicography

The fact that lexica that are to be used for analysing or producing real texts must also to some degree be *based* on real texts makes corpus analysis indispensable both in computational and traditional lexicography. During the last decades, corpus analysis has developed radically with more and more sophisticated tools at hand. The spectrum of available corpus facilities spans from purely stochastic tools that are capable of providing statistics of word occurrences and co-occurrences to more linguistically oriented tools that can provide automatic tagging of large corpora and even 'raw' syntactic analyses of texts. In the following sections an overview is given of the corpora and corpus tools that have been applied as a background for the linguistic work presented in this chapter.

### 5.2.1 Danish corpora

As a primary basis for the corpus analysis in the present work, the largest publicly available LGP (language for general purposes) text corpus for Danish has been applied, namely Bergenholtz DK87-90 Corpus (see Bergenholtz 1990). Furthermore, some use have been made of the very big LGP corpus produced in relation to the production of a new Danish Dictionary (see Andersson 1994), and finally, one LSP (language for specific purposes) corpus has been applied at a minor scale, namely The BIODA Corpus (Hauschildt et al. 1989) on genetic engineering. None of these corpora are tagged at the current stage, although hand tagging of some parts of the DDO corpus have currently been initiated.

The Bergenholtz Corpus is a LGP (language for general purposes) corpus which consists of approx. 4 mill. running words. The corpus was gathered over a period of 4 years: 1987-90. Text words have been gathered from 100 newly published novels, 50 magazines, and 50 newspapers. The general composition of text extractions is for each year:

novels and short stories: 50 %
newspapers: 25 %
magazines: 25 %

The Corpus of Den Danske Ordbog (DDO Corpus)(see Andersson 1994) consists of 40 mill. words and actually includes the Bergenholtz corpus as part of it. However, only 36 mill. words have been accessible to us, due to right restrictions. The corpus includes texts from the period 1983-92 distributed into 43.892 corpus entities. The following language types have received a special priority:

- Spoken language (private interviews, debates in the Danish Parliament, radio and television, etc.: 7 mill. words.

- Non-professional language, private diaries, school papers, etc: 4,5 mill. words.

Main media are:

- Books, magazines and newspapers: 28 mill. words.
- Radio and television: 3,8 mill. words.
- Small prints: 2 mill. words

The BIODA corpus contains approx. 1 mill. words. Even though being primarily an LSP corpus, The BIODA Corpus includes not only scientific texts on genetic engineering, but also political and journalistic accounts on the topic. As a consequence, the corpus is organised along 4 levels of communication:

1) expert to expert (18 % of the total amount of text)
2) expert to layman (38 %)
3) layman to expert (9 %)
4) layman to layman (35 %)

As can be seen from the percentages above, the three groups are far from equally represented in the corpus. This is due to the fact that it was difficult for the corpus collectors to find representatives for group 1 and 3. The first group is underrepresented because Danish is not the language used in scientific papers on genetic engineering (English is), whereas group 3 is underrepresented because this direction of communication is not very frequent in general.

### 5.2.2 Corpus tools for Danish

In this section we shall concentrate on the corpus tools which are especially tuned to Danish corpora by including word tagging and syntactic analysis to a certain degree. As an introductory remark, it can be stated that efficient tools for corpus handling of Danish are still called for. General 'search and statistic' tools are available in many variants, but tagging and analysis tools which are tuned to Danish are few in number and still not fully developed. In the following a scarce overview of three of the most substantial corpus tools for Danish is given.

**CALM - A lemmatizer**

The main purpose of CALM is to produce a lemmatized version of a given corpus (see Jensen & Tams 1993). A considerable task in this respect is the disambiguation of homographs, since these refer to two or more lemmas. A rule-based mechanism has been developed with the aim of deciding in each case, to which lemma a given homograph refers.

The CALM tool consists of four modules:

- A dictionary containing approx. 50.000 Danish words, each provided with a morphological reference.
- A morphology module assigning detailed morphological information to the word.
- Disambiguation rules for homographs, and finally
- A lemmatizing module which produces a list of lemmas together with an assignment of frequency

**VALP - A valency parser**

The VALP parser goes a bit further than CALM in the sense that it provides an analysis of valency patterns in large corpora (see Daugaard and Kirchmeier-Andersen 1993). The interim result, namely the word tagging, is however not directly available to the user. Also the program does not provide concordances but takes concordances as required input. The program consists of the following modules:

- A dictionary containing approx. 40.000 Danish words provided with morphological information.
- A morphology module assigning detailed morphological information.
- A syntactic module based on context free grammars providing an analysis of valency patterns.
- A module for statistical treatment of the analysis results
- A module for dictionary updates.

**CORPUS BENCH - A general corpus tool**

CORPUS BENCH is the only commercially developed product among the three described here, and this is reflected in both its user-friendliness and speed. The tool is suitable for very big corpora and is thus the tool used at Den Danske Ordbog for corpus work. The tool consists of the following facilities:

- A concordance tool.
- A frequency tool.
- A distributional tool, providing information on word use within criteria such as sex, age or topic.
- A tool for mutual information, which expresses the probability for word cooccurrences.

**The Xerox Tagger: work under development**

Finally, it should be noted that a fully automatic tagger for Danish is under development at Center for Sprogteknologi in collaboration with XEROX; however, at the current stage it is neither complete nor publicly available.

### 5.2.3 Corpus analysis in the present work

**Introduction**

Not all tools have been directly available for our work. However, both the VALP parser and CORPUS BENCH have been used substantially during the linguistic investigations performed.

As has been described earlier, the aim of using a corpus for linguistic research of the kind presented here, is to ensure that lexical descriptions are founded on real text and real distributional occurrences. Therefore, the goal of the corpus work is to get - apart from what can be conceived from traditional dictionaries[1] - an overview of the meanings a word can have and of the syntactic patterns in which it can occur. In other words, some categorisation of examples into syntactically and semantically coherent types was foreseen as an important task.

With such a categorisation as the primary aim, the corpus work can be seen as passing through several stages.

[1] As we have already mentioned Danish lacks a large-scale, up-to-date dictionary which could function as another satisfactory basis for our work.

### First stage of the corpus analysis

The first stage of the corpus analysis contained the following procedures:

- Extraction from a Danish medium-sized dictionary in electronic format[2] of approx. 100 of the most prototypic Danish motion verbs.
- A specification of the amount of material to be examined for each selected word. The choice was made to examine approx. 100 examples for each word. In cases of very polysemous words the number of examples was increased to 200 or even 300[3]. These examples are basically extracted from the Bergenholtz Corpus. In cases where the word is not frequent enough to occur in the corpus 100 times the DDO Corpus is applied also[4].
- A concordance for each word of interest was made[5].
- Each concordance list was run through the valency parser, VALP, in order to achieve a 'raw' analysis of the text.
- The results of the VALP parser were examined and corrected by hand.

### Testing the VALP parser

The original idea was to let the VALP results form a primary categorisation of the examples in order to ease the semantic categorisation to be made by hand. If we consider for example a motion verb like *dreje* (turn, rotate), a categorisation into intransitive, transitive and phrasal verb constructions would provide an excellent starting point for a more semantically oriented analysis. Also a recognition of passives and *der* constructions is highly relevant in the otherwise very cumbersome categorisation work.

However, as work proceeded, it turned out that the degree of correctness of the VALP results was too low as to let these function as a direct basis for a later semantic analysis by hand. In short, quite an amount of work was required

[2] *Retskrivningsordbogen* was applied.

[3] These limitations were set on the basis of a small sample test. Since we are not aiming at any degree of completeness but rather at a general approximation of the meaning and syntactic potential of the words in question, the number was only increased in a very few cases. This conforms to the consensus in traditional lexicographical work of around 100 concordances per word, see for instance Roberts & Montgomery (1996 p. 460)

[4] One fairly quickly gets an impression of whether further examinations of a word are necessary in order to get an overall picture of the meaning potential of a word.

[5] Two concordance programs have been used, namely 'Excerper program', which is a program that has this facility as its only one, and 'CORPUS BENCH'.

in order to adjust the syntactic results made by the machine and to discard wrongly generated analyses before semantic analysis could take place. Especially overgeneration proved to cause severe problems.

To give some examples, let us look at some parsing results from the concordance list for *bevæge* (move) and *forfølge* (follow/pursue). When these examples are run on the monovalent grammar of the system, the strategy of the system is to identify an NP that could candidate for a valency bound constituent. All sentences that contain only one such NP within the search domain are then considered to be examples of intransitive constructions. A major problem, however, lies in the identification of the correct search domain, as can be seen from the parsing result below:

(112) *når en* [search domain begin] *af klassekammeraterne*
(when one [search domain begin] of the class mates

*bevægede læberne* [search domain end]
moved their lips [search domain end])

In this example, the subject (or part of it) has been incorrectly overseen as 'noise', i.e. as not being part of the relevant search domain. The subject consists of a complex NP expressing a partitive relation, which has not been fully recognised by the system. Rather it has recognised *af klassekammeraterne* as an adverbial modifier. As is well-known, an adverbial clause situated before the main verb predicts inversion of the subject in Danish, so with such an analysis, *læberne* constitutes a perfect candidate for a subject, and therefore the construction is incorrectly considered to be intransitive.

(113) *det bliver spændende at se,* [search domain begin]
(it will be interesting to see [search domain begin]

*om politikerne tør forfølge sagen politisk* [search domain end]
whether the politicians dare pursue the case politically [search domain end])

In the second case, the conjunction *om* has been wrongly interpreted as a preposition (which is a correct interpretation in other cases), and *om politikerne* is therefore analysed as a prepositional phrase functioning as a modifier. Again, the object *sagen* is the only identified valency bound constituent, and therefore it is wrongly interpreted as the inverted subject of an intransitive verb.

The program is designed in such a way that the user is free to adjust the grammars in order to make them fit the purpose that s/he has in mind. However, most of the incorrect analyses that we have come across in our testing are due to ambiguity cases of different kinds which are very difficult, if at all possible, to correct without causing problems somewhere else. One must keep in mind that the dictionary used for parsing does not contain valency information at all, the valency pattern is guessed at on the basis of a 'raw' analysis and counting of the constituents in the given phrase. However, a refined wordtagging procedure might improve the results.

One thing became evident during the testing phase, though, and that was that the more valency bound constituents to be identified by a given grammar, the better the results. Thus, the recognition of intransitive constructions was by far the most complex task for the system, since relevant constituents were frequently overseen opening up for severe overgeneration. If we compare with the grammar for identifying ditransitive constructions, the degree of correct guesses increased considerably. Nevertheless, since we are mainly dealing with intransitive constructions of different kinds, we could hardly benefit from the fact that the reliability of the program is higher in other parts of the grammars.

### Applying the cooccurrence facility of CORPUS BENCH

For the semantic analysis, it was furthermore decided to apply some statistics on cooccurrence in order to identify typical subjects for the verbs in question. The cooccurrence facility of CORPUS BENCH was appropriate for this task.

When dealing with statistics, however, it is evident that a sample of 100 examples for each word of interest is not enough. Therefore, the DDO Guest Corpus of 36 mill. words was applied in order to ensure a maximal basis for calculation.

The result of the cooccurrence facility is a collocation value which expresses the probability of two words occurring together compared to the probability of the two words *not* occurring together. The higher the value, the stronger the two words attract each other (see Asmussen (1993)).

The result of the calculation is to be interpreted as follows:

- The first column shows the word that is closely related to the search word.
- The second column gives the collocation value (also labelled *mutual information*) as described above
- The third column shows the absolute number of times that the two words have been found together. The higher the value, the more reliable the result.

Consider below the test results of the motion verb *strømme* produced with the following options:

(114) *Mutual information of strømme*
Search domain: 4 words before search word
Minimum number of cooccurrences to be considered: 5

| word | mut.inf. | coocc. |
|---|---|---|
| *stride* | 1579.87 | 37 |
| *ferske* | 947.57 | 10 |
| *tårerne* | 336.61 | 20 |
| *cirkulære* | 235.28 | 5 |
| *energien* | 202.85 | 10 |
| *turisterne* | 170.50 | 7 |
| *regnen* | 145.98 | 13 |
| *vv* | 132.51 | 5 |
| *strømmer* | 129.90 | 8 |
| *flød* | 100.74 | 6 |
| *blodet* | 88.61 | 16 |
| *bølge* | 82.19 | 6 |
| *tårer* | 81,59 | 9 |
| *olien* | 78.43 | 5 |
| *journalist* | 77.50 | 12 |
| *elektriske* | 70.79 | 7 |
| *tusindvis* | 52.09 | 5 |
| *floden* | 49.84 | 5 |
| *vandet* | 45.63 | 29 |
| *kunderne* | 45.47 | 7 |
| *hav* | 34.05 | 6 |
| *sø* | 31.64 | 5 |
| *svage* | 28.58 | 5 |
| *hvorfra* | 28.30 | 5 |
| *publikum* | 27.33 | 13 |

As can be seen, some introspection is required in order to interpret the results. First of all, homographs must be identified and results for these must be discarded. In the case of *strømme*, the plural noun form (streams) coincides with the infinitive form of the verb. This explains some of the otherwise strange cooccuring words, such as *ferske* (fresh), *flød* (floated), *elektriske* (electric), *svage* (weak), *kolde* (cold), *varme* (warm).

It should also be noted that idiomatic expressions containing words that are not very frequent alone will be assigned very high scores. This is the case with *stride strømme* (as in: pour down) (which, in this case, should be discarded anyway because it is the noun reading).

Moreover, not all words occurring to the right of the search word are subjects. *hvorfra* (from where) is an illustrative example of this.

Having discarded such clearly irrelevant occurrences, an interesting and illustrative picture emerges. Typical concrete inanimate subjects to the verb *strømme* can now be identified as liquids, such as: *tårer* (tears), *blod* (blood), *regn* (rain), *olie* (oil), *vand* (vand). Animate subjects denoting groups of people are also frequent, such as: *turister* (tourists), *journalister* (journalists), *kunder* (customers) and *publikum* (audiens). Finally, semiabstract subjects occur, such as *energi* (energy) and *cirkulærer* (circulars).

However, not all motion verbs showed a correspondingly transparent pattern of cooccurrence. Very frequent words that are often seen in idiomatic-like or metaphoric expressions tended to give very disperse results. The first 5 words to cooccur with *komme* (come), for example, proved to be *tvivlen* (the doubt), *kø* (queue), *forsøgte* (tried), *umuligt* (impossible), and *vanskeligt* (difficult) ! In conclusion, the facility was used with some care and only in cases where traditional corpus inspection left a blurred impression of typical subjects.

### 5.2.4 Second stage of the corpus work: manual categorisation

In the second stage of the corpus work we considered which corpus tools really improved and eased our semantic categorisation work and which did not. On the basis of the results presented above, it was decided to examine the examples by hand and to perform the semantic and syntactic categorisation by manual introspection only. As for the cooccurrence statistics it was decided to use this as a statistically sound supplement to the manual examination of 100 examples per word.

Beginning from a more or less intuitive starting point, the semantic categorisation work took form into something following a rather explicit procedure:

1. Discard examples with homographs which are not of interest (ex. plural form of the noun *strøm* (stream).

2. Group examples into concrete and metaphoric senses, as well as into idiomatic expressions.

3. For all concrete examples: group examples into basic valency patterns and into phrasal constructions and non-phrasal occurrences (*han drejer bilen* (he turns the car) vs. *bilen drejer* (the car turns) vs. *han drejer af* (he turned off)).

4. Group examples into constructions with animate and inanimate subjects.

5. Group examples into those containing directional modifiers and those without.

For each group, typical syntactic behaviour was registered, such as negation, frequent modal constructions, expletives, typical modifiers (others than directional modifiers) etc.

For a verb like *bevæge sig* (move (intransitive)), this gave the following groupings (note that 200 examples were extracted for this verb; all transitive examples were then discarded since they are not considered to be part of our semantic domain; see Section 5.3.1 for further details):

(115) *Group 1*

| | |
|---|---|
| Number of examples: | 42 |
| Semantic characteristics: | Animate mover, no direction,<br>movement back and forth,<br>the ability/lack of ability to move,<br>movement in a special way. |
| Syntactic characteristics: | pure intransitive constructions,<br>several modal constructions,<br>many negated constructions.<br>Manner modifiers frequent. |
| Examples: | *jeg kunne ikke bevæge mig*<br>(I couldn't move)<br>*at de bevæger sig på en fræk måde*<br>(that they move in a shameless way) |

(116) *Group 2*

| | |
|---|---|
| Number of examples: | 35 |
| Semantic characteristics: | Animate mover, specific direction,<br>body parts or machines moving |
| Syntactic characteristics: | intransitive constructions with<br>directional modifiers. |
| Examples: | *Jeg bevægede mig ned i den skumle kælder*<br>(I moved down into the sinister cellar) |

(117) *Group 3*

| | |
|---|---|
| Number of examples: | 44 |
| Semantic characteristics: | Inanimate mover, no direction,<br>split group: many body parts<br>but also 'real' inanimate entities |
| Syntactic characteristics: | intransitive constructions |
| Examples: | *hans læber bevægede sig*<br>(his lips moved)<br>*skyerne bevægede sig for pålandsvinden*<br>(the clouds moved by the wind) |

(118) *Group 4*

| | |
|---|---|
| Number of examples: | 16 |
| Semantic characteristics: | Inanimate mover, direction,<br>body parts or machines moving |
| Syntactic characteristics: | intransitive constructions with<br>directional modifier |
| Examples: | *benene var tøet så meget at hun kunne*<br>*få dem til at bevæge sig ud i badeværelset*<br>(her legs had thawed so much that she<br>could make them move out into the<br>bathroom)<br>*en sky bevægede sig afsted højt oppe*<br>*på himlen*<br>(a cloud moved along high up in the sky) |

(119) *Group 5*

| | |
|---|---|
| Number of examples: | 12 |
| Semantic characteristics: | Abstract use, abstract mover or goal,<br>temporal entities frequent as mover |
| Syntactic characteristics: | intransitive constructions with or without<br>directional modifier<br>Often negated construction or modified<br>by a temporal modifier, i.e. *langsomt* |
| Examples: | *heller ikke tiden bevægede sig*<br>(not even time moved)<br>*de har dermed bevæget sig ind på et*<br>*typisk mandeområde*<br>(they have thereby moved into a typical<br>male's domain) |

Corpus analysis of the verb *vippe* (seesaw) shows the following patterns (note that all noun readings have been discarded from the 100 examples):

(120) *Group 1*

| | |
|---|---|
| Number of examples: | 25 |
| Semantic characteristics: | inanimate mover, back and forth movement, no protagonist control |
| Syntactic characteristics: | often combined with the modifier *frem og tilbage* (back and forth) or *op og ned* (up and down) |
| Examples: | *båden vippede pludselig kraftigt* (Suddenly the boat rocked fiercely) |

(121) *Group 2*

| | |
|---|---|
| Number of examples: | 7 |
| Semantic characteristics: | inanimate mover, move abruptly |
| Syntactic characteristics: | always combined with a directional particle like *op* (up), *ned* (down) or *frem* (forwards) |
| Examples: | *han lod den umagelige stol vippe baglæns* (he let the uncomfortable chair tip backwards) |

(122) *Group 3*

| | |
|---|---|
| Number of examples: | 7 |
| Semantic characteristics: | animate mover, back and forth movement |
| Syntactic characteristics: | all examples combine with a place modifier initiated by the preposition *på* (on) |
| Examples: | *han vippede faretruende på fodbalderne* (he rocked dangerously on his foot balls) |

(123) *Group 4*

| | |
|---|---|
| Number of examples: | 14 |
| Semantic characteristics: | animate makes body part move up and down |
| Syntactic characteristics: | all examples combine with the preposition *med* (with) or the preposition *på* (on) |
| Examples: | *han vippede med foden* (he rocked with his foot) |

(124) *Group 5*

| | |
|---|---|
| Number of examples: | 25 |
| Semantic characteristics: | animate makes something/ somebody move away abruptly from some place |
| Syntactic characteristics: | transitive phrasal verb constructions |
| Examples: | *jeg kan vippe dig af* (I can tip you off) |

(125) *Group 6*

| | |
|---|---|
| Number of examples: | 2 |
| Semantic characteristics: | abstract or metaphorical uses |
| Syntactic characteristics: | no specific characteristica |
| Examples: | *deres udtagelse til kampen vipper* ('their selection to the game rocks') (their selection to the game is uncertain) |

As can be seen, the groupings serve as a first impression of the meaning potential of the words in question.

## 5.3 An analysis of Danish motion verbs

### 5.3.1 Delimitation of verbs of interest

According to our examinations of the electronic version of *Retskrivningsordbogen* Danish has approx. 100 motion verbs, if we in the set of motion verbs include only simple motion verbs where an item moves, and where the causer of the movement is either the moving item itself or is not realised as a primary complement. The set typically includes intransitive verbs like *løbe* (run) and *rulle* (roll) and excludes causative transitive verbs like *stille* (put).

If we relate to Levin & Rappaport Hovav's coarse grouping of verbs into manner verbs and result verbs, we see that the Danish motion verbs exhibit a rather unequal distribution between the two groups (recall also Skautrup's examinations (Chapter 4)); approx. 20 categorised as result verbs - which for the motion domain corresponds to direction or path verbs or, as they are called by Levin & Rappaport: verbs of inherently directed motion - whereas approx. 80 are manner verbs. As is noted by Skautrup, an even more unequal picture emerges if we look into Danish dialects; Danish dialects exhibit an unlimited number of manner verbs[6], whereas there are indeed very few verbs of inherently directed motion. In our corpus examples it is thus also very striking that directed motion is mostly expressed in Danish by applying a manner verb together with a directional particle, a directional prepositional phrase or a combination of the two. We shall come back to this basic characteristic of Danish in the following sections. In Section 5.3.3 we give the full list of examined verbs.

### 5.3.2 Definition of the basic apparatus: frame elements, event structure, and noun typing

In this section we aim at defining which features to operate with within Frame Semantics and Event Structure in order to provide an appropriate analysis of Danish motion verbs. Furthermore we briefly present a partial semantic hierarchy for noun typing which we refer to when treating selectional restrictions.

#### Deciding on frame elements: maximal vs. minimal frame assignment

Recall that we in Chapter 3 chose to adopt a frame semantic approach to our analysis with the application of aspects of Levin & Rappaport Hovav's approach

[6]Especially, we see a huge number of verbs denoting derogative ways in which women walk or move about !

(1991) (forthcoming) and of Pustejovsky's (1995). However, as we have also indicated earlier, the question of exactly how to identify the relevant number of frame elements for a given situation type is not treated in Fillmore & Atkins (1994). Even when having performed extensive corpus analysis it does not appear to be crystal clear at all when a constituent should be considered a frame element and when it should not. The problem is well-known also in relation to valency descriptions at the syntactic level; which constituents are valency bound and which are not ? However, in valency theory there has emerged a large set of valid tests to apply which are agreed upon between most linguists (see for instance Schøsler & Kirchmeier-Andersen 1997, Chapter 1 for a survey of such tests for Danish).

The question is now whether Frame Semantics can borrow from the syntactic definitions or not. In other words, is there a one-to-one relationship between valency-bound constituents and frame elements such that constituents that are valency-bound are also necessarily considered to be central frame elements ? The answer is most likely positive; at least one would expect it to be rare that a valency-bound constituent referred to a peripheral meaning component and not to a core component. As a starting hypothesis, it is on the other hand also plausible that constituents that are considered to have a modifying status in syntax as well as lexically inherent features can prove to constitute a core semantic component.

When examining different applications of the Frame Semantic approach it becomes clear that there is not a fixed pre-ready recipe for identifying frame elements; rather the strategy is to start out with a very broad alignment of possible frame elements within a semantic class and then, as corpus analysis develops, to decrease the number of elements into the 'basic' ones, corresponding to the core meaning components.

If we look at Danish motion verbs and apply the broad set of frame elements presented in Atkins (1994) we could identify the same frame elements in the Danish corpus (frame elements are underlined) :

(126) *når <u>jeg</u> cykler en tur* (Mover)
(when I bike a round)

(127) *du må ikke cykle <u>på fortovet</u>* (Area)
(you are not allowed to bike on the sidewalk)

(128) *at vi en sommer cykler <u>tværs gennem landet</u>* (Path)
(that we one summer biked through the country)

(129) *han cykler til teatret <u>fra hjemmet i Skovshoved</u>* (Source)
(he bikes to the theatre from his home in Skovshoved)

(130) *somme tider cyklede han <u>til nabobyen</u>* (Goal)
(sometimes he biked to the neighbour village)

(131) *han skulle cykle de 2 en halv kilometer ned til stranden* (Distance)
(he was to bike the 2 and a half kilometres down to the beach)

(132) *de cykler med flade dæk* (Manner)
(they are biking with flat tires)

(133) *skramlende cyklede de afsted* (Sound)
(rattling they biked along)

(134) *de cykler roligt, for de..* (Speed)
(they bike slowly because they..)

(135) *når hun cyklede hjem på sin rustne cykel* (Vehicle)
(when she biked home on her rusty bike)

(136) *selv om jeg cykler langs kantstenen* (Medium)
(even if I bike along the kerb)

Also semantic parameters like Time and Cause could be candidates to frame elements as in:

(137) *om sommeren cyklede vi tit efter is* (Time)
(in summer we often cycled for ice-cream)

(138) *flaget vajede for vinden* (Cause)
(the flag waved by the wind)

In a computational dictionary for our purpose, however, the set of frame elements suggested by Atkins is far too unrestricted; we aim at identifying only the core frame elements that can be encoded in the dictionary and which are helpful in the disambiguation and translation phase. In other words, we are interested in distinguishing between *sense selecting* frame elements on the one hand and *sense modulating* frame element on the other, to refer to Cruse's terminology. Several of the frame elements suggested by Atkins illustrate the sense modulations rather than the sense selections which would trigger new translations in other languages. Since we do not intend to treat sense modulation in our model, we therefore aim at restricting our set of frame elements into an absolute minimum. We basically apply three criteria for such an identification of core elements:

1. *syntactic status:* valency bound constituents are good candidates for core frame elements;

2. *frequency:* semantic parameters that are very frequent in the corpus are good candidates for core frame elements;

3. *sense selecting parameters:* semantic parameters that trigger new senses are good candidates for core frame elements (to be tested by means of distributional tests).

On the basis of these criteria we include the following reduced set of frame elements in the motion event:

- *Mover* (Mo) refers to the item that moves in the motion event: the frame element is realised as subject and obviously extremely frequent in the corpus.
- *Direction* (Di) refers to the origin or goal of the movement which indicates that direction is involved: this parameter is weakly bound to the verb from a syntactic point of view and very frequent in the corpus; can also occur as an inherent feature with verbs of inherently directed motion; triggers sense selection as we have seen in different translations in several languages.
- *Manner* (Ma) refers to how movement takes place: mostly an inherent feature which is, however, inherently present in about 80 % of all motion verbs in Danish; can also be expressed by means of a modifier, relatively frequent in corpus.
- *Protagonist control* (Procon) refers to whether the movement is controlled by the moving item itself or by some external factor; very frequent in corpus, mostly realised inherently in the verb or via the subject but can also be expressed as a modifier; triggers sense selection in several cases.
- *Vehicle* (Ve) refers to the vehicle involved; it is frequent in corpus with a particular set of verbs and it is realised either as an object, a prepositional object or solely as an inherent feature to the verb; triggers sense selection in some cases.

The set of frame elements is not very homogeneous; Mover constitutes a basic requirement in order to be able to speak about a motion event at all whereas the four latter are optional in the motion event. Direction and Manner can be expressed as weakly bound or free modifiers respectively as in:

(139) *han svømmede hen til bredden* (Direction)
(he swam over to the bank)

(140) *han ankom til fods* (Manner)
(he arrived on foot)

as inherent features in the verb only, as in

(141) *toget ankom* (Direction)
(the train arrived)

(142) *han kravlede* (Manner)
(he crawled)

or both, as in:

(143) *for asylansøgere der ankommer til landet* (Direction)
(for refugees that arrive to the country)

(144) *de kravlede på hænder og fødder* (Manner)
(the crawled on hands and feet)

Manner and Direction differ from the rest in the sense that they can be further subdivided: the manner elements can for instance relate either to velocity or sound, as in

(145) *han spurtede* (he spurted)

compared to:

(146) *han listede* (he tiptoed)

and the direction element can relate to either source or goal, as in:

(147) *toget afgik* (the train departed)

compared to

(148) *toget ankom* (the train arrived)

The information given in such subdivision is closely related to the *idiosyncratic meaning* proposed by Levin & Rappaport Hovav (forthcoming) ('gap-filling' information in Tenny (1996)) and which is contrasted with the *structural meaning* exposed in the Manner/Direction distinction.

Protagonist control is mostly an inherent feature which typically comprises the proto-agent properties (in the sense of Dowty (1991)) of in particular *volition* of the moving item. Syntactically this can be realised by means of 'volitional' modifers as in:

(149) *han svømmede for langt ud <u>med vilje</u>* (he swam too far out on purpose)

(150) *han trampede <u>sig vej</u> gennem krattet* (he stomped his way through the brushwood)

Hence, such modifiers can also function as a distributional test for the assignment of this feature, for instance we cannot say:

(151) **jorden drejer rundt med vilje*
('the earth revolves on purpose')

(152) **kuglen trillede sig vej ud af æsken*
('the ball rolled oneself way out of the box')

Vehicle is an inherent frame element with all vehicle verbs but is also frequently expressed syntactically, as in:

(153) *han kørte i bil til stationen*
('he drove in car to the station')

Also verbs that are not prototypically vehicle verbs, such as *flyve* (fly) and *trille* (roll) can implicitly refer to a vehicle when combined with a human argument as shown below:

(154) *han fløj til Oslo for en dag*
(he flew to Oslo for a day)

(155) *klokken er tre om natten da Bente triller ind på en rasteplads*
('it is 3 am when Bente rolls into a pull-in')

contrary to their prototypical uses, as in:

(156) *fuglene fløj omkring i store flokke*
(the birds flew about in large flocks)

(157) *tårerne trillede ned ad kinderne på ham*
(the tears rolled down his cheeks)

#### Deciding on a simplistic event structure assignment

As we have already discussed in Chapter 2, an extended event structure analysis in the line of Pustejovsky (1995) is too ambitious for the purpose of lexical ambiguity in machine translation; on the other hand, it has been shown that some basic notions of aspect are indispensible when dealing with the elasticity in word meaning in relation to motion verbs. Therefore we opt for a simplistic event structure assignment where the traditional Vendlerian verb classes are used as features assigned to the particular events. For motion verbs this basically includes three aspectual types:

- *activity*
- *achievement*, and
- *accomplishment*

As we have seen in Chapter 2, activities can be seen as *simple* events, whereas achievements and accomplishments are *complex* events consisting of various subevents. We have also seen that result verbs such as those of directed motion inherently denote a final state of the event - and are therefore restricted in meaning potential - whereas manner verbs in their purest form denote a simple activity event and are therefore flexible with respect to extension into complex events.

On this basis we can now assign the following event types to the events denoted by the examples below:

(158) *han ankom til stationen* (achievement)
(he arrived to the station)

(159) *pendulet svinger* (activity)
(the pendulum swings)

(160) *han var god til at hoppe* (activicy)
(he was good at jumping)

(161) *han hoppede op på bordet* (accomplishment)
(he jumped up onto the table)

(162) *hun løber i parken* (activity)
(she runs in the park)

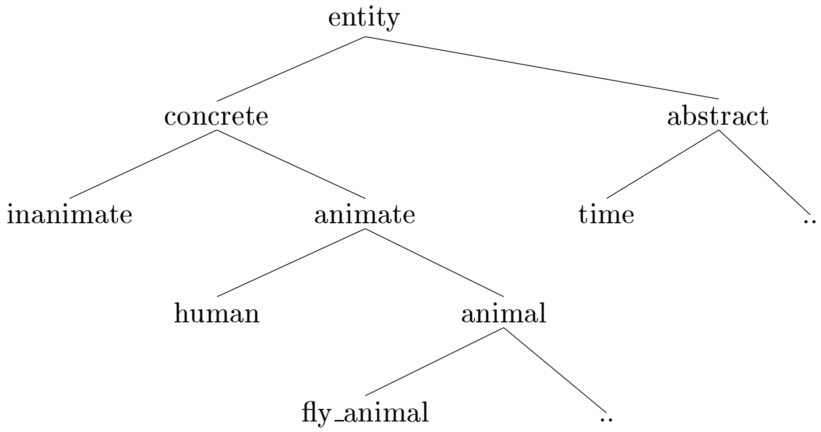
(163) *han svømmede ind i fjorden* (accomplishment)
(he swam into the fjord)

Note that iterative movements indicated by for instance *hoppe* (hop) and *swinge* (swing) are also assigned the simple aspectual type 'activity' even if they prototypically include several subevents.

### Deciding on a semantic hierarchy for noun typing

When performing a semantic description of verbs, it is inevitable to refer to selectional restrictions which again refer to some kind of semantic typing on nouns. In order to make such references clear we in this section present the partial semantic hierarchy that we are presuming in the descriptions of the following sections and chapters. The hierarchy is not to be understood as a complete hierarchy for noun typing, only we wish to place the labels that we apply henceforth in a hierarchical structure so that it should be clear exactly which sub- and supertypes are presumed.

(164) *Partial sort hierarchy for noun typing*




As can be seen, the partial hierarchy is not controversial in any way, however, it should be mentioned that it differs from the DISEM hierarchy proposed for Danish by Togeby (Boje & Schøsler (eds.) 1992) in the sense that we establish a label to generalise over humans and animals (animate); a generalisation which is not made in the DISEM hierarchy, but which for our purpose is indispensable since the characteristics of having a *body* is essential when treating motion.

### 5.3.3 A taxonomy of Danish motion verbs

#### Introduction

In Frame Semantics, a taxonomic approach to verbs is taken, dividing verbs into semantic classes and subclasses so that words that are considered semantic neighbours with similar sets of frame elements are treated coherently.

Traditionally, lexical taxonomical hierarchies are mostly discussed in relation to nouns which exhibit a clear pattern of hyponomy in examples like *cow* - *animal* and *rose* - *flower* (Lyons 1977, p.291-305). In contrast, verbs only to some extent exhibit relations of hyponomy in cases like *move* and *act* which can be seen as superordinate lexical items to other verbs.

This fact about verbs, however, does not prevent us from defining nodes in a taxonomy at each level where it is convenient for us to generalise over groups of verbs with similar meaning components. In our case, where the aim is to get hold of groups of verbs which exhibit the same regularities in meaning deviation, we therefore feel free to establish such nodes even if they are cases of what is often labelled *lexical gaps*. It should furthermore be noted that such lexical gaps can differ radically from language to language, what is labelled with a specific lexeme in one language may constitute a lexical gap in another.

However, as is stated by Cruse (1986 p.152), one should take into account that there is an extra measure of indeterminacy in the meaning of a verb out of context as compared to the nouns; a factor which makes the establishment of strict taxonomies of verbs out of context difficult. Being fully aware of this fact, we try in our taxonomy to capture only the most *prototypical* and 'pure' sense of the verbs in question leaving other 'derived' senses to be accounted for elsewhere.

In this context we shall not go deeper into the theoretical discussion of the cognitive processes that underlies mappings or transfers from prototypical senses of a word into 'derived' senses. We shall take for granted that such transfers take place (as described in e.g. Lakoff & Johnsen 1980) and that we have an intuitive understanding of what a prototypical sense of a word is (cf. Jackendoff 1983, Wierzbicka 1996 and others for uses of prototypes in the semantic definitions of words). Note however, that it is not always trivial to decide upon which senses to consider prototypical; for instance, one cannot get a clear answer from corpus analysis alone since frequent uses of a word cannot always be read as a sign of prototypicality. Examples of this are verbs like *løbe* (run) and *gå* (walk); where only approx. 5 % of the corpus examples showed prototypical uses of these verbs (i.e. moving by putting one foot in front of the other on the ground).

### The taxonomy

Taking into consideration the corpus data examined for Danish, a tentative taxonomy of Danish motion verbs can now be established on the basis of the set of semantic features defined in the sections above.

There are several central comments to be made on the taxonomy, for instance the uneven subdivisions made in direction verbs and manner verbs, respectively, which most presumably relates to the fact commented upon earlier, namely that there are very few direction verbs in Danish but many manner verbs denoting very different ways of moving.

This also relates to the assignment of frame elements. As can be seen below, these are assigned at the top nodes of the taxonomy whereas nodes lower in the hierarchy are not distinguished formally by means of frame elements; however, we can benefit from a further subgrouping in formalisation and implementation since we can refer to subgroups of verbs by means of types.

The frame element Mover, which is assigned to all motion verbs and which distinguishes the motion domain from other semantic domains, hardly requires further comments. Also Direction and Manner are more or less self-explanatory since they relate to the central inherent feature of direction verbs and manner verbs, respectively. Basically, what this distinction indicates is that motion verbs inherently denote *either* direction of motion *or* manner of motion as their main characteristics.

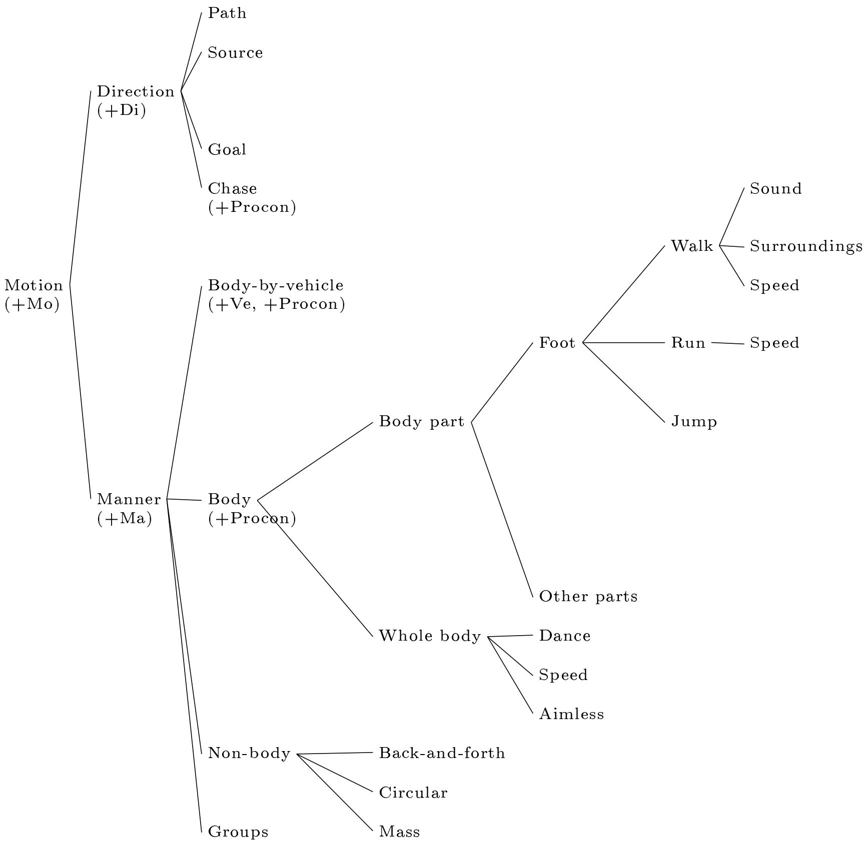
The Protagonist control feature, which we assign to all body- and body-by-vehicle verbs, on the other hand, deserves further discussion. By this assignment we claim that all motion verbs which prototypically denote ways in which bodies move about imply protagonist control. Of course bodies can also move unwillingly - without protagonist control, but in such cases we would often apply either a direction verb, like for instance *falde* (fall) or non-body verbs like for instance *dingle* (dangle) or *sitre* (quiver).

However, in the taxonomy, we implicitly claim that it is prototypical for *dingle* and *sitre* to be combined with inanimate arguments (since both are categorised as non-body verbs), and consequently, the possibility of these verbs occurring with an animate argument is not directly reflected by the taxonomy (but is commented upon in Sections 5.3.7 and 6.2.7).

It should be noted, however, that there are a few verbs in Danish which prototypically denote animate entities moving unwillingly because of some physical unbalance, as seen in *skælve* (tremble) and *spjætte* (jerk). It can be discussed to which degree these verbs belong to the classical motion domain; cf. Levin (1993) who categorises *tremble* as a verb of existence, and more specifically as: a verb of mode of being involving motion. Anyhow, these two verbs are excluded from our set of verbs.

As regards the last frame element *Vehicle* this is assigned inherently to all body-by-vehicle verbs, since a vehicle is involved with these verbs even if it is not mentioned explicitly.

(165) *Taxonomy of Danish motion verbs*




From a more general viewpoint, further comments can be made to the suggested categorisation. For instance, one could call into dispute the distinction between structural meaning and idiosyncratic meaning as put forward by Levin and Rappaport Hovav (forthcoming), Grimshaw (1993), Jackendoff (1990), Tenny (1995) and others. The top nodes in our taxonomy in many ways correspond to the levels of meaning referred to as structural meaning or templatic meaning; whereas the lower nodes correspond rather to the idiosyncratic meaning elements. However, it can be argued on the basis of the taxonomy that what is presumably to be understood as idiosyncratic information can to a large extent be generalised over; for instance the node *foot verbs* covers a large range of verbs denoting motion by moving ones feet in one way or another, and these verbs prove to exhibit a similar syntactic and semantic behaviour in several ways which we shall look closer at in the following section.

One thing that has been argued is that idiosyncratic meaning is closely related to *encyclopedic* world knowledge and thus not something that should necessarily be treated in linguistics. However, if this kind of information *is* reflected in language by means of regularities in distribution it can in our opinion be difficult to maintain this clearcut distinction. Another problem concerns the few superordinate verbs that occur in the hierarchy: does a verb like *bevæge sig* (move) have only structural meaning, for instance ? It is probably so.

### List of verbs

For the moment, however, we shall leave this discussion aside, and take a look at how our 100 Danish verbs are distributed in the taxonomy:

MOTION VERBS, + Mo
These verbs indicate an event where an item moves and where the causer of the movement is either the moving item itself or som external factor:

> *bevæge sig* (to move)

1. Direction verbs, +Mo, +Di
These verbs have inherently directed motion, even in structures with no directional modifier, all subgroups except for the chase verbs are telic in meaning; i.e. denote either achievements or accomplishments.

1.1 Path verbs, +Mo, +Di
Inherent element of direction: path, which is expressed by the object.

> *krydse* (cross)
> *passere* (pass)

1.2 Source verbs, +Mo, +Di
Inherent element of direction: source of the movement.

> *afgå* (depart)
> *forlade* (leave)
> *forsvinde* (disappear)
> *skride* (take off)

1.3 Goal verbs, +Mo, +Di
Inherent element of direction: goal of the movement; the movement can denote either vertical or horizontal direction; verbs of horizontal direction require more explicit goals in the context whereas verbs of vertical motion are 'given' wrt. direction. Note furthermore that the group is mixed wrt. event type; again vertical verbs always denote an activity as part of the event, whereas the horizontal group is mixed in this respect.

*ankomme* (arrive)
*dale* (descend)
*dratte (ned)* (flop down)
*falde* (fall)
*komme* (come)
*returnere* (return)
*stige* (ascend)
*styrte (ned)* (fall down)
*tage* (go)

1.4 Chase verbs, +Mo, +Di, +Procon
Inherent element of direction: goal is the chased object, which is expressed by the direct object.

*følge* (follow)
*forfølge* (pursue)
*jagte* (chase)
*jage* (chase)

2. Manner verbs, +Mo, +Ma
These verbs have an inherent notion of manner.

2.1 Body-by-vehicle verbs, +Mo, +Ma, +Procon, +Ve
Inherent manner element: moving by means of a vehicle. Vehicle can be unexpressed or expressed by means of an object or a prepositional object. The particular vehicle is for some verbs expressed in the verb itself.

*cykle* (cycle)
*drøne* (race)
*køre* (drive)
*ride* (ride)
*ro* (row)
*skøjte* (skate)
*sejle* (sail)
*stå på ski* (ski)
*surfe* (surf)

2.2 Body verbs, +Mo, +Ma, +Procon
Inherent manner element: moving by moving the body or part of the body.

2.2.1 Body part, +Mo, +Ma, +Procon
Inherent manner element: moving by moving part of the body

2.2.1.1 Foot verbs, +Mo, +Ma, +Procon
Inherent manner element: moving by putting one foot in front of the other on the ground.

2.2.1.1.1 Walk verbs, +Mo, +Ma, +Procon
Inherent manner element: moving fairly slowly by putting one foot in front of the other on the ground.

*gå* (walk)

2.2.1.1.1.1 Sound verbs, +Mo, +Ma, +Procon
Inherent manner element: walking and making little or much noise.

*liste* (tiptoe)
*snige sig* (sneak)
*trampe* (stomp)

2.2.1.1.1.2 Surroundings, +Mo, +Ma, +Procon
Inherent manner element: walking in specific surroundings, walking in step within a specific group of people, or wearing specific shoes or clothes.

*defilere* (march)
*marchere* (march)
*spadsere* (promenade)
*spankulere* (strut about)
*vade* (wade)
*vandre* (wander)

2.2.1.1.1.3 Speed, +Mo, +Ma, +Procon
Inherent manner element: walking slowly with uncertainty, aimlessly or clumsily.

*drysse* (saunter)
*jokke* (plod)
*sjokke* (shamble)
*sjoske* (shamble) *futte* (scurry)
*slentre* (stroll)
*traske* (plod)
*trisse* (trundle)
*tøffe* (chug)

2.2.1.1.2 Run verbs, +Mo, +Ma, +Procon
Inherent manner element: moving fairly quickly by putting one foot in front of the other on the ground.

*løbe* (run)

2.2.1.1.2.1 Speed verbs, +Mo, +Ma, +Procon
Inherent manner element: running relatively fast

*galoppere* (gallop)
*rende* (run)
*sprinte* (sprint)
*spurte* (spurt)
*trave* (trot)

2.2.1.1.3 Jump, +Mo, +Ma, +Procon
Inherent manner element: moving by shoving off with your foot/feet from the ground up into the air. Iterative movement can be involved.

*hinke* (hop on one leg)
*hoppe* (hop)
*springe* (jump)

2.2.1.2 Other parts, +Mo, +Ma, +Procon
Inherent manner element: moving by moving other parts of the body than the feet, like wings, knees and elbows, the stomach, arms and legs, or along a surface.

*flyve* (fly)
*kravle* (crawl)
*mave sig* (crawl forward on ones stomach)
*svømme* (swim)
*rutche* (slide)

2.2.2 Whole body, +Mo, +Ma, +Procon
Inherent manner component: This group denotes motion of the body that is not necessarily related to any specific body parts.

2.2.2.1 Dance verbs, +Mo, +Ma, +Procon
Inherent manner element: moving to music or rythms.

*danse* (dance)
*steppe* (tap dance)
*valse* (waltz)

2.2.2.2 Speed, +Mo, +Ma, +Procon
Inherent manner element: moving fast.

*fare* (rush)
*haste* (hurry around)
*ile* (hasten)
*piske* (dash)
*ræse* (rush)
*styrte* (rush)

2.2.2.3 Aimless, +Mo, +Ma, +Procon
Inherent manner element: moving carelessly around without aim, or moving by swinging ones hips. Mostly used in a derogatory manner.

*flakke* (wander about)
*føjte* (gad about)
*strejfe* (roam about)
*svanse* (swing your hips while moving)
*tulre* (blunder along)
*vimse* (bustle about)
*vrikke* (wriggle)

2.3 Non-body verbs, +Mo, +Ma
Inherent manner element: These verbs denote motion of non-body entities which is typically different from motion of body, for obvious reasons: there is no protagonist control of the motion, and the physical shapes differ radically from bodies. Therefore the manner elements in this group typically include iterative movements around an axis or a center of some kind.

2.3.1 Back and forth, +Mo, +Ma
Inherent manner element: moving back and forth with quick or slow movements.

*daske* (flap)
*dingle* (dangle)
*svaje* (sway)
*svinge* (swing)
*sitre* (quiver)
*ryste* (shake)
*vaje* (wave)
*vibrere* (vibrate)
*vippe* (rock)
*vugge* (rock)

2.3.2 Circular, +Mo, +Ma
Inherent manner element: moving in a circular way, often around an axis.

*dreje* (turn)
*kredse* (circle)
*rotere* (rotate)
*rulle* (roll)
*snurre* (spin)
*trille* (roll)

2.3.3 Mass, +Mo, +Ma
Inherent manner element: masses or liquids moving slowly or quickly.

*drive* (stream)
*vælde* (gush)
*sive* (ooze)
*strømme* (pour)

2.4 Groups, +Mo, +Ma
Inherent manner element: groups of entities moving about rather quickly in large groups.

*myldre* (teem/swarm)
*sværme* (swarm)
*vrimle* (teem/crawl)

### 5.3.4 Syntactic alternations of Danish motion verbs

In this section we examine which syntactic alternations Danish motion can undergo as a first step towards an overview of the semantic potential of each verb group[7]. Our working hypothesis is that verb behaviour is to a large extent determined by verb meaning; a hypothesis which has been largely discussed in linguistics during the last decades and which is dealt with most substantially by Levin (1993). As mentioned in Chapter 3, the hypothesis of semantic determination implies the following assumptions:

- If the members of a set of verbs share some meaning components, then the members of the set can be expected to exhibit the same syntactic behaviour, and
- if the members of a set of verbs exhibit the same syntactic behaviour, then the members of the set can be expected to share some meaning component.

[7]For an overview of syntactic alternations of Danish verbs in general, see Schøsler & Kirchmeier-Andersen 1997.

In other words, if the hypothesis holds, we have an important tool for the investigation of word meaning which supports our own introspection.

**Directional, causative, aspectual and expletive alternations**

On the basis of our corpus investigations, we have basically identified 5 kinds of syntactic alternations that Danish motion verbs can undergo:

- *directional* alternations,
- *causative* alternations,
- *aspectual* alternations,
- *expletive det*-alternations, and
- *expletive der*-alternations

These groups can be further subdivided into more specific types:

**Directional alternation**

Addition of directional modifier
Here we examine whether a directional modifier can be added to the clause. It is obvious that verbs with inherently directed motion (+Di) will allow this, but also many body verbs and transport verbs are frequently seen with a directional modifier, which in such cases adds an extended sense to the verb, which could be paraphrased as 'go by V-ing':

(166) *han lister* (he tiptoes)

(167) *han lister hen til huset* (he tiptoes over to the house)

There are also examples where this is not possible

(168) *flaget vajer* (the flag waves)

(169) **flaget vajer hen til huset* ('the flag waves over to the house')

Omission of directional modifier
A small group of transitive motion verbs do not allow omission of direction since it is expressed as object:

(170) *han forfulgte bilen* (he followed the car)

(171) **han forfulgte* ('he followed')

<u>Alternation between directional types</u>
This alternation tests whether the directional modifier can be expressed both as a pure PP and as a directional adverbial followed by a PP. In Danish, directional adverbs like hen (over), over (over) and ud (out) are extremely frequent in relation to motion verbs, as in:

(172) *han gik hen til stationen* (he went over to the station)

but can also be omitted as in:

(173) *han gik til stationen* (he walked to the station)

It seems that the directional adverbs are used with manner verbs (+Ma) when part of the focus is moved from the manner aspect to the directional aspect. Note that we prefer to apply *go* in the first translation, but *walk* in the second translation.

Manner verbs with very specific manner components (verbs placed low in the taxonomy) only combine with a directional modifier when there is a directional adverb:

(174) *han listede ud til hende* (he tiptoed out to her)

(175) **han listede til hende* ('he tiptoed to her')

Several direction verbs, in contrast, do not combine with directional adverbs at all:

(176) **han ankom hen til stationen* ('he arrived over to the station')

Whether this is due to the inherent direction component in *ankomme* or whether we are in this particular case dealing with incompatibility due to the prefix *an* is unclear to us.

Semantic aspects of the goal are also of relevance, typical destinations like *byen* (the town) are more willingly attached directly to a motion verb than concrete objects:

(177) *han gik til byen* (he walked to the town)

(178) **han gik til bordet* ('he walked to the table')

Furthermore, each directional adverb has its own restrictions, for instance *hen* can only be used when moving over short distances:

(179) **han rejste hen til Australien* ('he travelled over to Australia')

**Causative alternations**

<u>Ergative alternation</u>
The ergative alternation is possible for ergative verbs, which allow for the Mover (Mo) to be realised as either subject or object, as in:

(180) *vinden driver båden* (the wind drifts the boat)

(181) *båden driver* (the boat drifts)

For all unergative verbs where the subject denotes the direct cause of the eventuality described, this alternation is not possible:

(182) *jeg ror båden* (I row the boat)

(183) **båden ror* ('the boat rows')

<u>Induced action alternations</u>
This alternation is relevant for many body-by-vehicle verbs where the vehicle can be realised as both the subject and the object and where the causer of the movement causes both himself and the vehicle to move:

(184) *jeg kører bilen* (I drive the car)

(185) *bilen kører* (the car drives)

(186) *jeg kører* (I drive)

(187) *jeg kører i bilen* (I drive in the car)

This is not the case for other motion verbs, not even for ergative verbs:

(188) *jeg ryster cyklen* (I shake the bike)

(189) *cyklen ryster* (the bike shakes)

where it is not implied that the causator shakes.

<u>Passive alternation</u>
Passive constructions are rare with Danish motion verbs, which is obviously due to the fact that most of the verbs in this class are intransitive. However, some verbs do occur in passive *der*-constructions, typically when denoting something that groups of people are likely to do together on certain locations, as in:

(190) *der danses her* ('there is danced here')

Direction verbs and verbs with direct external cause do not allow for these constructions most presumably because these verbs lack the immediate cause relation between subject and verb:

(191) **der ankommes her* ('there is arrived here')

(192) **der vajes her* ('there is waved here')

Resultative alternation

In a resultative construction an adjectival phrase is added to describe a state achieved by performing the action described by the verb. Since this is only possible for transitive constructions, the motion verb constructions with resultatives are always found together with an object or a reflexive pronoun:

(193) *han gik sig træt* (he walked himself tired)

(194) *han gik gaderne tynde* (he walked the streets thin)

Verbs with inherently directed motion do not allow this:

(195) **han ankom sig træt* ('he arrived himself tired')

as well as non-body verbs:

(196) **flaget vajede sig slidt* ('the flag waved itself worn out')

As was the case for *der*-passives, the lacking immediate cause relation between subject and verb in the last two examples is a plausible explanation for their ungrammaticality.

**Aspectual alternations**

The *have-være* alternation

Two auxiliaries occur with Danish verbs, *have* (have) and *være* (be). Transitive verbs always take *have*, intransitive verbs can take either depending on the aspectual notions denoted by the verb or by the sentence as a whole (see Aa. Hansen 1967). As we have seen, also motion verbs differ with respect to aspect, motion verbs of inherent directed motion, being delimited in time - telic -, only take *være*, as in:

(197) *han er kommet hjem* ('he is come home') (he has come home)

(198) **han har kommet langt* ('he has come far')

Verbs which are not delimited in time - atelic - only take *have*:

(199) *den lap har dinglet hele tiden* (that patch has dangled all the time)

(200) **den lap er dinglet* ('that patch is dangled')

however, most motion verbs occur with both auxiliaries depending on the aspectual characteristics in the structure in which they are realised:

(201) *han er gået hjem* ('he is walked home')

(202) *han har gået 2 km* (he has walked 2 km)

The first sentence is telic due to the goal modifier *hjem*, the second is atelic due to the measure modifier *2 km*.

This alternation is often referred to as an important unergative/unaccusative diagnostics since exactly the atelic/telic distinction reflected in the *have/være* alternation is used as a test of unergativity vs. unaccusativity (*have* = atelic = unergative, *være* = telic = unaccusative). Thus, body verbs are in many languages considered to be flexible wrt. unaccusativity depending on the telicity of the construction in which they occur as we have already commented on in Chapter 4. The diagnostics, however, in several cases clashes with the more semantically motivated diagnostics of unergativity, namely the existence of proto-agent properties (in the sense of Dowty 1991) such as volition (see among several others Zaenen 1993 Levin & Rappaport Hovav 1996 and with specific attention to Danish: Ørsnes 1995:p.62-69 and Olsen 1997:34-57), since several *være* constructions can be combined with agentive modifiers as in:

(203) *han var kørt ind i ham med vilje*
('he had driven into him on purpose')
(he had hit him with the car on purpose)

Similarly, we see motion verbs combined with *have* whose typical subjects do not have typical proto-agent properties and which should therefore according to the semantic test be categorised as unaccusative:

(204) *hjulet havde drejet rundt i flere timer*
(the wheel had turned for several hours)

As has been stated earlier, we use Protagonist Control to refer to volitional properties whereas we assign the feature of telicity as an aspectual value. Thus, we leave aside the decision of whether a verb is unaccusative or not, still capturing - however - the same linguistic characteristics.

<u>Time modifier alternation</u>
Another aspectual alternation which is related to the *have-være* alternation con-

cerns the ability to attach a time modifier making the aspect of the whole sentence atelic:

(205) *han har gået i to timer* (he has walked for two hours)

Telic motion verbs do not allow this:

(206) **han er faldet i to timer* ('he is fallen for two hours')

(207) **han er kommet i to timer* ('he is come for two hours')

<u>Measure modifier alternation</u>
This alternation relates to the time modifier alternation since measure modifiers are mostly attached to atelic verbs. Many of the manner verbs (+Ma) allow this:

(208) *han har gået 2 km* (he has walked 2 km)

(209) *han har rejst 20 km* (he has travelled 20 km)

However, also some of the verbs with inherently directed motion can be seen with measure modifiers, but then a directional adverbial is attached as well, as in:

(210) *han er faldet to meter ned* (he has fallen two metres down)

(211) *han er kommet to meter længere hen*
(he has come two metres further along)

In these cases an inherently telic verb is 'coerced' into a construction of limited duration. Such a meaning coercion is not possible for all verbs in the group, in particular verbs of pure achievement are restricted in this respect:

(212) **toget er ankommet 2 meter længere hen*
('the train has arrived to metres further along')

<u>Conative alternation</u>
This alternation is relevant for verbs that can occur in transitive constructions and where the object of the verb can appear in an intransitive construction as the object to a prepositional phrase headed mainly by the preposition *på* (on, at):

(213) *jeg drejer hanen* (I turn the tap)

(214) *jeg drejer på hanen* (I turn at the tap)

(215) *jeg driver hesten* (I urge the horse)

(216) *jeg driver på hesten* (I urge at the horse)

The sentences containing the PP express an attempted action which is either not necessarily carried out or has extension over time.

Such attempted actions cannot be expressed with all kinds of motion verbs, as can be seen below:

(217) *jeg bevæger benene* (I move my legs)

(218) **jeg bevæger på benene* ('I move at my legs')

Cognate object alternation

Cognate objects, i.e. objects which partially or wholly repeat the content of the verb, are rare with motion verbs. Again only verbs that can denote extension over time are relevant. For instance, dance verbs are seen with cognate objects as in:

(219) *de dansede en vals* (they danced a waltz)

but also many body verbs and body-by-vehicle verbs allow the cognate object *tur* (tour):

(220) *han gik en tur* ('he walked a tour')

(221) *han kørte en tur* ('he drove a tour')

Also the immediate cause relation between subject and verb is a prerequisite for such constructions, for instance we cannot normally say:

(222) *?bolden rullede en tur* ('the ball rolled a tour')

**Expletive *det*-alternations**

*Det*-alternations are only seen with collective NPs as subjects and only with *group*-verbs and mass verbs (see also Shøsler & Kirchmeier-Andersen 1997, Chap. 2):

(223) *det vrimler med mennesker* (it swarms with people)

(224) *det sværmer med bier* (it swarms with bees)

(225) *det myldrer med insekter* (it crawls with insects)

(226) *det siver ned ad væggene med vand*
('it oozes down the walls with water')

Some of these verbs even occur in actual locative inversions, as in:

(227) *myrerne vrimler i køkkenet* (the ants crawl in the kitchen) vs.

(228) *køkkenet vrimler med myrer* (the kitchen crawls with ants)

Note that we can even have examples where no motion takes place and where the verb denotes group existence rather than group motion, as in the examples below:

(229) *motorrummet vrimlede med fine små detaljer*
(the engine room was teeming with fine little details)

(230) *huset vrimlede med ure* (the house was teeming with clocks)

### Expletive *der*-alternation

Active *der*-constructions are very frequent with Danish motion verbs but the rules for the use of existential *der* are complex and highly related to other aspects of the sentence, such as the subject NP which must be indefinite, as well as discourse phenomena (see among others Nimb 1993 and Paggio 1996). Independent of these aspects, only some motion verbs allow *der*-insertion as can be seen below:

(231) *manden gik forbi* (the man passed by)
*der gik en mand forbi* (there passed a man by)

(232) *hendes stemme rystede* (her voice shook)
**der rystede en stemme* ('there shook a voice')

The semantic notion of being able to appear and disappear has some significance here. This is implicit with the verb *gå* since the man is moving from one place to another, whereas the verb *ryste* cannot combine with direction and therefore cannot appear nor disappear as an act of the moving. However, in some cases, the notion of existence seems to override this constraint, for instance we *can* say:

(233) *der vajede et flag på husets tag*
('there waved a flag on the roof of the house')

which is almost equivalent to what is denoted by verbs of existence:

(234) *der stod/var/hang et flag på husets tag*
('there stood/was/hung a flag on the roof of the house')

### Test results

In order to get a more complete overview of the distribution of possible alternations throughout the taxonomy, the complete set of alternations for each verb has been examined. In the tables below, test results are given for each subgroup[8].

(235) *Direction verbs (*krydse, afgå, ankomme, følge..*)*

| | 1.1 Path | 1.2 Source | 1.3 Goal | 1.4 Chase |
|---|---|---|---|---|
| DIRECTIONAL | | | | |
| Addition | + | + | + | + |
| Omission | + | + | + | - |
| Alternation | - | - | mixed[9] | - |
| CAUSATIVE | | | | |
| Ergative | - | - | - | - |
| Induced action | - | - | - | - |
| Passive | - | - | - | + |
| Resultative | - | - | - | - |
| ASPECTUAL | | | | |
| *have-være* | mixed | - | - | - |
| Time | - | - | - | + |
| Measure | - | - | mixed | + |
| Conative | - | - | - | - |
| Cognate | - | - | - | - |
| *DET* | - | - | - | - |
| *DER* | + | + | + | - |

General comments: Direction verbs demonstrate homogeneous behaviour in a majority of the cases; an exception is the 'chase' group which differs with respect to transitivity and therefore to a number of syntactic tests. The group as a whole shows flexibility with respect to direction; these can be added or omitted; however, the group is mixed wrt. the *hen*-alternation, as can be seen with *komme* (come) and *tage* (go) which in contrast to the other verbs in this group allow a *hen*-construction:

(236) *småpigerne kom straks hen til mig*
(the little girls came over to me immediately)

(237) *jeg tog hen på hans arbejde*
(I went over to his work)

This difference in distribution may actually refer to aspectual deviations, *komme* and *tage* can be seen as pure accomplishments, whereas *ankomme* (arrive) and *afgå* (depart) are pure achievements.

[8]Thanks to Lina Henriksen who has assisted in part of the testing of verb alternations at an earlier stage of this work.

As regards other causative and aspectual alternations the verb group is indeed very restricted.

(238) *Body-by-vehicle verbs (*cykle, køre..*)*

| | 2.1 Body_by_vehicle |
|---|---|
| DIRECTIONAL | |
| Addition | + |
| Omission | + |
| Alternation | + |
| CAUSATIVE | |
| Ergative | mixed |
| Induced action | mixed |
| Passive | + |
| Resultative | + |
| ASPECTUAL | |
| *have-være* | + |
| Time | + |
| Measure | + |
| Conative | - |
| Cognate | + |
| *DET* | - |
| *DER* | + |

Comments: body-by-vehicle verbs are generally very flexible with respect to alternations of all kinds. A few verbs in the group show a larger flexibility than others, though, *køre* and *sejle* differ from the others in the sense that they denote both the motion that the vehicle performs in itself:

(239) *bilen kører* (the car drives)

(240) *skibet sejler* (the ship sails)

and the motion that bodies make vehicles do, so to speak:

(241) *jeg kører bilen* (I drive the car)

(242) *jeg sejler båden* (I sail the boat).

*Cykle* (cycle) and *skøjte* (skate) in contrast can only denote the motion that we make the vehicles do; we can therefore not say

(243) **cyklen cykler* ('the bike cycles')

or

(244) **skøjterne skøjter* ('the skates skate').

This has to do with the fact that the latter mentioned verbs repeat in their name what kind of vehicle is involved in the motion. The difference in flexibility is reflected in the schema by 'mixed' columns.

(245) *Foot verbs (*gå, løbe, hoppe*)*

| | 2.2.1.1.1 walk | 2.2.1.1.2 run | 2.2.1.1.3 jump |
|---|---|---|---|
| DIRECTIONAL | | | |
| Addition | + | + | + |
| Omission | + | + | + |
| Alternation | + | + | - |
| CAUSATIVE | | | |
| Ergative | - | - | - |
| Induced action | - | - | - |
| Passive | + | + | + |
| Resultative | + | + | - |
| ASPECTUAL | | | |
| *have-være* | + | + | + |
| Time | + | + | + |
| Measure | + | + | + |
| Conative | - | - | - |
| Cognate | + | + | + |
| *DET* | - | + | - |
| *DER* | + | + | + |

Comments: *gå* and *løbe* are seen as meta-verbs for a whole group of verbs (see the following figures treating sound, surrounding, speed and derogatory verbs). The verbs are very flexible with respect to directional and aspectual alternations and show very similar distributional patterns in general. One exception is the expletive *det*-alternation which can be used for *løbe* as in

(246) *det løber ned ad væggene med vand*
(it runs down the walls with water).

Jump verbs do not have a meta-status like *gå* and *løbe* and thus show a more restricted behaviour, these tend not to attach directional modifiers without a directional adverb for instance, as in

(247) *?haren hoppede til byen* (the hare hopped to the town).

whereas we can say:

(248) *haren hoppede ned ad bakken* (the hare hopped down the hill)

(249) *Subcategories to walk verbs: (*trampe, vandre, slentre, sjoske..*)*

| | 2.2.1.1.1.1 | 2.2.1.1.1.2 | 2.2.1.1.1.3 | 2.2.1.1.1.4 |
|---|---|---|---|---|
| | Sound | Surround. | Speed | Derogato. |
| DIRECTIONAL | | | | |
| Addition | + | + | + | + |
| Omission | + | + | + | + |
| Alternation | - | - | mixed | - |
| CAUSATIVE | | | | |
| Ergative | - | - | - | - |
| Induced action | - | - | - | - |
| Passive | + | + | + | + |
| Resultative | mixed | - | mixed | - |
| ASPECTUAL | | | | |
| *have-være* | + | + | + | + |
| Time | + | + | + | + |
| Measure | + | + | + | + |
| Conative | - | - | - | - |
| Cognate | mixed | mixed | mixed | mixed |
| *DET* | - | - | - | - |
| *DER* | + | + | + | + |

Comments: compared to *gå*, these verbs prove to be less flexible, probably due to their very specific meanings. For instance, resultatives are doubtful with most of these verbs although some are acceptable even if there is not evidence of all of them in the corpus:

(250) *han trampede kassen flad* (he stomped the box flat)

(251) *?han slentrede gaderne tynde* (he strolled the streets thin)

Cognate objects are also doubtful although again some are acceptable:

(252) *han listede en tur gennem huset*
(he tiptoed a tour through the house)

(253) *han slentrede en tur gennem byen*
(he strolled a tour through the town)

(254) *Subcategory to run verbs (*sprinte, spurte...*)*

| | 2.2.1.1.2.1 |
|---|---|
| | Speed |
| DIRECTIONAL | |
| Addition | + |
| Omission | + |
| Alternation | - |
| CAUSATIVE | |
| Ergative | - |
| Induced action | - |
| Passive | + |
| Resultative | - |
| ASPECTUAL | |
| *have-være* | + |
| Time | + |
| Measure | + |
| Conative | - |
| Cognate | mixed |
| *DET* | - |
| *DER* | + |

Comments: The subcategory of *løbe* is in most respects identical with the subcategory of *gå*. Compared to the speed category of walk verbs, these all denote fast movement, whereas the walk verbs denote slow movement. They exhibit the same mixed behaviour with respect to cognate objects; some combine successfully with a cognate object although such constructions are very rare in corpus.

(255) *?han spurtede en tur* (he sprinted a tour)

(256) *Other body parts (*kravle, klatre, svømme *..)*

| | 2.2.1.2 |
|---|---|
| | Other body parts |
| DIRECTIONAL | |
| Addition | + |
| Omission | + |
| Alternation | - |
| CAUSATIVE | |
| Ergative | - |
| Induced action | - |
| Passive | + |
| Resultative | - |
| ASPECTUAL | |
| *have-være* | + |
| Time | + |
| Measure | + |
| Conative | - |
| Cognate | mixed |
| *DET* | - |
| *DER* | + |

Comments: These verbs follow a pattern similar to the more specific walk and run verb: they require a directional particle in order to combine with a directional prepositional phrase as can be seen below:

(257) *han kravlede hen til kanten* (he crawled over to the edge)

(258) **han kravlede til kanten* ('he crawled to the edge')

Cognate objects are not very common (except for *svømme* (swim)) although they do not seem definitively wrong:

(259) *?han kravlede en tur*
('he crawled a tour')

(260) *Whole body (*danse, ile, føjte*)*

| | 2.2.2.1 | 2.2.2.2 | 2.2.2.3 |
|---|---|---|---|
| | Dance | Speed | Aimless |
| DIRECTIONAL | | | |
| Addition | + | + | - |
| Omission | + | + | + |
| Alternation | - | + | - |
| CAUSATIVE | | | |
| Ergative | - | - | - |
| Induced action | - | - | - |
| Passive | + | + | + |
| Resultative | + | - | - |
| ASPECTUAL | | | |
| *have-være* | + | + | - |
| Time | + | + | + |
| Measure | - | - | - |
| Conative | - | - | - |
| Cognate | + | - | - |
| *DET* | - | - | - |
| *DER* | + | + | + |

Comments: This group denotes motion of the body that is not necessarily related to any specific body parts, but apart from this common characteristic they do not share other semantic features. This can be read from the schema above: within each subgroup the verbs behave in a homogeneous way but if we compare the groups a rather blurred picture emerges of the syntax that they exhibit.

The aimless group, to take an example, denotes very general aspects of how people move about without aim or goal and therefore this group does not combine with directional modifiers:

(261) **han føjtede ned til byen*
('he gadded_about down to the town')

(262) **han flakkede ned til byen*
('he wandered_about down to the town')

(263) **han tulrede ned til byen* ('he blundered_along down to the town')

This fact is also reflected in the aspectual alternations; these verbs cannot undergo atelic - telic derivation:

(264) **han er føjtet hjem* ('he is wandered_about home')

(265) **han er flakket hjem til sin mor*
('he is wandered_about home to his mother')

The dance group on the other hand denotes a very specific kind of body motion, namely motion to music and rhythms; these verbs therefore show a pattern of more flexibility, for instance they can combine with directional modifiers as in:

(266) *de dansede ud på dansegulvet* (they danced out on the dance floor)

and occur with cognate objects as in:

(267) *de dansede en slags sejrsdans*
(they danced a sort of victory dance)

and maybe we can even create a resultative construction as in:

(268) *?hun dansede ham sønder og sammen*
(she danced him to pieces)

(269) *Non-body verbs (*dingle, rotere, sive..*)*

| | 2.3.1 | 2.3.2 | 2.3.3 |
|---|---|---|---|
| | Back&forth | Circular | Mass |
| DIRECTIONAL | | | |
| Addition | - | mixed | + |
| Omission | + | + | + |
| Alternation | - | - | - |
| CAUSATIVE | | | |
| Ergative | mixed | mixed | mixed |
| Induced action | - | - | - |
| Passive | - | - | - |
| Resultative | - | - | - |
| ASPECTUAL | | | |
| *have-være* | - | - | mixed |
| Time | + | + | + |
| Measure | - | - | - |
| Conative | mixed | mixed | - |
| Cognate | - | - | - |
| *DET* | - | - | + |
| *DER* | mixed | mixed | + |

Comments: Even if this group cannot be said to be very homogeneous internally (many boxes with 'mixed'), it differs radically from the other semantic groups in the following respects: it contains many verbs that cannot have a directional modifier added:

(270) **lappen daskede hen til væggen*
('the patch dangled over to the wall')

It contains many ergative verbs:

(271) *hjulet drejer/jeg drejer hjulet*
(the wheel turns/I turn the wheel)

The group does not allow *være* as an auxiliary, meaning that the verbs generally denote atelic events (apart from a very few cases):

(272) **kloden er drejet* ('the earth is rotated')

Finally, it contains verbs that cannot occur in *der*-constructions:

(273) **der vibrerede en ledning*
('there vibrated a cable')

(274) *Group verbs*

| | 2.4 Groups |
|---|---|
| DIRECTIONAL | |
| Addition | + |
| Omission | + |
| Alternation | - |
| CAUSATIVE | |
| Ergative | - |
| Induced action | - |
| Passive | - |
| Resultative | - |
| ASPECTUAL | |
| *have-være* | + |
| Time | + |
| Measure | - |
| Conative | - |
| Cognate | - |
| *DET* | + |
| *DER* | + |

Comments: The distributional patterns of group verbs differ radically from other motion verbs in almost all respects. In short, these verbs typically occur in expletive constructions, whereas other alternations are rare. In the corpus, most examples demonstrate cases where the moving items are expressed as the argument of the preposition *med*:

(275) *der vrimler med australske skuespillere*
(it crawls with Australian actors)

(276) *i de tidlige morgentimer vrimler græsplænen med spurveunger*
(in the early morning hours the lawn is teeming with young sparrows)

(277) *det vrimlede med unge mænd* (it was teeming with young guys)

Apart from these alternations, group verbs are indeed very restricted; although there are cases of addition of a directional phrase, as in:

(278) *gæsterne myldrede ind* ('the guests teemed in')

**Concluding remarks on the determination hypothesis**

Although each motion verb at the most detailed level of analysis proves to have its own individual syntactic potential, several general patterns of transparency emerge from the alternation tests described above. The following tables sum up the results.

(279) *Direction verbs*

| | | |
|---|---|---|
| Semantic Assignment: | +Mo, +Di | |
| Syntactic potential: | Directional alternations: | + |
| | Causative alternations: | - |
| | Aspectual alternations: | - |
| | *Det* alternations: | - |
| | *Der* alternations: | + |

As can be seen, direction verbs are rather restricted with respect to alternations, except for directional alternations, and *der* alternations.

(280) *Body-by-vehicle verbs*

| | | |
|---|---|---|
| Semantic Assignment: | +Mo, +Ma, +Procon, +Ve | |
| Syntactic potential: | Directional alternations: | + |
| | Causative alternations: | + |
| | Aspectual alternations: | + |
| | *Det* alternations: | - |
| | *Der* alternations: | + |

(281) *Body verbs*

| | | |
|---|---|---|
| Semantic Assignment: | +Mo, +Ma, +Procon | |
| Syntactic potential: | Directional alternations: | + |
| | Causative alternations: | + |
| | Aspectual alternations: | + |
| | *Det* alternations: | - |
| | *Der* alternations: | + |

Body-by-vehicle verbs and body verbs act in a similar way in the sense that both groups are very willing to undergo almost all kinds of alternations. Body-by-vehicle verbs are, however, by far the most flexible ones, for instance verbs from this group can undergo the induced action alternation. As regards the body verbs, it should be noted that the ability to undergo alternations decreases considerably when examining more specific verbs (i.e. verbs that are placed low in the taxonomy). Thus, verbs like *gå* (walk) and *løbe* (run) are found in many kinds of constructions whereas *spadsere* (promenade) and *liste* (tiptoe) are more restricted with respect to syntactic potential.

(282) *Non-body verbs*

| | | |
|---|---|---|
| Semantic Assignment: | +Mo, +Ma | |
| Syntactic potential: | Directional alternations: | only a few |
| | Causative alternations: | only ergatives |
| | Aspectual alternations: | only duratives |
| | *Det* alternations: | only mass verbs |
| | *Der* alternations: | not all verbs |

Non-body verbs demonstrate a syntactic behaviour which differs radically from other motion verbs, even if the group seen in isolation is rather heterogeneous. Many verbs in this group cannot occur with a directional modifier, (**flaget vajer hen til huset* ('the flag waves over to the house')) an ability that almost all verbs from the other groups have. It is in this group of verbs that ergative alternations are found (*jeg snurrer hjulet rundt, hjulet snurrer* (I turn the wheel around, the wheel turns)); only durative time modifiers are found with these verbs (*hjulet snurrede i en halv time* (the wheel turned for half an hour)), and finally many verbs in this group cannot occur in *der* constructions (**der ryster et tog* ('there shakes a train')), a characteristic that is not found with other motion verbs.

(283) *Group verbs*

| | | |
|---|---|---|
| Semantic Assignment: | +Mo, +Ma | |
| Syntactic potential: | Directional alternations: | + |
| | Causative alternations: | - |
| | Aspectual alternations: | + |
| | *Det* alternations: | + |
| | *Der* alternations: | + |

Also group verbs exhibit a characteristic syntactic behaviour which distinguish them from the rest; one of the basic characteristics of this group is the very frequent use of expletive constructions.

To sum up, we can conclude that the taxonomy of motion verbs and the assignment of prototypical semantic features to the verbs are indeed supported by syntactic evidence. In other words, the hypothesis of semantic determination has proved to be worth pursuing since we have now acquired new and relevant information about the correspondences which exist between the two levels of analysis of Danish motion verbs: at least for this semantic class we can observe that the mapping between syntax and semantics is to some extent transparent.

Obviously, more empirical work is required in order to draw more far-reaching conclusions, but as a methodology for further examination of the nature of lexical knowledge the systematic comparison between syntactic alternations and assigned meaning components seems fruitful.

### 5.3.5 Semantic potential of Danish motion verbs

We have now looked at Danish motion verbs from a taxonomical viewpoint considering in the taxonomy the most prototypical meanings of the verbs in question and establishing the set of frame elements to be assigned to each subgroup. We have furthermore on the basis of corpus work examined which syntactic alternations each subgroup from the taxonomy can undergo. These examinations enable us to consider now the *regularities* that may exist in the ways in which Danish motion verbs can undergo meaning changes.

First of all, it is interesting - although not surprising, to see that there is a clear relationship between possible syntactic alternations and possible meaning potential as will be demonstrated in the following. Thus, direction verbs which are very restricted in syntax hardly ever undergo systematic meaning changes, whereas body verbs, body-by-vehicle verbs, and non-body verbs are more elastic in meaning.

Secondly, however, we need to stress that not all the syntactic alternations described in the preceding section trigger meaning changes that should be accounted for by means of different semantic senses in the lexicon. Following the line of Cruse (1991) and Copestake & Briscoe (1995) we therefore wish to distinguish between *sense modulation* and *sense extension*, since this distinction is crucial for the establishment of senses in the dictionary (see Chapter 3). Our working hypothesis is that the minimal set of frame elements that we have established for motion verbs, together with a set of distributional tests, can be applied for the identification of these two kinds of sense change. In other words, if a new sense in the corpus evokes another set of frame elements than the one evoked prototypically (i.e. on the basis of the taxomony) then we are dealing with sense selection which should also be reflected in distribution. Cases of sense selection should definitely be represented in the dictionary as a derived sense since they prove to have translational relevance.

**Delimitations**

Are we then in this thesis trying to deal with all kinds of meaning changes that motion verbs can undergo? The answer is no; several delimitations are set beforehand. First of all we are not dealing with free metaphors since we consider such meaning changes to be too dynamic and world-knowledge dependent to be captured in a computational model and by an apparatus of the dimensions that we are proposing here. Fully conventionalised metaphors on the other hand are partially treated as will be seen below, the *time metaphor* applied in the motion frame for instance is so conventionalised and frequent in the corpus (approx. 5 % of the analysed set of corpus examples of the verb *gå* demonstrated time metaphor use) that we do not hesitate to talk about a specific derived sense that should be represented in the dictionary, as in the case below:

(284) *der vil formentlig gå en til to uger*
('there will most probably walk one to two weeks')
(it will most probably take one or two weeks)

Furthermore, we exclude idiomatic expressions from our work, exemplified below:

(285) *han går under navnet 'Fede Kaj'*
('he walks under the name 'Fat Kaj")

(he goes by the nickname 'Fat Kaj')

(286) *han går til klaver*
('he walks to piano') (he receives piano lessons)

Although idiomatic expressions are of high translational relevance we claim that they should receive a treatment fully independent of the semantic model that we propose here. In short, the meaning of idiomatic expressions cannot be treated by identifying semantic parameters of each constituent in a given expression, rather it should be seen as a multi-word unit only. A similar problem occurs when motion verbs are realised as support verbs to deverbal nominals (see Pedersen 1994), as in:

(287) *firmaet gik efter få måneder i betalingsstandsning*
('the firm after a few months walked into suspension of payment')

Phrasal verbs where the particle has lost its directional sense are also excluded from this work since such constructions in most cases share the same properties as multi-word units: the sense of such verbs cannot be interpreted compositionally, as seen in the examples below and should therefore be treated as specific senses in the dictionary which are not directly related in meaning to the prototypical sense of the verb:

(288) *han gik af da han blev 60*
('he walked off when he was 60')
(he was pensioned when he was 60)

(289) *han gik ud fra at spillet var tabt*
('he walked out from that the game was lost')
(he took for granted that the game was lost)

(290) *det gik pludselig op for mig at der var noget galt*
('it suddenly walked up for me that there was something wrong')
(It suddenly appeared to me that something was wrong)

(291) *jeg var ved at gå til*
('I was about to walk to')(I was cracking up)

In contrast, we do treat phrasal motion verbs where the particle has a directional sense, also in cases where this infers with the meaning of the verb stem as a whole:

(292) *han drejede af uden at blinke*
(he turned off without giving signal)

Last but not least, another delimitation to our work is that we are particularly interested in meaning changes that exhibit some degree of *systematicity*. With systematicity we mean sense changes that are shared between a group of verbs and where we are therefore enabled to *generalise* over groups of verbs.

**Four kinds of systematic polysemy**

From our corpus examinations we have been able to identify basically four kinds of systematic deviation from the prototypical sense of the verbs:

1. *Meaning extension*, where a new semantic dimension is added to the already existing meaning components.
2. *Meaning reduction*, where the prototypical meaning component of a verb is reduced to only part of that meaning.
3. *Co-composition*, where a new related meaning is derived in composition with a non-prototypical argument to the verb[10].
4. *Meaning focusing*, where focus is added to one specific meaning component of the verb frame.

These deviations prove to be highly systematic in the sense that verbs belonging to the same subgroup tend to be sensitive to the same changes in meaning (but there are exceptions to this as will be illustrated below).

Examples of each meaning shift are given in the figures below.

(293) *Meaning extension*:

| Prototypical sense | Extended sense |
|---|---|
| Class description: to move your body in a certain manner | Class description: to go in a certain manner |
| Frame elements: +Mover (anim) +Manner, +Protagonist control | Frame elements: +Mover (anim) +Manner, +Protagonist control +Direction |
| Example: *han kunne gå da han var 10 måneder gammel* (he could walk when he was ten months old) | Example: *han gik hen til skranken* (he went to the counter) |
| Syntactic constraint: none | Syntactic constraint: directional modifier |
| Event type: activity | Event type: accomplishment |

[10]Note that we adopt a more restricted interpretation of the concept than Pustejovsky (1995).

Subgroups from the taxonomy sensitive to this permutation are body verbs, such as *liste* (tiptoe) and *svømme* (swim), body-by-vehicle verbs, such as *cykle* (cycle), and some non-body verbs, such as *trille* (roll).

In a very few cases, body verbs can undergo a similar meaning change without a directional modifier added. Consider the example below where the event type has changed into an accomplishment (*være* is used) even if no directional modifier is used:

(294) *jeg er løbet* ('I am runned')(I'm leaving now)

We cannot even be sure that this change is marked by an auxiliary, as is seen below:

(295) *jeg går nu*
('I walk now')(I'm leaving now/I'm walking now)

In the last case, disambiguation in relation to the activity sense requires a contextual analysis.

(296) *Meaning reduction*

| Prototypical sense | Reduced sense |
|---|---|
| Class description: to move iteratively up and down or back and forth | Class description: to move suddenly up or down, forward or backward in a single movement |
| Frame elements: Mover (mostly non-hum) + Manner | Frame elements: Mover (mostly nhum) + Manner + Direction |
| Example: *han lod salmiakstangen vippe i munden* (he let the stick of liquorice flip up and down in his mouth | Example: *han lod den umagelige højryggede stol vippe baglæns* (he let the uncomfortable high-backed chair tip backwards) |
| Syntactic constraint: none | Syntactic constraint: directional particle - phrasal verb construction |
| Event type: activity | Event type: accomplishment |

Subgroups from the taxonomy sensitive to this permutation are back-and-forth verbs, such as *svaje* (sway), and *svinge* (swing), as well as jump verbs such as *hoppe* (hop).

(297) *Co-composition (a)*

| Prototypical sense | Co-compositional sense |
|---|---|
| Class description: Animals that move by means of their wings or vehicles that move by means of a motor | Class description: humans moving by means of a vehicle |
| Frame elements: Mover (animal or vehicle) Manner + Protagonist control | Frame elements: Mover (human) + Manner + Vehicle + Protagonist control |
| Examples: *nu ser jeg bare mågerne flyve ind og ud af solstriben* (now I just see the gulls fly in and out of the sunbeam)<br><br>*bilerne må gerne køre i gaden, men i listetempo* (the cars are allowed to run in the street, but in tiptoe speed) | Examples: *Allan var fløjet til Oslo for kun to dage siden* (Allan had flown to Oslo only two days ago)<br><br>*Folk er efterhånden rædselsslagne for at køre i området* (eventually, people are horror-struck by driving in the area) |
| Selectional restrictions: Mover is an animal that can fly or a vehicle with a motor | Selectional restrictions: Mover is human |

Similar changes of co-composition are seen with verbs that have the ability to express induced action when combined with a human Mover, such as for instance *køre* (run/drive), where we can say *bilen kører* (the car runs), but also *hun kører* (she drives) meaning that both she and the car move. Also some verbs from the non-body group, such as *trille* (roll) and *rulle* (roll) can undergo this meaning change, meaning that you transport yourself in something with wheels.

(298) *Co-composition (b)*

| Prototypical sense | Co-compositional sense |
|---|---|
| Class description: non-human entities moving in a certain manner without protagonist control | Class description: humans moving about casually without aim or goal |
| Frame elements: Mover (non human) +Manner | Frame elements: Mover (human) + Manner + Protagonist Control |
| Example: *de tørrede gedder havde hængt og dasket vinteren over* (the dried pikes had been dangling the whole winter)<br><br>*sodede skyer driver afsted under ildhimlen* (sooty clouds drift away under the burning sky) | Example: *jeg havde god tid til at daske rundt i byen* (I had time enough to saunter around in town)<br><br>*han gik og drev ovre hos dem der sad på flæsket* (he hung around with the top guys) |
| Selectional constraint: Mover is nhum | Selectional constraint: Mover is hum |

Subgroups from the taxonomy sensitive to this permutation are several non-body verbs.

(299) *Co-composition (c)*

| Prototypical sense | Co-compositional sense |
|---|---|
| Class description: to move your body in a certain manner | Class description: time entities passing by |
| Frame elements: Mover (animate) + Manner + Protagonist control | Frame elements: Mover (time entities) + Manner |
| Example: *han kunne gå da han var 10 måneder gammel* (he could walk when he was 10 months old) | Example: *når der er gået 7 dage efter at den sidste pille blev taget, begyndes på en ny pakke* (when 7 days have passed by since the last pill was taken a new package is initiated) |
| Selectional restriction: Mover is human | Selectional restriction: Mover is a time entity |

Subgroups from the taxonomy sensitive to this sense derivation: general body verbs, i.e. body verbs placed high in the hierarchy, excluding verbs like *liste* (tiptoe), and *trave* (trot), for instance. Also the direction verb *komme* (come) allows for this sense change, as in:

(300) *den tid kommer snart, hvor vi skal til at tænke på julegaver.*
('that time soon comes when we are to think about Christmas presents')

We generally rubricate this kind of co-composition under *conventionalised metaphors*. Likewise with the locative interpretation of the same motion verbs when combined with locatives, as we saw in Chapter 4 (examples are repeated here for clarity reasons):

(301) *mit hus kommer først bag den tre etagers bygning du ser derhenne*
('my house comes not-until the three floor building you see over- there')

(302) *hvor går den vej hen ?*
('where walks that road over ?')
(where does that road lead to ?)

(303) *smøgen løber herfra ned til havnen*
(the passage runs from here down to the harbour)

(304) *Meaning focusing (a)*

| Prototypical sense | Focused sense |
|---|---|
| Class description: to move your body in a certain manner | Class description: to move your body in a certain manner within a certain and well-defined process |
| Frame elements: Mover (animated) + Manner + Protagonist control | Frame elements: Mover (animated) + Manner + Protagonist control |
| Example: *Terence havde aldrig lært svømme*<br>(Terence had never learned to swim) | Example: *Han tænkte på om han skulle svømme en tur*<br>('he thought about whether he should swim a tour')<br><br>*Jeg svømmede rundt over stedet, indtil jeg måtte holde op*<br>(I swam about above the place until I had to stop) |
| Syntactic constraint: none | Syntactic constraint: cognate object or particle *rundt* |

(305) *Meaning focusing (b)*

| Prototypical sense | Focused sense |
|---|---|
| Definition: to move about in large groups | Definition: to move (or be) in large groups with focus on location |
| Frame elements: Mover (no restriction) + Manner | Frame elements: Mover (no restriction) + Manner |
| Example: *vagtler, agerhøns og vilde kaniner vrimlede mellem stenene* (quails, partridges, and wild rabbits were running around between the stones) | Example: *Næsgaards jorder kommer atter til at vrimle med blå spejdere* (the fields of Næsgaard will once again be teeming with blue scouts) |
| Syntactic constraint: none | Syntactic constraint: location is mapped to subject |

Subgroups from the taxonomy sensitive to this permutation are group verbs, such as *myldre* (crawl/teem), *sværme* (swarm/teem), and *vrimle* (crawl/teem)).

### General patterns of sense selection

If we look at the frame element assignment in the first three kinds of meaning change (meaning extension, meaning reduction, and co-composition) it becomes clear that these all present deviation in frame element assignment between the prototypical sense and the derived sense. In contrast, meaning focusing does not present such a deviation in frame elements between the two senses described, here we only have a difference in focus.

This leads to the intermediate conclusion that we must in the latter example be dealing with sense modulation, whereas the other three examples represent cases of sense selection. Only cases of sense selection will be treated henceforth, since only these are considered to be cases of 'real' ambiguity.

Drawing a conclusion on the examples given above, the following relation between subgroups in the taxonomy and possible cases of sense selection can be identified:

- Direction verbs are generally very rigid in meaning and hardly candidates to any systematic meaning changes whatsoever.
- Body verbs and body-by-vehicle verbs are very frequent candidates to meaning extension and co-composition, although the general pattern is that the lower in taxonomy (i.e. the more specific in meaning) the less flexible with regard to meaning change. Jump verbs can also undergo meaning reduction.
- In the non-body group, back-and-forth verbs and circular verbs are generally candidates to all kinds of meaning shifts: meaning extension, mean-

ing reduction as well as co-composition. However, back-and-forth verbs which denote very quick movements back and forth (*vibrere* (vibrate), *sitre* (quiver) are very restricted. Likewise with mass verbs which can only undergo meaning extention.

- Group verbs are relatively restricted in meaning, but do undergo meaning extension. They can also change into denoting rather group existence than group motion.

These results relate nicely to Levin & Rappaport Hovav's (forthcoming) analysis (as referred to in Chapter 3), since exactly direction verbs are 'saturated' in meaning from an aspectual viewpoint because they inherently denote a final state of the event and therefore can't be varied, whereas manner verbs (including body verbs, body-by-vehicle verbs, non-body verbs, and group verbs) denote activities (i.e. simple, non-finite events) and are therefore open to variation.

### 5.3.6 Sense selection and distributional tests

According to Cruse, sense selection should be sensitive to distributional tests, so ideally it should now be possible to establish distributional tests that would demonstrate the ambiguity of these three cases.

However, it seems that verbs do not exhibit the same ability to apply to distributional tests as do nouns and adjectives. We believe that this has to do with the fact that it is often impossible to identify synonyms and antonyms to verbs; phenomena which are very often involved in distributional tests, as we also saw in Chapter 3. The zeugma test (see Chapter 2), which explores the fact that two senses of a word cannot be brought into play simultaneously without oddness, therefore seems to be the most reliable test to apply in this context although it is not in our opinion a waterproof test.

Not surprisingly, sense deviations that have different selectional restrictions on its arguments evoke the required oddness when occurring simultaneously:

(306) **pigen og tiden gik hurtigt*
('the girl and the time walked quickly')

(307) **lappen og jeg daskede rundt i nogle timer*
('the patch and I dangled around for a few hours')

Meaning extension, in contrast, does not require another type of argument and is therefore not so willing to undergo the zeugma test. However, if we in the test apply modifiers that are typically seen with each of the two senses we do get a zeugma:

(308) **jeg gik i to timer og ned til stranden*
('I walked for two hours and down to the beach')

The oddness here is due to the fact that the two senses have a different event structure which is exemplified by the two different kinds of modifiers, as we have also seen earlier.

For meaning reduction it seems even harder to establish convincing distributional tests, but again, if we use a durative time modifier which combines only with the prototypical sense together with a directional particle which triggers a reduced sense, we get a zeugma:

(309) **stolen vippede baglæns i flere timer*

### 5.3.7 Exceptions to elasticity in meaning

We have now identified some patterns of elasticity and we have located these patterns to specific sets of verbs. In other words, we believe to have identified some *productive* mechanisms in polysemy; however, the right term to use would be *semi*-productive since there are always exceptions to these.

To take an example, we can look at the verbs denoting relatively slow movements back and forth which should be candidates to the meaning reduction shift. There are 6 verbs in the group: *daske* (flap), *dingle* (dangle), *svaje* (sway), *svinge* (swing), *vaje* (wave), *vippe* (rock), and *vugge* (rock). However, at least one verb in the group, namely *vaje*, is very restricted in use and cannot undergo this meaning shift:

(310) **flaget vajede frem* ('the flag waved forwards').

The same holds for the co-composition based meaning shifts that certain non-body verbs undergo when they combine with a human mover. In the cases of *daske* (flap) we saw that the human mover obtains protagonist control of the movement, even if he moves without aim or goal, as can be seen below:

(311) *Jeg daskede roligt afsted ad de nye spadserestier*
(I slouched quietly along by the new footpaths)

Other verbs from this group do not assign a frame element of protagonist control when combined with a human mover and should therefore not, according to our criteria, be treated as a new sense. An example of this is *dingle* (dangle) in

(312) *de medvirkende måtte hænge og dingle i timevis efter optagelserne*
('the participants had to hang and dangle for hours after the recordings')

(313) *fra lygtepælene dinglede tyve, grever og embedsmænd side om side.*
(from the lamp posts dangled thieves, counts, and officials side by side)

Here the human Mover is dead or in a state where he can definitely not control the movement.

In such cases we would not want to trigger a productive mechanism because we are not dealing with a new sense. One solution to this is to elaborate a blocking mechanism that can be assigned to lexemes that cannot undergo default sense change of a given type. It would be most satisfactory from a linguistic point of view if such an assignment could take place by means of a general mechanism working on the basis of a semantic principle, such as for instance *preemption by synonymy* as described in Briscoe et al. (1995). They give animal nouns as an example of nouns that can typically undergo what they call a *grinding* sense change where they become mass nouns referring to the meat of edible flesh of that animal (e.g. *lamb, haddock, rabbit*)(1995:p.275). But this process is blocked by the existence of a synonymous form as in *pork, pig* and *beef, cow*.

The question is whether similar mechanisms exist in relation to verbs which describe events and not entities. Preemption by synonymy most presumably works for verbs as well at least to some degree, but it proves much harder to establish exactly *when* such a synonym exists since events are generally of a more complex composition than nouns.

We do not at the current stage of our investigations see a clear pattern of preemption in relation to Danish motion verbs and although further research is called for on the topic, we are tempted to conclude that many cases of blocking can only be explained on the basis of idiosyncratic etymological and social conventions.

## 5.4 Summary

In this chapter we have focused on the analysis of Danish motion verbs with the primary aim of identifying the elasticity in meaning that Danish motion verbs demonstrate so that we would be able to treat verb ambiguity in the monolingual analysis dictionary in a satisfactory way.

We started out by looking at our corpora and described the procedure by which the corpus work has been undertaken. On the basis of the corpus analysis, a set of frame elements were identified, namely *Mover, Direction, Manner, Vehicle*, and *Protagonist Control*; likewise a set of possible aspectual values for motion events were established: *activity, accomplishment* and *achievement*, and finally we presented a partial sort hierarchy for encoding selectional restrictions on motion verbs.

Based on these features, we made a taxonomy of the prototypical meanings of Danish motion verbs and, in order to clarify their syntactic potential, each subgroup from the taxonomy was tested against a set of distributional tests. There proved to be a more or less transparent relation between syntactic potential and semantic assignments and we concluded that the hypothesis of semantic determination put forward by Levin was therefore worth pursuing since we acquired new and valuable information about the correspondences which exist between the two levels of analysis of Danish motion verbs, and last but not least: we believed to have found an important tool for supporting the investigation of meaning which would now not have to build on introspection alone.

A description of the regularities of polysemous motion verbs was the last topic treated in this chapter. We performed a grouping of systematic polysemy into four kinds of deviations from the prototypical meanings of a motion verb: *meaning extension, meaning reduction, co-composition*, and *meaning focusing*. However, we concluded that meaning focusing should be interpreted as sense modulation in Cruse's terminology since no new frame elements where evoked, whereas the three former where all cases of sense selection which should be treated as discrete (although related) senses in the dictionary. These assumptions where supported by distributional tests, although it became clear that such tests are not at all trivial to establish when dealing with verb ambiguity.

Not surprisingly, we saw a pattern of redundancy between those verbs that had a small syntactic potential and those that were restricted in meaning: direction verbs were found to be very rigid in meaning, whereas body verbs and body-by-vehicle verbs could undergo both meaning extension and co-composition. Also non-body verbs exhibited elasticity in meaning since they could undergo both meaning extention, meaning reduction and co-composition. These results were partially explained on the basis of Levin & Rappaport Hovav's (forthcoming) analysis, where it is claimed that direction verbs are 'saturated' in meaning from an aspectual viewpoint because they inherently denote a final state of the event and therefore can't be varied, whereas manner verbs denote activities and are therefore open to variation.

Last but not least, we stressed that we can only speak of *semi*-productive mechanisms in polysemy; although there are clear patterns of regularity, the corpora always expose cases of exceptions to these patterns. We claimed tentatively that such exceptions could not be described systematically, but that they were rather to be seen as idiosyncratic cases of etymological and social conventions.

# Chapter 6

# Formalisation and Implementation

## 6.1 Introduction

This chapter contains three main sections:

- a first section (6.2) where we propose a formalisation of the examples of systematic polysemy that we identified in the previous chapter; i.e. meaning extension, meaning reduction, and co-composition. We start out by presenting some basic accounts of the formalism applied, namely an HPSG-inspired one, then we proceed to the actual lexical entries and last but not least we account for the set of lexical rules that we suggest could be applied as a productive device in the lexicon;

- a second section (6.3) where we rudimentarily provide suggestions as to how a frame element analysis can be explored in transfer in relation to lexical divergences;

- and finally a third section (6.4) where we provide documentation of a sample implementation of our system integrated in the Danish analysis grammar LSGRAM-DK, which is written in the ALEP-formalism.

## 6.2 Formalisation in an HSPG-inspired lexicon

### 6.2.1 Standard HPSG and lexical semantics

In the previous chapters we have become acquainted with various grammatical theories which account for the different aspects of lexical semantics and polysemy. Some of these, for instance the Generative lexicon (Pustejovsky 1995) and Conceptual Semantics (Jackendoff 1983, 1990)) strive towards a formalised model whereas others are more descriptive in nature (Levin (1993) among others).

Also within Frame Semantics initiatives have been taken as regards formalisation of lexical information in a typed feature structure formalism (Heid 1997) as depicted below (Heid 1997:p.100) where, for each frame element, a grammatical function as well as a phrase type is assigned:

(314) *TFS representation of Frame Semantics*

$$
\begin{bmatrix}
\text{LEMMA: ``...''} \\
\text{FRAME\_ELEMENT\_GROUP:} \left\langle \begin{bmatrix} \text{FRAME\_ELEMENT} & [...] \\ \text{GRAM\_FUNC} & [...] \\ \text{PHRASE\_TYPE} & [...] \end{bmatrix} \begin{bmatrix} \text{FRAME\_ELEMENT} & [...] \\ \text{GRAM\_FUNC} & [...] \\ \text{PHRASE\_TYPE} & [...] \end{bmatrix} \right\rangle \\
\text{INHERENT CHARACTERISTICS } [...]
\end{bmatrix}
$$

We have chosen for our formalisation to apply the more widely known HPSG notation (see Pollard and Sag (1987) and (1994)) which is a formalism which combines the application of typed feature structures with a general theory of grammar. An encoding in this formalism enables us to test directly the performance of the lexicon in the analysis phase.

HPSG constitutes one of the typed feature structure-based formalisms which in recent years have had most impact within the area of computational linguistics; however, the theory does not focus specifically on lexical semantics and polysemy and therefore several aspects of semantic information are left somewhat unspecified in the standard theory. For instance, there is no commitment in standard HPSG to a specific theory of role assignment and there is no consistent labelling of predication types or of argument classes, rather arguments are labelled with names directly related to the predicate name, thus for *walk* for instance the argument is labelled *walker*, for *give* they are labelled *giver*, *given* and *gift* and so forth.

This neutrality with regards to commitment to a specific lexical-semantic theory opens for the possibility of assigning and further develop the levels of semantic information that one finds appropriate for a particular application.

### 6.2.2 Conflation of syntactic, semantic and pragmatic information in HPSG

As has hopefully become clear from the previous chapters, one of our aims is to account in a practical way for the relations that exist between syntax and semantics in Danish motion verb constructions. In relation to this, HPSG also presents some interesting characteristics, since exactly the *conflation* of syntactic, semantic and pragmatic information (each represented in a specific subsign of the lexical and phrasal signs: CATEGORY, CONTENT and CONTEXT, respectively) is given much attention in this theory[1].

More specifically, two basic elements of the HPSG lexicon are the subcategorisation lists (SUBJECT and COMPLEMENTS) in which what is traditionally called the valency-bound constituents of a predicate are expressed by means of listed elements in accordance with a so-called *obliqueness hierarchy*.

The idea of organising relational functions in an obliqueness hierarchy stems from some observations made by among others Keenan and Comrie (cited by Pollard & Sag (1987)). They claim that for instance in the case of relativisation, any language allowing relativisation of a particular relational function in the hierarchy will also allow relativisation of all functions to the left of that relational function. To give an example, a language that allows for a sentence like:

(315) *the man [who Kim likes _]*

would also allow for a sentence like:

(316) *the man [who _ likes Kim]*

due to the fact that they obey the restrictions of the following functional hierarchy:

(317) SUBJECT → DIRECT OBJECT → INDIRECT OBJECT → OBLIQUES → GENITIVE → OBJECTS OF COMPARISON

[1] This is to be seen in contrast to transformation-based theories.

The hierarchy is represented in the subcategorisation lists by letting less oblique constituents precede more oblique constituents. The subject which is the least oblique element has its own list, SUBJECT.

Structure sharing indexes link elements on the subcategorisation lists with semantic and pragmatic chunks of information. Such a conflation of highly structured information expressed directly in the lexical sign enables the linguist to express information of a rather complex nature in the lexicon, aspects which in several other formalism must be expressed in grammar. In other words, HPSG is a well-known example of a *lexicalist theory* which (in theory) leaves for the grammar specification only some very basic linguistic principles.

### 6.2.3 Frame Semantics in the lexical sign

The Frame Semantic level of analysis relates rather directly to what in standard HPSG is labelled the *PSOA* (parametrised state of affairs) which constitutes a part of the CONTENT sign (henceforth CONT), as depicted in the attribute-value matrix (AVM) below[2]:

(318) *The content sub-sign*

$$\begin{bmatrix}\text{SYNSEM} & {}_{\mathit{tsynsem}}\begin{bmatrix}\text{CONT} & {}_{\mathit{tcont}}\begin{bmatrix}\text{PSOA}\end{bmatrix}\end{bmatrix}\end{bmatrix}$$

Since neither predicate name labelling (as in standard HPSG: *giver, given* etc.) nor frame elements can be generalised over easily - for instance we cannot generalise over deep subjects as a whole, we suggest to divide PSOA into a general argument assignment part (ARG_PSOA) where an intermediate indexing of arguments is applied (in the line of Badia and Colominas 1995) and a frame element psoa (FE_PSOA) where the frame element assignments which are specific for each semantic frame are assigned. Such a division is depicted below (for abbreviation purposes types are omitted henceforth, except for types on ARG_PSOA and FE_PSOA):

[2]Some clarifying notes on data types and inheritance which will be exemplified during this chapter: in accordance with standard HPSG, the basic data type of our lexical entries is typed feature structures; consequently, each feature structure is assigned a type (placed as a subscript to the brackets enclosing that feature). Each feature consists of an attribute (CAPITAL LETTERS) and a value (*emphasised*), and each value can be either of the form atomic, a typed feature structure, a list or a set. Furthermore, we presuppose orthogonal multiple inheritance hierarchies in that we make use of partitions between parental nodes in order to avoid inheritance of conflicting information. Types themselves are specified by means of feature declarations which list the attributes which a particular type can have and the type of their values.

(319) *The PSOA subsign*

$$\begin{bmatrix} \text{CONT} & \begin{bmatrix} \text{PSOA} & \begin{bmatrix} \text{ARG\_PSOA} & {}_{\textit{arg12\_psoa}}\begin{bmatrix} \text{ARG1} & \text{(CONTENT)} \\ \text{ARG2} & \text{(CONTENT)} \end{bmatrix} \\ \text{FE\_PSOA} & {}_{\textit{tmotion}}\begin{bmatrix} \text{MOVER} & \text{(CONTENT)} \\ \text{MANNER} & \text{(an atomic value)} \\ \text{VEHICLE} & \text{(CONTENT or an atomic value)} \\ \text{PROCON} & \text{(CONTENT or an atomic value)} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Where relevant, arguments and frame elements are structure shared with the corresponding complements on the subcategorisation lists (structure sharing indexes are marked with numbers in boxes). In some cases, more than one frame element can be structure-shared with the same complement. Frame elements that are exclusively realised as inherent features are assigned atomic values. This constitutes an important difference between ARG_PSOA and FE_PSOA, since arguments in ARG_PSOA are always CONTENTs and not atomic values. Thus, the so-called 'shadow-arguments' (Pustejovsky) or 'inherent' arguments referring to arguments which are not expressed in syntax are not expressed in the ARG_PSOA subsign, as depicted below in the case of the vehicle verb *cykle* (cycle):

(320) *cykle* (cycle)

$$\begin{bmatrix} \text{SYNSEM} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{AUX} & \textit{have} \end{bmatrix} \\ \text{SUBJ} & \langle \text{NP}\boxed{1} \rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \text{PSOA} & \begin{bmatrix} \text{ARG\_PSOA} & {}_{\textit{arg1\_psoa}}\begin{bmatrix} \text{RELN} & \textit{cykle} \\ \text{ARG1} & \boxed{1} \end{bmatrix} \\ \text{FE\_PSOA} & {}_{\textit{tbodybyvehicle}}\begin{bmatrix} \text{MOVER} & \boxed{1} \\ \text{MANNER} & \textit{by\_vehicle} \\ \text{VEHICLE} & \textit{bike} \\ \text{PROCON} & \boxed{1} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

### 6.2.4 The treatment of directionals in the lexical sign

As regards directional particles and directional prepositional phrases we consider these to be weakly bound when combined with a motion verb. This means that they occur on the COMPS list but that they have a special semi-adjunct status as marked in ARG_PSOA by the attribute ARG_ADJUNCT. Contrary to strongly bound prepositions (of which the preposition *om* in *kredse om* (circle around) is an example) weakly bound prepositions cannot be predicted uniquely in the verb entry and therefore do not occur as a restriction on the COMPS list since they usually comprise a set of prepositions. Instead, the set of possible directional prepositions are assigned as restrictions to the DIRPP constituent.

As is seen below, directional particles which are not followed by a PP also have ARG_ADJUNCT status as well as a frame element status:

(321) *gå* (walk)

$$\left[\begin{array}{ll} \text{SYNSEM} & \left[\begin{array}{ll} \text{CAT} & \left[\begin{array}{l} \text{HEAD}\ \left[\text{AUX}\quad \textit{være}\right] \\ \text{SUBJ}\ \langle \text{NP}\boxed{1}\rangle \\ \text{COMPS}\ \langle\text{DIRPART}\boxed{2}\rangle \end{array}\right] \\ \text{CONT} & \left[\begin{array}{ll} \text{PSOA} & \left[\begin{array}{ll} \text{ARG\_PSOA} & \left[\begin{array}{ll} \text{REL} & \textit{gå} \\ \text{ARG1} & \boxed{1} \\ \text{ARG\_ADJUNCT} & \boxed{2}\end{array}\right] \\ \text{FE\_PSOA} & {}_{\textit{twalk}}\left[\begin{array}{ll} \text{MOVER} & \boxed{1} \\ \text{MANNER} & \textit{on\_foot} \\ \text{DIRECTION} & \boxed{2} \\ \text{PROCON} & \boxed{1}\end{array}\right]\end{array}\right]\end{array}\right]\end{array}\right]\end{array}\right]$$

The entry applies in an example like the following:

(322) *han er gået ud*
('he is walked out')
(he has gone out)

In this respect our analysis may differ from several other accounts of phrasal verbs which would lexicalise the particle as part of the lexeme (see for instance Weilgaard 1997). However, only in cases where the particle has lost its concrete sense would we suggest full lexicalisation since in such cases there seems to be no systematic predictability wrt. to the meaning change that takes place, as seen below:

(323) *han gik af da han blev 60*
('he walked off when he was 60')
(he was pensioned when he was 60)

When directional particles are combined with a prepositional phrase, as in *han gik hen til huset* (he went over to the house), the particle *in combination with* the PP (labelled PARTPP) is considered to be the ARG_ADJUNCT.

### 6.2.5 Typing of ARG_PSOA and FE_PSOA

As regards the *typing* of ARG_PSOA we adopt a simple subtyping along the dimensions of *arity* in the following fashion (for simplicity reasons we only give the ARG_PSOA assignments relevant to motion events):

(324) *The typing for ARG_PSOA*

```
TYPE: arg1_psoa
ATTRIBUTES: ARG1 = type[content[]].

TYPE:arg12_psoa
ATTRIBUTES: ARG1 = type[content[]],
ARG2 = type[content[]].

TYPE:arg1_adj_psoa
ATTRIBUTES: ARG1 = type[content[]],
ARG_ADJ = type[content[]].
```

FE_PSOA on the other hand is typed according to the taxonomy of motion verbs established in Chapter 5. The taxonomy is directly converted into a type hierarchy with the frame elements being the attributes assigned to each type (in order not confuse types with frame elements types for FE_PSOA are initiated with a t):

(325) *Type hierarchy for FE_PSOA*

```
TYPE: tfe_psoa
ATTRIBUTES: none
SUBTYPES: tmotion, (..tperception, tcommunication..)

TYPE: tmotion
ATTRIBUTES: MOVER = type[content [ ]]
SUBTYPES: tdirection, tmanner

TYPE: tdirection
ATTRIBUTES: DIRECTION = type[content [ ]] ; atom
(direction can either be a composite data structure or an atomic
data structure)

TYPE: tmanner
ATTRIBUTES:  MANNER = atom
SUBTYPES: tbody, tbody_by_vehicle, tnon_body, tgroups
```

```
TYPE: tbody
ATTRIBUTES: PROCON = type[content [ ]]
SUBTYPES:  tbody_part, twhole_body

TYPE: tbody_by_vehicle
ATTRIBUTES: PROCON = type[content [ ]]
            VEHICLE = type[content[ ]] ; atom

TYPE: tbody_part
ATTRIBUTES: none
SUBTYPES: tfoot, totherparts

TYPE: tfoot
ATTRIBUTES: none
SUBTYPES: twalk, trun, tjump

TYPE: twalk
ATTRIBUTES: none
SUBTYPES: tsound, tsurroundings, tspeed,

TYPE: trun
ATTRIBUTES: none
SUBTYPES: tspeedr

TYPE: tspeedr
ATTRIBUTES: none

TYPE: tjump
ATTRIBUTES: none

TYPE: tsound
ATTRIBUTES: none

TYPE: tsurroundings
ATTRIBUTES: none

TYPE: tspeed
ATTRIBUTES: none

TYPE: twhole_body
ATTRIBUTES: none
SUBTYPES: tdance, tspeedwb, taimless

TYPE: tdance
ATTRIBUTES: none
```

```
TYPE: tspeedwb
ATTRIBUTES: none

TYPE: taimless
ATTRIBUTES: none

TYPE: tnonbody
ATTRIBUTES: none
SUBTYPES: tback_forth, tcircular, tmass

TYPE: tgroups
ATTRIBUTES: none
```

Frame elements can now be inherited via the type hierarchy; in the case of *trampe* (stomp), for instance, which is a body part verb, the frame element Procon (Protagonist control) is assigned via its supertype tbody, whereas Manner and Mover are assigned via the supertypes of tbody, namely tmanner and tmotion.

As can be seen, subtypes are assigned at any depth even if no new frame elements are assigned. This enables us to refer to subgroups of verbs even if they are not unique wrt. their attributes.

### 6.2.6 A simplistic event structure analysis in the lexical sign

In Chapter 5 we decided upon a simplistic event structure analysis, comprising three aspectual types for motion events, namely *activity, achievement* and *accomplishment*. In order to incorporate this information we need to further subdivide PSOA since we are now dealing with a new type of PSOA:ASP_PSOA (aspect psoa) (see Theofilidis et al. 1994 for a similar treatment). ASP_PSOA contains the attribute ASPECT which in relation to motion events can take each of the three types as its value, as illustrated below:

(326) *The ASP_PSOA*

$$\left[\text{PSOA}\left[\text{ASP\_PSOA} \mid \text{ASPECT} \quad \textit{activity} \wedge \textit{achieve} \wedge \textit{accompl}\right]\right]$$

For a verb like *ankomme* this would give the following PSOA representation:

(327) *ankomme* (arrive)

$$\left[\text{CONT}\left[\text{PSOA}\left[\begin{array}{l}\text{ARG\_PSOA}\begin{bmatrix}\text{REL } \textit{ankomme} \\ \text{ARG1} \quad \boxed{1}\end{bmatrix} \\ \text{ASP\_PSOA } |\text{ASPECT } \textit{achieve} \\ \text{FE\_PSOA}\begin{bmatrix}\text{MOVER } \boxed{1} \\ \text{DIRECTION} \quad \textit{goal}\end{bmatrix}_{\textit{tgoal}}\end{array}\right]\right]\right]$$

Note that there is a relation between aspectual value and the auxiliary which is expressed as a head feature (AUX) in the CAT subsign; achievements and accomplishments which are intransitive require *være* (examples are *ankomme* (arrive) and *tage* (go)) whereas transitive constructions expressing accomplishments require *have* (examples are *krydse* (cross) and *passere* (pass)). Activities on the other hand require *have* (e.g. *svømme* (swim)).

### 6.2.7 Selectional restrictions in the lexical sign

Semantic types on nouns (eg. *concrete*) are assigned in the content part of nouns as values to a relation feature. The feature functions as a selectional restriction on the subcategorisation lists in the lexical entry for the verb, following the abbreviated notation shown below:

(328) *REL_SORT as selectional restriction*

$$\left[\text{SUBJ } \langle \text{NP}\left[\text{REL\_SORT } \textit{concrete}\right]\rangle\right]$$

Note that we presuppose a preference mechanism of some kind which ensures that type errors on relation sort do not result in a failure of the parse. If no specific preference mechanism is available in the system this can be done by elaborating default lexical rules with no relation sort restrictions. Such rules can fire in cases where more specific rules fail to fire due to a type mismatching between the REL_SORT of the actual complement and the REL_SORT expected by the head (see also Chapter 2 for a brief introduction to preference systems).

To give an example where such a default rule would come into action, we can mention again the examples from Chapter 5 with the verb *dingle* (dangle):

(329) *de medvirkende måtte hænge og dingle i timevis efter optagelserne*
('the participants had to hang and dangle for hours after the recordings')

(330) *fra lygtepælene dinglede tyve, grever og embedsmænd side om side.*
(from the lamp posts dangled thieves, counts, and officials side by side)

These examples are not to trigger a sense change device since no new senses of *dingle* are evoked. On the other hand, in our lexical entry for *dingle* the prototypical relation sort is encoded as a selectional restriction (REL_SORT = inanimate). This doesn't match the arguments in the examples above which are of the relation sort animate. Such exceptions to selectional restrictions which do however not trigger new senses are so frequent in our corpus that they cannot be ignored. Actually for some verbs the non-prototypical arguments are just as frequent as the prototypical arguments, for instance in the case of *strømme* (pour) which prototypically has liquids as its argument as seen here:

(331) *blodet strømmede ud af næsen* (the blood poured out of the nose)

the corpus shows a very frequent use of non-liquids as arguments, as in the examples below:

(332) *bidragene begyndte at strømme ind*
(the contributions began to pour in)

(333) *folk begyndte at strømme ind*
(people started pouring in)

### 6.2.8 The lexical sign as a whole

In the following AVM it is illustrated what our lexical entry looks like comprising all the levels of analysis presented in the sections above (parentheses indicate optionality):

(334) *ankomme* (arrive)

$$\begin{bmatrix} \text{SYNSEM} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} \begin{bmatrix}\text{AUX} & \textit{være}\end{bmatrix} \\ \text{SUBJ } \langle \text{NP}\begin{bmatrix}\text{REL\_SORT } \textit{concrete}\end{bmatrix}\boxed{1}\rangle \\ \text{COMPS } \langle\ (\text{DIRPP } \boxed{2})\ \rangle \end{bmatrix} \\ \text{CONT} & \begin{bmatrix}\text{PSOA} & \begin{bmatrix} \text{ARG\_PSOA} & \begin{bmatrix}\text{REL} & \textit{ankomme} \\ \text{ARG1} & \boxed{1} \\ (\text{ARG\_ADJUNCT} & \boxed{2}) \end{bmatrix} \\ \text{ASP\_PSOA} \mid \text{ASPECT } \textit{achieve} & \\ \text{FE\_PSOA}_{\textit{tgoal}} & \begin{bmatrix}\text{MOVER } \boxed{1} & \\ \text{DIRECTION} & \boxed{2} \wedge \textit{goal}\end{bmatrix}\end{bmatrix}\end{bmatrix}\end{bmatrix}\end{bmatrix}$$

The entry accounts for the verb *ankomme* and would apply to both phrases below due to the optional pp-complement:

(335) *han ankom ved 8-tiden* (he arrived at around 8)

(336) *han ankom til stationen* (he arrived to the station)

Note that the frame element DIRECTION is not optional. Irrespective of the presence of a directional modifier the frame element is evoked because of inherent characteristics of the verb.

### 6.2.9 Productive devices in the lexicon

In order to implement the systematic sense changes that we accounted for in Chapter 5, we need to make some choices wrt. the lexicon. Our aim is to elaborate some handy mechanism in which derived senses can be produced on the basis of prototypical entries in a convenient fashion.

Productive devices in the lexicon can be expressed in at least two ways:

- either by means of underspecification, defining then lexical entries as open to contextual specification of their properties on the basis of constraints in the type hierarchy;
- or by means of an external module of lexical rules operating on full lexical entries and producing from these another set of full lexical entries.

Lexical rules have been exposed to some criticism due to their somewhat problematic status from a theoretical point of view, since they permit an actual violation of the type assignments and thereby to some extent mess up the monotonicity of the type hierarchy (see for instance Schmidt 1994 and Bredenkamp, Markantonatou & Sadler 1996). On the other hand, productive devices expressed by underspecification require that there exists a uniform relation between the underspecified entry and the specific contextual interpretations; a relation that in several cases would require sophisticated principles of inheritance, such as default inheritance.

We claim that sense changes do not always show such uniform vertical relations[3] and we are therefore most inclined to embrace the suggestion put forward by Copestake and Briscoe (1995). In the line of Cruse (1991), they argue that it makes sense to distinguish between two different kinds of systematic polysemy,

[3]One sense change that would however be a good candidate to the underspecification technique is meaning extension.

namely vagueness and ambiguity (in Cruse's terminology sense modulation and sense selection). Vagueness is in their opinion best treated by means of under-specification in the lexical entry wrt. the specific usage of the words whereas ambiguity should be handled by means of lexical rules.

Let us look at two examples in their paper which illustrate this differentiation: the English word *reel* is ambiguous wrt. to being either a *fishing reel* or a *film reel*. In a specific context one of these senses will be selected, but this ambiguity (or rather vagueness) can neatly be expressed by means of underspecification in the lexical entry. In contrast, the English word *lamb* which can refer either to a living animal or to a dish should rather be treated as two predictably related senses expressed by means of lexical rules. Here the method of underspecification is infeasible since an underspecified lexical entry would in such a case be far too general.

As we have already put forward we are in this context primarily interested in cases of sense selection, i.e. real ambiguity, and therefore we shall only treat these in this chapter. In other words, meaning extension, meaning reduction and co-composition are therefore candidates for a treatment by means of lexical rules. In the following we shall primarily refer to these as *sense change rules*.

A sense change rule can be described formally as depicted below:

(337) *Sense change rule*

$$\left[\begin{array}{l}\mathrm{I}\ \left[\mathrm{LEXSIGN}\right]\\ \mathrm{O}\ \left[\mathrm{LEXSIGN}\right]\end{array}\right]_{\mathit{tsensechangerule}}$$

If a lexical entry matches the input part of the lexical rule, given by the path I, the particular rule will be applied, resulting in an instantiated output of the rule, given by the path O.

### 6.2.10 Sense change rules for meaning extension, meaning reduction and co-composition

In the subsections below we give examples of rules which cover the sense changes that we presented in Chapter 5.3.5.

#### Meaning extension

The rule below takes all lexical entries of type tbody that unify with the input part of the rule and generates an output lexical entry which has an extended

GO-meaning which we assign the type *tbodydir*. Note that the Manner element must be present in the input sign but that it is unrestricted wrt. its value. This enables us to generalise over all kinds of tbody verbs. At the syntactic level, the output lexical entry is assigned an extra complement, namely a directional prepositional phrase; at the semantic level, an extra frame element is assigned, namely Direction. Furthermore, Direction is structure shared with the complement on the complement list. Finally, the event type is changed from an activity into an accomplishment which is also reflected in the head feature AUX which changes from *have* to *være*.

As can be seen below, we structure share the subject with ARG1 in the derived sense, even if according to the Unaccusative Hypothesis the subject is to be interpreted as an unaccusative construction and thereby as an internal argument (ARG2), in Chomsky's terms. As we mentioned in Chapter 5, we prefer not to apply this analysis but rather to stick to a model where aspects of unaccusativity are reflected in aspect (telic vs. atelic) and auxiliary (*have* vs. *være*) as well as in the feature Procon.

(338) *A MOVE-GO sense extension rule*

$$
\textit{tMOVEGOrule}\begin{bmatrix}
\text{I} & \begin{bmatrix} \text{SYNSEM} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{AUX} & \textit{have} \end{bmatrix} \\ \text{SUBJ} & \langle \text{NP}\begin{bmatrix} \text{REL\_SORT} & \textit{animate} \end{bmatrix}\boxed{1}\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \text{PSOA} & \begin{bmatrix} \text{ARG\_PSOA} & \begin{bmatrix} \text{ARG1} & \boxed{1} \end{bmatrix} \\ \text{ASP\_PSOA} \mid \text{ASPECT} & \textit{activity} \\ \text{FE\_PSOA} & {}_{\textit{tbody}}\begin{bmatrix} \text{MOVER} & \boxed{1} \\ \text{MANNER} & \_ \\ \text{PROCON} & \boxed{1} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\
\text{O} & \begin{bmatrix} \text{SYNSEM} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{AUX} & \textit{være} \end{bmatrix} \\ \text{SUBJ} & \langle \text{NP}\begin{bmatrix} \text{REL\_SORT} & \textit{animate} \end{bmatrix}\boxed{1}\rangle \\ \text{COMPS} & \langle \text{DIRPP}\boxed{2}\rangle \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \text{PSOA} & \begin{bmatrix} \text{ARG\_PSOA} & \begin{bmatrix} \text{ARG1} & \boxed{1} \\ \text{ARG\_ADJ} & \boxed{2} \end{bmatrix} \\ \text{ASP\_PSOA} \mid \text{ASPECT} & \textit{accompl} \\ \text{FE\_PSOA} & {}_{\textit{tbodydir}}\begin{bmatrix} \text{MOVER} & \boxed{1} \\ \text{MANNER} & \_ \\ \text{DIRECTION} & \boxed{2} \\ \text{PROCON} & \boxed{1} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}
\end{bmatrix}
$$

Similar rules are foreseen for verbs of type tbody_by_vehicle, tnon_body and tgroup.

### Meaning reduction

The meaning reduction rule is given below:

(339) *A reduction sense change rule*

$$
{\small\textit{treduction}}\left[\begin{array}{ll}
\text{I} & \left[\text{SYNSEM}\left[\begin{array}{ll}
\text{CAT} & \left[\begin{array}{l}
\text{HEAD}\left[\text{AUX} \quad \textit{have}\right]\\
\text{SUBJ } \langle \text{NP}\left[\text{REL\_SORT } \textit{inanimate}\right]\boxed{1}\rangle\\
\text{COMPS } \langle\rangle
\end{array}\right]\\
\text{CONT} & \left[\text{PSOA}\left[\begin{array}{ll}
\text{ARG\_PSOA} & \left[\text{ARG1} \quad \boxed{1}\right]\\
\text{ASP\_PSOA} \mid \text{ASPECT } \textit{activity} & \\
\underset{\textit{tback\_forth}}{\text{FE\_PSOA}} & \left[\begin{array}{ll}\text{MOVER} & \boxed{1}\\ \text{MANNER} & \_\end{array}\right]
\end{array}\right]\right]
\end{array}\right]\right]\\
\text{O} & \left[\text{SYNSEM}\left[\begin{array}{ll}
\text{CAT} & \left[\begin{array}{l}
\text{HEAD}\left[\text{AUX} \quad \textit{være}\right]\\
\text{SUBJ } \langle \text{NP}\left[\text{REL\_SORT } \textit{inanimate}\right]\boxed{1}\rangle\\
\text{COMPS } \langle \text{DIRPART}\boxed{2}\rangle
\end{array}\right]\\
\text{CONT} & \left[\text{PSOA}\left[\begin{array}{ll}
\text{ARG\_PSOA} & \left[\begin{array}{ll}\text{ARG1} & \boxed{1}\\ \text{ARG\_ADJ} & \boxed{2}\end{array}\right]\\
\text{ASP\_PSOA} \mid \text{ASPECT } \textit{accompl} & \\
\underset{\textit{tsuddenmove}}{\text{FE\_PSOA}} & \left[\begin{array}{ll}\text{MOVER} & \boxed{1}\\ \text{MANNER} & \_\\ \text{DIRECTION} & \boxed{2}\end{array}\right]
\end{array}\right]\right]
\end{array}\right]\right]
\end{array}\right]
$$

The rule takes verbs of back-and-forth movement and changes them into verbs of a single sudden movement by adding a directional particle. Again the event type and the auxiliary change correspondingly. A new type is generated, namely tsuddenmove.

### Co-composition

Co-composition can be treated in a similar way, by relating the prototypical sense and the co-compositioned sense to each other by means of a lexical rule. Note, however, that this deviates from the generative way in which Pustejovsky (1995) produces co-compositioned senses. In his work co-compositioned senses are produced in a fully generative way mainly on the basis of the Qualia Structure of the argument and the verb in composition. Thus, in his interpretation, multiple elements within the phrase behave as functors, whereas in our formalisation we keep the verb as the functor. This means that in our formalisation the full meaning of the new sense is not predicted since we have no Qualia Structure on which to base such calculations (recall our discussion in Chapter 3). Some calculations could be done on the basis of the frame element group, though, for instance many non-body verbs tend to get protagonist control when combined with a human mover. This fact influences other elements of meaning which however are not accounted for in our model.

The sense change rule below illustrates the sense change that several non-body verbs undergo when combined with a human mover.

(340) *A co-composition rule (a)*

$$
{}_{\textit{tCOCOMPrule}}\begin{bmatrix}
\text{I} & \begin{bmatrix} \text{SYNSEM} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix}\text{AUX} & \textit{have}\end{bmatrix} \\ \text{SUBJ} & \langle \text{NP}\begin{bmatrix}\text{REL\_SORT} & \textit{inanimate}\end{bmatrix}\boxed{1}\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \text{PSOA} & \begin{bmatrix} \text{ARG\_SPOA} & \begin{bmatrix}\text{ARG1} & \boxed{1}\end{bmatrix} \\ \text{ASP\_PSOA} \mid \text{ASPECT} & \textit{activity} \\ \underset{\textit{tnon\_body}}{\text{FE\_PSOA}} & \begin{bmatrix}\text{MOVER} & \boxed{1} \\ \text{MANNER} & \_\end{bmatrix} \end{bmatrix}\end{bmatrix}\end{bmatrix}\end{bmatrix} \\
\text{O} & \begin{bmatrix} \text{SYNSEM} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix}\text{AUX} & \textit{have}\end{bmatrix} \\ \text{SUBJ} & \langle \text{NP}\begin{bmatrix}\text{REL\_SORT} & \textit{human}\end{bmatrix}\boxed{1}\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \text{PSOA} & \begin{bmatrix} \text{ARG\_SPOA} & \begin{bmatrix}\text{ARG1} & \boxed{1}\end{bmatrix} \\ \text{ASP\_PSOA} \mid \text{ASPECT} & \textit{activity} \\ \underset{\textit{taimless}}{\text{FE\_PSOA}} & \begin{bmatrix}\text{MOVER} & \boxed{1} \\ \text{MANNER} & \_ \\ \text{PROCON} & \boxed{1}\end{bmatrix} \end{bmatrix}\end{bmatrix}\end{bmatrix}\end{bmatrix}
\end{bmatrix}
$$

The rule takes non-body verbs and produces lexical entries of the type taimless which are assigned the additional feature PROCON (Protagonist Control). Actually the derived senses are so close to the verbs that are prototypically assigned the type taimless that they can almost function as synonyms to these:

(341) *jeg havde god tid til at daske/drive/tulre rundt i byen*
(I had plenty of time to dangle/drift/blunder around in town)

A lexical rule for co-composition with animal verbs is depicted below:

(342) *A co-composition rule (b)*

$$
\left[\begin{array}{ll}
\text{I} & \left[\begin{array}{ll}\text{SYNSEM} & \left[\begin{array}{ll}
\text{CAT} & \left[\begin{array}{l}\text{HEAD}\left[\text{AUX} \quad \textit{have}\right]\\ \text{SUBJ}\ \langle \text{NP}\left[\text{REL\_SORT}\ \textit{fly\_animal}\right]\boxed{1}\rangle\\ \text{COMPS}\ \langle\rangle\end{array}\right]\\
\text{CONT} & \left[\begin{array}{ll}\text{PSOA} & \left[\begin{array}{ll}\text{ARG\_PSOA} & \left[\text{ARG1}\quad\boxed{1}\right]\\ \text{ASP\_PSOA} & \mid\ \text{ASPECT}\ \textit{activity}\\ \text{FE\_PSOA} & \left[\begin{array}{ll}\text{MOVER} & \boxed{1}\\ \text{MANNER} & \textit{inair}\\ \text{PROCON} & \boxed{1}\end{array}\right]_{\textit{totherparts}}\end{array}\right]\end{array}\right]
\end{array}\right]\end{array}\right]\\
\text{O} & \left[\begin{array}{ll}\text{SYNSEM} & \left[\begin{array}{ll}
\text{CAT} & \left[\begin{array}{l}\text{HEAD}\left[\text{AUX} \quad \textit{have}\right]\\ \text{SUBJ}\ \langle \text{NP}\left[\text{REL\_SORT}\ \textit{human}\right]\boxed{1}\rangle\\ \text{COMPS}\ \langle\rangle\end{array}\right]\\
\text{CONT} & \left[\begin{array}{ll}\text{PSOA} & \left[\begin{array}{ll}\text{ARG\_SPOA} & \left[\text{ARG1}\quad\boxed{1}\right]\\ \text{ASP\_PSOA} & \mid\ \text{ASPECT}\ \textit{activity}\\ \text{FE\_PSOA} & \left[\begin{array}{ll}\text{MOVER} & \boxed{1}\\ \text{MANNER} & \textit{inair}\\ \text{VEHICLE} & \textit{aircraft}\\ \text{PROCON} & \boxed{1}\end{array}\right]_{\textit{tbodybyvehicle}}\end{array}\right]\end{array}\right]
\end{array}\right]\end{array}\right]
\end{array}\right]_{\textit{tCOCOMPrule}}
$$

This rule covers those other_part verbs which prototypically have animals as subjects, such as *flyve* (fly) which changes meaning when combined with a human argument, involving then a vehicle. Similar rules are foreseen for other verbs which allow for this kind of change, like *rulle* (roll) and *trille* (roll).

Note that a derived sense may again trigger a sense change rule. For instance, the derived sense of *flyve* matches the input part of a tMOVEGOrule so this rule is applied to the output of tCOCCOMP triggering again an extended direction meaning as expressed in:

(343) *Allan var fløjet til Oslo for en dag*

(Allan had flown to Oslo for a day)

Finally, we show below a co-composition rule which takes care of the conventionalised time metaphor which is so frequent with body verbs:

(344) *A co-composition rule (c)*

$$
{}_{\textit{tCOCOMPrule}}\begin{bmatrix}
\text{I} & \begin{bmatrix} \text{SYNSEM} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{SUBJ } \langle \text{NP}\begin{bmatrix}\text{REL\_SORT } \textit{animate}\end{bmatrix}\boxed{1}\rangle \\ \text{COMPS } \langle\rangle \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \text{PSOA} & \begin{bmatrix} \text{ARG\_PSOA}\begin{bmatrix}\text{ARG1} & \boxed{1}\end{bmatrix} \\ \text{ASP\_PSOA} \mid \text{ASPECT } \textit{activity} \\ \text{FE\_PSOA} \; {}_{\textit{tbody}}\begin{bmatrix} \text{MOVER } \boxed{1} \\ \text{MANNER } \_ \\ \text{PROCON } \boxed{1} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\
\text{O} & \begin{bmatrix} \text{SYNSEM} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{SUBJ } \langle \text{NP}\begin{bmatrix}\text{REL\_SORT } \textit{time}\end{bmatrix}\boxed{1}\rangle \\ \text{COMPS } \langle\rangle \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \text{PSOA} & \begin{bmatrix} \text{ARG\_PSOA}\begin{bmatrix}\text{ARG1} & \boxed{1}\end{bmatrix} \\ \text{ASP\_PSOA} \mid \text{ASPECT } \_ \\ \text{FE\_PSOA} \; {}_{\textit{ttime}}\begin{bmatrix} \text{MOVER } \boxed{1} \\ \text{MANNER } \textit{time} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}
\end{bmatrix}
$$

Here a new type is defined, namely ttime. A similar rule is foreseen for locational metaphors which apply motion verbs.

### 6.2.11 Blocking of sense change rules

Applying sense change rules to treat systematic polysemy may solve the problem of sense enumeration, but it also provides us with another severe problem because of the strong generative nature of this device: it introduces the problem of overgeneration. When we talk about semi-productive mechanisms in polysemy, we indicate exactly the fact that there are always exceptions to these. This fact was exemplified in Chapter 5 with verbs like *vaje* and *dingle* which, according to the corpora, did not display the same elasticity in meaning as their semantic neighbours although one should in theory expect them to do so on the basis of the frame elements assigned. In Chapter 5, we also concluded that such exceptions did not seem to follow any systematic patterns which could be described on the basis of some heuristic principles. We have therefore decided to establish an idiosyncratic blocking mechanism which prevents lexical entries which are restricted in meaning compared to their semantic neighbours to trigger specific sense change rules.

Technically, this is effectuated by means of a flag inserted in a technical part of the relevant lexical signs, RULERESTR (rule restrictions), as depicted below:

(345) *Insertion of rule restriction*

$$
\left[\begin{array}{ll}
\text{RULERESTR} \neg \textit{reduction} & \\
\text{SYNSEM} & \left[\begin{array}{ll}
\text{CAT} & \left[\begin{array}{l}
\text{HEAD} \left[\text{AUX} \quad \textit{have}\right] \\
\text{SUBJ} \ \langle \text{NP}\left[\text{REL\_SORT} \ \textit{inanimate}\right]\boxed{1} \rangle \\
\text{COMPS} \ \langle \ \rangle
\end{array}\right] \\
\text{CONT} & \left[\begin{array}{ll}
\text{PSOA} & \left[\begin{array}{l}
\text{ARG\_PSOA} \left[\begin{array}{ll} \text{REL} & \textit{vaje} \\ \text{ARG1} & \boxed{1} \end{array}\right] \\
\text{ASP\_PSOA} \mid \text{ASPECT} \ \textit{activity} \\
\begin{array}{l}\text{FE\_PSOA} \\ \textit{tback\_forth}\end{array} \left[\begin{array}{ll} \text{MOVER} & \boxed{1} \\ \text{MANNER} & \textit{back\_and\_forth} \end{array}\right]
\end{array}\right]
\end{array}\right]
\end{array}\right]
\end{array}\right]
$$

The flag is applied in the lexical rules as a restriction on the input sign. Lexical signs that contain the negated name of a given rule in the RULERESTR cannot trigger the sense change rule of that particular name. This is shown in the reduction rule below where a flag is introduced (note that since the rule applies unification (and not matching) we are not forced to assign the feature RULERESTR *reduction* to all other entries):

(346) *Blocking in lexical rules*

$$
{}_{\textit{treduction}}\begin{bmatrix}
\text{I} & \begin{bmatrix} \text{RULERESTR} & \textit{reduction} \\
\text{SYNSEM} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{AUX} & \textit{have} \end{bmatrix} \\ \text{SUBJ} & \langle \text{NP}\begin{bmatrix} \text{REL\_SORT} & \textit{inanimate} \end{bmatrix}\boxed{1} \rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \\
\text{CONT} & \begin{bmatrix} \text{PSOA} & \begin{bmatrix} \text{ARG\_PSOA} & \begin{bmatrix} \text{ARG1} & \boxed{1} \end{bmatrix} \\ \text{ASP\_PSOA} \mid \text{ASPECT} & \textit{activity} \\ \text{FE\_PSOA} & {}_{\textit{tback\_forth}}\begin{bmatrix} \text{MOVER} & \boxed{1} \\ \text{MANNER} & \_ \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\
\text{O} & \begin{bmatrix} \text{SYNSEM} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{AUX} & \textit{være} \end{bmatrix} \\ \text{SUBJ} & \langle \text{NP}\begin{bmatrix} \text{REL\_SORT} & \textit{inanimate} \end{bmatrix}\boxed{1} \rangle \\ \text{COMPS} & \langle \text{DIRPART}\boxed{2} \rangle \end{bmatrix} \\
\text{CONT} & \begin{bmatrix} \text{PSOA} & \begin{bmatrix} \text{ARG\_PSOA} & \begin{bmatrix} \text{ARG1} & \boxed{1} \\ \text{ARG\_ADJ} & \boxed{2} \end{bmatrix} \\ \text{ASP\_PSOA} \mid \text{ASPECT} & \textit{accompl} \\ \text{FE\_PSOA} & {}_{\textit{tsuddenmove}}\begin{bmatrix} \text{MOVER} & \boxed{1} \\ \text{MANNER} & \_ \\ \text{DIRECTION} & \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}
\end{bmatrix}
$$

Such a solution can be seen as a compromise between a purely generative approach on the one hand (as proposed by Pustejovsky) and a purely idiosyncratic approach on the other (as seen in traditional machine translation dictionaries) and in our view it benefits from both approaches since it reflects some of the regularities in polysemy but still captures the inevitable idiosyncracies of words.

## 6.3 Exploiting Frame Semantics in transfer

A basic assumption behind our work is that the problem of correct translation to a large extent depends on a sufficiently nuanced disambiguation in the monolingual analysis phase. Therefore we have concentrated on the composition of the Danish dictionary and have only briefly in Chapter 2 sketched out particular transfer problems.

In order not to totally ignore the work to be done in translation, in the present section we shall account briefly for two aspects of transfer:

- On the one hand we want to illustrate how simple lexical transfer of derived senses can take place in a rather straightforward fashion,
- and on the other, we want to illustrate how the assignment of frame elements can be exploited in so-called head-switching operations of structural transfer as required in the treatment of the well-known PATH-MANNER divergence between Romance and Germanic languages.

### 6.3.1 Lexical mapping of derived senses

If we look at the sense change rules established in the previous sections and relate them to our test languages English and Spanish, there is no doubt that the sense changes captured by the proposed system are of translational relevance. In other words, most of the sense changes described trigger translations in English and Spanish that differ from the translation of the prototypical senses, as seen in the examples below:

(347) a) *stolen vippede* → *the chair <u>rocked</u>*
b) *stolen vippede baglæns* → *the chair <u>tipped</u> backwards*

(348) a) *pigen går* → *la chica <u>camina</u>*
b) *tiden går* → *el tiempo <u>pasa</u>*

As long as structural transfer is not involved (we shall come back to this in the next section), transfer of derived senses can take place in the same way as transfer of prototypical senses, namely by means of simple lexical mapping rules with reference to the motion type involved.

A transfer rule can be depicted in a way similar to sense change rules namely be means of a complex feature structure with an input structure and an output structure:

(349) *Transfer rule*

$$_{\textit{ttransferrule}}\begin{bmatrix} \text{I} \begin{bmatrix}\text{LEXSIGN}\end{bmatrix} \\ \text{O} \begin{bmatrix}\text{LEXSIGN}\end{bmatrix} \end{bmatrix}$$

It should be stressed, however, that the rules are interpreted differently: transfer rules *replace* an input structure with an output structure which is not fully instantiated before look-up in the target dictionary has taken place. In contrast, sense change rules *produce* an additional entry on the basis of an input structure but do not change the input.

Below are depicted the lexical mapping rules to be applied in the translation examples just given.

Lexical mapping rule from Danish to English of the prototypical sense of *vippe* (rock):

(350) *Lexical mapping rule (a)*

$$\left[\begin{array}{l}\mathrm{DA}\,|\,\mathrm{SYNSEM}\left[\begin{array}{l}\mathrm{CAT}\,|\,\mathrm{HEAD}\,|\,\mathrm{DALU}\ \textit{vippe}\\ \mathrm{CONT}\,|\,\mathrm{PSOA}\,|\,\mathrm{FE\_PSOA}\ \textit{tback\_forth}\end{array}\right]\\ \mathrm{EN}\,|\,\mathrm{SYNSEM}\,|\,\mathrm{CAT}\,|\,\mathrm{HEAD}\,|\,\mathrm{ENLU}\ \textit{rock}\end{array}\right]_{\textit{ttransferrule}}$$

Lexical mapping rule from Danish to English of the derived sense of *vippe* (rock):

(351) *Lexical mapping rule (b)*

$$\left[\begin{array}{l}\mathrm{DA}\,|\,\mathrm{SYNSEM}\left[\begin{array}{l}\mathrm{CAT}\,|\,\mathrm{HEAD}\,|\,\mathrm{DALU}\ \textit{vippe}\\ \mathrm{CONT}\,|\,\mathrm{PSOA}\,|\,\mathrm{FE\_PSOA}\ \textit{tsuddenmove}\end{array}\right]\\ \mathrm{EN}\,|\,\mathrm{SYNSEM}\,|\,\mathrm{CAT}\,|\,\mathrm{HEAD}\,|\,\mathrm{ENLU}\ \textit{tip}\end{array}\right]_{\textit{ttransferrule}}$$

Lexical mapping rule from Danish to Spanish of the prototypical sense of *gå* (walk):

(352) *Lexical mapping rule (c)*

$$\left[\begin{array}{l}\mathrm{DA}\,|\,\mathrm{SYNSEM}\left[\begin{array}{l}\mathrm{CAT}\,|\,\mathrm{HEAD}\,|\,\mathrm{DALU}\ \textit{gå}\\ \mathrm{CONT}\,|\,\mathrm{PSOA}\,|\,\mathrm{FE\_PSOA}\ \textit{tbody}\end{array}\right]\\ \mathrm{SP}\,|\,\mathrm{SYNSEM}\,|\,\mathrm{CAT}\,|\,\mathrm{HEAD}\,|\,\mathrm{SPLU}\ \textit{caminar}\end{array}\right]_{\textit{ttransferrule}}$$

Lexical mapping rule from Danish to Spanish of the time metaphor sense of *gå* (pasar):

(353) *Lexical mapping rule (d)*

$$\left[\begin{array}{l}\mathrm{DA}\,|\,\mathrm{SYNSEM}\left[\begin{array}{l}\mathrm{CAT}\,|\,\mathrm{HEAD}\,|\,\mathrm{DALU}\ \textit{gå}\\ \mathrm{CONT}\,|\,\mathrm{PSOA}\,|\,\mathrm{FE\_PSOA}\ \textit{ttime}\end{array}\right]\\ \mathrm{SP}\,|\,\mathrm{SYNSEM}\,|\,\mathrm{CAT}\,|\,\mathrm{HEAD}\,|\,\mathrm{SPLU}\ \textit{pasar}\end{array}\right]_{\textit{ttransferrule}}$$

The actual identification of a prototypical or a derived sense is taken care of in the monolingual analysis phase.

### 6.3.2 Head-switching rules of structural transfer for the treatment of PATH-MANNER divergences

We saw in Chapter 2 that *divergences* frequently occur between language pairs; that is in cases where source and target language apply different structures for expressing the same information. When dealing with motion events, one specific divergence has proven to be of central importance: namely the PATH-MANNER divergence that we typically see between Germanic and Romance language. For the sake of clarity we repeat below the example from Chapter 2 which illustrates this divergence:

(354) *han svømmede ind i grotten → entró en la cueva nadando*
(he swam into the cave)

Since this divergence is so frequent and relates to a large group of verbs it would be highly convenient if it could be treated by a general mechanism exploiting the semantic information of the FE_PSOA. The information that we want to exploit is the following: in the FE_PSOA it is identified when a manner verb is used in the extended GO-meaning; a sense change which triggers the PATH-MANNER divergence when translating into for instance Spanish. What happens is that the manner element should be either omitted or expressed by means of a modifier, for instance a gerund (*nadando*). The lexical information of this modifier can be taken from the main verb in the Danish structure.

The problem now lies in the calculation of the path verb to be applied in Spanish where we can roughly choose between the following:

(355) Horizontal motion:
*entrar* (enter)
*salir* (exit)
*ir* (go)
*cruzar* (cross)

Vertical motion:
*subir* (ascend)
*bajar* (descend)

In Danish the path is expressed by means of the prepositional phrase and in particular by means of the directional particle and we can roughly establish the following relations of correspondence between directional particles in Danish and path verbs in Spanish:

(356) Horizontal motion:
*ind* (in) → *entrar*
*ud* (out) → *salir*
*hen* (over) → *ir*
*over* (over) → *cruzar*

Vertical motion:
*op* (up) → *subir*
*ned* (down) → *bajar*

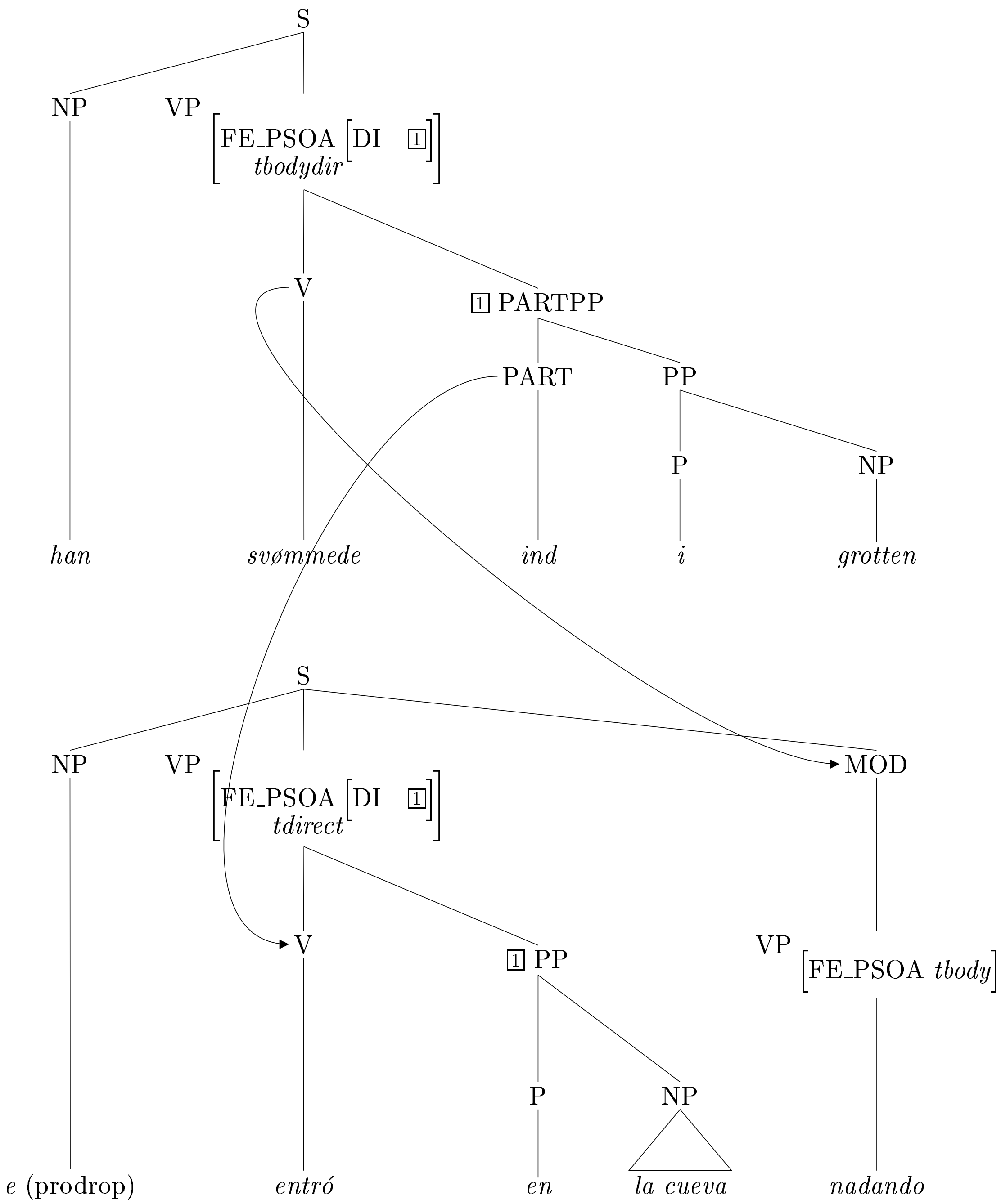
In machine translation, the operation that we are foreseeing here is traditionally called *head-switching*. Head-switching means that there is a switch between two languages with regards to the syntactic head of the phrase in question. The meaning element constituting the syntactic head in Danish (the manner verb) becomes an adverbial phrase in Spanish and the type of direction which is expressed by a particle in Danish is expressed by the head, a direction verb, in the Spanish.

In transfer-based machine translation systems, head-switching is handled by means of structural transfer rules where the syntactic structure of the source language is reorganised into an appropriate structure of the target language. Thus, in contrast to lexical transfer, structural transfer operates on phrasal signs.

Since we have not elaborated a complete grammar for our purpose and therefore have not decided on all relevant aspects of the phrasal sign[4] we shall not provide a full description of the syntactic restructuring which is to take place in transfer. Instead, we shall tentatively sketch out the head-switching operation by means of a Danish and a Spanish tree structure and two arrows of correspondence which roughly illustrate the operation of structural transfer:

[4] For instance, one should consider how to treat manner modifiers that are not subcategorised for by the verb; if these occur over long distances or in embedded constructions it can be problematic to ensure the correct propagation of information (see among several others Kaplan & Wedekind 1993 for a discussion of problematic issues in relation to head-switching).

(357) *Head switching in relation to PATH-MANNER divergence*




In order not to get word specific phrasal transfer rules, which as we noted in Chapter 2 are inconvenient for several reasons, we shall instead refer to the types of FE_PSOA *tbodydir* and *tbody_by_vehicledir* in the phrasal rules and let simple lexical transfer do the actual mapping from the Danish manner of motion verb (*svømme*) to the corresponding Spanish manner of motion verb (*nadar*). Likewise with the calculation of the correct path verb in Spanish, which can also take place by means of simple lexical mapping as sketched out above: (*ind* → *entrar* etc.).

A similar strategy could be applied on causative motion events which exhibit a related divergence between PATH and CAUSE, as we saw in Chapter 4:

(358) *jeg klippede papiret af pakken*
(I cut the wrapper off the package)

which in Spanish translates into:

(359) *quité el papel del paquete cortándolo*

as well as

(360) *han sparkede døren op*
(he kicked the door open)

which translates into:

(361) *abrió la puerta de una patada.*

## 6.4 Sample implementation in LSGRAM-DK written in ALEP

### 6.4.1 Platforms suitable for implementation of our lexicon

We have seen it as an interesting aspect of this work to try and integrate our lexicon of 100 Danish motion verbs into an already running grammar and dictionary of Danish. The elaboration of dictionaries is generally an extremely expensive task, and especially the elaboration of dictionaries containing semantic information has a very high cost. Therefore, we see it as a fruitful process to try and adapt our work to already existing standards in the field.

The ideal test platform for our dictionary would obviously have been that of a transfer-based machine translation system which would have enabled us to test also aspects of transfer. However, as we have stressed throughout this thesis, we consider the monolingual analysis phase to constitute by far the most important part of the disambiguation phase.

Furthermore, a full test of our model in a transfer-based machine translation system would not have been feasible anyway considering the existing systems translating from Danish into a foreign language. Of the ones that apply the transfer-method (for instance the PATRANS system), none supports typed feature structures or type hierarchies; a fact that would have complicated a test in one of these systems considerably.

In fact, Danish grammars applying typed feature structures and type hierarchies are very few. We have chosen one of these, the Danish LSGRAM grammar, henceforth LSGRAM-DK (c.f. Music & Navarretta (1996)) elaborated in the ALEP formalism (Advances Linguistic Engineering Platform) within an EU project.

LSGRAM-DK constitutes an adequate platform for an adaption of our work for several reasons:

- it is based on a standard platform, ALEP, which allows for the usage of typed feature structures, type hierarchies etc.
- it is HPSG-inspired and thus relates more or less directly to the formalisation that we propose,
- it is well-tested using standard test suites developed in the LSGRAM project as whole
- it includes Danish core grammar phenomena.

However, ALEP also presents some central deviations and limitations from the formalism that we are presupposing in the previous sections of this chapter and which have considerable impact on the implementation:

- in contrast to standard HPSG, ALEP (and thereby LSGRAM-DK) applies a stratificational approach, implying that morphological, syntactic and semantic information are not treated simultaneously, but rather handled at different levels of analysis, as will be explained further below,
- ALEP does not support multiple inheritance,
- and last but not least: ALEP does not support the use of lexical rules.

In our case the last mentioned deviation is by far the one with the most severe consequences for the strategy that we propose in relation to systematic polysemy.

The functionality of lexical rules can be captured by means of structure building rules working on lexical signs[5] , however, because of the stratificational approach

[5]This functionality can be seen in LSGRAM-DK, for instance in the treatment of passives, where the passive versions of the Danish verbs are generated via a structure building rule in the grammar.

where semantic information is added *after* parsing this solution cannot be adapted for our task.

Instead, we have been forced to simulate the functionality of lexical rules by means of *macro-calls* in the particular lexical entries. Macros are expanded at compile time and are in essence an abbreviation which allows us to bring all instances of a particular construct together such that only one global statement exists.

In practice, this means that the semantic representation of a derived GO-sense need only be stated once and can then be referred to by a macro-call in all entries for which this sense change is relevant. It should be noted that lexical rules are also expanded at compile time and that there is therefore no difference between the two approaches from an efficiency point of view. However, contrary to lexical rules, a macro approach enforces us to actually enumerate the sense changes that a verb can undergo; a strategy which we in the previous sections have put much effort into avoiding since it is not so satisfying from a theoretical point of view.

In the following subsections, a presentation of our implementation is given in more detail.

### 6.4.2 Information structure in LSGRAM-DK

As already mentioned, LSGRAM-DK applies a stratificational approach by structuring the linguistics information into the following four levels:

1. Textual level (splits the text up into units)

2. Morphological level (builds morphological structures applying a two-level morphology approach)

3. Syntactic level (builds syntactic structures)

4. Semantic level (assigns semantic features to structures; for instance argument structure is assigned at this level)

In ALEP, level 3 is the analysis level where actual parsing takes place, whereas level 4 is solely a refinement level where additional features are assigned in a monotonic fashion to the structures produced in analysis (see Simpkins 1994). LSGRAM-DK has chosen to assign semantic information at refinement level.

### 6.4.3 The lexical sign in LSGRAM-DK

HSPG-inspired in its approach, the structure of the lexical sign in LSGRAM-DK is close to the one that we propose in Section 6.2, only it is more complex since the lexicon runs with a grammar taking care of a large set of grammatical phenomena and with a similar lexical structure for all word classes[6]; information which we have not considered in the previous sections. A semantically refined lexical sign in LSGRAM-DK for a token like *krydsede* (crossed) looks like the following in the ALEP notation[7]:

[6]For instance the lq_cont in the figure below makes room for quantifiers which are relevant in relation to nouns. Also the sem subsign is introduced in order to cover both content and context; although context is not further specified in the grammar.

[7]The sign is slightly simplified; for instance morphographemic information is omitted.

(362) *The lexical sign in LSGRAM-DK in ALEP notation*

$$
\left[\begin{array}{l}
\text{synsem} \Rightarrow
\left[\begin{array}{l}
\text{tsynsem} \\
\text{syn} \Rightarrow
\left[\begin{array}{l}
\text{tsyn} \\
\text{cat} \Rightarrow
\left[\begin{array}{l}
\text{tsubstcat} \\
\text{head} \Rightarrow
\left[\begin{array}{l}
\text{thead} \\
\text{major} \Rightarrow
\left[\begin{array}{l}
\text{tvfininfin\_fin} \\
\text{pos} \Rightarrow \text{v} \\
\text{modal} \Rightarrow \text{n} \\
\text{voice} \Rightarrow \text{act} \\
\text{type} \Rightarrow \text{past}
\end{array}\right] \\
\text{marking} \Rightarrow \text{unmarked}
\end{array}\right] \\
\text{subj} \Rightarrow
\left[\begin{array}{l}
\text{tsynsem} \\
\text{syn} \Rightarrow \left[..\right] \\
\text{sem} \Rightarrow \left[..\right]
\end{array}\right] \\
\text{compls} \Rightarrow
\left[\begin{array}{l}
\text{tsynsem} \\
\text{syn} \Rightarrow \left[..\right] \\
\text{sem} \Rightarrow \left[..\right]
\end{array}\right]
\end{array}\right]
\end{array}\right] \\
\text{sem} \Rightarrow
\left[\begin{array}{l}
\text{tsem} \\
\text{cont} \Rightarrow
\left[\begin{array}{l}
\text{lq\_cont} \\
\text{rd\_cont} \Rightarrow
\left[\begin{array}{l}
\text{r\_psoa} \\
\text{psoa} \Rightarrow
\left[\begin{array}{l}
\text{arg2\_psoa} \\
\text{reln} \Rightarrow \text{krydse} \\
\text{arg1} \Rightarrow \left[\begin{array}{l} \text{lq\_cont} \\ .. \end{array}\right] \\
\text{arg2} \Rightarrow \left[\begin{array}{l} \text{lq\_cont} \\ .. \end{array}\right]
\end{array}\right]
\end{array}\right]
\end{array}\right]
\end{array}\right]
\end{array}\right]
\end{array}\right]
$$

This corresponds to the following AVM in the HPSG notation that we have applied in the previous sections:

(363) *The lexical sign in LSGRAM-DK in HPSG notation*

$$
\begin{bmatrix}
\text{SYNSEM} & {}_{tsynsem}\begin{bmatrix}
\text{SYN} & {}_{tsyn}\begin{bmatrix}
\text{CAT} & {}_{tsubstcat}\begin{bmatrix}
\text{HEAD} & {}_{thead}\begin{bmatrix}
\text{MAJOR} & {}_{tvfininfin\_fin}\begin{bmatrix} \text{pos} & v \\ \text{modal} & no \\ \text{voice} & act \\ \text{type} & past \end{bmatrix} \\
\text{MARKING} & unmarked
\end{bmatrix} \\
\text{SUBJ} & {}_{tsynsem}\begin{bmatrix} \text{SYN} & {}_{tsyn}\begin{bmatrix} .. \end{bmatrix} \\ \text{SEM} & {}_{tsem}\begin{bmatrix} .. \end{bmatrix} \end{bmatrix} \\
\text{COMPLS} & {}_{tsynsem}\begin{bmatrix} \text{SYN} & {}_{tsyn}\begin{bmatrix} .. \end{bmatrix} \\ \text{SEM} & {}_{tsem}\begin{bmatrix} .. \end{bmatrix} \end{bmatrix}
\end{bmatrix}
\end{bmatrix} \\
\text{SEM} & {}_{tsem}\begin{bmatrix}
\text{CONT} & {}_{lq\_cont}\begin{bmatrix}
\text{RD\_CONT} & {}_{r\_psoa}\begin{bmatrix}
\text{PSOA} & {}_{arg2\_psoa}\begin{bmatrix} \text{RELN} & krydse \\ \text{ARG1} & {}_{lq\_cont}\begin{bmatrix} .. \end{bmatrix} \\ \text{ARG2} & {}_{lq\_cont}\begin{bmatrix} .. \end{bmatrix} \end{bmatrix}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

### 6.4.4 The lexica of morphology and syntax

Level 2, 3 and 4 each require a separate lexicon, containing morphological, syntactic and semantic information, respectively.

As regards the lexica for morphology and syntax we have not performed any changes to the LSGRAM-DK structure whatsoever, only we have extended the already existing lexica with entries for motion verbs. For illustration, a sample of the morphological lexicon is given below:

(364) *Samples from the morphological lexicon*

```
m_TLM_v[ ankomme     ankom,    m_vinfl5[], m_tgrf_gem[]].
m_TLM_v[ cirkulere,  cirkuler, m_vinfl1[]].
m_TLM_v[ cykle,      cykl,     m_vinfl1[]].
m_TLM_v[ dale,       dal,      m_vinfl1[]].
m_TLM_v[ danse,      dans,     m_vinfl1[]].
m_TLM_v[ daske,      dask,     m_vinfl1[]].
m_TLM_v[ defilere,   defiler,  m_vinfl1[]].
m_TLM_v[ dingle,     dingl,    m_vinfl1[]].
m_TLM_v[ dratte,     drat,     m_vinfl1[], m_tgrf_gem[]].
m_TLM_v[ dreje,      drej,     m_vinfl1[]].
```

Macro-calls to a set of two-level morphology rules here ensure that the verbs are properly segmented and recognised. Correct segmentation is dependent on inflectional paradigms (infl..) as well as to information on gemination and vowel shift.

As regards the syntactic lexicon of motion verbs we have systematised the entries according to the following principles:

- Intransitive motion verbs which cannot take directional arguments
  (e.g. *ryste* (chake), *vibrere* (vibrate))
- Intransitive motion verbs which can take weakly bound directional prepositional phrases (optional)
  (e.g. *gå* (walk), *ankomme* (arrive))
- Intransitive motion verbs which can take strongly bound directional prepositional phrases (optional)
  (e.g. *kredse (om)* (circulate (around)), *vrimle (med)* (teem (with)))
- Transitive motion verbs
  (*krydse* (cross), *forlade* (leave))

The syntactic lexical entry for *gå* (walk) is depicted below; note that tsynsem_opt_compl is a subsign which is equal in structure to tsynsem but which is treated as optional by the grammar:

(365) *The syntactic entry for gå*

```
'v_gå'~
sign:{
  procinfo=>mLEXSPECana[v,LU],
  synsem=>tsynsem:{
    syn=>tsyn:{
          str=>tv:{lu=>LU=>'gå',lemma=>(gå/gik) - altx},
          cat=>tsubst_cat:{
                head=>tsubst_head:{
                    major=>
                        (tv_major:{perf_aux=>(have/være)}/
                         tv_fininfin:{type=>past}) - altx},
                subj=>[
                  tsynsem:{
                      syn=>tsyn:{
                            cat=>tsubst_cat:{
                                  head=>tsubst_head:{
                                      spr=>sat,
                                      major=>tnom_major:{
                                          pos=>(n;pron),
                                          prd=>no}}}}}],
```

```
compls=>[
  tsynsem_opt_compl:{
     syn=>tsyn:{
        cat=>tsubst_cat:{
                 head=>tsubst_head:{
                   major=>tadv_major:{
                     pos=>adv,
                     adv_lu=>(hen;over;
                     ind;ud;op;ned),
                     prd=>no}},
                 compls=>[]}}},
  tsynsem_opt_compl:{
     syn=>tsyn:{
           cat=>tsubst_cat:{
                 head=>tsubst_head:{
                     major=>tp_major:{
                         pos=>p,
                         prd=>no}},
                 compls=>[]}}}]}}}}.
```

### 6.4.5 The semantic lexicon

As regards the semantic lexicon, we have added information on frame elements, aspect and selectional restrictions since none of these features where part of the LSGRAM-DK implementation (see Music & Navarretta 1996 for a description of the coverage of LSGRAM-DK). The features are added according to the same principles as schetched out in the previous sections, by adding two new kinds of PSOA to the sign and by introducing a rel_sort feature on nouns for semantic typing:

(366) *The extended semantic sign*

$$
\text{cont} \Rightarrow \left[\begin{array}{l}
\text{lq\_cont} \\
\text{rd\_cont} \Rightarrow \left[\begin{array}{l}
\text{r\_psoa} \\
\text{psoa} \Rightarrow \left[\begin{array}{l}
\text{arg\_psoa} \Rightarrow \left[\begin{array}{l}
\text{arg2\_psoa} \\
\text{reln} \Rightarrow \text{krydse} \\
\text{arg1} \Rightarrow \left[\begin{array}{l} \text{lq\_cont} \\ .. \end{array}\right] \\
\text{arg2} \Rightarrow \left[\begin{array}{l} \text{lq\_cont} \\ .. \end{array}\right]
\end{array}\right] \\
\text{asp\_psoa} \Rightarrow \left[\begin{array}{l}
\text{tasp\_psoa} \\
\text{aspect} \Rightarrow \text{achieve}
\end{array}\right] \\
\text{fe\_psoa} \Rightarrow \left[\begin{array}{l}
\text{tdirection} \\
\text{mover} \Rightarrow \left[\begin{array}{l} \text{lq\_cont} \\ .. \\ .. \mid \text{rel\_sort} \Rightarrow \text{concrete} \\ .. \end{array}\right] \\
\text{direction} \Rightarrow \left[\begin{array}{l} \text{lq\_cont} \\ .. \end{array}\right]
\end{array}\right]
\end{array}\right]
\end{array}\right]
\end{array}\right]
$$

### The type hierarchies for fe_psoa and rel_sort

The fact that ALEP does not support multiple inheritance obviously limits the application of complex type hierarchies to a considerable degree. Adapting the type hierarchy for motion verbs as it stands, however, proves not to be problematic, since the hierarchy seen in isolation only uses simple inheritance. If we imagine a more complex semantic hierarchy including also other semantic verb classes and other components of meaning the case would not be quite as simple. In an extended semantic system, a platform allowing multiple inheritance would therefore be called for.

Below is given the ALEP version of the type hierarchy for motion verbs as well as the attribute assignments:

(367) *Type hierarchy for fe_psoa*[8]

```
tfe_psoa > {
          tmotion > {
               tdirection > {
                        tpath,
                        tsource,
                        tgoal,
                        tchase},
               tmanner > {
                        tbody_by_vehicle,
                        tbody > {
                             tbody_part > {
                                tfoot > {
                                     twalk > {
                                         tsound,
                                         tsurroundings,
                                         tspeed
                                         },
                                     trun > {
                                         trspeed},
                                     jump},
                                tother_parts},
                             twhole_body > {
                                       tdance,
                                       twspeed,
                                       taimless}},
                        tnon_body > {
                              tback_and_forth,
                              tcircular,
                              tmass},
                        tgroup}},
%% derived types:
tsuddenmove, tbodydir, tnonbodydir, tgroupdir, tbodybyvehicledir, ttime}.
%% Type declarations
type(
  tfeg:{}, 'frame element group').
```

[8]Further clarification notes on the ALEP notation:' > ' indicates that the type to the left is a supertype of the type to the right. In type declarations (as well as in rules) types are followed by a ':' and the features in '{}' which are included by the particular type. Attribute names are assigned values by means of an arrow: '⇒'. Values can be either variables or constants: Variables are in capital letters, constants in minor.

```
type(
  tmotion:{
    mover => type({ content:{} })
            }, 'motion verbs').
type(
   tdirection:{
    directio => type({ content:{} }),
    direction  => atom}, 'directional motion verbs' ).
type(
  tpath:{}, 'path verbs').
type(
  tsource:{}, 'source verbs').
type(
  tgoal:{}, 'goal verbs').
type(
  tchase:{}, 'chase verbs').
type(
   tmanner:{
     manner => atom}, 'manner motion verbs').
type(
   tbody_by_vehicle:{
     vehic => type({ content:{} }),
     vehicle => atom,
     pro_con => type({ content:{} })
                    },'bodybyvehcile verbs').
type(
   tbody:{
    pro_con => type({ content:{}})},'body verbs').
type(
  tbody_part:{}, 'body part verbs').
type(
  tfoot:{}, 'foot verbs').
type(
  twalk:{}, 'walk verbs').
type(
  tsound:{}, 'sound verbs').
type(
  tsurroundings:{}, 'surrounding verbs').
type(
  tspeed:{}, 'speed verbs').
type(
  trun:{}, 'run verbs').
type(
```

```
  trspeed:{}, 'run speed verbs').
type(
  tjump:{}, 'jump verbs').
type(
  tother_parts:{}, 'other body part verbs').
type(
  twhole_body:{}, 'whole body verbs').
type(
  tdance:{}, 'dance verbs').
type(
  twspeed:{}, 'speed verbs').
type(
  taimless:{}, 'aimless verbs').
type(
   tnon_body:{
     directio => type({ content:{} }),
     place => type({ content:{} })
             }, 'non-body verbs' ).
type(
  tback_and_forth:{}, 'back and forth verbs').
type(
  tcircular:{}, 'circular verbs').
type(
  tmass:{}, 'mass verbs').
type(
  tgroup:{}, 'group verbs').
%% Types for derived senses that trigger new types:
type(
  tsuddenmove:{mover => type({ content:{} }),
              manner => atom,
              directio => type({ content:{} })}).
type(
  tbodydir:{mover => type({ content:{} }),
            manner => atom,
            directio => type({ content:{} }),
            pro_con => type({ content:{} })},'bodydir').
type(
  tnonbodydir:{mover => type({ content:{} }),
            manner => atom,
            directio => type({ content:{} })},'non-bodydir').

type(
  ttime:{mover => type({ content:{} }),
```

```
        manner => atom},'time').
type(
  tbodybyvehicledir:{mover => type({ content:{} }),
                     manner => atom,
                     directio => type({ content:{} }),
                     vehic => type({ content:{} }),
                     vehicle => atom,
                     pro_con => type({ content:{} })},
                     'bodybyvehciledir').
```

Note that an attribute can only be assigned one datatype. Therefore we are forced to establish two different attributes for the frame element Direction, namely 'direction' and 'directio'.

Also the partial sort hierarchy for nouns presented in Chapter 5 can be encoded as a type hierarchy in ALEP:

(368) *Type hierarchy for rel_sort*

```
rel_sort > {
                entity > {
                     concr > {
                         inanim,
                          anim > {
                             human,
                             animal > {
                              fly_animal,
                              nfly}}},
                   abstr > {
                         time,
                         ntime}}}.

%% Declarations for rel_sort

type(rel_sort:{}).
type(entity:{}).
type(concr:{}).
type(inanim:{}).
type(anim:{}).
type(human:{}).
type(animal:{}).
type(fly_animal:{}).
type(nfly:{}).
```

```
type(abstr:{}).
type(time:{}).
type(ntime:{}).
```

### A semantic noun entry

Since the original LSGRAM-DK lexicon includes no semantic typing on nouns; and since we have no interest in recoding the dictionary of several thousands nouns with semantic typing, we have chosen to limit ourselves to a coding of approx. 30 nouns typically occurring as subjects in our corpus excerpts.

Human nouns are covered by the following lexical sign at refinement level; nouns of the other semantic sorts are coded in a similar fashion[9].

(369) *Extended noun entry*

```
sign:{
  procinfo=>mLEXSPECref[n,'human'],
  synsem=>tsynsem:{
    syn=>tsyn:{
          str=>tn:{lu=>LU=>(dame/asylansøger/demonstrant/ung/
                            detailhandelsekspert/gæst/
                            chauffør/ingeniør/dreng/pige/
                            adel/menneske)},
          cat=>tsubst_cat:{
                head=>tsubst_head:{major=>tn_major:{prd=>no}},
                compls=>[]}},
    sem=>tsem:{
          content=>lq_cont:{
                rd_cont=>r_npro:{
                    index=>INDEX=>ind_index:{},
                    possessor=>POSSESSOR,
                    restr=>[inst_zero_psoa:{
                              inst=>INDEX,
                              rel=>rel:{rel_name=>LU,
                                        rel_sort=>human:{}}}]}}}}}.
```

[9]Note that '/' expresses alternation between the different LU values of the sign: at compile time the sign is expanded out into a set of signs comprising a sign for each lexical unit.

### A semantic verb entry

A lexical verb entry of a motion verb at refinement level looks as follows:

(370) *Lexical entry for krydse*

```
'v_krydse' ~
sign:{
  procinfo=>mLEXSPECref[v,LU],
  synsem=>tsynsem:{
    syn=>tsyn:{
           str=>tv:{lu=>LU=>krydse},
           cat=>tsubst_cat:{
               subj=>[tsynsem:{sem=>tsem:{
                    content=>ARG1=>lq_cont:{
                         rd_cont=>r_npro:{
                          restr=>[inst_zero_psoa:{
                          rel=>rel:{rel_sort=>concr:{}}}]}}}}],
               compls=>[tsynsem:{sem=>tsem:{content=>ARG2}
               }]},
    sem=> tsem:{
          content=>lq_cont:{
               rd_cont=>r_psoa:{
                          arg_psoa=>arg2_psoa:{
                                  rel=>rel:{rel_name=>LU},
                                  arg1=>ARG1,
                                  arg2=>ARG2},
                          aspect_psoa=>aspect_psoa:{
                                  aspect=>achiev},
                          fe_psoa=>tdirection:{
                                  mover=>ARG1,
                                  directio=>ARG2}}
                     }}.

%% Definition: to move or travel to the other side of a certain
%% area (road, country etc.)
%% Example: han kunne næppe formodes at have set en lille pige
%% krydse gaden
%% English: cross
```

### Sense change macros

As can be seen, each entry is provided with a definition of the prototypical sense of the verb, an example from the corpus of the prototypical sense and eventually some examples of the potential derived senses. Finally, an English translation is given.

A macro which expands a sign into an extended GO sense is depicted below:

(371) *Macro for the extended GO sense*

```
macro(
  mSENSECHANGEmovegob[REL,ARG1,ARG_ADJ],
  tsem:{
    content=>lq_cont:{
               rd_cont=>r_psoa:{
                          arg_psoa=>arg2_psoa:{
                                  rel=>rel:{rel_name=>REL},
                                  arg1=>ARG1,
                                  arg_ADJ=>ARG_ADJ},
                          aspect_psoa=> aspect_psoa:{
                                  aspect=>accompl},
                          fe_psoa=>tbodydir:{
                                  mover=>ARG1,
                                  manner=>body,
                                  directio=>ARG_ADJ}}
                    }}).
```

Call of a macro consists simply of a call assertion together with a set of parameters in square brackets, in the given case 'mSENSECHANGEmovegob [REL,ARG1,-ARG_ADJ]'. The macro mSENSECHANGEmovegob can thus be inserted in all entries where we want this sense extension to apply. This means that we turn around the problem of blocking, since on this approach we simply avoid macro calls in restricted lexical entries, such as *vaje* (wave).

A lexical entry including a move-go sense change looks as follows:

(372) *Lexical entry with sense change macro*

```
'liste'~
sign:{
  procinfo=>mLEXSPECref[v,'liste'],
  synsem=>tsynsem:{
```

```
    syn=>tsyn:{
           str=>tv:{lu=>LU=>liste},
           cat=>tsubst_cat:{
                 subj=>[tsynsem:{sem=>tsem:{
                    content=>ARG1=>lq_cont:{
                         rd_cont=>r_npro:{
                          restr=>[inst_zero_psoa:{
                          rel=>rel:{rel_sort=>anim:{}}}]}}}}],
               compls=>[tsynsem_extracted:{},
                       tsynsem_extracted:{}]}},
   sem=> tsem:{
       content=>lq_cont:{
               rd_cont=>r_psoa:{
                          arg_psoa=>arg1_psoa:{
                                  rel=>rel:{rel_name=>LU},
                                  arg1=>A1},
                          aspect_psoa=> aspect_psoa:{
                                  aspect=>activity},
                          fe_psoa=>tbody:{
                                  mover=>A1,
                                  manner=>body_part,
                                  pro_con=>A1}}
                    }}}}.
sign:{
 procinfo=>mLEXSPECref[v,'liste+adv+pp'],
 synsem=>tsynsem:{
  syn=>tsyn:{
           str=>tv:{lu=>liste},
           cat=>tsubst_cat:{
                  subj=>[tsynsem:{sem=>tsem:{
                    content=>ARG1=>lq_cont:{
                         rd_cont=>r_npro:{
                          restr=>[inst_zero_psoa:{
                          rel=>rel:{rel_sort=>anim:{}}}]}}}}],
                  compls=>[tsynsem:{
                          syn=>tsyn:{
                            cat=>tsubst_cat:{
                              head=>tsubst_head:{
                                major=>tadv_major:{
                                  pos=>adv}}}}},
                 tsynsem:{sem=>tsem:{content=>ARG_ADJ}}]}},
  sem=>mSENSECHANGEmovegob[LU,ARG1,ARG_ADJ]}}./*macro call*/
%% Definition: to walk silently
```

```
%% Examples: han halvvejs listede og loeb/han vaagnede ved
%% 5-tiden og listede ud
%% i koekkenet
%% English: tiptoe
```

The first sign constitutes the prototypical lexical sign with a prototypical frame element assignment, whereas the second sign describes the syntactic constraints which trigger a sense extension macro (namely *liste* combined with a particle and a PP).

The time metaphor co-composition sense change macro is given below:

(373) *Macro for time metaphor*

```
macro(
  mSENSECHANGEcocompt[REL,A1],
  tsem:{
    content=>lq_cont:{
               rd_cont=>r_psoa:{
                          arg_psoa=>arg1_psoa:{
                                  rel=>rel:{rel_name=>REL},
                                  arg1=>A1},
                          aspect_psoa=>  aspect_psoa:{
                                  aspect=>activity},
                          fe_psoa=>ttime:{
                                  mover=>A1,
                                  manner=>time
                                  }}
                     }}).
```

Since *liste* is also sensitive to this sense change, this macro is also called in the entry for *liste* together with the corresponding semantic constraint on the subject (rel_sort $\Rightarrow$ time):

(374) *Entry calling co-composition macro*

```
sign:{
  procinfo=>mLEXSPECref[v,'liste'],
  synsem=>tsynsem:{
    syn=>tsyn:{
           str=>tv:{lu=>LU=>liste},
           cat=>tsubst_cat:{
```

```
                  subj=>[tsynsem:{sem=>tsem:{
                     content=>ARG1=>lq_cont:{
                         rd_cont=>r_npro:{
                          restr=>[inst_zero_psoa:{
                          rel=>rel:{rel_sort=>time:{}}}]}}}}],
                compls=>[tsynsem_extracted:{},
                         tsynsem_extracted:{}]}},
  sem=>mSENSECHANGEcocompt[LU,ARG1]}}.
```

We refer to Appendix A for a printout of the dictionary[10] and to Appendix B for a printout of the macros. Note that the dictionary in Appendix A for abbreviation purposes apply heavy use of macros apart from the sense change macros. These are, however, also represented in Appendix B.

### 6.4.6 Modifications of the LSGRAM-DK grammar

Practically no changes have been done in order to make the original LSGRAM-DK grammar run with our semantically extended entries. In fact, only minimal assertions in order to ensure the correct percolation of features from daughter to mother have been performed, as illustrated in a refinement rule corresponding to the content principle in HPSG, which defines the semantic head[11]:

(375) *Content Principle rule*

```
sign:{
  procinfo=>mPSPECref['Cont.Princ.(2) h_adj,
                { ~(quant;n)+head}, r_psoa'],
  synsem=>tsynsem:{
    syn=>tsyn:{
        str=>tphrasal:{constr=>h_adj},
        cat=>tsubst_cat:{head=>tsubst_head:{major=>MAJOR}}},
    sem=>tsem:{content=>lq_cont:{quants=>QUANTS,
                                 rd_cont=>r_psoa:{
        % name change %          arg_psoa=>PSOA,
        % inserted %             aspect_psoa=>ASPECT,
                                 restr=>[RESTR|RESTRS],
```

[10]The dictionary contains the verbs listed in Section 5.3.3. Excluded are however three reflexive verbs (*bevæge sig* (move) and *mave sig* (crawl forward on ones stomach) and *snige sig* (sneak)) since multi-word units are not covered by LSGRAM-DK.

[11]Some further modification, however, is foreseen in relation to the rel_sort percolation which at the current stage of the grammar does not work in relation to argument bearing nouns. Therefore, such nouns are excluded from the test corpus in subject position.

```
        % inserted %             fe_psoa=>FEG}}}}}
     <
     { sign:{
         synsem=>tsynsem:{
           syn=>tsyn:{
               str=>tphrasal:{},
               cat=>tsubst_cat:{
                   head =>tsubst_head:{
                       major=>tsubst_major:{
                           pos=>(~(quant;n)),
                           selects=>tmodifiee:{
                               synsem=>TSYNSEM}}}}},
           sem=>tsem:{content=>lq_cont:{rd_cont=>rd_cont:
                         {restr=>[RESTR]}}}}},
       sign:{
         procinfo=>tprocinfo:{parsehead=>y},
         synsem=>TSYNSEM=>tsynsem:{
           syn=>tsyn:{
                 str=>tphrasal:{},
                 cat=>tsubst_cat:{
                     head=>tsubst_head:{
                         major=>MAJOR=>tsubst_major:
                          {pos=>(~(n;pron))}}}},
           sem=>tsem:{
                content=>lq_cont:{
                   quants=>QUANTS,
                   rd_cont=>r_psoa:{
   % name change %     arg_psoa=>PSOA,
   % inserted %        aspect_psoa=>ASPECT,
                       restr=>RESTRS,
   % inserted %        fe_psoa=>FE_PSOA}}}}}
                    }.
```

In other words we have chosen to change as little as possible in existing grammars and lexica of LSGRAM-DK, although there are some aspects which we have found inconvenient and which should probably be changed on a longer term:

- optional arguments which are not realised in a given structure are present in the ARG_PSOA but are not structure shared with anything: we would prefer not to have them there even if it meant a complication of the grammar,

- when prepositional phrases occur as complements only the NP is structure shared with the arguments; this means that the preposition is 'neutralised', so to speak, in the content part of the sign. We would prefer to structure share with the whole prepositional phrase in cases where the preposition is weakly bound.

### 6.4.7 Testing the lexicon in interaction with the grammar

We have constructed a small test corpus of 42 sentences (see appendix C) in order to test the lexicon in interaction with the grammar. The test corpus is based on the corpus examples that have been extracted for each motion verb and is constructed in such a way that most subtypes of motion verbs are tested.

Due to the limited grammatical and lexical coverage of LSGRAM-DK considerable simplification and limitation of the corpus examples have been necessary in order to make the sentences parse; for example coordinated verbs are frequent with some Danish motion verbs when expressing durative events (as in *de hang og dinglede* ('they hang and dangled')) but such constructions have been excluded or simplified since verb coordination is not part of the coverage of LSGRAM. The coverage of subordinate clauses is also very limited at the current stage of the grammar.

Likewise with respect to lexical coverage: LSGRAM-DK contains approx. 3000 lexical entries which is obviously a very limited vocabulary when working with authentic corpus examples; however, the application of default entries, where all unknown words are interpreted as nouns with some default attributes related to them has increased the actual potential of the system, since sentences can now be analysed even if not all nouns are recognised. In such cases part of the analyses are only rudimentary; however, as long as unknown nouns do not function as subjects this does not affect our semantic analysis too much.

#### A sample of the test results

The test results were generally very promising in the sense that we achieved the correct analyses in 39 of the 42 examples. By correct analyses we mean that we achieved the correct structure sharing between frame elements and complements and last but not least the correct triggering of systematic meaning shifts. In order to exemplify the test results we list some illustrative examples of correct and incorrect analyses below.

Examples of correct analyses:

> Test sentence 6: *Mange asylansøgere ankommer til landet uden papirer*
> (many refuge applicants arrive to the country without papers)

Test sentence 6 exemplifies a direction verb with a directional modifier which is structure shared with the frame element Direction. Note that no new meaning is triggered since direction verbs do not undergo meaning shift when directional modifiers are added. Screen print of partial analysis[12] after refinement:

[12]Unfortunately, only partial analyses can be given due to space limitations; the node that we have chosen to reproduce is the CONT subsign on the S-node, where the three PSOAs are represented.

Test sentence 17: *Ingeniøren kørte altid i bil*
('the engineer always drove in car')

In test sentence 17, we see an example of a syntactically realised mode of transport which is therefore structure shared with the frame element Vehicle.

Test sentence 18: *Pigerne red ned til stranden på de fleste hverdage* (the girls rode down to the beach on most weekdays)

Sentence 18 is case of a MOVE-GO sense change which is triggered on a body-by-vehicle verb because of the directional prepositional phrase. Note that the frame element Vehicle (the horse) is not realised syntactically here but rather implicitly denoted by the verb.

Test sentence 22: *Der gik nogle uger*
('there walked some weeks')(some weeks passed by)

Sentence 22 is an example of a time metaphor which has triggered a co-composition sense change macro due to the semantic sort of the subject (time).

Test sentence 33: *Pigerne daskede rundt i byen i nogle timer*
(the girls slouched along in the town for some hours)

Here we see another example of co-composition triggered by the subject *pigerne* (the girls).

Examples of incorrect analyses:

> Test sentence 23: *Ingeniøren er gået for idag*
> ('the engineer is walked for today')(the engineer has left for today)

Here we have a case of a MOVE-GO sense with no prepositional phrase. Therefore, as can be seen, the sense change has not come into action and a wrong aspectual type has therefore also been assigned. It would be rather unproblematic to correct this mistake because we have the auxiliary as a marker in this example, however, this is not always the case, as we saw in Section 5.3.5.

Test sentence 32: *Gedderne daskede fra hustagene*
(the pikes dangled from the house roofs)

Sentence 32 demonstrates a very general problem, namely the fact that animals behave in two different ways in language: either as animate entities which have bodies and move about or as inanimate entities meant for eating, selling or the like. 32 is an example of the latter; we are dealing with a kind of food (namely dried fish which is a common dish in several Northern countries) which can definitely not slouch around although this is the interpretation which has wrongly been triggered because of the 'animate' typing which all animals are assigned.

### General evaluation of the test results

In none of the cases did the sense change mechanism result in overgeneration[13]; thus 'full' lexical disambiguation of motion verbs was obtained in all examples primarily due to the syntactic and semantic constraints expressed directly in the lexical entries.

The promising results should be seen in the light of a somewhat tuned test corpus. Although based on original text, we have chosen examples that we knew would test the elaborated macros in order to perform a general debugging of the lexicon at first run. Obviously, a more unrestricted test would not produce test results of the same correctness. However, we believe that the identified sense changes would only go wrong in some very few cases, for instance in cases where directionals are not correctly identified because they occur far away from the verb or because they are expressed in an uncommon way. Another problem concerns new unforeseen meaning changes which would most presumably occur in unrestricted text. Since we only aim at treating the most systematic meaning deviations, such cases are not foreseen by our implementation.

## 6.5 Summary

In this chapter we have focussed on the formalisation and implementation of the lexical entries to be included in the Danish analysis module. We have argued that the HPSG formalism represents a suitable platform for our purpose since it combines the application of typed feature structures (previously applied within Frame Semantics) with the conflation of syntactic and semantic information in the lexical sign. This enables us to combine syntactic restrictions with semantic interpretations in the same sign.

We have furthermore adopted an approach based on sense change rules to cases of systematic polysemy since this device enables us to avoid the traditional sense enumeration technique where each sense is atomic and considered unrelated in meaning. It has been argued that the lexical rule device is more suitable for this task than the alternative technique of lexical *underspecification*, since the relation between the senses that we describe here is not in all cases uniform. Seen in the light of the conclusion that we reached at in Chapter 5, namely that exceptions to sense changes could not be described systematically, but that they were rather to

[13] It should be noted, though, that the syntactic module of LSGRAM-DK overgenerates wrt. valency-boundness, since the developers have chosen not to let valency-bound interpretations of PPs override non-valency bound interpretations but rather to produce all PP attachment possibilities. We do not, however, consider this to be a case of lexical ambiguity, but rather a case of structural ambiguity.

be seen as idiosyncratic cases of etymological and social conventions, we have also introduced a lexical blocking mechanism to be applied to block the application of lexical rules. By this approach we believe to have captured two in essence opposite directed characteristics: namely the regularities *and* the idiosyncracies of verb polysemy.

We have also in this chapter looked into transfer in order to illustrate how frame semantic information can be exploited specifically in relation to the PATH/ MANNER divergence which is so frequent when translating motion events from Germanic languages into Romance languages (and vice versa). We have rudimentarily shown how the head-switching operation foreseen in transfer between these languages can be elaborated in a general fashion avoiding word-specific structural transfer rules.

Last but not least, we have implemented our lexicon in a running grammar of Danish in order to perform a minimal test of our lexicon model in the analysis phase. Although the ALEP formalism does not support the use of lexical rules, the Danish core grammar developed within this formalism proved to be the most suitable for our task since it applied typed feature structures as well as an HPSG-inspired theory of signs. 100 Danish motion verbs have been encoded in the system and the LSGRAM-DK has been extended to include information on Frame Semantics, aspect, and semantic typing on nouns. By means of macros the functionality of lexical rules has been simulated and sense change macros have been established for the three kinds of sense changes described in Chapter 5: sense extension, sense reduction and co-composition. A test of the system comprising a test corpus of 42 sentences involving motion events gave very promising results as regards the triggering of correct sense changes; a more unrestricted test of the system is still lacking though, primarily due to limitations in lexical and grammatical coverage.

# Chapter 7

# Conclusions

The primary aim of this thesis has been to investigate new methods for the treatment of lexical ambiguity in machine translation in order to achieve a more adequate treatment of systematic polysemy in the lexicon.

A secondary aim was the more empirical one of examining a group of not very well-studied verbs in Danish in order to identify the polysemous regularities that these verbs exhibit.

Considering recent work on the topic within the area of lexical semantics, we have primarily been inspired by two approaches which treat verb polysemy in particular:

- Cognitive Frame Semantics (Atkins & Fillmore 1994) which views the problem of verb polysemy from the traditional lexicographer's point of view and suggests to leave the traditional 'dictionary sense' aside in favour of what is defined as *situation types* described by means of frame elements.
- The alternation approach suggested by Levin (1993) and Levin & Rappaport Hovav (forthcoming) which focuses on the *syntagmatic effects* of polysemy in order to identify the relation between syntactic potential and *elasticity in meaning*.

Also the Generative Lexicon (Pustejovsky 1995), which treats systematic polysemy by means of a set of generative devices operating on core lexical entries, has to some extent influenced this work; for instance we have taken over the concept of *co-composition* and the idea of new senses being derived on the basis of core senses. However, we generally found Pustejovsky's model far too ambitious and not sufficiently constrained for our purpose.

In contrast, the taxonomical and very textually anchored view point adopted by the two former approaches proved directly useful to us. The idea of identifying, for each subgroup of verbs, on the one hand the syntactic potential and on the other the patterns of polysemous regularities - in Frame Semantics described by means of associated frame elements, was indeed very fruitful in the treatment of Danish motion verbs.

Thus, having established a basic apparatus consisting of a small set of frame elements, a limited set of aspectual types, as well as a restricted sort hierarchy for nouns, a taxonomy was established for Danish motion verbs and an examination of the syntactic potential of Danish motion verbs was performed. These examinations exposed a transparent relationship between syntactic potential and shared meaning components, in other words: words that were considered to share one or more meaning components also demonstrated a similar syntactic behaviour.

Realising this relation between syntax and semantics helped us identify, for each subgroup of verbs, the different categories of systematic meaning deviation that the particular verbs could undergo. Four kinds of meaning deviation were identified, namely *meaning extension*, *meaning reduction*, *co-composition* and *meaning focusing*.

The three former were categorised as cases of *sense selection* or *real ambiguity* in Cruse's terminology (1991) since the derived senses proved to evoke other sets of frame elements than the prototypical senses; in contrast, the latter - meaning focusing - did not evoke new frame elements but rather changed the focus between the given frame elements and was thus considered to be a case of *sense modulation* which should not, according to Cruse, be treated as a case of meaning change.

As regards the systematicity of meaning change, the following patterns were exposed:

- Direction verbs proved to be rigid in meaning and hardly candidates to any systematic meaning changes whatsoever,
- Body verbs and body-by-vehicle verbs are very frequent candidates to meaning extension and co-composition, although the general pattern was that the lower in taxonomy (i.e. the more specific in meaning) the less flexible with regard to meaning change,
- Non-body verbs are generally candidates to all kinds of meaning shifts: meaning extension, meaning reduction as well as co-composition.
- Group verbs are relatively restricted in meaning but do undergo meaning extension.

These results proved to correspond well to Levin & Rappaport Hovav's (forthcoming) analysis of verb elasticity, since exactly direction verbs are 'saturated' in meaning from an aspectual viewpoint because they inherently denote a final state of the event and therefore can't be varied, whereas manner verbs (including body verbs, body-by-vehicle verbs and non-body verbs) denote activities and are therefore open to variation.

Finally, it was stressed that we can only speak of *semi*-productive mechanisms in polysemy: within each subgroup there will always occur verbs that are restricted in some way or another; and we claimed tentatively that these exceptions can hardly be described on the basis of heuristic principles; rather they should be seen as idiosyncratic cases of etymological and social conventions.

In the formalisation and implementation phase, we applied an HPSG-inspired formalism (Pollard & Sag 1987, 1994) which is based on Typed Feature Structures and which allows for the conflation of syntactic and semantic information in the lexical sign. It was discussed whether to implement the productive mechanisms in polysemy by means of underspecification in the type hierarchy or rather by means of lexical rules. Since we did not find a uniform relation between prototypical and derived senses in all cases - especially in the case of co-composition the two senses could deviate quite a lot from each other, heavily influenced by the semantics of the noun (cf. Pustejovsky 1995), the lexical rules approach was adopted (as suggested by Copestake and Briscoe 1995).

Exceptions to systematic polysemy were treated by the insertion of a blocking mechanism in the lexical entries and in the lexical rules; a solution which can be seen as a compromise between a fully generative approach, as seen in Pustejovsky 1995, and a purely idiosyncratic approach, as seen in traditional machine translation. We concluded that we by this approach had captured two in essence opposite directed characteristics of verbs: namely the regularities *and* the idiosyncrasies of verb polysemy.

Furthermore, we have shown how a frame semantic analysis can be exploited in transfer in relation to the PATH-MANNER divergence which we claim can be accounted for by a general head-switching operation avoiding the highly inconvenient word-specific complex transfer rules.

Finally, in implementation, 100 Danish motion verbs have been encoded in the ALEP formalism and integrated in the Danish analysis grammar LSGRAM-DK. In fact, at the time of implementation, LSGRAM-DK proved to be the only large-scale analysis module available for Danish which applied Typed Feature Structures and which used an HPSG-inspired approach. However, lexical rules are not supported in ALEP, and therefore, the functionality of lexical rules has been partially simulated by means of macros which can be seen as abbreviations that allow to bring all instances of a construct together such that only one global statement exists. Macros for the semantic interpretation of derivated senses are thus called in all entries where a particular sense change comes into play. Correspondingly, in restricted lexical entries - which would be 'blocked' in a lexical rules approach - such macro calls are avoided.

A test on an adjusted suite of 42 corpus examples was performed and the results were very positive: 39 of the parsed sentences achieved the correct disambiguation of motion senses. A large scale test on unrestricted text is still lacking, though, and cannot be performed with the present lexical and grammatical coverage of LSGRAM-DK.

In conclusion, we have demonstrated that the approach adopted is both operational and sufficiently nuanced as to capture the central semantic parameters of motion events which are relevant when dealing with Danish polysemous verbs of motion in machine translation. In this specific domain, the notions of Manner, Direction, and Protagonist Control proved to be the distinctive features which triggered new terms or even distinct constructions in the target languages applied, namely English and Spanish.

We also believe that the methodology proposed can be extended to other semantic domains. The taxonomical approach where each semantic domain - on the basis of corpus data - is examined both from the syntactic and from the semantic point of view opens for a more detailed and fruitful analysis than what is seen in more general approaches to semantic roles and aspectual types. In other words, we assume that each semantic domain contains its own critical divergences between languages and that these can best be captured by a domain-oriented analysis.

In the perception domain, for instance, a crucial semantic parameter relates to the degree of activity involved in the perception process. Again, in Danish, we tend to have a small vocabulary which semantically covers a large range of meanings which in English, as well as in several other languages, would trigger different lexemes (see Braasch 1994 for a general account of Danish perception verbs).

For illustration, consider the examples below where we in Danish apply the perception verb *se* (see) for denoting both passive and active visual perception:

(376) *vi kunne ikke undgå at se hende*
(we couldn't avoid seeing her)

(377) *vi så på de fine stoffer*
(we looked at the fine fabrics)

(378) *han så fjernsyn hver dag*
(he watched television everyday)

These features of vagueness and ambiguity of Danish verbs, which seem to recur in different semantic domains, stress the importance of an approach which captures the regularities in meaning change. In fact, some treatment of systematic polysemy seems indispensable if we want to deal with these verbs in natural language processing.

# References


Alonge, Bes, Braasch, Calzolari, Dini, Emele, Fillmore, Heid, Monachini, Ostler, Spanu & Vliegen, (1993). 'Some Prominent Approaches to Syntactic and Lexical-Semantic Modelling in the View of DELIS Descriptive work', in: *Report from LRE-project: DELIS*. European Commission, Luxembourg.

Andersson, H. (1994) 'Den Danske Ordbogs tekstkorpus', in: H. Andersson, B. Henriksen, J. Asmussen, K.T. Thomsen: *Udarbejdelsen af Den Danske Ordbog*. Selskab for Nordisk Filologi: Årsberetning 1992-1993.

Asmussen, J. (1993) 'Corpus Bench - et korpusværktøj', in: *UDOG Rapport nr. 1, Skriften på Skærmen, nr. 7*, Århus Business School.

Atkins, S. (1994) 'Word Meaning in the Lexicographical Process'. *Course material from EURALEX 1994*, Amsterdam.

Atkins, S. & C. Fillmore, (1994). 'Starting where the dictionaries stop: the challenge of corpus lexicography', in: Atkins, B. & Zampolli, A. (eds.) *Conceptual approaches to the lexicon*. Oxford University Press, Oxford.

Badia, T. & C. Colominas (1995). 'Predicate-Argument Structure', Version 2.0, *MLAP 93-15*, European Commission (Unpublished).

Barnett, J., I. Mani, E. Rich, C. Aone, K. Knight, J.C. Martinez (1993). 'Capturing Language-Specific Semantic Distinctions in Interlingua-Based MT', in: S. Nirenburg (ed.) *Progress in Machine Translation*, IOS Press, Amsterdam, Oxford, Washington.

Bech, A. (1992) 'PaTrans-projektet', in: H. Holmboe (ed.) *Proceedings from Datalingvistisk Årsmøde nr. 2*, 1992, Århus Business School.

Bergenholtz, H., (1990). 'DK87-DK90: Dansk korpus med almensproglige tekster', in: M. Kunøe & og Erik Larsen (eds.) *3. Møde om Udforskning af Dansk Sprog*, Aårhus Universitet

Boje, F. & L. Schøsler (eds) (1992). 'DISEM, a Semantic MT Component', in: *CST Working Papers no.1*, Center for Sprogteknologi, Copenhagen.

Braasch, A. (1994). 'There's no Accounting for Taste - Except in Dictionaries', in: *EURALEX Proceedings, 1994*, Amsterdam.

Braasch, A., B. Maegaard, J. Nordborg & A: Neville (1994). 'Domain Restricted Translation', in: H. Bergenholtz, A. Lykke Jakobsen, B. Maegaard, H. Mørk & P. Skyum-Nielsen (eds) *Translating LSP Texts, OFT Symposium*, Copenhagen Business School.

Bredenkamp, A., S. Markantonatou, L. Sadler (1996). 'Lexical Rules: What are they ?', in: *COLING '96 Proceedings, 16th International Conference on Computational Linguistics*, Copenhagen.

Briscoe, T., A. Copestake, A. Lascarides (1995). 'Blocking', in: P. Saint-Dizier, E. Viegas (eds.): *Computational Lexical Semantics*, Cambridge University Press.

Christoffersen, E. (1995). 'Testning af fagsproglige valensrammer i et maskinoversættelsessystem', in: *Udforskning af dansk ordforråd og grammatik, UDOG report no.4*, Center for Sprogteknologi, Copenhagen.

Copeland, C., J. Durand, S. Krauwer & B. Maegaard (ed.) (1991). 'The Eurotra Formal Specifications', in: *Studies in Machine Translation and Natural Language Processing vol. 2*, Commision of European Communities.

Copestake, A. & T. Briscoe (1995). 'Semi-Productive Polysemy and Sense Extension', in: *Journal of Semantics 12*, p.15-67, Oxford University Press.

Cruse, D.A. (1991). *Lexical Semantics*, Cambridge Textbooks in Linguitics, Cambridge University Press.

Daugaard, J., S. Kirchmeier-Andersen (1993). 'Korpusværktøjet VALP', in: *UDOG rapport nr. 1, Skriften på Skærmen.* Århus Business School.

Dini, L. & V. Di Tomaso (1995). 'Linking Theory and Lexical Ambiguity: The Case of Italian Motion Verbs', in: *Quaderni del Laboratorio di Linguistica 9.*

Dorr, B.J. (1993). *Machine Translation: a View from the Lexicon*, The MIT Press, Cambridge, Massachusetts, London.

Dowty, D.R. (1991). 'Thematic Proto-Roles and Argument Selection', in: *Language* 67:547-619.

Durst-Andersen, P. & M. Herslund (1992). 'Prepositional Objects in Danish', in: *Papers from the XIII Scandinavian Conference of Linguistics.* Roskilde.

Fillmore, C. (1966). 'Deictic Categories in the Semantics of 'Come'', in: *Journal of Language and Philosophy vol.2 No.1*, D. Reidel Publishing Company, Holland.

Fillmore, C., (1985). 'Frames and the Semantics of Understanding', in: *Quaderni di semantica VI:2.* University of Bologna.

Goodman, K. & S. Nirenburg (ed.) (1991). *The KBMT Project: Knowledge-Based Machine Translation*, Morgan Kaufmann Publishers, Inc. San Mateo, CA.

Grimshaw, J. (1993). 'Semantic Structure and Semantic Content in Lexical Representation', unpublished ms., Rutgers University, New Brunswick, NJ. (Apud Levin & Rappaport forthcoming).

ten Hacken, P. (1990). 'A Manual for Reading Distinction'. *EUROTRA document*, unpublished.

ten Hacken, P., D. Maas, B. Maegaard (1991). 'Dictionaries in Eurotra', in: C. Copeland, J. Durand, S. Krauwer & B. Maegaard (eds.) *The Eurotra Linguistic Specifications, Studies in Machine Translation and Natural Language Specifications, Vol. 1*, European Commission, Luxembourg.

Hansen, E. (1972). 'Modal interessens - Nu bør det (komme) frem', in: *Danske Studier*, University of Copenhagen.

Hansen, Aa. (1967). *Moderne Dansk, vol I-III.* Det Danske Sprog- og Litteraturselskab, Grafisk Forlag, Copenhagen.

Hauschildt, B., Ole Lauridsen, T. Riiber & H. Søndergaard (1989). 'Rapport over oprettelsen af et dansk og et tysk maskinlæsbart tekstkorpus inden for området genteknik'. *Inititativområdet Fagsprog og Fagsproglig Kommunikation*, Århus Business School.

Heid, U. (1997). *Zur Structurierung von einsprachigen und kontrastiven elektronischen Wörterbüchern*, Lexicographica, Series Maior 77, Max Niemeyer Verlag, Tübingen.

Heid, U. & K. Krüger (1996). 'A Multilingual Lexicon Based on Frame Semantics', *proceedings of AISB 1996 Workshop*, Stuttgart.

Heim, I. (1982). *The Semantics of Definite and Indefinite Noun Phrases*, PhD. thesis, University of Massachusetts-Amherst.

Herslund, M. (1993). 'Transitivity and the Danish Verbs', in: *LAMBDA no. 18*, Copenhagen Business School, Copenhagen.

Hjortø, K. (1936). *Skrift og Tale*, Nyt Nordisk Forlag, Copenhagen.

Hutchins, W.J., & H.L. Somers (1992). *An Introduction to Machine Translation*, Academic Press, London.

Jackendoff, R. (1983). *Semantics and Cognition*, Massachusetts.

Jackendoff, R. (1990). *Semantic Structures*, MIT Press, London.

Jensen, N.O., T. Tams (1993). 'Lemmatiseringsprogrammet CALM', in: H.

Holmboe (ed.) *UDOG Rapport nr. 1, Skriften på Skærmen, nr. 7*, Århus Business School.

Jensen, P.A., C. Vikner (1994). 'Lexical Knowledge and the Semantic Analysis of Danish Genitive Constructions', in: Hansen & Wegener (eds.) *Topics in Knowledge-based NLP Systems*, Copenhagen Business School.

Kamp, H. (1980). 'A Theory of Truth and Semantic Representation', in J. Groenendijk, T. Janssen, and M. Stokhof (eds.) *Formal Methods in the Study of Language* pp. 277-322, Vol. 136, Dordrecht.

Kaplan, R. & J. Wedekind (1993).'Restriction and Correspondence-based Translation', in: *Proceedings of the Sixth European Conference of the Association of Computational Linguistics, 1993*, p.193-202, Utrectht.

Kirchmeier-Andersen, S., B.S. Pedersen, L. Schøsler (1994). 'Combining Semantics and Syntax in Monolingual Dictionaries. Attacking the Enemy from both Flanks', in: *Proceedings from Euralex '94*, Amsterdam.

Lakoff, G. & M. Johnson (1980). *Metaphors We Live By*, The University of Chicago Press, Chicago.

Levin, B., (1993). *English Verb Classes and Alternations.* University of Chicago Press.

Levin, B. & M. Rappaport (1991). 'The Lexical Semantics of Verbs of Motion: the Perspective from unaccusativity' in: Roca, I. (ed.): *Thematic Structure.* Berlin.

Levin, B. & Rappaport Hovav (1995). *Unaccusativity. At the Syntax-Lexical Semantics Interface.* The MIT Press, Cambridge, Massachusetts, London.

Levin, B. & M. Rappaport Hovav (forthcoming).'Building Verb Meanings', in: M. Butt & W. Geuder (eds.) *The Projection of Arguments: Lexical and Compositional Factors*, CSLI Publications, Stanford, C.A.

Lyons, J. (1977). *Semantics*, Cambridge University Press.

Lyons, J. (1995). *Linguistics Semantics, an Introduction*, Cambridge University Press, Cambridge.

Mahesh, K., S. Nirenburg & S. Beale (1997). 'If You Have It, Flaunt It: Using Full Ontological Knowledge for Word Sense Disambiguation', in: R. Mitkov, N. Nicolov & N. Nikolov (eds.) *Proceedings from Recent Advances in Natural Language Processing*, Tzigov Chark, Bulgaria 1997.

Melby, A. (1995). *The Possibility of Language, a Discussion of the Nature of Language, with Implications for Human and Machine Translation*, John Benjamins Publishing Company, Amsterdam and Philadelphia.

Miller, G.A. & P.N. Johnson-Laird (1976). *Language and Perception*, Harvard University Press, Cambridge, Massachusetts.

Montague, R. (1974). *Formal Philosophy*, Yale University Press, R. Thomason (ed.).

Music, B. & C. Navarretta (1996). 'Documentation of the Danish Lingware', *Deliverable E-D8-DK*, European Commission.

Nimb, S. (1993). *En formel semantisk analyse af danske 'Der'-sætninger'*, Master's Thesis, University of Copenhagen.

Nirenburg, S. (ed.) (1987). *Machine Translation, Theoretical and Methodological Issues*, Carnegie Mellon, Cambridge University Press, Cambridge, London, New York.

Olsen, S. (1997). *Salido el gato, bailan los ratones, En undersøgelse af den syntaktiske manifestation af inakkusativitet på spansk.* Master's Thesis, University of Copenhagen.

Paggio, P. (1996). *The Treatment of Information Structure in Machine Translation.* PhD dissertation, University of Copenhagen.

Pedersen, B.S. (1993). 'Lexical Semantics and Preference', in: P. Bennett & P. Paggio (eds.) *Preference in Eurotra. Studies in Machine Translation and Natural Language Processing, vol. 3*, Luxembourg.

Pedersen, B.S. (1997). 'Danish Motion Verbs: Syntactic Alternations and the Hypothesis of Semantic Determination', in: U. Nikkanne (ed.)*Nordic Journal of Linguistics*, Oslo.

Pedersen, B.S. (1997a). 'Lexical Ambiguity in Machine Translation: Using Frame Semantics for Expressing Regularities in Polysemy', in: R. Mitkov, N. Nicolov & N. Nikolov (eds.) *Recent Advances in Natural Language Processing*, Tzigov Chark, Bulgaria.

Pedersen, S.N. (1994).'Overvejelser om franske og danske reciprokke støtteverbumskonstruktioner', in: H. Kromann (ed.) *Studier i fagsproglig grammatik og pragmatik, ARK vol. 17*, Copenhagen Business School.

Perlmutter, D.M. (1978). 'Impersonal Passives and the Unaccusative Hypothesis', in: J.J.Jaeger et al. (eds.), *Proceedings of the 4th Annual Meeting of the Berkeley Linguistics Society*, UC Berkeley, pp.157-189. (Apud Levin & Rappaport 1991).

Pollard, C. & I. Sag (1987). *In formation-Based Syntax and Semantics*, CSLI, Stanford.

Pollard, C. & I. Sag (1994). *Head-Driven Phrase Structure Grammar.* University of Chicago Press.

Pustejovsky, J. (1995). *The Generative Lexicon*, The MIT Press, Cambridge, MAssachusetss, London.

Roberts, R. & C. Montgomery (1996). 'The Use of Corpora in Bilingual Lexicography', in: *EURALEX proceedings 1996*, Göteborg.

Schmidt, P. (1994). *Basic Decisions on Type and Feature System and on the Core Formalism*, MLAP 93-15 deliverable 1, European Commission.

Schneider, T. (1993). 'The METAL system. Status 1991', in: S. Nirenburg (ed.) *Progress in Machine Translation*, IOS Press, Amsterdam, Oxford, Washington DC.

Schøsler, L. (1994). 'Feature Analysis of Danish Pronominal Paradigms with a View to a Danish Application of the Pronominal Approach', in: *International Journal of Lexicography vol.7*,no.2:118-141.

Schøsler, L. & S. Kirchmeier-Andersen (1997). *Studies in Valency 2. The Pronominal Approach Applied to Danish*, RASK Supplement vol.5, Odense University Press.

Simpkins, N.K.(1994). *ET-6/1 Linguistic Formalism, User Guide*, Cray Systems, Luxembourg.

Skautrup, P. (1927). *Et Hardsysselmål, Ordforråd 1-2*, Krohns Bogtrykkeri, Copenhagen.

Slobin, D. (1996). 'Typology and Rethoric: Verbs of Motion in English and Spanish', in: Shibani and Thomsen (eds.) *Grammatical Constructions: Their Forms and Meanings*, Oxford University Press.

Talmy, L. (1975). 'Semantics and Syntax of Motion', in: J.Kimball (ed.) *Syntax and Semantics vol.4*, Academic Press, London.

Talmy, L. (1985). 'Lexicalization Patterns: Semantic Structure in Lexical Forms', in: T. Shopen (ed.) *Grammatical Categories and the Lexicon vol. 3*, Press Syndicate of the University of Cambridge.

Tenny, C. (1995). 'How motion verbs are special: The interaction of semantic and pragmatic information in aspectual verb meanings', in: M. Dascal (ed.) *Pragmatics and Cognition Vol. 3 (1) p. 31-73*, John Benjamins Publishing Co.

Tenny, C. (1995a). 'Modularity in Thematic versus Aspectual Licensing: Path amd Moved Objects in Motion Verbs', in: *Canadian Journal of Linguistics* 40(2).

Teo, E., C. Ting, H. Lee, L. Peh (1997). 'Probalistic Word Sense Disambiguation: a Portable Approach Using Minimal Knowledge', in: R. Mitkov, N. Nicolov & N.

Nikolov (eds.) *Proceedings from Recent Advances in Natural Language Processing*, Bulgaria 1997.

Theofilidis, A., M. Verhagen & T. Badia (1994). 'Specifications for a Common Semantic Representation Format'. *Project Report: D-WP1a, LRE 61029*, Luxembourg.

Vikner, C. & P.A. Jensen (1994). 'Knowledge Bases and Generative Lexicons', paper presented at *Workshop on Lexical Semantics, November 11*, University of Copenhagen.

Weilgaard, L. (1997). 'Danske partikelverber', in: B. Maegaard & B. Pedersen (ed.) *UDOG no. 6*, Center for Sprogteknologi, Copenhagen.

Wierzbicka, A. (1985). *Lexicography and Conceptual Analysis*, Karoma Publishers, Inc.

Wierzbicka, A. (1996). *Semantics, Primes and Universals*, Oxford University Press, Oxford, New York.

Zaenen, A. (1993). 'Unaccusativity in Dutch: Integrating Syntax and Semantics' in: J. Pustejovsky (ed.) *Semantics and the Lexicon.* Kluwer Academic Publishers, Dordrecht.

Yarowsky, D. (1992). 'Word-Sense Disambiguation Using Statistical Models of Roget's Categories Trained on Large Corpora', in: *Proceedings of the 14th International Conference on Computational Linguistics (COLING92)*, p.454-460, Nantes.

Ørsnes, B. (1995). *The Derivation and Compounding of Complex Event Nominals in Modern Danish - an HPSG Approach with an Implementation in Prolog.* PhD Dissertation. University of Copenhagen.

Dictionaries:

*Collins Cobuild Dictionary* (1987). Collins ELT, London.

*Dansk-Engelsk Ordbog* (1985). Vinterberg & Bodelsen, Gyldendal.

*Engelsk-Dansk Ordbog*, (1985). Kjærulff Nielsen, Gyldendal.

*Nudansk Ordbog* (1986). Politikens Forlag, Copenhagen.

*Ordbog over Det Danske Sprog*, bind 1-29, (1929-1992). Gyldendal, Copenhagen.

*Retskrivningsordbogen* (1986 and 1997). Dansk Sprognævn, Gyldendal, Copenhagen.

# Appendix A: Printout of the dictionary of motion verbs

For appendix A, see ALEP files.

# Appendix B: Printout of lexical macros

For appendix B, see ALEP files.

# Appendix C: The test corpus

For Appendix C, see ALEP files.