Universiteit
Leiden
The Netherlands

# Under construction : cognitive and computational aspects of extended lexical units

Poß, M.

**Citation**
Poß, M. (2010, May 11). *Under construction : cognitive and computational aspects of extended lexical units*. *LOT dissertation series*. Utrecht. Retrieved from https://hdl.handle.net/1887/15570

| | |
|---|---|
| Version: | Not Applicable (or Unknown) |
| License: | Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden |
| Downloaded from: | https://hdl.handle.net/1887/15570 |

**Note:** To cite this publication please use the final published version (if applicable).

# Under Construction

## Cognitive and Computational Aspects of Extended Lexical Units

Published by
LOT
Trans 10
3512 JK Utrecht
the Netherlands

Phone: +31 30 253 6006
Fax: +31 30 253 6000
e-mail: lot@let.uu.nl
http://wwwlot.let.uu.nl/

Cover illustration: Industrial wasteland, Landschaftspark Duisburg Nord. Copyright © 2010 Nicole Karepin.

ISBN: 978-94-6093-027-0
NUR: 616

Copyright © 2010 Michaela Poß. All rights reserved.

# Under Construction
## Cognitive and Computational Aspects of Extended Lexical Units

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. P.F. van der Heijden,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 11 mei 2010
klokke 11.15 uur

door

Michaela Poß

geboren in 1976 te Duisburg, Duitsland

Promotiecommissie

| | |
|---|---|
| Promotores: | Prof. dr. Arie Verhagen |
| | Prof. dr. Ivan A. Sag (Stanford University) |
| Co-promotor: | Dr. Ton van der Wouden |
| Overige leden: | Prof. dr. G.E. Booij |
| | Dr. C.L.J.M. Cremers |
| | Prof. dr. L. Steels (Vrije Universiteit Brussel) |

*Trotz und für Jonathan*

# Contents


**1 Introduction** **1**
- 1.1 Idiosyncrasies everywhere . . . . . 1
- 1.2 Theoretical prerequisites . . . . . 3
- 1.3 What's in the book? . . . . . 5

**2 ELUs and grammar** **9**
- 2.1 Constructions . . . . . 10
  - 2.1.1 Semantic contribution of constructions . . 12
  - 2.1.2 Phrasal or lexical constructions . . . . . 15
- 2.2 Sign-Based Construction Grammar . . . . . 19
  - 2.2.1 Concepts . . . . . 20
  - 2.2.2 SBCG and CxG . . . . . 26
  - 2.2.3 SBCG and HPSG . . . . . 28
- 2.3 Delilah . . . . . 30
- 2.4 Summary . . . . . 33

**3 NCoN and NPN in Dutch** **35**
- 3.1 Introduction . . . . . 35
- 3.2 Zero Semantics . . . . . 36
- 3.3 The NCoN construction . . . . . 38
- 3.4 The NPN construction . . . . . 48
- 3.5 Conclusion . . . . . 73

**4 The Way-Construction** **77**
- 4.1 A construction in its own right? . . . . . 78
  - 4.1.1 Formal properties . . . . . 79
  - 4.1.2 The elements of the way-construction . . . 90

4.1.3 Semantic properties . . . . . 97
4.2 The TLC . . . . . 98
4.3 The *weg*-construction without *weg* . . . . . 99
4.3.1 Non-referentiality . . . . . 103
4.3.2 A matter of aspect . . . . . 105
4.4 Formal analysis . . . . . 109
4.5 Conclusion . . . . . 116

**5 The Dative Alternation 119**
5.1 The data . . . . . 120
5.1.1 Dative alternations . . . . . 120
5.1.2 The distribution of the alternating patterns 122
5.1.3 The method . . . . . 123
5.1.4 The results . . . . . 124
5.2 Theoretical considerations . . . . . 129
5.2.1 Polysemy versus monosemy . . . . . 129
5.2.2 Categorical versus gradient grammar . . . 132
5.2.3 Factors that influence the alternation . . . 134
5.3 Analysis of the corpus results . . . . . 137
5.3.1 Dutch . . . . . 137
5.3.2 English . . . . . 146
5.3.3 German . . . . . 148
5.4 Discussion . . . . . 152
5.4.1 Polysemy versus monosemy . . . . . 155
5.4.2 Categorical versus gradient grammar . . . 160
5.5 Formal analysis . . . . . 162
5.6 Conclusion . . . . . 169

**6 Conclusion and Outlook 173**

# Acknowledgements

Since every single Leiden dissertation starts with the remark that it is a pity that we are not allowed to thank our supervisors in the acknowledgements, I will give in to the pressure and say the following: it is a pity that we are not allowed to thank our supervisors in the acknowledgements. But wait. We *are* allowed to thank our supervisors in the acknowledgements now, which is a very sensible development in the ancient protocol.

But I have been warned not to exaggerate too much. So I kept it simple, which was not simple enough, so I will keep it even shorter: Ton, Ivan, Arie: thank you.

Then there is one person who earned very special thanks, Crit Cremers. Crit, the the ideas you've implanted and the support that you've given, the hours of discussion and the beer should have catapulted you in the *eredivisie*.

I also want to thank Geert Booij and Luc Steels for reading the manuscript and providing valuable comments. I had inspiring and long discussions with Adele Goldberg, Stefan Müller, Manfred Sailer, and Gertjan Postma, which maybe not always convinced me to change my opinion, but which eventually enabled me to formulate my ideas.

A handful of people could be described as my 'stable linguistic and nightlife peer group' in the Netherlands. Frank, Hilke, Diana, Eric, Pepijn, Mark, Felix, and Martin, good times.

Olga and Machteld, Gea, Jos, Lisa, Claartje, Boban, Káča, Anikó, Chris, Birgit, Sylvia, Liesbeth, Jenny, Egbert, Stella, Maarten, Juliette, Alwin, Kristin, Michael, Chris, Johan, Thilo, Niels, Rinus, Luis, Jenneke, Jeroen, Katrien, Jesse, Bob, Michel, Adrienne, Tom, and Vivien should not remain unmentioned, ei-

ther. So shouldn't the people that I most likely forgot on this list.

The list of Stanford and San Francisco people would be comparably long, but I restrict myself to Joan Bresnan, Beth Levin, Elizabeth Traugott, Penny Eckert, John, Hal, Neal and Natalie, Laura, Anubha, Diane, Sabine, Lynn, and Mark. Not to forget Elena, Stuart, and Andy from the coffee house, where approximately 90% of this dissertation were written. Also thanks to the Dead Tongues, the most sophisticated rock cover band in Linguistics, as well as Joseph Magdalena, the piano bar man, for letting me sing when I was drunk.

Back in Germany, there are a couple of people who suffered from my approach to writing a dissertation, hence I want to address them as well. Hendrik and Lea Cyrus, Judith and Rüdiger Weidenfeld-Kramer, Inge Harmes, Anne Haves and Frank Menken, Prof. Amand Berteloot and Prof. Wolf Paprotté, thank you for embedding graphics in the document, kicking my ass when necessary, babysitting, and, first and foremost, convincing me that Linguistics is something I want to do.

I want to thank Nicole and Rolf Karepin for allowing me to use this great picture for the front page. Years ago, Ton and I had decided to take a photo of my dog's tail, but as it goes with pugs, the dog decided differently.

Then there are my fantastic paranimfs, Hilke Reckman and Sigrid Jansson. Thank you two for the good times we had in the past, those that are yet to come, and particularly thank you for standing next to me during the defense (and doing all the other stuff, too). I'm excited about the picture with the short dark-haired girl between the two long blondes, it couldn't be more perfect.

Oh, before I forget: not a single decision I made in my life was as good as going to Leiden to work with Ton. Thanks.

Last but not least I want to thank my family. Needless to say that parents are most crucial on the way from the cradle to the defense. Mom, dad, and my sisters Heidi and Claudia, thanks for all the opportunities I were given. And then there is this new reading of 'family', my nuclear one. Ralf and Jonathan,

thank you for being there. Although having a baby does not exactly speed up the process of finishing a dissertation, it was Ralf who made it possible for me to spend hours typing without hearing 'lalalalalalalala' the entire time. Thanks for that. And all the rest. Germany, Leiden, Stanford, Leiden, Germany. My life could not have been better.

# Chapter 1

# Introduction

## 1.1 Idiosyncrasies everywhere

Natural language does, unfortunately, not always behave as regular as it would be convenient for grammar writers. There are irregularities and idiosyncrasies in all possible areas. To give a few examples:

- Phonology: The British town Worcestershire is pronounced [ˈwʊstəʃə˞], although we would expect, based on the morphemic parts of the word, that it is pronounced as [ˈwɜːsəstəˌʃaɪə]
- Morphology/composition: In German, the *Feuerwehr* (fire-fight-off, 'fire department') is the institution that fights off fire, whereas the *Bürgerwehr* (civilians-fight-off, 'neighbourhood watch') is not the institution that fights off civilians.
- Lexicon: The two most-cited examples for idiomatic expressions are *kick the bucket* and *spill the beans*. Although both can have a literal interpretation, they tend to mean *die* and *tell a secret*
- Syntax: In English, coordination is only possible if both conjuncts are of the same type. Still, there are examples

like *by and large*, where an adverb and an adjective, two different types, are coordinated.

These cases are to exemplify the range of language phenomena where pure knowledge about words and rules is not sufficient. In order to know how to pronounce Worcestershire, or what the connections are between the compound parts in *Feuerwehr* and *Bürgerwehr*, the speaker has to have some extra-knowledge about the words or structures.

The search for general structure is not the only thing in linguistics that yields interesting results. There are many idiosyncratic phenomena that are worthwile investigating, and it is not only in recent years that researchers started analyzing and writing about 'weird things'. For Dutch, Paardekooper (1956) was one influential paper that showed that there is systematicity beyond regular structures. Many others followed in that tradition, among others Postma (1996), Verhagen (2003), or van der Wouden (2005), (2001), and (2007), who all conducted extensive research on little phenomena and thereby helped getting a better idea of the bigger picture. This dissertation is supposed to follow this tradition as well. It does not aim at explaining the big picture, but it tries to shed some light on three single phenomena that we find in the Dutch language.

The research presented here has been carried out as part of the NWO Vidi project "Dutch as a Construction Language",[1] that was granted to Dr. Ton van der Wouden. The aim of the project was twofold: creating a corpus-based inventory of all the linguistic building blocks of Dutch that are larger than words and provide a grammatical description of the found items on the one hand, and researching the possibilities for an implementation of these "constructions" in a lexicon-based parser/generator on the other. For the results of the former part of the project, see Van van der Wouden (to appear 2010), the results of the implementation part of the project are presented in this dissertation.

[1] NWO grant 276-70-003.

## 1.2 Theoretical prerequisites

Let me first say a few things about the understanding of language that forms the basis of this research. From the point of view that is taken here, linguistic knowledge, including the abovementioned types of idiosyncrasies, is entirely acquired and stored. In other words, the speaker learns about the special meaning or function or combinatoric properties of a linguistic item and subsequently makes this information part of her lexical knowledge. In that sense, the term 'lexical knowledge' is used rather widely. It does not just refer to the knowledge of bounded lexical items, but I use it as the term for everything that is acquired and stored, be it the meaning of a word, or how to interpret a certain sentence pattern, like the double-object construction, or something inbetween, like the idiom *kick the bucket*. All these items have to be acquired at one point, and subsequently, they need to be stored. As I will line out in the next chapter, in the constructionist view that I adopt, there are no boundaries between lexicon and grammar, because lexical and grammatical knowledge are based on the same kind of learning. Everything is stored in the lexicon, be it small lexical items or larger building blocks.

The linguistic items I will tackle, I consider to be 'Extended Lexical Units' (ELUs, see Poß & van der Wouden (2005)). The idea behind this term is quite simple: ELUs are all kinds of linguistic items that we consider acquired and stored, and that are larger than just one word. ELUs are wider than, e.g., multiword units, since they also include abstract patterns that have no clearly fixed lexical items. In the three case studies that I will present in this work, I will deal with lexically fixed items as well as schematic and abstract patterns. Nevertheless, all the phenomena described qualify as ELUs from my perspective, as they all are assumed to be part of the lexicon.

This view is compatible with and clearly influenced by the so-called constructionist approaches, a research stream that emerged in the last 20 years, and that blossoms in many different flavours these days. One of the hallmarks of all the different constructionist approaches (I will go into a bit more detail in

chapter 2), is exactly this idea: the building blocks of language, be it a word or an abstract, schematic pattern, are stored in a network of constructions that can combine with each other. In that sense, constructionist grammars are not words-and-rules approaches, but constructions-and-rules approaches, since the main linguistic items, constructions, adhere to rules that concern the combinatorics of the constructions.

One question that will not be dealt with in this dissertation is where the knowledge about linguistic structures and idiosyncrasies comes from. I will not take an issue with the argumentation of my point of view, but I will follow that stream of research within cognitive linguistics that supports the idea that not necessarily language-specific learning mechanisms are at stake when it comes to acquisition. A detailed discussion of concrete ideas will be beyond the scope of this work, but I can refer the reader to Tomasello (2003) for a usage-based approach to language acquisition.

Constructionist approaches emerged as an attempt to solve the problem of idiosyncrasies in language. The seminal paper by Fillmore *et al.* (1988) dealt with the *let alone*-construction (as in: *He isn't smart, let alone handsome!*), showing that it is important to give room in a grammar model to cases where syntax and semantics have a clear mismatch. The same was done 11 years later by Kay & Fillmore (1999), when they tackled the *What's X doing Y*-construction (as in: *What's the fly doing in my soup?*), and more empirical studies support the idea that dealing with idiosyncrasies is crucial for understanding language on a larger scale.

In the literature, estimates concerning the quantitative importance of ELUs differ greatly. Sprenger (2003) reports that some 10% of the content words in a corpus of Dutch newspaper text was part of some larger lexical entity, whereas Altenberg (1998:102) writes: "A rough estimation indicates that over 80 percent of the words in the corpus form part of a recurrent word-combination in one way or another." Bybee (2005) takes a middle position, citing Erman & Warren (2000) who found that what they call prefabricated word combinations constitute about 55%

of both spoken and written discourse". Sag *et al.* (2001) offer a comparable estimate.

But not only from a cognitive perspective an integrated approach to ELUs is desirable, also for Natural Language Processing, a straightforward solution for idiosyncracy problems helps gaining wider coverage of systems. If the amount of mismatches and ELUs is so enormous, how could (rule-based) NLP applications ever reach full coverage if they cannot tackle this large subset of possible and naturally occuring structures?

Hence, it seems natural to take a grammar model that is designed to cope with idiosyncrasies as the basis of computational language analysis. But this is where it gets problematic. The constructionist approaches are not one homogeneous framework, yet, but appear to be rather scattered. They are a family of models based on a small set of common tenets, rather than a well worked-out theory. On top of that, many researchers within the constructionist frameworks come from a more cognitive background and are not very concerned with formal grammar models. In itself, this is not doing any harm to the framework, since it is not necessarily impossible to have deeper insights into languages without applying a formal analysis. On the other hand, the lack of a formal model makes it difficult to translate these insights to NLP, and to utilize them in order to reach wider coverage.

## 1.3 What's in the book?

What I want to achieve in this dissertation is twofold. First of all, I will tackle a number of theoretical questions and outline the grammar model that I will apply. All this will happen in chapter 2. The questions I will raise are topics that I consider important for a constructional analysis and that are either not common ground in mainstream construction grammars, or that are still debated amongst researchers. So I will lay out what I consider a construction, how semantics and syntax are related to each other within a construction, and I will argue that the constructions that I discuss are lexical in nature.

Then, I will turn to the question what will be an appropriate formal grammar to model ELUs and constructions. I do not need to reinvent the wheel in order to answer that question, but I adopt the model of Sign-Based Construction Grammar (SBCG) (see Sag *et al.* (2003), Sag (2007b)). What makes this model extremely suitable is the fact that it, like the strongly related Head-Driven Phrase Structure Grammar (HPSG), is a mathematically solid theory that makes use of structure sharing and unification as the main mechanism of structure building. HPSG has successfully been implemented for many languages, and a lot of computational issues of the model have been solved. Therefore, it can serve extremely well as the basis for other models that aim at an implementation. Since the focus of this research lies on constructions that are organized on lexical level, I will zoom into SBCG on that part only.

Since I did practical research with the semantic parser/generator Delilah (Cremers (2004)), I will shortly introduce that system as well. Delilah's architecture, I will argue, allows for a model of linguistic phenomena that is in accordance with a formal constructionist approach. The main characteristic that I consider important is the use of a lexical type hierarchy, which is part of both formal construction grammar and the Delilah architecture. Delilah was developed by Crit Cremers and Maarten Hijzelendoorn at Leiden University, and it profited greatly from the work by Hilke Reckman (see Cremers (2004) and Reckman (2009) for more detailed information. A stable version of this system can be tried out at www.delilah.eu). I will give more detailed information on the Delilah system in the next chapter.

In the more practical part of this dissertation, I will tackle a number of linguistic phenomena in order to shed some new light on their behavior. I chose three (families) of constructions that have been treated in the literature already. The first set of related constructions, presented in chapter 3, I refer to as the NCoN and NPN-construction. Since both phenomena have been tackled for Dutch by Postma (1996) in a Government and Binding framework, I use these constructions as a means to oppose

my lexical approach to an existing, generative-transformational one. I base my analysis of the NPN construction on a dataset that I extracted from the Corpus Gesproken Nederlands (CGN), the corpus of spoken Dutch.[2] The use of this chapter will be to provide a new analysis of the data for Dutch on the one hand, and to show that, for NCoN and NPN, a lexical approach is more appropriate to capture the empirical facts.

In chapter 4, I will turn to a construction that has been one of construction grammarians' favourite: the *way*-construction. Much has been written about it, be it for English or Dutch, so it almost seems superfluous to dedicate yet another book chapter to this phenomenon. But my reasons are valid, since I will present an entirely new analysis of the Dutch version of the *way*-construction. All previous approaches take the same semantic content as a starting point, but I will provide arguments that there are syntactic and semantic reasons to assume a different semantic structure of this pattern.

Chapter 5 will be different from the previous one in two ways: first of all, it is much more data-intensive than the two other case studies. And secondly, the chapter does not only deal with Dutch, but it takes a contrastive perspective and compares Dutch, English, and German. The constructions that will be tackled are the so-called Dative Alternation, which are rather well-researched for English and Dutch, and barely for German. This is not very surprising, as the prepositional construction is very marginal in German. Nevertheless, even though it is scarce, it does naturally occur, and therefore it is important to include German in this study.

The empirical side of this chapter is particularly interesting. I looked for comparable datasets in the three languages, found it in the Europarl corpus (Koehn 2002), and extracted all instances with the verb *give* and its German and Dutch counterparts, *geben* and *geven*, from the corpora. A lot of manual work has been involved in this process, and the categorization and annotation is not very finegrained. Still the empirical difference between the distribution of the two constructions is significant across the

[2]See Oostdijk (2000).

three languages. To my knowledge, this is the only empirical contrastive study yet.

For all three case studies, I will argue for a words-and-constructions model. Eventually, I will mainly stick to the following principles: the lexicon consists of lexical and constructional entries, and constructional entries are the combination of a mother and a daughter, both of which are specified for their semantic and syntactic properties. The properties of the mother are a function of the properties of the daughters by default, but they can be overridden by construction-specific properties (like constructional semantics, see Sag *et al.* (2003:480)).

All three constructions will receive a lexical analysis. In this respect, this dissertation deviates from many constructionist approaches to the topics under discussion, since at least the *way*-construction and the dative alternation have been discussed in terms of phrasal constructions before.

Additionally, I will present the results of the implementation of these constructions in the semantic parser/generator *Delilah*. Since it is one goal of this dissertation to show that computational systems can profit from the implementation of constructionist mechanisms, the *Delilah* code of every construction discussed will be provided, as well.

# Chapter 2

# ELUs and grammar

Grammatical theories often lack a proper and principled way of treating ELUs like idioms, prefabs, etc., which is very unsatisfying, as an entire domain of naturally occuring language cannot be explained and modeled. But lately, much has been written about the importance of all kinds of collocations, idioms and multi-word units, from a theoretical and an applicational point of view. ELUs not only need to be given a place in grammar theories, they also challenge Natural Language Processing applications. One of the major obstacles in NLP is the treatment of multiword units (or expressions). As Sag *et al.* (2001) put it explicitly: they are a "pain in the neck" for NLP. Wide coverage by analytical systems can only be reached if the coverage of 'core' grammar is broadened to include 'peripheral', more idiosyncratic patterns as well. A proper theory of language should not confine itself to core phenomena and thereby neglect a large part of naturally and frequently occurring phenomena.

One recent theoretical stream that has emerged in various styles and places and that tries to integrate ELUs into a grammatical theory is the constructionist approach. As it is impossible to talk about one homogeneous framework, I will outline the major points which I consider important in the following section.

## 2.1 Constructions

Since the 1980's, linguistics has seen the rise of a number of grammatical theories in which the classical notion of construction has a central position. As there are different constructionist approaches on the market, it is not reasonable to speak of Construction Grammar as a homogeneous framework. Langacker (2003) distinguishes three major streams within CxG: Cognitive Grammar (Langacker (1987), (1991)), Construction Grammar (Goldberg (1995), Fillmore *et al.* (1988)), and Radical Construction Grammar (Croft (2001)). It is not the objective of this book to work out the details of and the differences between these approaches, as much has been written about that topic (see, e.g., Goldberg (2003), Langacker (2003), or Croft (2004) for overviews of the wide field of Construction Grammar approaches). Nonetheless, all these approaches share a bundle of basic assumptions, which are, among others, nonderivationality, monostratality, unity of grammar and lexicon, a cline from schematic to specific constructions that are all stored in the lexicon, the linking of constructions in an inheritance network, and unification as the method for the composition of structure.

A clear definition of the notion construction, on the other hand, is less easy to find. For Langacker, who, at least in his Cognitive Grammar model, does not use the specific term *construction*, the fundamental unit in language is the *symbolic assembly*, a form-meaning pairing of varying schematicity with a semantic and a phonological pole, stored in a structured inventory of conventional symbolic units (Langacker (1987)). Also in Croft's (2001) Radical Construction Grammar (RCG), the construction is the basic unit in the grammar, and everything *is* a construction. Grammatical constructions "consist of pairings of form and meaning that are at least partially arbitrary" (Croft (2001:18)). But this arbitrariness holds for any, even the most abstract, configurations, for in Croft's model, all structure building and interpretation is conventionalized.

As representative for the third branch of constructionist ap-

proaches, next to Cognitive Grammar and Radical Construction Grammar, I take the model proposed by Goldberg. In her 1995 book, she gives a definition of the notion 'construction' that reads as follows:

"C is a CONSTRUCTION iff$_{def}$ C is a form-meaning pair <$F_i$, $S_i$>, such that some aspect of $F_i$ are some aspect of $S_i$ is not strictly predictable from C's component parts or from other previously established constructions." (Goldberg (1995:4))

In her 2006 book, her definition becomes broader, with a more cognitive touch, but even less formal:

"All levels of grammatical analysis involve constructions: learned pairings of form with semantic or discourse function" (Goldberg (2006a:5))

Here, all that is left to a construction is that it is a learned form-meaning pair. Cognitively, that is useful, as it adopts insights from language acquisition work that has been done in this framework, like Tomasello (2003) and (2006). From a formal perspective, this definition is nevertheless not satisfying, as it is too broad and too unrestrictive.

A formally more adequate way to define constructions has been proposed in Sign-Based Construction Grammar (SBCG), as described among others in chapter 16 of Sag *et al.* (2003), Sag (2007a), and (2007b). This version of construction grammar tries to incorporate constructionist ideas into the (formally worked out) framework of Head-Driven Phrase Structure Grammar (HPSG, see Pollard & Sag (1994) and Sag *et al.* (2003)). A construction in SBCG is a constraint on the combination of signs (which can be lexical or phrasal), and it specifies features of the daughters (or the input) and the mother (the output).

The SBCG architecture of grammar is slightly different from the one in HPSG, but the changes made allow the grammar to deal with more idiosyncratic phenomena more straightforwardly

than an HPSG does, since it allows for more fine-grained combinatoric mechanisms.

But before I go into the details of the model, I want to turn to a couple of issues that need to be made explicit when working with a constructionist approach.

### 2.1.1 Semantic contribution of constructions

In this section, I want to discuss a series of problems that are of central importance to a constructional approach. In a nutshell, the question is whether it is the lexical item that provides the critical information about its arguments and associated structures, or whether abstract phrasal entities are part of a speaker's knowledge. From an acquisition point of view, that means: Does the speaker acquire knowledge about, let's say, verbs and the configurations they can occur with, or does she acquire knowledge about the function and use of sentence patterns.

One assumption that has been debated by several researchers is the issue of whether—if abstract phrasal schemas are constructions as well and therefore are a form-function pairing—phrase structure alone must carry meaning. Similar ideas have also been formulated in transformational frameworks (e.g., by Postma (1996), see next chapter). As Goldberg (1995:1) puts it, "[a] central thesis of this work is that basic sentences of English are instances of *constructions*—form-meaning correspondences that exist independently of particular verbs. That is, it is argued that constructions themselves carry meaning, independently of the words in the sentence." But this is misleading. Phrasal patterns alone are not the source of semantic interpretation. Only in combination with more (constructional) information, a structure can be interpreted. Meaning is part of the whole. Phrasal schemas are generally far too unspecific to be able to function as the bearer of meaning by itself. Let me make that clearer with an example discussed by Goldberg (1995) and Goldberg & Jackendoff (2004).

In her 1995 book (chapter 8), Goldberg defends the claim that a separate resultative construction needs to be posited in

order to be able to account for the syntactic and semantic idiosyncrasies of sentences like *He hammered the metal flat*, independently of the verbs that appear in the construction. Her representation of the resultative construction thus looks like this (Goldberg 1995:189):

(1) Goldberg's Resultative Construction

$$\left[\begin{array}{lccccccc}
\text{Sem:} & & \text{CA.-BEC.} & < & \text{agt} & \text{pat} & \text{result-goal} & > \\
 & & |R & & | & | & | & \\
\text{R: instance,} & & & & & & & \\
\text{means} & & \text{PRED} & < & & & & > \\
 & & \downarrow & & \downarrow & \downarrow & \downarrow & \\
\text{Syn:} & & \text{V} & & \text{SUBJ} & \text{OBJ} & \text{OBL}_{AP/PP} &
\end{array}\right]$$

I assume that this schema does not mean to express that the structure V SUBJ OBJ OBL carries resultative semantics. Also, very little information is given about the actual phrase structure of this construction, and nothing is said about word order. Instead, the notation of a construction should be understood as follows: certain argument roles occur together, with each role fulfilling a particular function, and with an overall meaning of the construction. Note that there is a clear difference between saying V SUBJ OBJ OBL *means* X and V SUBJ OBJ OBL, with each part fulfilling function A, B, C, and D, respectively, *leads to an interpretation* Y. The former could not distinguish between the sentences

(2) The coach yelled the player off the stage (resultative)

and

(3) The boy finished his dinner in a hurry (not resultative)

as both sentences are of the pattern SUBJ OBJ OBL. The difference between the two is that in the non-resultative version, the OBL does not map to the semantic role of result/goal. Therefore, sentence (3) cannot be an instance of the resultative construction. We might say that patterns can carry multiple meanings,

but the less specific a pattern is (e.g. NP VP), the more meanings it can carry. It does not seem helpful for the hearer/speaker to rely on the meaning of a pattern by itself.

As I said earlier, Goldberg makes no claims about linearization within a construction, at least not in the ones discussed here. This seems to be different in Verhagen's work. His representation of the Dutch *way*-construction (for detailed discussion, see Chapter 5), e.g., looks like this (Verhagen 2003:34):

(4) The Dutch *way*-construction in Verhagen (2003):

$$\left[\begin{array}{lccccc} \text{Sem:} & \text{creator} & \text{create-move,} & \text{for-self} & \text{created-way,} & \text{path} \\ & | & |\text{means} & | & | & | \\ \text{Syn:} & [\text{SUBJ}_i & [\text{V} & [\text{REFL}_i & [\textit{een weg}] & \text{OBL}]]] \end{array}\right]$$

Verhagen applies a bracketed structure on the Syn side, which looks as if he makes claims about the internal ordering of arguments.[1] But pinning down the formal side in such a way comes with a series of problems, some of which are the special treatment that would be needed for adjunction, insertion of modals, control structures, topicalization and any type of extraposition. Especially for languages with freer word order, pinning down the linear structure of a construction rigidly is not useful. Problems arise from two sides: on the one hand, this adoption needs a refined way of dealing with alternate word orders, and on the other hand, a detailed mechanism needs to be worked out, how one construction can interact with other constructions, e.g. passive, fronting, or adjunction. One possible way of dealing with the second issue is to establish constructions like the passive resultative additional to the active resultative, but this leads to an extreme inflation of the inventory, which is very inconvenient and redundant. Wherever there are generalized processes, i.e. interactions that hold for an entire set of entities, the mechanism

[1] Verhagen (p.c.) explains that this schema is not meant to make claims about linearization, but only about the hierarchical structuring of items. Still, intuitively, this representation looks a lot like a linearization pattern and should be rethought. If an internal structure is added to the notation that is used, this leads to the assumption that there is an additional layer of analysis.

should be part of the system rather than enumerating the results.

But let us go back to the resultatives. In their 2004 paper, Goldberg and Jackendoff analyze the resultative as a family of constructions, and their representation of the 'causative property resultative', the one that licenses sentences like *He hammered the metal flat*, looks as follows (Goldberg & Jackendoff 2004:19):

(5) The causative property resultative

| | |
|---|---|
| Syntax | $NP_1$ V $NP_2$ $AP_3$ |
| Semantics | $X_1$ CAUSE [$Y_2$ BECOME $Z_3$] |
| | MEANS: [VERBAL SUBEVENT] |

In this notation, the co-indexation is meant to be responsible for the correct mapping of form and function. And again, as in Goldberg's notation, it is not the pattern NP V NP AP that establishes the construction, but the combination of the syntactic pole with the semantic pole.

To sum up, it is not the phrase structure of an expression that carries meaning, but it is the construction as a whole, the pairing of form and meaning, that is the semantic source. The construction can, but not necessarily has to pin down the syntactic structure of an expression precisely.

### 2.1.2 Phrasal or lexical constructions

A different point that has been discussed on various occasions deals with the question whether constructions are phrasal or lexical in nature. Müller (2006a) argues that a phrasal approach is not superior to a strictly lexical one, and he shows in Müller (2006a) and (2007) that the lexical approach can easily capture the phenomena that are often argued to be problematic under lexical treatments.

The concrete question that Müller raises is the following. In cases like resultatives, there is an additional semantic component that is not contributed by the verb alone, at least not in its default interpretation.

"The question of interest here is: Where does this additional meaning come from? There are two main ways of answering this question.

**Answer 1** It is there since the NP[nom], NP[acc], Pred and V are used in a certain phrasal configuration.

**Answer 2** It is there since a special lexical item selects for NP[nom], NP[acc], Pred and contributes the appropriate meaning." (Müller 2007:375)

Hence, the crucial question here is whether to think of the special additional semantics of the resultative (the causation) as being part of a (phrasal) construction, or the interpretation of a verb that has resultative meaning and selects for subject, object and predicate.

In the lexical approach, this verb meaning is the result of a lexical rule which is applied to a verb without the resultative meaning and which maps this verb to a new one that has the appropriate argument structure and semantic contribution online. In HPSG, using lexical rules is a method to allow for generalizations over a class of lexical items by capturing the relations between the members of this class. Flickinger (1987) distinguishes between horizontal and vertical lexical rules, with the horizontal ones expressing a "systematic relationship holding between two word classes, or more precisely, between the members of one class and the members of another class" (Flickinger (1987:105), cf. Meurers (1997)), and the vertical ones expressing common properties of all words of a certain class. The expected lexicon inflation that Goldberg fears[2] is not problematic, as the lexical rules do not actually create lemmas of the type "paint$_{resultative}$" that are part of the lexicon and which Goldberg calls "implausible verb senses", but they map certain paradigms of verbs onto

[2]As explained in a talk given at the Workshop on Constructions in Grammatical Theory in Stanford, 2007. If all the possible verbs that are created by applying lexical rules, were stored in the lexicon, this would indeed inflate the lexicon to a large degree. Also, it does not seem convenient to have a lemma *sneeze* (plus all the other verbs that can occur in this construction) which has a caused motion argument frame (as in *She sneezed the foam off the cappuchino).*

resultative counterparts when they occur in a suitable environment (for a discussion, see Müller (2007), (To Appear 2010), and the references cited therein.)

One advantage of the lexical approach is that the interaction with other processes in the lexicon, like passive or nominalization, works smoothly with a lexicon-based approach, but would need to be accounted for in a phrasal approach in more detail, since there is no clear proposal that describes the necessary machinery, yet. If there is a resultative (phrasal) construction, how do we arrive at a passive resultative sentence without having to assume transformations? The same holds for passivizable idioms. The idiom *spill the beans* can be passivized ('the beans were spilled yesterday'), or the verb can occur in adjectival position ('the listeners take advantage of the spilled beans').

As Müller (2007) shows, the phrasal approach has certain shortcomings as soon as the more intuitive level is left behind. The problems he discusses and that do not form an obstacle in his lexical approach, are, among others, the insertion of material like modals or adverbs into the structure, discontinuous constituents, and control structures, in other words, anything that makes the pattern deviate from the one that is captured in the construction. He shows that none of the grammatical interactions has a problem in the lexical-rule account.

In Müller (2006b), it is shown that for three German constructions, a lexical approach is to be preferred over a phrasal one. Müller discusses the verbal complex, resultatives, and particle verbs, three phenomena that are analyzed as being phrasal constructions in CxG. Müller demonstrates that for all three constructions, there are recurring problems: the phrasal approach is not able to handle argument order rearrangement, as is very frequent in German as well as in Dutch, passives, fronting, adjunction and derivation (like *Leerfischung*, 'empty-fishing') properly. For the German resultative, he demonstrates, a phrasal analysis would need six different separate constructions in order to account for the word order alternations, eight more construction for the analysis of traceless fronting, plus all of those can be passives, which needed separate constructions as well, plus there

are the derivations of resultative verbs, which needed a specific construction, too. It is obvious that this is not a very elegant solution. In Müller (2006b) and (2007), the lexical rule-analyses for these constructions are sketched, and it is clear that most interactions with other parts of the grammar automatically follow from the lexical rules.

Müller (2007:388) acknowledges that a major point of critique with regard to the Goldberg-style approaches is the fact that analyses have not been fully worked out, yet, and clearly not for a large-scale description. Still, some of the technical problems he points out are more than just a challenge to a formal grammar model, like the failing default inheritance that Goldberg suggests in her work (see Müller (2006a)).

The debate about phrasal versus lexical constructions, as intriguing as it is, seems to be caused mainly by the lack of a common ground between the two approaches. Usually, the CxG oriented studies try to capture insights from cognitive linguistic research and language acquisition in a model that claims cognitive reality as its goal. And in that sense, Goldberg has good reasons to assume that the existence of argument structure constructions needs to be accounted for. In her own experimental research (Casenhiser & Goldberg (2005), Goldberg (2006b), Goldberg *et al.* (2004)) as well as in acquisition research presented in, e.g., Tomasello (2003) and (2006), it has been shown that phrasal configurations can prime and facilitate the interpretation and acquisition of language. Nevertheless, it is true that her model lacks formality and testing and is an intuitive way to formulate observations rather than a theoretic model. Also, it is not clear whether those priming effects really reveal insights about the semantic contribution of phrase structure, or whether it is the result of a hearer's knowledge of the interpretation of a (nonsense-)verb's argument structure.

Müller, on the other hand, emphasizes the point that lexical approaches readily provide an analysis of the phenomena that are usually mentioned as the pillar arguments for a constructionist or phrasal approach. For the resultative construction, he refers to existing analyses (Müller 2007). He does not tackle, however,

the question of cognitive reality, as this is not, at least not explicitly, part of his approach. Müller aims for a coherent and elegant analysis of the relevant data, in which he succeeds, and Goldberg aims at a representation of insights gained in cognitive and experimental research. She wants to model processes rather than data. Though she does not completely succeed in modelling the facts she wants to line out appropriately, she does not pursue a lexical analysis as it contradicts her basic assumption that speakers have abstract knowledge of phrasal configurations, and not only of argument frames and processes that are possible on verbs. Müller certainly has a strong argument when he points out that a lexical rule approach can model the data more elegantly. Still, not everything in nature is necessarily elegant or most efficient. And not all attempts to model natural processes manage to use the same strategies that naturally occur to reach their goals.[3]

Nevertheless, the focus of this dissertation lies on the lexical entries and lexical organization of grammar. In the case studies that I will present in the next three chapters, I will mainly discuss the organization of lexical entries. I will make use of the advantage that with a lexical treatment, interactions with other parts of the grammar naturally follow from the analysis. Whether that resembles some sort of cognitive reality, will remain undecided.

## 2.2 Sign-Based Construction Grammar

One approach that is trying to bridge the gap between the two positions discussed above, modelling the data properly on the one hand, and working towards a cognitively reasonable theory on the other, is the model of Sign-Based Construction Grammar (SBCG), that is mainly developed in Sag (2007b), (2008), (to appear 2010), Kay & Sag (2009), and other work in Boas & Sag (To

[3] I owe Crit Cremers for an enlightening example of this fact. As he pointed out, everything in nature that can fly uses wings that move up and down in order to do so. Most flying tools that man created, like airplanes or helicopters, manage to copy the act of flying without moving anthing up and down.

Appear 2010). In principle, SBCG is a version of HPSG which allows for phrasal entries (phrasal signs) in the lexicon, in order to be able to deal with expressions that cannot be captured on a lexical level. SBCG is more lexical than the abovementioned constructionist approaches are, since it provides lexical analyses for phenomena like the resultative, passives, the Dative Alternation, or the *way*-construction. And it is more phrasal than HPSG, since it provides phrasal analyses for expressions like for example tag-questions (in order to be able to model sentences like *Sears is open, aren't they*, where there is an unusual agreement relationship). However, I want to point the reader to the synopsis given in Sag (to appear 2010), to gain a better picture of the entire approach. In the following paragraphs, I will only go into those concepts that play a role for my analyses, and I will not go into deeper discussions regarding the notation, or the way of modelling the feature structures. I do not aim at a theory-neutral (whatever that may be) description of my results, but since there is always room for other notations (in a sense, the notation and model used in Delilah is at many places a notational variant of the notation and model used in SBCG work), I will not dive into any questions regarding notation or model-dependent details.

### 2.2.1 Concepts

Sag (2007b:1) sets out to "expand the empirical coverage of HPSG, while at the same time putting BCG[4] on a firmer theoretical footing" . The central object in this framework is the sign, which can be lexical or phrasal. Signs are functions that specify values for the following features (ibid.:17):

[4]Berkeley Construction Grammar, in the sense of Fillmore *et al.* (1988), Kay (2002), among others.

(6) The sign:

$$\begin{bmatrix} \textit{sign} & \\ \text{PHONOLOGY} & \textit{phonological phrase} \\ \text{FORM} & \textit{list(formative)} \\ \text{ARG-ST} & \textit{list(sign)} \\ \text{SYNTAX} & \textit{syn-obj} \\ \text{SEMANTICS} & \textit{sem-obj} \\ \text{CONTEXT} & \textit{context} \end{bmatrix}$$

A construction is a local constraint on sign combination. It specifies a MOTHER (MTR), which is a sign, and DAUGHTERS (DTR), which are lists of signs (Sag 2007b:2):

(7) The construction:

$$\textit{cxt} \Rightarrow \begin{bmatrix} \text{MOTHER} & \textit{sign} \\ \text{DTRS} & \textit{list}\left(\textit{sign}\right) \end{bmatrix}$$

Since a construction specifies information about the mothers and the daughters, it provides a powerful tool to override compositional semantics when necessary. This is one of the characteristics of SBCG that are comparable to the Delilah-model and that is made repeated use of.

Actual language material, an expression or a word, is modeled as a construct, licensed by either a construction of the language, or a lexical entry. Which signs are actually possible in a SBCG is determined by the Sign Principle:

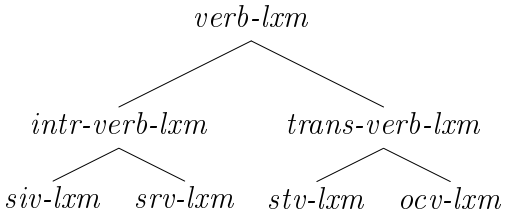
(8) **The Sign Principle**:
Every sign must be lexically or constructionally licensed, where:

a sign is lexically licensed only if it satisfies some entry in the lexicon, and
a sign is constructionally licensed only if it is the mother of some construct

The lexical items in SBCG are divided in two groups: lexical entries and lexical class constructions. The latter define all distinctive properties the members of the class have in common (like a lexical rule in HPSG does), while the lexical entry puts further constraints like for instance the FORM and SEM values on the lexical class construction.

Lexical class constructions are organized in hierarchical lexeme types, with each subtype placing further constraints on its supertype.

(9) Lexeme type hierarchy




On top of the (of course partial) hierarchy is the most general *verb lexeme*, that defines constraints that every single verb needs to obey. Intransitive verb lexemes and transitive verb lexemes are subtypes of the supertype verb lexeme, and intransitive verb lexemes have the subtypes strict-intransitive (siv) and subject-raising (srv) verb lexemes, transitive verb lexemes have the subtypes strict-transitive (stv) and object-control (ocv) verb lexemes, and so on.

The constraints that are placed on the supertype verb lexeme are these:

(10) Verb lexeme[5]:

$$\textit{verb-lxm} \Rightarrow \begin{bmatrix} \text{ARG-ST} & \langle X, \ldots \rangle \\ \text{SYN} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \textit{verb} \\ \text{XARG} \quad X \end{bmatrix} \\ \text{MARKING} & \textit{unmk} \end{bmatrix} \end{bmatrix}$$

[5] All examples are from Sag (to appear 2010).

This feature structure captures the information that the lexical class construction *verb-lxm* is unmarked[6], of the category verb, and that its first argument is an external argument (by identifying the first position of the argument structure list with the value for the XARG, the external argument, which ensures that it is the subject).

Subclasses of the general verb lexeme put more constraints on the number and nature of the internal arguments of a verb. The strict-intransitive lexeme, for instance, does not allow for any more arguments than the subject:

(11) Strict-intransitive lexeme:

$$\textit{siv-lxm} \Rightarrow \begin{bmatrix} \text{ARG-ST} & \langle \text{NP}_i\boxed{..} \rangle \\ \text{SYN} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \textit{verb} \\ \text{SELECT} & \langle\rangle \\ \text{XARG} & \text{NP}_i\boxed{..} \end{bmatrix} \\ \text{MARKING} & \textit{unmk} \end{bmatrix} \end{bmatrix}$$

The lexical class construction of strict-intransitive lexemes is licensed by lexical items that obey the constraints that are formulated in the class description. The strict-intransitive verb *laugh*, for instance, licenses the lexical class of siv-lexemes, but it adds specific information about its particular constraints.

(12) Lexical entry for *laugh*:

$$\begin{bmatrix} \textit{siv-lexeme} \\ \text{FORM} & \langle \text{laugh} \rangle \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & s \\ \text{FRAMES} & \left\langle \begin{bmatrix} \textit{laugh-fr} \\ \text{SIT} & s \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

[6]MARKING is a feature that helps distinguishing between expressions like *Kim laughed* (unmarked) and *that Kim laughed* (marked). The default value of a sign is unmarked, possible values are *that*, *whether*, *than*, etc. See Sag (to appear 2010), section 3.3.3, for more detail.

This way the lexical entry for the verb *laugh* is represented. It is a member of the lexical class construction of strict-intransitive verbs that opens a laugh-frame. Constraints like argument structure are not stated in this entry, since they are determined in the description of the lexical class construction of siv-lexemes.

In order for a lexeme to be able to combine with other material, it must give rise to a word, which is the minimal element of syntactic construction. This is done by combining the lexical entry with the inflectional construction, the pairing of a mother and daughters, where the mother is a word and the daughter is a nonempty list of lexemes:

(13) Inflectional constructions (Sag 2007b:32):

$$\textit{infl-cxt} \Rightarrow \begin{bmatrix} \text{MTR} & \textit{word} \\ \text{DTRS} & \textit{nelist}\left(\textit{lexeme}\right) \end{bmatrix}$$

'Combining' is not meant to express concatenation here, but a way of applying a set of constraints in order to build a word from a lexeme. A preterite (inflectional) construction, e.g., builds the word *laughed* from the lexeme <laugh>. Building a new lexeme from a lexical sign, for instance the lexeme *unhappy* from the lexeme *happy*, is handled via derivational constructions.

(14) Derivational constructions:

$$\textit{deriv-cxt} \Rightarrow \begin{bmatrix} \text{MTR} & \textit{lexeme} \\ \text{DTRS} & \textit{nelist}\left(\textit{lex-sign}\right) \end{bmatrix}$$

While the daughter of in inflectional construction is a lexeme, the daughter of a derivational construction is a lexical sign, i.e., a word or a lexeme. In order to 'build' concrete lexical material that can enter an expression, we need to combine a lexical entry with a word-building construction.

Phrasal constructions are organized the same way that lexical constructions are. They are combinatoric constructions that license expressions built from more than one building block. Just like a inflectional construction builds a word from a lexeme, a

phrasal construction builds a phrase from a list of expressions, that are either words or phrases.

(15) The phrasal construction:

$$\textit{phr-cxt} \Rightarrow \begin{bmatrix} \text{MOTHER} & \textit{phrase} \\ \text{DTRS} & \textit{list}\left(\textit{expressions}\right) \end{bmatrix}$$

Just like lexical-class constructions, the combinatoric constructions are organized in a type hierarchy. Again, subtypes inherit the features of their supertypes and add new constraints to the descriptions. To illustrate higher level phrasal constructions, look at those two examples:

(16) Predicational Head-Complement Construction:[7]

$$\textit{pred-hd-comp-cxt} \Rightarrow \begin{bmatrix} \text{MTR} & \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{VAL} & \langle X \rangle \\ \text{MRKG} & \textit{M:unmk} \end{bmatrix} \end{bmatrix} \\ \text{DTRS} & \left\langle H \right\rangle \oplus \textit{L}_2\textit{:nelist} \\ \text{HD-DTR} & \textit{H}, \begin{bmatrix} \textit{word} \\ \text{SYN} & \begin{bmatrix} \text{VAL} & \left\langle X \right\rangle \oplus \textit{L}_2 \\ \text{MRKG} & \textit{M:} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

(17) The Subject-Predicate construction:[8]

$$\textit{subjpred-cxt} \Rightarrow \begin{bmatrix} \text{MTR} & \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{MRKG} & \textit{M} \\ \text{VAL} & \langle\rangle \end{bmatrix} \end{bmatrix} \\ \text{DTRS} & \left\langle X, \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{CAT} & \left[\text{VF } \textit{fin}\right] \\ \text{MRKG} & \textit{M:unmk} \\ \text{VAL} & \left\langle X \right\rangle \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

[7] Sag (to appear 2010:42).
[8] Ibid.

The Head-Complement construction licenses constructs, where the MARKING value of the mother, which is unmarked, matches that of the head daughter. Also, the head daughter must be followed by all its valents, except the first argument, the subject. In other words, this construction can license VPs built of verbs and their arguments. The VP combines with the subject under the constraints specified in the Subject-Predicate construction.

A simple declarative clause is a well-formed sign if it obeys to these constraints: the mother has an empty valence list, and there are two daughters, the second of which is a finite verbal sign which selects the first one on its valence list.

Generally, constructions in SBCG are descriptions of local trees. The feature structures place constraints on the combinatorics of lexical and phrasal signs, and the same architecture is used throughout the model, be it for atomic lexical entries, or complex clausal constructs.[9]

### 2.2.2 SBCG and CxG

The idea to give lexical analyses to a wide range of expressions is not in accordance with mainstream constructionist ideas. But to what extend is SBCG also a constructionist approach? Let us take another look at the characterization of construction grammar in Kay (1995): "Construction Grammar (CG) is a non-modular, non-derivational, monostratal, unification-based grammatical approach, which aims at full coverage of the facts of any language under study without loss of linguistic generalizations, within and across languages."

The first point, the non-modularity, distinguishes most grammatical frameworks from generative (Chomskyan) approaches (most of the named characteristics are also found in other members of the constraint-based group of theories, e.g., HPSG, LFG, CCG, etc.), at least if we understand modularity in the narrow sense of using independent modules for independent processes, like syntax and semantics. While the generative models assume

[9]For a further read, I recommend the Sag (to appear 2010) paper, for a short overview, one might read Michaelis (2009).

different, temporally divided modules for structure building and interpretation, this is not the case for constructionist approaches (alongside with all constraint-based grammars). Form and function are two sides of a coin and are not separable in the process of production or processing. In SBCG, like in HPSG, all information about the combinatorics, semantics, phonology and context is coded in one entity, a sign, and they are not to be interpreted in a particular temporal order, but synchronously.

Non-derivationality and monostratality are other characteristics that are directly opposite to the generative transformational approaches. In CxG, no structure is assumed to be derived from another structure, and there is a single syntactic representation for any given expression. For example, a passive sentence is not held to be derived from a more basic type, the active sentence. Instead, a passive derivational construction licenses passive participles, and these can project a VP in exactly the same way that active verb forms can. There is no underlying common structure, and there is no mechanism that derives one construct from another.

As in HPSG, the major mechanism of structure building in SBCG is the combination of signs that are equipped with constraints. Every sign consists of features and their values, and only those signs can be combined that are specified (or underspecified) for the same features and values.

One feature of constructionist grammars that Goldberg (2003) emphasizes is the idea that it is "constructions all the way down". Any linguistic item is (an instantiation of) a construction, be it a schematic, lexically flexible pattern or a morpheme or word. In SBCG, there is still a distinction between a construction on the one hand, and a lexical entry on the other hand. The question remains whether this is only a terminological issue or a real difference, since lexical entries could be modelled as mother/daughter pairing as well.

In a nutshell, the lexical level in SBCG looks as follows: there are three different types of lexical categories: lexical entries, lexical-class constructions, and lexical constructs (Sag 2007b:25). A lexical entry is a "constraint relating form, syntactic cate-

gory, and meaning" (ibid.). Lexical-class constructions specify the common features of classes of lexical entries (lexemes or words) and place constraints on them. Lexical constructs are either words or lexemes. They are licensed by (post-)inflection or derivation. Their daughters and mothers are lexical signs.

### 2.2.3 SBCG and HPSG

The differences between SBCG and HPSG are mainly architectural ones. But there are some basic (cognitive) assumptions that differ in both approaches. To begin with, HPSG is a 'word and rules' model. There are lexical entries that are stored in the lexicon, a signature that constraints possible features and their values for all types, and a set of principles and rules that steer the combination of and structure building with lexical items. In SBCG, these principles and rules have been replaced by constructions. So instead of, e.g., a Head-Feature-Principle, presented in (18) (see Sag *et al.* (2003:73)), which is a general constraint on trees and makes sure that all the features of the head-daughter of a node percolate up to the higher node, we now have a Head-Feature construction, presented in (19), which takes the same form as any other construction (see Sag *et al.* (2003:480)):

(18) Head Feature Principle
In any headed construct, the HEAD value of the mother and the HEAD value of the daughter must be identical

(19) Headed Construction

$$\textit{hd-cxt} \Rightarrow \begin{bmatrix} \text{MTR} & \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{HEAD} & \boxed{1} \end{bmatrix} \end{bmatrix} \\ \text{HD-DTR} & \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{HEAD} & \boxed{1} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Therefore, the tripartition of the model is abandoned in favor of a system that contains lexical and constructional entries that combine with each other.

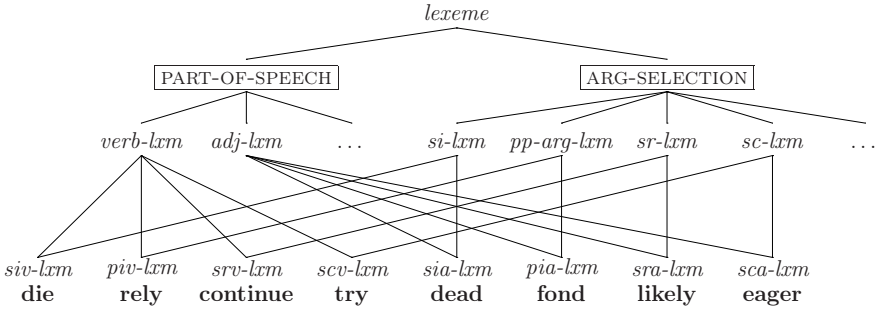
The second difference between HPSG and SBCG is the result of a rearrangement of the system. While in HPSG all structures

must be projected by lexical items, SCBG has an inventory of phrasal constructions that combine with lexical material. Still, the central mechanisms are similar. While HPSG interprets a sentence like *She sneezed the foam off the cappuccino* as being the result of a lexical rule which is applied to the verb *sneeze* and alters its valency in a way that licences an expression like this, SBCG has a lexical entry, a lexeme, combine with a derivational construction that builds a new lexeme *sneeze* with the appropriate argument structure and the properties of a resultative verb.

While this still looks like a minor difference, SBCG, different from HPSG, has a way of incorporating phrasal constructions as well. These are needed to license constructs that cannot follow from the more compositional combinatoric constructions or from lexical class constructions. These constructions can be stated directly, while in HPSG, it is not straightforward to model any type of nonlocal constraints.

A third difference (possibly depending on the flavor of HPSG one is using) is the hierarchical organization of types in the lexicon. Often, the inheritance network for HPSG is a default inheritance hierarchy, while for SBCG, it is a multiple inheritance hierarchy. In SBCG, any node in the tree of types may inherit its features from more than one mother, which allows the lexicon to express generalizations that hold for more than a one-dimensional relation without having to state constaints redundantly. An example of a multiple inheritance network is given in (20) (Sag *et al.* 2003:471):

(20) Multiple inheritance hierarchy



## 2.3 Delilah

For testing and evaluation of my analyses, I used an existing parser/generator. Delilah is a computational system that parses and generates Dutch language expressions. It provides a syntactic and semantic representation of sentences and phrases, and its focus lies on the semantic output. Delilah uses a version of Combinatory Categorial Grammar (CCG) for parsing, and the system is written in Prolog. The information is modeled as typed feature structures.

The model is completely lexicon-based, i.e., all the information regarding the combinatorics of items and the semantic composition of expressions is captured in lexical types. Where SBCG uses lexical-class constructions, Delilah uses 'lexical templates'. Principally, both do the same thing. They constitute a lexical type hierarchy, where subtypes inherit the constraints of their supertypes and add more specific information.

Delilah's semantics are in first-order logic, with a neo-Davidsonian event analysis. For a fully specified parse tree, Delilah gives us three different kinds of semantic output: Stored Logical Form (SLF), Normal Logical Form (NLF), and Flat Logical Form (FLF).[10] The first output is the SLF, which is derivational in the sense that it is built by means of unification of the sem values of the templates and lemmas. It remains underspecified for interpretations like scopal relations and semantic dependencies, and leaves ambiguities unresolved in a Cooper "storage" mechanism. These specifications are resolved by the conversion of the stored lambda terms in the two kinds of semantic descriptions that are derived from the SLF in a second step, namely NLF and FLF, which are computed by a specialized algorithm. The major difference between NLF and FLF is that "in NLF, matters of scope and semantic dependency are encoded globally and implicitly, as in standard predicate logic. In FLF, scope and semantic dependency are compiled out and made explicit at local levels." (Cremers & Reckman (2008:8))

Delilah allows for the implementation of constructions, since

[10] See Cremers & Reckman (2008) for more details.

it can specify constructional semantics directly under the sem value of the mother node. It builds its analyses by inserting the information specified in a lemma into lexical templates, which are similar to the lexical-class constructions in SBCG. See here a lexeme entry for the lexeme *lezen* ('read'):[11]

(21) Lexeme entry for *lezen*

```
lemma(lezen,
      verb,
      [trans_v, trans_v_sc, trans_ssub, trans_qsub],
      [arg( ID+ID1+1 ):synsem:theta:theme_of,
      arg( ID+_ID3+10 ):synsem:theta:agent_of,
      head:synsem:etype:event, head:phon:lezen,
      head:concept:read, head:sem:read],
      [pastsing:las, pastplur:lazen,
      participle:gelezen]  )  .
```

The lexical entry for a lexeme always specifies five types of information: 1) the name of the lemma ('lezen'), 2) its syntactic category ('verb'), 3) the (list of) templates (types) that this lexeme licenses ('trans$_v$, trans$_{vs}$c, ...'), 4) the constraints that the lexeme adds to the template, once they are unified (argument structure, in this example first object and subject, since this is the transitive *lezen*, the phonological form of the word, and the semantic concept, which is READ), and eventually 5) a (possibly empty) list of irregular inflected forms ('pastsing:las, ...').

Every lemma is specified for the lexical-class constructions (templates) it licenses. The first template on the list of *lezen* is the trans$_v$ template, the lexical-class construction that licenses simple monotransitive verbs:

[11] All examples are taken from Reckman (2009). This dissertation, together with Cremers (2004) and Cremers & Reckman (2008), is where the reader is referred to in order to get a complete description of the system and particularly its semantic component.

(22) Template for simple transitive verbs:

```
template(trans_v, verb,
   [ id: Top+ID1,
   synsem:[cat:vp, tense:untensed, predtype:nonerg,
          eventvar:EV, extth:Stheta\tilde[Top+ID, A]],
   sem: {{ {[SemS*(ID+ID1) #A, SemO*(ID+ID2)#B,}
        {\lambda}EStructure.{\exists}E.Main(E) &
        Etype(E) & EStructure(Top+ID) #EV],{},{}},
        {\lambda}Time. Stheta(EV,A) & Otheta(EV,B) &
        attime(EV, Time)},
   head:[phon: _X,
        synsem:[v_type:transacc, flex:infin,
                etype:Etype],
        sem: Main],
   arg( ID1+ID1+10 ):[phon:_Subj,
                      synsem:[theta:Stheta,
                              obj:subject_of(Top+ID)],
                      sem:SemS],
   arg( ID1+ID2+1 ):[phon:_Obj,
                     synsem:[obj:dirobject_of(Top+ID),
                             theta:Otheta, cat:np,
                             dir:left(1), mode:0,
                             case:obliq], sem:SemO]
   ] ).
```

The lexical-class construction $\text{trans}_v$ licenses expressions that instantiate a member of the group of verbs that take two arguments (ID+ID1+10, the subject, and ID1+ID2+1, the object). Its syntactic features like syntactic category, tense, ergativity, event type and external theta are expressed under synsem, and the binding of semantic variables is defined under sem. Here, the semantic values of all arguments are provided with a variable, and a lambda term is defined that computes the semantics of an expression that is licensed by this template. Note that all variables need to be matched with the values on the lemmas that instantiate this pattern, for instance the phon value, the type of theta roles for the arguments, and the event types. Also, the pattern is specified for infinite verb forms. In order to create a

finite expression, the flex values are matched with the flex values of a finite verb form. Just like in SBCG, Delilah uses an inflectional construction in order to build a word from a lexeme. These constructions add information about form, tense, person, and number to a lexical item.

It is possible to model constructions the way they are described in SBCG in the semantic parser/generator Delilah. The method of structure building is comparable, and both systems adhere to a lexical approach to constructions. The lexicon contains basic lexical entries that specify the characteristics of single lexemes, and lexical templates that obey to semantic and syntactic constraints. Delilah is, just like SBCG, a model that is based on signs, with each construction (or template) being the pairing of a mother and a daughter. Both poles of a construction, mother and daughter, come with their own constraints, and it is a matter of specifying the semantics of the mother node in order to allow for constructional semantics.

## 2.4 Summary

In this chapter, I set the theoretical ground for the rest of this dissertation. In the introduction, I showed that—at the moment—construction grammar seems to be the most dedicated and lively grammatical framework available, if one wants to deal with idiosyncratic syntax or semantics. Though in theory this is true, adopting a constructionist view still demands a lot of explanations. In section 2.1, I positioned myself with respect to certain points that do not seem to be common ground in the field.

To begin with, I briefly discussed the phrasal-or-lexical constructions discussion. I argue not to subscribe to the idea that phrasal patterns carry meaning. Instead, I agree with Goldberg's and Jackendoff's view that form and meaning of a construction are inseparable and cannot be seen as derived from one another.

A related question concerns the way that the lexicon is organized in the broader sense. Do we want to assume phrasal entries to be part of our knowledge, or do we think that all structures

are raised by a lexical item? Müller (2007) argues for a strictly lexicalist approach to a variety of constructions, the defenders of constructionism argue for phrasal lexicon entries. The constructions that I will analyze in the next chapters will all be analyzed lexically, although that does not mean that I am against any kind of phrasal construction. Instead, expressions should be analyzed lexically if possible, and phrasally if the lexical analysis fails. This way, it is made sure that interactions in grammar can be accounted for lexically, while the more idiosyncratic language material that fails to yield sensible lexical analyses, can be accounted for phrasally.

Then I turned to the model that is one of the most formal constructionist models: Sign-Based Construction Grammar. I first sketched the most important concepts like sign and construction. Eventually, I compared SBCG with the models that it is based on: CxG and HPSG, and I showed a couple of similarities and differences.

Finally, I sketched Delilah, the parser/generator that I used for implementation purposes. Even though Delilah was not developed as a constructionist system in the first place, it allows for SBCG-like analyses. This is grounded in the architecture of the model, as it is a consequence of the way constructions are designed. Just as in SBCG, they are descriptions of signs, which are formalized as feature structures that place constraints on the input and the output, or the daughters and the mother, of a sign. Also, all grammatical and combinatoric information is captured in the lexicon, it is a words-and-constructions model.

# Chapter 3

# NCoN and NPN in Dutch

## 3.1 Introduction

In this and the following two chapters, I will turn away from the theoretical discussions to hands-on work on a set of data. This more practical part of this dissertation will deal with three different case studies, and I will present them in the order from small to big, or from concrete to abstract.

I will start with two related phenomena that are entirely concrete on the one hand, and fairly abstract on the other. The construction that licenses the first group, I will refer to as NCoN, the construction that licenses the latter, I will call NPN. An example of the NCoN construction, you see in (23), and an example of the NPN construction, is given in (24):

(23) *het schip verging met man en muis*
the ship went-down with man and mouse
'the ship sank with everyone (on it)'

(24) *de auto's stonden bumper aan bumper*
the cars stood bumper to bumper
'the cars stood bumper to bumper'

There is a second step I want to focus on in the realm of this chapter. Expressions of the type NCoN and NPN have already been analyzed by Postma (1996), but in a Government and Binding framework. I will contrast his analysis of Dutch NCoN and

NPN with my own analysis, and I hope to be able to point the reader to some shortcomings of an analysis of the Dutch data that necessarily has to claim universality and relies on the semantics attributed to phrase structure.

Constructional[1] approaches to grammar (e.g. Fillmore *et al.* (1988), Goldberg (1995), Croft (2001), Culicover (1999), Ginzburg & Sag (2000), Culicover & Jackendoff (2005) and work cited therein) tend to put great emphasis on the fact that they can capture linguistic idiosyncrasies as well as core phenomena of language(s). Usually, the terms nuts and core are used in order to oppose the own ideas to those of generative mainstream approaches. However, also in "Chomskyan" generative frameworks, some work has been done on constructions that qualify nicely as "nuts". In this chapter, I will sketch one generative approach to two idiosyncratic phenomena, namely Postma's model of *Zero Semantics*, and I will contrast it with a constructionist, lexicon-based analysis. But let me start with a condensed overview of Postma's analysis.

## 3.2 Zero Semantics

Zero Semantics (Postma 1996) is an interpretive model of syntax that aims at explaining semantic processes by purely syntactic means, based on the assumption of the independence of (morpho-)syntax. Postma assumes two levels of semantic contribution: lexical semantics, made available by terminal nodes (lexical items), and structural semantics, made available by syntax. Postma calls this *quantificational semantics*, as he reduces all the structures he investigates to quantification. For the bigger picture, it seems valid to choose the less specific term *structural semantics* for our purposes. Lexical semantics can be overridden (or "de-activated") by structural semantics, and as a consequence, the lexical material ends up in what Postma calls *zero semantics*

[1]Thanks to Crit Cremers, Egbert Fortuin, Pepijn Hendriks, Frank Landsbergen, and, of course, Ton van der Wouden for discussion.

(ZS).[2] Zero semantics is the result of *move* $\alpha$, where particular terminal nodes are moved out of the domain of lexical closure, where lexical items can be interpreted. In other words, structural semantics, the configuration of a tree, can override lexical semantics and make the lexical material eventually invisible. The only item that contributes to the interpretation of an expression is its phrase structure.

Postma (1996:19) describes the model he proposes as an interpretive theory of grammar, in which primitive semantic insertion is not linked to lexical insertion, and in which—apart from the lexical level—complex trees can be interpreted in a primitive way, provided that they fulfill specific requirements on interpretability. Quantificational interpretation is tied to particular morphosyntactic patterns rather than to the nature of the terminal strings that lexicalize the pattern.

In order to illustrate this model, he applies it to different types of noncompositional processes:[3]

(25) (a) *Er was geen kip in de stad*
there was no chicken in the town
'there was nobody in town'

(b) *Jan doet geen vlieg kwaad*
John does no fly evil
'John does not hurt anybody'

(c) *Ik begrijp er geen snars van*
I understand there no SNARS of
'I do not understand anything of it'

(d) *Jan heeft kind noch kraai*
John has child nor crow
'John has nobody at all'

[2] In the remainder, I will refer to the model Zero Semantics with capital letters, and to the state of being in zero semantics with small letters.

[3] Examples taken from Postma (1996:1), emphasis in the original. For alternative approaches to the interpretation of negative polarity items, see, e.g. van der Wouden (1997).

(e) *Het schip verging met man en muis*
the ship went-down with man and mouse
'the ship sank with everyone (on it)'

In all the examples, the semantics of the underlined words does not contribute to the semantics of the sentence, at least not in a compositional way. Sentence (a) is not about chickens, sentence (b) is not about flies, the word *snars* in (c) is according to Postma a nonce creation, (the WNT (1864-1998), though, does give both a meaning and an etymology for the lemma *snars*, so, at least diachronically, it is not completely unmotivated, as Postma suggests), and the coordinations in (d) and (e) also do not contribute any semantics to the utterance directly. According to Postma, they are in zero semantics.

## 3.3 The NCoN construction

In this section I will focus on the "dummy coordination" (dummy in the sense that the coordinated words don't contribute any semantics) of the type presented in (25) (d) and (e). Expressions of this type are interesting, as they all exhibit certain semantic and syntactic features that need to be accounted for. There is a set of structural restrictions on these expressions, which Postma identifies as follows: the nouns must be bare, and the expressions only receive an idiomatic reading if the nouns are singular.[4] On the semantic side, Postma makes the following generalization: all bare noun coordination constructions receive a universal quantificational interpretation, although this cannot be assigned to the semantics of one particular lexical item. He captures this in the following definition:[5]

(26) Interpretation of coordinative bare singulars
Let $\gamma$ be a coordinative construction : $N_1P_{sg}$ & $N_2P_{sg}$ with $N_1P$ and $N_2P$ distinct bare singulars, then $\gamma$ is in

[4] I want to add that the nouns are typically count nouns. Coordination of bare singular mass nouns is a regular process of Dutch with a compositional semantics, as in *bloemkool en kaassaus* ('cauliflower and cheese sauce').

[5] Postma (1996:4), emphasis in the original.

zero semantics ([ZS & ZS]), and is interpreted as a *collective* universal quantification ('everyone', 'everything', 'entirely', 'extremely').

Although this is not within the scope of this chapter, I am hesitant to agree on the notion of collective universal quantification. If we accept that expressions such as *met man en muis* function as a universal quantifier, it is more in accordance with traditional views to analyze the quantification here as distributive. One argument would be the observation that there is no group action involved, as e.g. in *three men are building a house*, with the collective reading of *together*. In the case of distributive universal quantification, every member of the set has to be able to claim the predication for himself, which is not the case for collective quantification. Consider this example:

(27) *Peter, Jan en Frits hebben een huis gebouwd*
Peter, Jan and Frits have a house built
'Peter, Jan and Frits have built a house'

This sentence is ambiguous with regard to its quantificational force. Under collective quantification, the three men have built one house together, and no single man can claim that he built a house. Under distributive quantification, each man has built one house himself, and each man can claim for himself that the predication expressed is applicable.

Another indicator that we are dealing with distributive quantification in this type of construction is the following: in a clause like *het schip verging met man en muis*, the predication is applicable to all members of the set. There is no situation I can think of where the following sentence is a meaningful utterance:

(28) **Het schip verging met man en muis, behalve*
the ship went-down with man and mouse, except
*de kapitein*
the captain
'The ship sunk with man and mouse, but the captain survived'

In a collective reading, that should not be excluded, as, e.g., in this example:

(29) *Alle mannen droegen hoeden, alleen Peter niet*
all men wore hats, only Peter not
'all men wore hats, only Peter did not'

Postma proposes that the entire coordination is moved out of the domain of lexical closure. The coordination of bare singular count nouns carries the semantics of universal quantification, and whatever lexical item is inserted into these structures, ends up in zero semantics.

There are several questions that follow from this proposal, some of which I want to mention here. First of all, Zero Semantics cannot account for the fact that the construction is not productive in that sense that it cannot produce novel utterances. *Met man en muis* is felicitous, *met muis en man* is not, and neither is *met muis en kraai*. If the process involved is indeed a purely structural one that does not look at the terminal nodes after encountering a particular syntactic structure, one would expect that process to be fully productive. In fact, all the examples given in Postma (1996) exhibit that constraint: both lexical choice and word order are fully determined.

A second, probably even more challenging problem for the Zero Semantics approach is the following: Postma defines the semantics induced by coordinative bare singulars as a (collective) universal quantification, and paraphrases this with *everyone*, *everything*, etc. The question at hand is, then: how is it determined which of those different quantifiers is picked? *Man en muis* can only be paraphrased with *everyone*, because the expression can only be used in contexts where the antecedent(s) are animate. The same holds for *kind noch kraai* ('child nor crow'). (But note that here, the quantification is positive for *met man en muis*, and negative for *kind noch kraai*). *Have en goed* ('have and good(s)') on the other hand can only refer to inanimate antecedents, and coordinations like *met hand en tand* ('with hand and tooth') even have a very specific semantics, namely 'with all possible

means'.[6] The negative quantification in *kind noch kraai* can be explained by the presence of *noch* ('nor'), but in the end, the meaning of *met hand en tand* can hardly be attributed to one or more lexical items involved. Even if it could, all these explanations demand the ability to look into the lexical items proper, without putting them into zero semantics where they are invisible for interpretation.

Another bare singular count noun coordination that cannot be explained with the Zero Semantics approach is illustrated in the following example:[7]

(30) *De vier prinsessen, de dochters van Juliana, zijn*
the four princesses, the daughters of Juliana, are
*voor galg en rad opgegroeid*
for gallows en wheel grown-up
'the four princesses, the daughters of Juliana, grew up for a life in crime'

In the expression *voor galg en rad opgroeien*, there is no trace of universal quantification, although it exhibits the same pattern as *met man en muis*. This means that there would have to be a mechanism that exempts certain instantiations of the same abstract pattern from the structural semantics of the construction. Even if *voor galg en rad* is stored in the lexicon as exception to the rule, that would not be in accordance with the idea that the lexical level is invisible to interpretation. It would have to be visible to be exempted, in the first place. This effect is difficult to address within Zero Semantics.

Some critical points have already been mentioned, but apart from that, Postma's approach seems to have no means of dealing with collocational force, with subtle semantic differences, or with selectional restrictions, as the interpretive device can not look into the semantic properties of the lexical items. Under a

[6] We can observe a tendency here that, if one of the coordinates refers to a human being, the set that is quantified over, consists of humans as well. Whether this holds for all instances will have to be left to future research.

[7] http://www.gaykrant.nl/index.php?id=9&a=bericht&bericht=642 (01-30-2007).

Zero Semantics analysis, all instances of this construction are treated the same. Another unexplained phenomenon is the unproductivity of the NCoN-constructions. If the interpretation of these configurations is a built-in part of UG, why would the pattern not be productive? The prediction would be that the insertion of any bare singular count noun would render a valid interpretation, but that is not the case.

**Analysis NCoN**

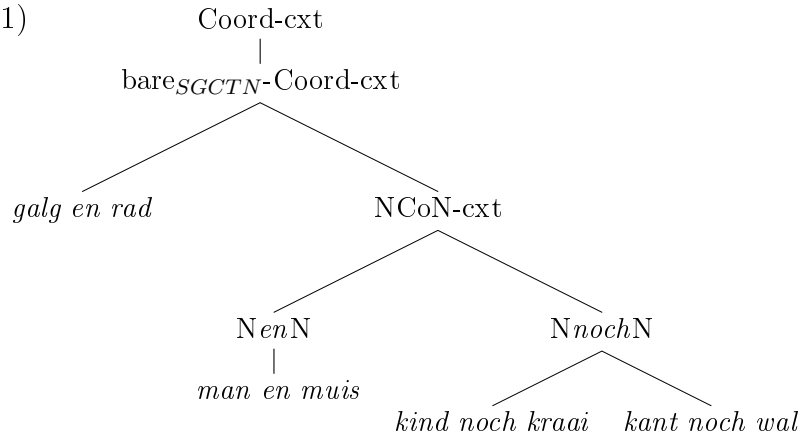
The analysis that I propose here for NCoN actually goes further than the *Zero Semantics* approach to this construction, in that it can capture more data. At the same time, it reaches less far, in that it doesn't claim any universal applicability. I agree with Postma's claim that expressions of the type *bare singular (count) noun – coordinator – bare singular (count) noun* exhibit (at least some kind of) universal quantification. But this observation is too unspecific, as different instantiations exhibit different properties which we want to be able to capture.

We assume an abstract pattern NCoN in the lexicon, a non-productive subtype of the more general coordination construction. But—as this pattern is not productive—it can not be instantiated as such, but only functions as the common mother to a set of daughter constructions. These are again abstract patterns, but this time the coordinator is specified, and with that the choice between positive or negative quantification is made. The lexical class of N*en*N lexemes imposes quantification of the type *all*, the N*noch*N lexemes impose quantification of the type *no*. These constructions again do not function as a productive schema, but only as two types of lexemes that are fully specified for lexical material on the one hand (they are, in fact, string-listed, i.e., the entire string is listed as "word with spaces" in the lexicon as one single word), and for their specific constraints like, e.g., the type of head that can be modified, and their semantic contribution on the other.

The constructional network for this family of constructions is given in (31). Note that all the terminal constructions are

lexically specified. This ensures that no new expressions can be formed on the basis of the more abstract patterns. Furthermore, it is not visible from this network that all terminal constructions have certain valencies which are specified in their lexical entries. Typically, the constructions only occur with a particular set of verbs (such as *opgroeien* in the case of *voor galg en rad*), hence they need to be on the valence list of the verb, or they are subcategorized by a particular preposition (such as *met* in the case of *man en muis*).

(31)




Since NCoN is an instance of the coordination construction, I will start with the coordination construction as it is described in Sag *et al.* (2003:485). This construction licenses expressions like the following:

(32) A coordination construct[8]

$$\begin{bmatrix}
\textit{coord-cx} & & \\
\text{MTR} & \begin{bmatrix}
\text{PHON} & \langle\ \text{Kim, sleeps, and, Pat, works}\ \rangle \\
\text{SYN} & \begin{bmatrix}
\text{CAT} & \textit{verb} \\
\text{VAL} & \begin{bmatrix} \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix}
\end{bmatrix} \\
\text{SEM} & \begin{bmatrix} \ldots \end{bmatrix}
\end{bmatrix} \\
\text{DTRS} & \left\langle \begin{bmatrix}
\text{PHON} & \langle\text{Kim, sleeps}\rangle \\
\text{SYN} & \begin{bmatrix}
\text{CAT} & \textit{verb} \\
\text{VAL} & \begin{bmatrix} \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix}
\end{bmatrix} \\
\text{SEM} & \begin{bmatrix} \ldots \end{bmatrix}
\end{bmatrix}, \right. \\
& \begin{bmatrix}
\text{PHON} & \langle\text{and}\rangle \\
\text{SYN} & \begin{bmatrix} \text{CAT} & \textit{conj} \end{bmatrix} \\
\text{SEM} & \begin{bmatrix} \ldots \end{bmatrix}
\end{bmatrix}, \\
& \left. \begin{bmatrix}
\text{PHON} & \langle\text{Pat, works}\rangle \\
\text{SYN} & \begin{bmatrix}
\text{CAT} & \textit{verb} \\
\text{VAL} & \begin{bmatrix} \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix}
\end{bmatrix} \\
\text{SEM} & \begin{bmatrix} \ldots \end{bmatrix}
\end{bmatrix} \right\rangle
\end{bmatrix}$$

In this construct, two daughters of the identical category are coordinated by a conjunction. The constraint that only likes can be coordinated is expressed by reentrancy of the DTRS|SYN values, which is stated in the combinatoric coordination construction:

[8]The HEAD-feature has been replaced by CAT.

(33) The coordination construction[9]

$$
\textit{coord-cx} \Rightarrow \left[\begin{array}{ll}
\text{MTR} & \left[\begin{array}{ll}
\text{SYN} & \left[\begin{array}{ll}\text{HEAD} & \left[\text{FORM } \boxed{1}\right]\\ \text{VAL} & \boxed{2}\\ \text{GAP} & \boxed{A}\end{array}\right]\\
\text{SEM} & \left[\text{IND } s0\right]
\end{array}\right]\\
\text{DTRS} & \left\langle \left[\begin{array}{ll}
\text{SYN} & \left[\begin{array}{ll}\text{HEAD} & \left[\text{FORM } \boxed{1}\right]\\ \text{VAL} & \boxed{2}\\ \text{GAP} & \boxed{A}\end{array}\right]\\
\text{SEM} & \left[\text{IND } s_1\right]
\end{array}\right], \ldots, \right.\\
 & \left[\begin{array}{ll}
\text{SYN} & \left[\begin{array}{ll}\text{HEAD} & \left[\text{FORM } \boxed{1}\right]\\ \text{VAL} & \boxed{2}\\ \text{GAP} & \boxed{A}\end{array}\right]\\
\text{SEM} & \left[\text{IND } s_{n-1}\right]
\end{array}\right],\\
 & \left[\begin{array}{ll}\text{HEAD} & \textit{conj}\\ \text{IND} & s_0\\ \text{RESTR} & \langle \left[\text{ARGS } \langle s_1 \ldots S_n \rangle\right] \rangle\end{array}\right],\\
 & \left.\left[\begin{array}{ll}
\text{SYN} & \left[\begin{array}{ll}\text{HEAD} & \left[\text{FORM } \boxed{1}\right]\\ \text{VAL} & \boxed{2}\\ \text{GAP} & \boxed{A}\end{array}\right]\\
\text{SEM} & \left[\text{IND } s_n\right]
\end{array}\right]\right\rangle
\end{array}\right]
$$

All NCoNs are coordinations, but they have far more specific properties than the construction above. Since the NCoN-pattern is unproductive, all the instances of it are maximally specific. On top of that, the internal structure is completely fixed. Hence,

[9] From (Sag *et al.* 2003:485). Again, note that in the more recent literature on SBCG, the feature HEAD has been eliminated.

there is no abstract productive NCoN-construction, there are only fully specified lexical entries, all of which are licensed by this construction:

(34) *NCoN*-lexeme

$$\textit{NCoN-cxt} \Rightarrow \begin{bmatrix} \textit{coord-cx} & & & \\ \text{MTR} & \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{CAT} & \textit{noun} \\ \text{CASE} & \textit{acc} \\ \text{COUNT} & - \\ \text{AGR} & \text{3sing} \\ \text{DET} & - \end{bmatrix} \end{bmatrix} & & \\ \text{DTRS} & \left\langle \begin{bmatrix} \text{SYN} & \boxed{1} \begin{bmatrix} \text{CAT} & \textit{noun} \\ \text{CASE} & \textit{acc} \\ \text{COUNT} & + \\ \text{AGR} & \text{3rd sg} \\ \text{DET} & - \end{bmatrix} \end{bmatrix}, \right. & & \\ & \left. \begin{bmatrix} \text{PHON} & \langle \text{noch} \rangle \\ \text{SYN} & \begin{bmatrix} \text{CAT} & \textit{conj} \end{bmatrix} \end{bmatrix}, \right. & & \\ & \left. \begin{bmatrix} \text{SYN} & \boxed{1} \end{bmatrix} \right\rangle & & \end{bmatrix}$$

This construction is licensed by lexical items that contribute their own semantics to the mother node. The semantics of the daughters in the NCoN construction is overridden. The lexical entry for the expression *kind noch kraai* looks as follows:

(35) Lexicon entry for *kind noch kraai*

$$\begin{bmatrix} \text{FORM} & \langle \text{kind noch kraai} \rangle \\ \text{SEM} & \text{nobody} \end{bmatrix}$$

The thing to see here is that, first of all, the semantics of the mother is completely constructional in the sense that it is not a product of the combination of the semantics of the daughters.

The second thing to see is that, while the SYN features of the two nominal daughters are identical by means of reentrancy, the SYN feature of the mother is different. This is due to the fact that the coordinated nouns are unmarked singular count nouns, but the entire construct does not exhibit the combinatorics of a count noun. The mother is marked for accusative case, though, so it cannot appear in subject position.

The second example of this NCoN-group is *met man en muis*, which behaves slightly differently due to the preposition that selects for the NCoN. In opposite to *kind noch kraai*, *met man en muis* selects for the preposition *met*.

(36) Lexicon entry for *man en muis*

$$\begin{bmatrix} \text{FORM} & \langle \text{man en muis} \rangle \\ \text{SEM} & \text{everybody} \end{bmatrix}$$

Note that, since *man en muis* always is selected by the preposition *met*, it needs to be selected and therefore, appear on the valence list of *met*.

(37) Lexicon entry for *met*

$$\begin{bmatrix} \text{FORM} & \langle \text{met} \rangle \\ \text{SYN} & \left[ \text{ARG-ST} \left[ \text{FORM} \quad \langle \text{man en muis} \rangle \right] \right] \end{bmatrix}$$

The difference between this (lexical) constructional approach and the phrasal Zero Semantics approach is obvious from the two examples above: since the lexical approach allows for specification of the daughters as well as the mother, we can state the combinatoric features directly. This way, we do not lose the generalization that can be formed over all occurrences of the NCoN-construction. The semantics of the actual construct comes from the sem value that is stated in the particular lexical entries. There it is determined whether the semantic contribution of a lexical item is, e.g., all, or everybody, or everything.

To sum up, the analysis of expressions of the type NcoN presented here is situated somewhere between string-listing and interpretive. Even though all expressions of that type are assumed

to be listed in the lexicon, they are brought together as subtypes of a more abstract supertype construction which is unproductive, in order to be able to capture their commonalities. The stipulation that a small number of expressions of the type NCoN is acquired rather than built in, does seem rather intuitive, which is not necessarily a bad label. For production purposes, even in the Zero Semantics model, the lexical material that can go into the slots has to be acquired in the first place. Saying that the semantic contribution of the expressions is acquired (learned and stored) alongside the lexical items, is not a big step, but it helps explaining some facts that have to remain mysteries from Postma's point of view, like, e.g., why the pattern is not productive.

## 3.4 The NPN construction

The second type of construction Postma claims to involve universal quantification are duplicated bare singular noun constructions. I will refer to this type of construction as NPN.[10] Postma gives the following examples:

(38) *We stonden bumper aan bumper*
We stood bumper to bumper
'Our bumpers touched each other / we stood with one bumper at the other'

(39) *De kinderen liepen hand in hand*
The children walked hand in hand
'The cildren walked hand in hand'

For this construction, Postma postulates a distributive universal quantification. His definition of NPN says the following:

(40) Interpretation of duplicated bare singular NP p NP constructions
Let $\delta$ be a construction : [NP p NP ] with the NPs identical bare singulars and one noun a full copy of the other

[10] For recent work on the English NPN construction see Jackendoff (2008).

> one, [FS p ZS], then $\delta$ is interpreted as involving a *distributive* universal quantification with NP as the restrictive set.

For this analysis, the only element in this construction that is in zero semantics, is one of the two nouns. The other noun and the preposition contribute their full semantics to the interpretation.

Compared to the NCoN type of expressions, the NPN clearly has a higher degree of schematicity. That is challenging in the sense that more freedom with respect to lexical choice and more strictness with respect to the structural pattern (by coindexation of the two nouns rather than what seemingly looks like a free choice) might look like support for a Zero Semantics approach. And indeed, the pattern identified by Postma, even in its abstractness, is a good clue to the correct construction. But, then, another complication arises that might not be visible at first glance, but once you turn to natural data, the problem gets obvious.

The common syntactic part of this construction, as identified by Postma, is the structure *bare singular noun – preposition – bare singular noun*, with the two nouns being coindexed, and the semantics is universal quantification over the NP. But, as a look at the distribution of NPNs in the Spoken Dutch Corpus (CGN) reveals, the semantics show more variation than described by Postma. In order to get a clearer picture of the use and semantics of NPNs in Dutch, I extracted all examples of the NPN construction from the 9 million corpus of spoken Dutch. In a second step, I grouped the examples with respect to semantic overlap.[11] Unlike Postma, I also included NPNs that are preceded by *van* ('of'), as they exhibit a distribution similar to the bare NPNs and will help to arrive at a more concrete picture of the properties of NPNs. The categories I will describe and the prepositions that occur in them are shown in (41):

[11] A list of all the expressions found is given in the appendix.

(41) Categories of NPN

| **Cat** | **Pattern** |
|---|---|
| Totalization | *N voor/per N* |
| Temporal Succession | *(van)N aan/na/op N* |
| Spatial Succession | *van N naar/tot N* |
| Juxtaposition | *N in/naast/over/aan N* |
| Spatial extension | *van N tot N* |
| Transition | *van N tot/naar N* |

As we can see here, there are certain regularities within the distinct subconstructions, which I will describe in more detail in the next section. Nevertheless, there are sufficiently many common properties among all of those categories to establish a superordinate abstract NPN-construction, which is specified for the following properties:

(42) Abstract NPN-construction

$$
\textit{NPN-cxt} \Rightarrow \left[\begin{array}{ll}
\text{MTR} & \left[\begin{array}{ll}
\text{SYN} & \left[\text{CAT} \left[\begin{array}{ll} \textit{noun} & \\ \text{COUNT} & - \\ \text{MRKG} & \textit{unmk} \end{array}\right]\right] \\
\text{SEM} & \textit{cx-sem}
\end{array}\right] \\
\text{DTRS} & \left\langle \boxed{1} \left[\begin{array}{ll}
\text{SYN} & \left[\text{CAT} \left[\begin{array}{ll} \textit{noun} & \\ \text{FORM} & \textit{phon} \\ \text{COUNT} & + \\ \text{AGR} & \text{3rd sg} \\ \text{MRKG} & \textit{unmk} \end{array}\right]\right] \\
\text{SEM} & \left[\ldots\right]
\end{array}\right], \left[\begin{array}{ll} \textit{prep} & \\ \text{SEM} & \left[\ldots\right] \end{array}\right], \boxed{1} \right\rangle
\end{array}\right]
$$

The NPN-construction is the set of a mother and three daughters, where the mother is a phrase with a nominal head and

constructional semantics that need to be further specified in the separate sub-constructions. The daughters that need to be combined are two identical singular bare count nouns and a preposition. The difference with the NCoNs is this: the latter are basically string-listed. They are not abstract constructions that can license new expressions ad hoc, but they are all maximally specified constructs, lexical items, which are grouped under an abstract supertype. The NPNs, on the other hand, are a family of productive, formally similar constructions that (often) only differ with respect to their semantic content. In other words, they are a group of abstract patterns that all exhibit the same set of properties, as opposed to a set of concrete strings that exhibit an overlapping set of properties.

The subtypes that we distinguish for the NPN construction contribute more specific information at various places. Mainly, they add the concrete contribution of the mother's SEM-value, if applicable, they give the paradigm of possible prepositions, and they give information about whether or not the entire NPN is to be selected by the preposition *van*.

The other important feature of the sub-constructions is that they can occur in (and are restricted with respect to) different grammatical functions. There are constructions that can only occur in argument position, and there are some that can only occur in adjunct position. Since some of them can occur in argument as well as adjunct position, this question will not be tackled within the sub-construction itself, but by combining with the appropriate functional construction. Since the model makes use of multiple inheritance, this can be solved. A construct then inherits its properties from the NPN-subconstruction as well as, e.g., an adjunction construction.

In the following sections, we will see what the specific subtypes of NPN look like.

### Totalization

The first sub-construction of NPN I will call Totalization. The NPNs that are licensed by this category have in common that

they express (distributive) universal quantification over an item that is not the N itself, but that is completely made up of a finite set of Ns. Consider the following examples:

(43) *kolom voor kolom werken we even de*
column by column work we shortly the
*studiewijzer door van week zeventien*
course-catalogue through of week seventeen
'column by column, we quickly work through the course catalogue of week seventeen'[12]

(44) *hij heeft tientallen jaren lichamen ontleed vezel voor*
he has tens-of years bodies analyzed fibre by
*vezel om de loop van aderen en spieren*
fibre in-order-to the course of veins and muscles
*en de plaats van de botten te kunnen bepalen*
and the place of the bones to can determine
'he analyzed dozens of bodies fibre by fibre in order to be able to determine the course of the veins and the muscles and the place of the bones'

In example sentence (43), the course catalogue is construed as being subdivided into columns. The entire catalogue has to be worked through, and that is done by working through the columns consecutively. In example sentence (44), the scope of what is analyzed is the body, which is made up of fibres that are again analyzed consecutively.

The alternative assumption is mere quantification over N, but that does not go far enough. In (43), the speaker does not mean to say that whoever has to read the course catalogue will have to read every column, but rather that she has to read the entire course catalogue, or, at least, that she has to read every column *of the course catalogue*. In the first alternative, the scope of the quantification is the direct object. This is, with very few exceptions, the standard case for this category: the NPN functions as an adverbial, and the scope of quantification is the direct object,

[12] All examples in this chapter are from the CGN.

whereas in the second alternative, the scope of quantification is N in relation with the direct object. But whichever one chooses, mere quantification over N is not enough.[13]

In the Dutch data, this category of NPNs only allows for two prepositions, namely *voor* and *per*. Interestingly, *per* is felt to be infelicitous by some Netherlandic Dutch speakers, and only appears in sentences that stem from the Belgian part of the corpus.[14]

The NPN totalization construction looks like follows:[15]

(45) NPN-totalization

$$\textit{npn-tot-lxm} \Rightarrow \begin{bmatrix} \textit{NPN-cx} & \\ \text{MTR} & \left[\text{SEM } \textit{totalization}\right] \\ \text{DTS} & \begin{bmatrix} \textit{prep} \\ \text{PHON} \quad \langle \text{voor/per} \rangle \end{bmatrix} \end{bmatrix}$$

The lexical class construction npn-totalization is a subtype of the less constrained NPN-construction, and it adds to the information of the supertype the information about the semantics of the construction, which is stored under the SEM value of the MOTHER, and the possible form of the DAUGHTER that is the preposition, namely *voor* and *per*.

Note that even though there are no abstract nouns in the CGN data, the nouns that can be inserted into this construction cannot be further specified with regard to the value of abstractness/concreteness. A quick Google search provided us with ex-

[13] Jackendoff (n.d.:6) defines a comparable construction for English, which he calls N*by*N. Examples are expressions like *piece by piece* and *day by day*. The semantics of N*by*N, he describes as 'succession'. In my analysis, those two expressions would fall into two categories, namely Totalization (*piece by piece*) and Temporal Succession (*day by day*). Of course, this is the mere result of classifying items along different dimensions.

[14] This type of regional restrictions on the construction is also a point that might be interesting to capture in a model of Dutch. In the realm of this analysis, though, this information has been left out.

[15] The notation is simply a shortcut, where only the information that is not already present from the supertype is given.

amples like (46), where the direct object *het* ('it') is made up of words, sentences, and (clearly abstract) thoughts:[16]

(46) *een nauwlettend lezer van het stuk, die zich*
an accurate reader of the piece, who himself
*door Vondels oratorische betoogtrant niet wil*
through Vondel's oratorical argumentation not wants
*laten meeslepen, maar het juist woord voor woord,*
let drag-along, but it just word for word,
*zin voor zin, gedachte voor gedachte, wil*
sentence for sentence, thought for thought, wants
*proeven [...]*
prove [...]
'an accurate reader of the piece, who doesn't want to be dragged along by Vondel's oratorical argumentation, but who just wants to prove it word by word, sentence by sentence, thought by thought, [...]'

The external combinatorics of the totalization-NPN are steered via the valence list of transitive verbs. Since the totalization-NPN modifies mono-and ditransitive verbs, these groups of verbs have to be able to select for this kind of NPN.

In a Delilah implementation, the template for the lexical class construction totalization-NPN looks as follows:

(47) Lexical-class construction totalization-NPN

```
template( npn-tot-cx,
[ID+ID1,
head:[phon:Phon],
synsem:[cat:np, aggr:count, def:indef]
sem: {{[SemN$Y#N ], [], []},
     Z@some^Y^and( quant(Y, all), N,
     entails( Y, decr), Z, entails(Y, decr))}
arg(ID1+ID2):[synsem:[cat:n],
              head:[phon:Phon]
arg(ID1+ID3+1):[phon:_PPPPhon,
```

[16] http://www.dbnl.org/tekst/vond001aenl01/vond001aenl01_002.htm (05.01.2007).

```
          synsem:[cat:pp, head:[phon:voor/per],
          case:obliq,
          dir:right(1), flag:0], sem:SemN],
arg(ID2+ID3):[synsem:[cat:n], aggr:count,
   head:[phon:Phon],sem:SemN]],

          ]).
```

This template licenses structures where there are two identical bare count nouns connected by a preposition *voor* ore *per*. Under sem, a variable is bound that has to be identified with the object that is quantified over.

### Temporal succession

The next category captures expressions of temporal succession. The scope of quantification in this construction is the set of events or referents designated by the noun. The relative order of the referents is organized along a temporal axis. Examples for this type of NPN are given in (48) and (49):

(48) *en zo tussen vrees en verlangen gevangen ging*
and so between fear and desire caught went
*Tony avond na avond naar het café om*
Tony evening after evening to the bar in-order-to
*zich tot vergetelheid te drinken*
himself to oblivion to drink
'and caught between fear and desire, Tony went to the bar in order to drink himself into oblivion night after night'

(49) *hier in deze kale hoek verslijt ik broek na*
here in this unadorned corner wear-out I pants after
*broek*
pants
'here in this unadorned corner, I wear out one pair of pants after the other'

Semantically related is the version with *van*-NPN. The examples are scarce, but they clearly fall into the same semantic category as the bare Temporal Succession NPNs. An example is given in (50)

(50) *laat die het allemaal van dag tot dag in de*
let them het all from day to day in the
*krant zetten*
newspaper put
'let them put all this into the newspaper everyday.'

Beck & von Stechow (2006:1) also tackle the bare version of this construction in their approach to pluractional adverbials. They analyze a sentence pair as

(51) 1. These three dogs entered the room one after another
2. They entered the room dog after dog

informally like this:

(52) 1. D3 → D2 → D1
2. "x → y" = x enters the room after y

under the following truth conditions:[17]

(53) These three dogs entered the room, and the entering can be divided into a sequence of subevents in each of which one of the dogs enters, and the dogs can be divided into a sequence of individual dogs each of which entered in one of the subevents

But we need to differentiate along an additional axis, namely along the axis of function of the expression. Beck and von Stechow only deal with adverbials, but the Temporal Succession NPN can also occur in argument position. In (54), you see the abstract construction for temporal succession:

[17]For *dog after dog*, there remains the problem that there must be one dog which does not follow another dog. For the problem of 'the first dog', see also Beck & von Stechow (2006).

(54) NPN temporal succession

$$\textit{npn-temp-suc-cxt} \Rightarrow \begin{bmatrix} \textit{NPN-cx} & \\ \text{MOTHER} & \begin{bmatrix}\text{SEM} & \textit{temp.suc.}\end{bmatrix} \\ \text{DTS} & \begin{bmatrix}\textit{prep} & \\ \text{PHON} & \langle \text{aan/na/op} \rangle\end{bmatrix} \end{bmatrix}$$

In (55) and (57), we see the spelled out Delilah templates. The first one describes the template for a prepositional adjunct, the second one describes the template for an totalization-NPN in argument position:

(55) Temporal succession *van* N *tot* N, prepositional adjunct

```
template( pp_adj_temp_succ,
        [id: _Top+ID1,
        head:[phon:van, sem:_Sem],
        synsem:[autom:prepadj,  category:temporal,
           island:WH, cat:Cat, exsem:ExSem ,
           exaggr:ExAggr],
sem: {{ [{{[SemN$X&ID1+ID2#N1], [], []}, ,
           Z@some^X^and(quant(X, every), N1,
           entails( X, every), Z, entails( X, incr))}
           &(ID1+ID2)#NP1,
             {{[SemN$X&ID4+ID5#N2], [], []}, ,
           P@some^X^and(quant(Y, some), N2,
           entails( Y, decr), P,
           entails( Y, incr))}
           &(ID4+ID5)#NP2|QSt],PSt, RSt},
               and(Body,and( between~[NP1, NP2, EV]))}
arg( ID1+ID2+1 ):[phon:Phon,
       synsem:[cat:n, aggr:count, refmode:time,
       case:obliq, dir:right(0), flag:0], sem:SemN],
arg( ID1+ID3+1 ):[phon:_PPPPhon,
      synsem:[cat:pp, head:[phon:tot],
      case:obliq, dir:right(1), flag:0],
      sem:SemN],
arg( ID4+ID5+1 ):[phon:Phon,
       synsem:[cat:n, aggr:count,
```

```
       refmode:time, case:obliq,
       dir:right(2), flag:0],
       sem:SemN],
arg( ID1+_ID3+2 ):[phon:_SPhon,
                   synsem:[cat:Cat, eventvar:EV,
        head:[etype:event], exsem:ExSem,
        exaggr:ExAggr, dir:left(1), flag:8],
        sem:{{QSt, PSt, RSt}, Body}]
        ] ).
```

The template prepadj-temp-succ is a subtype of the type PP adjuncts (prepadj), of which it inherits all the appropriate constraints, and it is headed by a word that has the form *van*. It is a temporal adjunct, and it behaves like an island when it comes to extraction. It selects for a noun which has to be a count noun and needs to be a member of the paradigm of temporal reference (like 'month', 'day', 'minute', etc.). The noun contributes its own semantics to the whole construction. Again, it selects for a PP, this time headed by the preposition *tot*. Then, *tot* selects for a second noun, which has to have the same phon-value as the noun to the left of the preposition. This template can only be selected by verbs that are specified as being eventive.

The semantics of the entire pattern and the semantic combinatorics of the particular lexical items that enter the slots are determined under the sem value of the first node. It reads like this:

(56) For every X (a variable that is bound by the first noun), there is a Y (a variable that is bound by the second noun) that follows it, and for the period between X and Y, the event designated by the body applies.

For sentences like (49), we need to combine the abstract NPN construction with the information that the NPN functions as an argument. In order to do so, a theta role must be assigned.

(57) Temporal succession, NPN, nominal argument (shortened)

```
template( np_arg_temp_succ,
[ID+ID1,
head:[phon:Phon],
synsem:[theta:tempus, cat:np, aggr:count, def:indef]
sem: {{ [{{[SemN$X&ID1+ID2#N1], [], []}, ,
           Z@some^X^and(quant(X, every), N1,
           entails( X, every), Z, entails( X, incr))}
           &(ID1+ID2)#NP1,
             {{[SemN$X&ID4+ID5#N2], [], []}, ,
           P@some^X^and(quant(Y, some), N2,
           entails( Y, decr), P,
           entails( Y, incr))}
           &(ID4+ID5)#NP2|QSt],PSt, RSt},
           and(Body,and( between~[NP1, NP2, EV]))}
arg( ID1+ID2):[synsem:[cat:pp],
arg(ID2+ID3):[synsem:[cat:n], aggr:count,
   head:[phon:Phon],sem:SemN]],
             ]).
```

This template captures basically the same information as the previous one, only that it builds an NP without *van* rather than PP adjunct, which is captured under the synsem predicate of the mother node.

### Spatial succession

Whereas the members of the former category are consecutively ordered along a temporal axis, the members of this group are ordered along a spatial one. The spatial succession category only allows for *van*-NPNs, nevertheless I want to include it to get a more complete picture.[18]

[18]Note that the examples are all taken from a spoken corpus, and in speech, interruptions, reorganizations, etc., indicated with xxx in the examples, are of course usual.

(58) *ik zou xxx droom van mij is van dorp tot dorp*
I would xxx dream of mine is from village to village
*een keer he 'k geloof dat een Nederlander dat*
a time eh I believe that a Dutchman that
*gedaan heeft van dorp tot dorp uh rijden met de*
done has from village to village eh drive with the
*wagen*
wagon
'My dream is to go from village to village at one point; I think that a Dutchman has done that before: go by car from village to village'

The goal of driving around here is to visit all and every village in the country, and that is accomplished by driving from one village to the next. It is universally quantified over the set of villages.

(59) NPN spatial succession

$$\textit{npn-spat-suc-cxt} \Rightarrow \begin{bmatrix} \textit{van-NPN-cx} & \\ \text{MOTHER} & \begin{bmatrix}\text{SEM} & \textit{spat.suc.}\end{bmatrix} \\ \text{DTS} & \begin{bmatrix}\textit{prep} & \\ \text{PHON} & \langle\text{tot}\rangle\end{bmatrix} \end{bmatrix}$$

This lexical class is a subtype of the *van*-NPN construction. It adds the information that the instantiated preposition is *tot* ('to'), plus the constructional meaning. The semantics can be expressed like

(60) For every X, there is a Y that follows it, and the action described in the body is applied from X to Y.

This is expressed under the sem value of the according template:

(61) Spatial succession, NPN, prepositional adjunct

```
template( pp_adj_spat_succ,
        [id: _Top+ID1,
        head:[phon:van, sem:_Sem],
        synsem:[autom:prepadj,  category:spatial,
```

```
        island:WH, cat:Cat, exsem:ExSem ,
        exaggr:ExAggr],
        sem: {{ [{{[SemN$X&ID1+ID2#N1], [], []}, ,
             Z@some^X^and(quant(X, every), N1,
             entails( X, every), Z,
             entails( X, incr))
             &(ID1+ID2)#NP1,
             {{[SemN$X&ID4+ID5#N2], [], []}, ,
               P@some^X^and(quant(Y, some), N2,
               entails( Y, decr), P,
               entails( Y, incr))}
               &(ID4+ID5)#NP2|QSt],
               PSt, RSt}, and(Body,
               and( from~[NP1, EV],
               to~[ NP2, EV]))}
arg( ID1+ID2+1 ):[phon:Phon,
   synsem:[cat:n, aggr:count,
   material:concrete, case:obliq, dir:right(0),
   flag:0], sem:SemN],
arg( ID1+I42+1 ):[phon:_PPPPhon,
   synsem:[cat:pp, head:[phon:tot].
   case:obliq, dir:right(1),
   flag:0], sem:SemN],
arg( ID4+ID5+1 ):[phon:Phon,
   synsem:[cat:n, aggr:count,
   material:concrete, case:obliq, dir:right(2),
   flag:0], sem:SemN],
arg( ID1+_ID3+2 ):[phon:_SPhon,
   synsem:[cat:Cat, eventvar:EV, head:[etype:event],
   exsem:ExSem, exaggr:ExAggr, dir:left(1), flag:8],
   sem:{{QSt, PSt, RSt}, Body}]
        ] ).
```

### Juxtaposition

This category contains expressions which Jackendoff groups under the name of *juxtaposition*. As he puts it, "[a]nother meaning of *N to N* has to do with close contact or juxtaposition of similar parts of similar objects, particularly body parts". (Jackendoff

2008:16) Although this group is clearly productive in Dutch, as it is English, most of the candidates found in the CGN show a high token frequency and give the impression that they actually stored rather than built. The most prominent member of this category is *hand in hand*, which is a good candidate for storage, just as *neus aan neus* in (62), but other examples, like (63), are clearly built rather than stored and retrieved.

(62) *en nu sta je neus aan neus met de beer en*
and now stand you nose to nose with the bear and
*dat is natuurlijk een totaal andere beleving*
that is of-course a totally different experience
'and now you stand nose to nose with the bear, and that is of course a totally different experience'

(63) *ja ik ben dan zelf ook aan de Côte d'Azur heb*
yes I am then myself also at the Côte d'Azur have
*ik dat ook meegemaakt dat je daar echt camping*
I that ook experienced that you there really camping
*naast camping aan de kust hebt liggen*
next-to camping at the coast have lying
'I have seen it myself that at the Côte d'Azur, there is one camping lot next to the other'

The prepositions that are found in the examples from the CGN are *in*, *aan*, *over*, and *naast*, all prepositions of spatial ordering. Hence, one question arises with respect to the status of the preposition. Two different approaches are possible:

- The prepositions in question function as an anchor for the construction. In the lexicon entry for *aan*, e.g., there is either a pointer that directs to the template of the construction, or there is a separate lexicon entry for *aan* combined with the template, or
- the selection of prepositions is governed in the template of the construction. This can either be done with the help of the PHON feature, namely by giving it a list rather than

an atomic value, or by giving the preposition a feature of the type category, which then has to have the value spatial.

I opt for the latter option. Information in the lexicon about the type of preposition is useful for all kinds of purposes, and as prepositions are a rather small and rather closed class, this can be accomplished without major difficulties.

(64) NPN juxtaposition

$$\textit{npn-juxtapos-cxt} \Rightarrow \begin{bmatrix} \textit{van-NPN-cx} & \\ \text{MOTHER} & \begin{bmatrix}\text{SEM} & \textit{juxtaposition}\end{bmatrix} \\ \text{DTRS} & \begin{bmatrix}\textit{prep} & \\ \text{SEM} & \text{spatial}\end{bmatrix} \end{bmatrix}$$

This lexical class construction is a subtype of the *van*-NPN construction and adds the information about its particular semantics and the information that the preposition must be out of the set of spatial prepositions.

Since this lexical class construction can again occur in two different positions, namely as a nominal argument or as a nominal adjunct, there are again two Delilah templates:

(65) NPN juxtaposition, nominal argument

```
template( np_arg_juxt,
[ID+ID1,
head:[phon:Phon],
synsem:[theta:spatial, cat:np, aggr:count, def:indef]
sem: {{[SOMEN1&(ID1+ID2)#NP1, SOMEN2#NP2|QSt],
          PSt, RSt},
          EV@and(Body,and at~[ NP1, EV],
          at~[NP2, EV])}
arg( ID1+I2+1 ):[phon:_PPPPhon,
   synsem:[cat:pp, head:[category:spatial].
   case:obliq, dir:right(1), flag:0], sem:SemN],
arg(ID2+ID3):[synsem:[cat:n], aggr:count,
   head:[phon:Phon],sem:SemN]],
           ]).
```

(66) NPN juxtaposition, nominal adjunct

```
template( np_adj_juxt,
[ID+ID1,
head:[phon:Phon],
synsem:[cat:np, aggr:count, def:indef]
sem: {{[SOMEN1&(ID1+ID2)#NP1, SOMEN2#NP2|QSt],
          PSt, RSt},
          EV@and(Body,and at~[ NP1, EV],
          at~[NP2, EV])}
arg( ID1+I2+1 ):[phon:_PPPPhon,
   synsem:[cat:pp, head:[category:spatial].
   case:obliq, dir:right(1), flag:0], sem:SemN],
arg(ID2+ID3):[synsem:[cat:n], aggr:count,
   head:[phon:Phon],sem:SemN]],
          ]).
```

The semantics under the sem value of the mother node say the following:

(67) There is some X and there is some Y, and the action described in the body is applied, which leads to the fact that X is at Y.

### Spatial extension

In this category, the construal emphasizes the extension of something between two or more concrete poles, which are expressed in the nouns. In example (68), the water extended over the area between two dykes.

(68) *dat 't water van dijk tot dijk stond*
that the water from dyke to dyke stood
'that the water extended between the two dykes'

Just as in the spatial succession category, all instances found are of the type *van* NPN, and the only preposition possible is *tot*.

Actually, this is a case of NPN which I consider closest to a regular, compositional expression. The two Ns are both involved in the event, and an entity extends between the two of them. The

question remaining is, why the nouns are determinerless count nouns. If we compare (68) with a version along the temporal axis, like, e.g., (69), which necessarily is compositional, we see that semantically, both patterns behave alike:

(69) *dat de Quiddich wedstrijd van maandag tot*
that the Quidditch match from Monday until
*maandag duurde*
Monday took
'that the Quidditch match took from Monday until Monday'

Then, compare (69) with (70):

(70) *dat de Quidditch wedstrijd van maandagochtend*
that the Quidditch match from Monday-morning
*tot dinsdagavond duurde*
until Tuesday-evening took
'that the Quidditch match took from Monday morning until Tuesday evening'

Example (70) is completely non-idiosyncratic, both semantically and syntactically. Example (69) is just as non-idiosyncratic, as the only possible reading is one where the match starts on a Monday in one week, and terminates on the Monday the following week. This explains why there is no such category as Temporal extension. The construal of an expression like *negen tot negen* ('nine till nine') will necessarily be 'from one day nine o'clock until the next day nine o'clock'. Structurally, the pattern is not suspicious, either, as the nouns involved are not count nouns.

The construction for an NPN spatial extension looks as follows:

(71) NPN spatial extension

$$\textit{npn-spat-ext-cxt} \Rightarrow \begin{bmatrix} \textit{van-NPN-cx} & \\ \text{MOTHER} & \begin{bmatrix}\text{SEM} & \textit{spat. extension}\end{bmatrix} \\ \text{DTS} & \begin{bmatrix}\textit{prep} \\ \text{PHON} \quad \langle \text{tot} \rangle\end{bmatrix} \end{bmatrix}$$

Let us take a look at the lexical templates that license expressions with the spatial extension NPN. Again, in Delilah I wrote two distinct templates, one for a spatial extension NPN that functions as an adjunct (72), and one that functions as an argument (73).

(72) Spatial Extension adjunct

```
template( pp_adj_spat_ext,
     [id: _Top+ID1,
     synsem:[autom:prepadj,  category:locative,
     cat:Cat, exsem:ExSem , exaggr:ExAggr],
     sem: {{[SOMEN1&(ID1+ID2)#NP1, SOMEN2#NP2|QSt],
             PSt, RSt},
             and(Body,and( between~[NP1, N2, EV],
             at~[ NP1, EV],
             at~[NP2, EV])}
     head:[phon:van],
arg( ID1+ID2+1 ):[phon:Phon,
   synsem:[cat:n, aggr:count,
   material:concrete, case:obliq,
   dir:right(0), flag:0], sem:SemN],
arg( ID1+I42+1 ):[phon:_PPPPhon,
   synsem:[cat:pp, head:[phon:tot].
   case:obliq, dir:right(1), flag:0], sem:SemN],
arg( ID4+ID5+1 ):[phon:Phon]],
   synsem:[cat:n, aggr:count,
   material:concrete, case:obliq,
   dir:right(2), flag:0], sem:SemN],
arg( ID1+_ID3+2 ):[phon:_SPhon,
   synsem:[cat:Cat, eventvar:EV,
   exsem:ExSem, exaggr:ExAggr,
   dir:left(1), flag:8],
   sem:{{QSt, PSt, RSt}, Body}]
       ]  ).
```

(73) Spatial Extension argument

```
template( pp_arg_spat_ext,
        [id: _Top+ID1,
        head:[phon:van, sem:_Sem],
        synsem:[theta:locative,
               cat:pp, exsem:ExSem , exaggr:ExAggr],
        sem: {{[SOMEN1&(ID1+ID2)#NP1, SOMEN2#NP2|QSt],
               PSt, RSt},
             EV@and(Body,and( between~[NP1, N2, EV],
             at~[ NP1, EV], at~[NP2, EV])}
arg( ID1+ID2+1 ):[phon:Phon,
   synsem:[cat:n, aggr:count,
   material:concrete, case:obliq,
   dir:right(0), flag:0], sem:SemN],
arg( ID1+I42+1 ):[phon:_PPPPhon,
   synsem:[cat:pp, head:[phon:tot].
   case:obliq, dir:right(1), flag:0], sem:SemN],
arg( ID4+ID5+1 ):[phon:Phon],
   synsem:[cat:n, aggr:count,
   material:concrete, case:obliq,
   dir:right(2), flag:0], sem:SemN],
        ] ).
```

The information captured under sem is the following:

(74) There is some X and there is some Y, and the action described in the body is applied in the space between X and Y.

Hence, for a sentence like *het water stond van dijk tot dijk* (literally 'the water stood from dyke to dyke') the event described in the body is that the water stands, which is between dyke 1 and dyke 2.
The adjunct version of this type of NPN is needed for sentences as example (75):

(75) *hij bouwde het dorp van dijk tot dijk*
he built the village from dyke to dyke
'he built the village (extending) between the two dykes'

In this template, the construction is not assigned a thematic role, but rather is categorized as a prepositional adjunct of the category 'locative'. The internal syntax and the semantics are the same as in the argument template.

### Transition

This category, again, encodes universal quantification. Here, transition of something from one member of the set described by the noun to the other is expressed. It includes that all members of the set described by N participate.[19]

(76) *dat gerucht ging zo van man tot man . van oor tot oor*
the rumour went so from man to man . from ear to ear
'the rumour just went from man to man, and from ear to ear'

This category has a lot in common with the succession one, including quantification over the set described by the noun, and it adds the factor of transition. Therefore, it is no surprise that the template for Transition adjuncts looks a lot like the Temporal Succession one:

(77) NPN transition

$$\textit{npn-transition-cxt} \Rightarrow \begin{bmatrix} \textit{NPN-cx} & \\ \text{MOTHER} & \left[\text{SEM } \textit{transition}\right] \\ \text{DTS} & \begin{bmatrix} \textit{prep} \\ \text{PHON} \quad \langle \text{tot} \rangle \end{bmatrix} \end{bmatrix}$$

The npn-transition class is a subtype of the *van*-npn construction and adds to it the information about its specific semantics and the choice of the preposition, namely *tot* ('to').

[19] There is a homonymous expression *van man tot man*, always used with a verb (or deverbal form) like *praten* ('talk') or *zeggen* ('say'). This expression is not a form of any category of NPN, but, just as *oog om oog* ('an eye for an eye'), it falls into the rest category of stored, unsystematic items.

(78) Transition NPN, prepositional adjunct

```
template( pp_adj_transition,
        [
        id: _Top+ID1,
        head:[phon:van, sem:_Sem],
        synsem:[autom:prepadj,
        category:temporal, island:WH,
        cat:Cat, exsem:ExSem , exaggr:ExAggr],
        sem: {{ [{{[SemN$X&ID1+ID2#N1], [], []},
        Z@some^X^and(quant(X, every), N1,
        entails( X, every),
        Z, entails( X, incr))}&(ID1+ID2)#NP1,
        {{[SemN$X&ID4+ID5#N2], [], []},
        P@some^X^and(quant(Y, some), N2,
        entails( Y, decr),
        P, entails( Y, incr))}&(ID4+ID5)#NP2|QSt],
        PSt, RSt},
        and(Body,and( from~[NP1, EV], to~[ NP2, EV]))}
arg( ID1+ID2+1 ):[phon:Phon,
   synsem:[cat:n, aggr:count,
   case:obliq, dir:right(0), flag:0], sem:SemN],
arg( ID1+I42+1 ):[phon:_PPPPhon,
   synsem:[cat:pp, head:[phon:tot].
   case:obliq, dir:right(1), flag:0], sem:SemN],
arg( ID4+ID5+1 ):[phon:Phon,
   synsem:[cat:n, aggr:count,
   case:obliq, dir:right(2), flag:0],   sem:SemN],
arg( ID1+_ID3+2 ):[phon:_SPhon,
   synsem:[cat:Cat, eventvar:EV,
   head:[etype:event], exsem:ExSem,
   exaggr:ExAggr, dir:left(1), flag:8],
   sem:{{QSt, PSt, RSt}, Body}]
        ] ).
```

Since the transition NPN can also occur in argument position, there is a lexical template for that case as well:

(79) Transition prepositional argument

```
template( pp_arg_transition,
        [id: _Top+ID1,
        head:[phon:van, sem:_Sem],
        synsem:[theta:spatial, autom:prepadj,
               island:WH,
        cat:Cat, exsem:ExSem , exaggr:ExAggr],
        sem: {{ [{{[SemN$X&ID1+ID2#N1], [], []},
             Z@some^X^and(quant(X, every), N1,
             entails( X, every),
             Z, entails( X, incr))}&(ID1+ID2)#NP1,
             {{[SemN$X&ID4+ID5#N2], [], []},
             P@some^X^and(quant(Y, some), N2,
             entails( Y, decr),
             P, entails( Y, incr))}
             &(ID4+ID5)#NP2|QSt],
             PSt, RSt},
             and(Body,and( from~[NP1, EV],
             to~[ NP2, EV]))}
arg( ID1+ID2+1 ):[phon:Phon,
   synsem:[cat:n, aggr:count,
   case:obliq, dir:right(0), flag:0], sem:SemN],
arg( ID1+I42+1 ):[phon:_PPPPhon,
   synsem:[cat:pp, head:[phon:tot].
   case:obliq, dir:right(1), flag:0], sem:SemN],
arg( ID4+ID5+1 ):[phon:Phon,
   synsem:[cat:n, aggr:count,
   case:obliq, dir:right(2), flag:0],
   sem:SemN],
arg( ID1+_ID3+2 ):[phon:_SPhon,
   synsem:[cat:Cat, eventvar:EV,
   head:[etype:event], exsem:ExSem,
   exaggr:ExAggr, dir:left(1), flag:8],
   sem:{{QSt, PSt, RSt}, Body}]
        ] ).
```

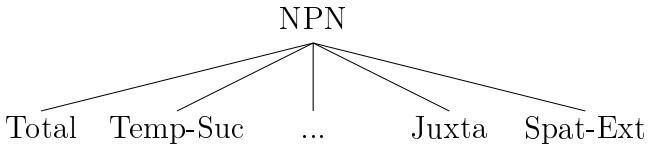
The difference with the spatial succession template is that the entire construction is not marked for the category spatial, and the nouns do not need to be concrete. The semantics turns out right, as the from–to relationship is combined with the event described in the body. Typically, the verb is something like *transmit*, and then the first argument of the verb is transmitted from X to Y, and for every X there is a Y that follows it.

## Analysis NPN

The different NPN lexical class constructions I proposed were rather straightforward results of the instances we observe in the CGN. The instances fall into different semantic categories, and every category comes with its particular constraints on the choice of possible prepositions, which is mostly a lexical, but in one case a semantic constraint (although I do not want to say that lexical constraints are not semantic in nature. They are, but mostly, the range of possible candidates is small enough to just state them directly.) But since the corpus is relatively tiny, this doesn't mean that the list is necessarily exhaustive. Nevertheless, I will base the analysis presented here purely on the data from the CGN. If the analysis of a larger data set reveals additional categories, an extension of the model is of course possible.[20]

Just as for the NCoN family of constructions, I propose an approach to NPN constructions that is lexicon-based rather than interpretive. I argue for an abstract construction (or lexical entry) of the form NPN, with N being a bare singular count noun, and that is the mother of a set of daughters, all of which have been described in the previous section.

(80) NPN hierarchy




The difference with the NCoN construction in (31) lies in the

[20]For a taxonomy of English NPN constructions, see Jackendoff (2008)

productivity of the constructions. In the case of NPN, they do function as pattern that can be used for the interpretation of new expressions. The (productive) subconstructions of NPN are Totalization, Temporal and Spatial Succession, Juxtaposition, Spatial Extension, and Transition. All of these inherit the constraints from the abstract supertype and add their own specific features, generally the type of preposition that can be instantiated, and the semantics of the particular lexical class construction.

On top of that, there are NPN constructions that are selected by *van*, some are not, some occur sometimes with and sometimes without *van*. This information has to be captured in the type hierarchy as well. In order to do so, we need to posit two distinct abstract supertypes, the *van*-NPN and the bare-NPN construction. The *van*-NPN is a version of the bare-NPN construction, with the difference that the NPN complex is selected by the preposition *van*. Different from the bare-NPN construction, the *van*-NPN construction projects a PP rather than an NP.

In the type hierarchy pf NPN constructions, the Totalization and Juxtaposition constructions are subtypes of the bare-NPN construction, the Spatial Succession, Spatial Extension and the Transition construction are subtypes of the *van*-NPN construction, and the Temporal Succession construction is a subtype of both higher level constructions, which licenses NPN expressions with and without the preposition *van*.

A second dimension that plays a role in the model of NPN constructions is the grammatical function of the NPN. Since most of them can occur both in argument and adjunct position, this needs to be accounted for. Again, this can be solved within a type hierarchy. The temporal succession NPN is both a subtype of a PP adjunct construction as well as a complementation construction. For architectural reasons, the Delilah templates described two distinct types for the argument and the adjunct position. This is not desirable, since it would resemble the constructional approach better if subtype and supertype were seperated, but eventually, this is only a matter of elegance that can easily be altered.

## 3.5 Conclusion

The NCoN and the NPN construction that I discussed in this chapter have a lot in common. Obviously, both constructions instantiate determinerless singular count nouns, which need to be accounted for, since they are syntactically odd in Dutch. Also, both types of constructions have semantics that cannot be predicted. The main difference between NCoN and NPN lies in the specificity of the patterns. In the case of the NCoN construction, all instances are lexically fixed, the construction itself is unproductive. Nevertheless, there is some need to posit an abstract type of NCoN, since there are regularities across all instances of the construction that we want to be able to capture. Hence, there is a supertype NCoN which captures the common constraints of all possible expressions that are licensed by this pattern.

The NPN subconstructions, different from the NCoN constructions, are still productive. Hence, the organization of the constructions needs to be different than that of the NCoN-types. Instead of lexical items that interact with the constructional network of types, we have a set of different types of abstract lexical-class constructions specified, which license novel expressions.

The different types of NPN constructions were categorized with the help of a corpus study. I extracted all instances of the pattern NPN from the Spoken Dutch Corpus (CGN), and grouped the matches in six semantic categories, which I described shortly. The main purpose of this categorization was to show some of the shortcomings of the Zero Semantics approach to these types of constructions. I argued that the ZS-analysis fails to account for subtle semantic differences between instances of the type NPN.

As I tried to show in this chapter, the major difference between a Zero Semantics analysis and a constructionist analysis lies in the universality of the pattern. Postma claims the form-function pairing to be an innate grammatical principle rather than a learned and stored mechanism, which leads to the complication that he covers more ground than empirical analysis seem to motivate. The problem that the constructionist analysis

suffers from, is the reverse: if the construction is not built in, then why does it look the way it looks, and not differently? In order to be able to answer a question like this, the two possible approaches are: 1) looking at the earliest stages of the particular pattern and its development, in order to explain synchronically odd features from a diachronic perspective, and 2) looking at the "odd" features in other environments. As we have seen in the part about NPN, some subconstructions are so close to compositionality that it will be a good topic for future research to investigate whether it is possible to arrive at a compositional analysis if we have a better understanding of the elements involved. For the two constructions discussed, this would clearly have to begin with a closer look at the properties of bare singular count nouns in Dutch.[21] However, Postma is right in his observation that both constructions involve quantificational processes, and there are reasons to assume that this is located in the bare singular noun. The nature of the relation between bare singular noun and quantification asks for further research.

Although I generally agree with Postma that both the NCoN and the NPN construction are remarkable and ask for a analysis that relates semantic to structural properties, Postma's claims go too far. By proposing the structural semantics of the constructions to be universal, his model makes predictions which fail to hold consistently in Dutch itself, let alone across languages.

If we consider other languages, we do not necessarily see the discussed constructions in action. Interestingly, French expresses *hand in hand* with *la main dans la main*, i.e. with determiners. With other nouns, it occurs without the determiners, like in *face à face* ('face to face') and *vis à vis* ('face to face'). Spanish, another Romance language, does not exhibit the construction in question at all. *Hand in hand* is expressed with a PP, namely *de la mano* ('of the hand'). Czech also lacks a comparable construction. This does not exactly come unexpected, as—being determinerless—Czech would not have any means to signal

[21] Work has been done on bare singulars, e.g. de Swart & Zwarts (n.d.), de Swart *et al.* (2005), but either the questions addressed or the data were such that a straightforward analysis would not have been easily possible.

the fact that this is an instance of the NPN construction, and that therefore, the semantics are not compositionally built. On the other hand, Russian, comparable to Czech to a large extend, does exhibit this construction, in expressions like *ruka ob ruku* ('hand in hand') and *plecho k plechu* ('shoulder to shoulder'), but then again, in this type of NPN, the semantics of the expression are not very odd.

For English, German, or Dutch, the signalling power of an NPN pattern is indeed very effective, as it structurally deviates from the non-idiosyncratic cases, at least when the nouns are common count nouns. As this is not a pattern that can be built with the general structures, a special semantics for the special syntax is not surprising. Special meaning is encoded in special structures. But, then again, how special is the meaning of structures like *hand in hand* and *bumper aan bumper*? If Paul and Mary walk hand in hand, there are two hands involved that are somehow intertwined. One hand could be seen as embracing the other, so the other hand would end up *inside* the first one. The same holds for *bumper aan bumper*. If two cars stand bumper to bumper, there are two bumpers, that are touching, and if 1000 cars stand bumper to bumper, it is always two bumpers, the front bumper of one car and the rear bumper of another car, that are touching each other. The question arising from this is then: why do we have to assume that one of the two similar nouns is moved out of the domain of lexical closure, if both nouns play a role in the framed situation?[22] If we accept that the semantics are built in a way which lets both nouns contribute their share to the interpretation, we can account for the fact that Russian has a comparable pattern, although it does not have determiners. If the interpretation can be built compositionally, there is no need for a pattern that stands out in order to signal particular semantics.

[22]Beck & von Stechow (2006) give a compositional analysis of expressions like *dog after dog* and *piece by piece*, in the context of categorizing pluractionals. They conclude, that all pluralization is sensitive to a division of appropriate subparts, and that N P N constructions tell us which units are contained in the cover (p. 53).

As Postma points out, the NCoN constructions as well as the NPN constructions are not random idiosyncratic expressions, but they seem to obey certain rules that hold for (almost) the entire group of expressions.[23] But if we attribute a structural semantics A to a phrasal configuration B, this mechanism cannot account for differences between distinct examples of this configuration.

[23]In a personal conversation, Postma explained that, in an ideal case, the Zero Semantics approach should be applicable for the entirety of idioms and noncompositional expressions. A model like this is very hard to imagine. First of all, the majority of idioms does exhibit a standard sentence pattern, like, e.g., transitive. But, even if we assumed that idiomatic transitive sentences have a phrase marker that is different from an unidiomatic transitive one, it could not take care of the different meanings that different idioms have. To assume that *to spill the beans* and *kick the bucket* have structures that are not only distinct from compositional transitive expressions, but also from each other, would create an inventory of different phrase markers that is probably at least as large as the inventory of idioms.

# Chapter 4

# The Way-Construction

In this chapter, I am going to discuss a construction type of a rather high degree of flexibility, namely the Dutch version of what is commonly known as the *way*-construction (for earlier analyses of the English *way*-construction, see, among others, Jackendoff (1990), Goldberg (1995) and (1996), Marantz (1992), and Levin & Rappaport Hovav (1995). For earlier analyses of this construction in Dutch, see Verhagen (2003), Poß (2005), and van Egmond (2006)).[1]

Examples of the construction under discussion are given below:

(81) *Braid virus baant zich een weg door*
Braid virus *banen-3rdSG* REFL a way through
*email.*
email.
'Braid virus makes its way through email.'

(82) *Twee bussen boren zich een weg naar het hart van*
Two busses drill REFL a way to the hart of
*Istanbul.*
Istanbul.
'Two buses drill their way to the hart of Istanbul.'

[1]Goldberg (1995:217), referring to personal communication with Annie Zaenen, states that there is no such thing as a *way*-construction in Dutch, which is apparently not true. But that has been mentioned in other papers.

(83) *De flits baant zich een gloeiend heet*
The lightning *banen-3rdSG* REFL a red hot
*pad door de lucht.*
path through the air.
'The lightning makes his red hot path through the air.'

Below, I will first discuss the question why the Dutch *way*-construction has constructional status, followed by a discussion of the formal and semantic properties.

## 4.1 A construction in its own right?

As has been pointed out by all of the aforementioned authors, the English as well as the Dutch *way*-construction have rather peculiar features, syntactically and semantically. Therefore, it needs to be accounted for as a construction in its own right. Both patterns have a variety of properties that cannot be captured with the general systematic patterns and regularities of the respective languages. Jackendoff (1990:218) describes the English construction as "a fairly outrageous example of mismatch between syntactic and conceptual structure". He describes three different layers of mismatch, namely that the relative embedding of syntactic and conceptual constituents is different, that the Path argument and the *way*-NP are neither syntactic nor semantic arguments of the main verb, and that the conceptual functions GO (event function) and AFF (for 'affect'), which he attributes to the semantic structure of the *way*-construction, are not expressed by any lexical item in the sentence (ibid., p. 218). In this respect, it seems valid to claim that the *way*-construction is not a sentence pattern that compositionally falls out of the existing inventory of contructions, but needs to be accounted for separately in the grammar.

The Dutch *way*-construction was first analyzed by Verhagen (2003). He found that it differs from its English counterpart in a number of respects. A contrastive study of the English and the Dutch construction has been described by Verhagen (2007).

In the following part, I will give an overview of the most important properties of the Dutch *way*-construction.

### 4.1.1 Formal properties

The morphosyntactic properties of the *weg*-construction[2] differ from the English one in that in Dutch, it instantiates a sentence pattern with two objects. The first object is always a reflexive pronoun coindexed with the subject. The second object is the NP *een weg* ('a way') by default, although semantically similar nouns as *pad* ('path') or *baan* ('way') do occur as well.

Verhagen (2003:34) presents a schematic description of the Dutch *way*-construction ((85)[3], which is based on (84), Goldberg's description of the English construction (Goldberg 1995:207)):

(84)

| | | | | | | |
|---|---|---|---|---|---|---|
| Sem: | CR-MOVE | < | creator-th. | createe-way, | path | > |
| | \|means | | \| | ⋮ | ⋮ | |
| | PRED | < | | | | > |
| | ↓ | | ↓ | ↓ | ↓ | |
| Syn: | V | | $\text{Subj}_1$ | $\text{Obj}_{way1}$ | Obl | |

(85)

| | | | | | |
|---|---|---|---|---|---|
| Sem: | creator | cr.-move, | for-self | created-way, | path |
| | \| | \|means | \| | \| | \| |
| Syn: | [$\text{SUBJ}_\text{i}$ | [V | [$\text{REFL}_\text{i}$ | [*een weg*] | OBL]]] |

What both figures show is the following: the (abstract) construction consists of a number of meaning components (under Sem) that must occur in all particular constructs. The presence of syntactic elements (under Syn) is obligatory as well: for Dutch, that is a ditransitive pattern with an oblique argument (i.e., a PP,

[2] In the remainder, I am going to use the term *weg*-construction for the Dutch *way*-construction.

[3] I added a bracket that is missing in the original.

usually). The two sides of the description have no distinctive power in themselves, i.e., they do not point to this one construction uniquely. The semantic structure could also be represented by an utterance that does not make use of the *weg*-construction, and a ditransitive sentence with an oblique argument is nothing particularly special, either.[4] What makes it unique is the linking of those two layers (depicted in the lines between the semantic and syntactic structure, showing which semantic component is expressed by which constituent, or vice versa).

The difference between Goldberg's schema in (84) and Verhagen's schema in (85) is that Verhagen seems to make claims about phrase structure, while Goldberg does not. In chapter 2, I discussed at length the shortcomings of a constructional approach that makes itself too dependent on phrase structure, or, even worse, that assumes that phrase structure carries meaning. I am not sure whether Goldberg intends to express this idea, but from her formalization, this does not follow. She merely sums up the argument positions that are part of the *way*-construction, and with which grammatical role they are going to be expressed in a sentence. The actual order of the elements is not expressed and might follow from higher, more general word order properties of English. Still, the construction needs to rule out passive and other lexical processes.

Verhagen's schema of the Dutch construction is much more rigid than Goldberg's. By bracketing the Syn-side of the construction, he pins down the phrase structure for the *weg*-construction. Again, I want to refer to chapter 2 for a discussion why that is not desirable.

In Verhagen (2002:414), an interesting point is raised. Verhagen observes that the Dutch *weg*-construction is a productive instance of an unproductive pattern, namely the ditransitive. He calls it paradoxical that—while English has a productive ditransitive and Dutch does not—the Dutch *weg*-construction instantiates the ditransitive pattern and the English construction does not. I don't think that this is very paradoxical. If a language

[4]Below, I will argue that the PP is not an adjunct on the sentence level, but a modifier of the *weg*-NP.

has a productive pattern X, a novel utterance can easily be interpreted as an instantiation of that pattern. In the case of the ditransitive, a sentence like *Hij graaft zich een weg naar buiten* ('he digs himself a way to outside'), the possible ditransitive interpretation would be that he digs a way outside (the theme) and that he is the recipient or benefactive of that. This interpretation is perfectly imaginable, but it lacks the specific semantics of the *weg*-construction as well as some formal properties. It would be hard, for example, to account for the fact that this sentence does not have a passive counterpart that is productive to a normal degree.[5] If a language does not have a productive ditransitive, the ditransitive interpretation is blocked, so the expression must be an instance of another construction, in this case the *weg*-construction. If a language does have a productive ditransitive, the double-object pattern would not be a very good option for the *weg*-construction, since assigning a literal reading would be too easy. Hence, the fact that a language with an unproductive ditransitive uses a ditransitive pattern as the formal schema for another construction while a language with a productive ditransitive does not, is not very surprising. This does not mean, of course, that constructional ambiguity is uncommon, let alone impossible. See, e.g., the overlapping formal structures of the Caused-Motion construction and the prepositional alternative to ditransitives, as in *He sent a letter to London* and *He sent a letter to grandma*. Nevertheless, using an unusual pattern for the expression of unusual semantics helps interpreting the construction correctly.

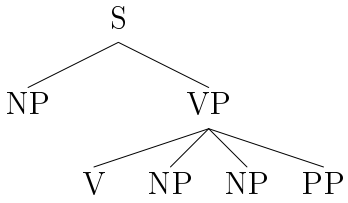
A last point that I want to discuss concerns the hierarchical structure of the pattern. In analogy with the English construction, Verhagen claims that the PP adjunct is directly dominated by the VP (as in (86)). In Verhagen (2002:416), he argues that the *weg*-construction in Dutch is not a subordinate construction of the ditransitive, but that it rather forms an island in the Dutch network of constructions. Under his own analysis, it would actu-

[5]Some might say that the *weg*-construction does not have a productive passive. And indeed, examples are scarce, but they exist. So I chose not to exclude passives entirely.

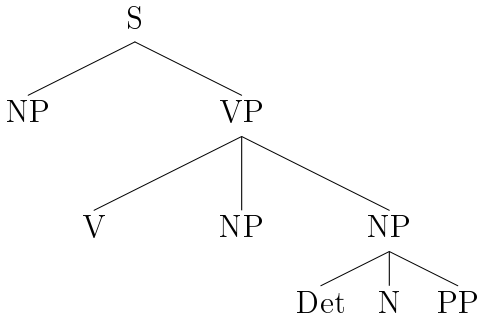
ally be much harder to argue that it is a subconstruction than that it is not, as the extra argument (the OBL) would have to be explained. Only under an analysis where the OBL modifies the *weg*-NP, as in (87), the *weg*-construction would qualify as a candidate for a subconstruction of the ditransitive.

The Dutch *weg*-construction is not a more specific Ditransitive with an extra argument position, but the PP is a modifier of *weg* and therefore forms a complex constituent with the *weg*-NP (as in (87)).[6]

(86)




(87)




Van Egmond (2006, chapter 3.7) argues for the analysis in (87), and the strongest points in her argumentation will be repeated here.[7] One of her arguments stems from general constraints on phrase structures. She argues that we never find a case where there are three VP internal arguments, so it is not reasonable to posit this for the *weg*-construction:

[6]The trees presented here do not aim at representing a particular theoretical framework. Therefore, there are no intermediate levels, or bar levels, and ternary branching is allowed for. In fact, these trees encode dependencies, rather than actual phrase structures.

[7]Boban Arsenijevic (p.c.) already suggested in 2004 that the PP is not a daughter of the verb but a modifier of the noun. This remark unfortunately remained unheard in Poß (2005) and Poß & van der Wouden (2005).

"[I]f it were that the *weg* noun that denotes the path that is created and the PP denotes the path that is travelled (as suggested by Verhagen), the *weg*-construction would contain three internal arguments: the reflexive, the *weg* NP (created path) and the PP (travelled path). This is clearly undesirable, because verbs do not normally take more than two internal arguments. This also means that the creation of the path and the travelling of the path are not necessarily co-extensive, as they are expressed by separate arguments. However, the creation and the travelling of the path are always co-extensive in the *weg*-construction, because the path is travelled by creating it, when the subject stops creating the path, he will stop traversing the path. Consequently, there should only be one argument that expresses the path." (Van Egmond 2006: 72f.)

So, in addition to the general claim that VPs don't usually select three internal arguments (or, in more constructionist words, if constructions don't usually license patterns with three internal arguments), she claims a semantic inseperability of the creation and the travelling of the path, as has been suggested by Verhagen (2002), which forms another reason not to argue for two different arguments, but one.

Those arguments are easy to cast aside. The first one is unstable as there are cases of four-place predicates, as in *John exchanges his book for a CD with Mary* or *I bet you four dollars that* ..., so it might be rare, but it is not impossible to encode three internal arguments. Van Egmond's second, semantic, argument only is applicable under an analysis that adopts the semantics of the *weg*-construction as creating and traversing a path. Below, I will discuss an alternative analysis that does not encode two distict events. Therefore, this argument is not suitable under the new analysis.

Nevertheless, there is evidence for the constituency of NP and PP, for instance the case of passivization. Even though passives of the *weg*-construction are rather hard to find, to a degree that

lets one assume that they are only marginally grammatical, the examples that I found in a Google search front the entire complex.[8]

(88) *De weg naar fantasievollere mannenmode werd*
The way to more-fanciful man-fashion was
*gebaand door de Beatles [...]*
GEBAAND by the Beatles [...]
'The way to more fanciful fashion was paved by the Beatles'[9]

(89) *De weg naar de kanselruil werd gebaand door*
The way to the pulpit-swap was GEBAAND by
*een verklaring [...]*
a declaration [...]
'The way to the pulpit swap was paved by a declaration'[10]

Notice that the reflexive is missing in both examples. Therefore, we cannot speak of clear passive *weg*-sentences. Nevertheless, simply the observation that NP and PP can be fronted as one whole in this closely related type of expression, is further evidence that the entire complex forms one constituent.

Note that we find a definite article here instead of an indefinite. This unusual behaviour, in combination with the sparsity of passives of the *weg*-construction proper and the comparably large amount of 'for-someone' *weg*-sentences makes it likely that there is no productive passive of the *weg*-construction proper, and that the few examples found are instances of the 'for-someone' sentences, that usually take a definite article. Another reason for

[8]The search for *werd gebaand* ('was GEBAAND') resulted in a mere 135 hits, the search for *gebaand werd* ('GEBAAND was') in 83. It is evident, though, that most of the sentences are not instantiations of the *weg*-construction proper, but of the pattern where the way is made 'for someone/something'.

[9]URL: www.cultuurwijzer.nl/cultuurwijs.nl/cultuurwijs.nl/i000665.html (10.12.2007)

[10]URL: archief.trouw.nl/artikel?REC=454dfe240925604c46b1b1cf27c8-c982 (10.12.2007)

the definite article could be the fact that Dutch does not allow for indefinite NPs in sentence-initial position.

Again, under the analysis presented below, it is not surprising that passives are infelicitous. First of all, the verb is necessarily construed as an intransitive verb, and those do not feed passive, and secondly, the NP is considered nonreferential, which does not make it a good candidate for being the subject of a passive sentence.

A test that is applicable for examining the constituency of a complex, the cleft sentence test, is not supported by natural data. Neither in the available corpora nor in a Google search did I find examples of clefted *weg*-constructions. Also, my own intuition does not allow a clefted *weg*-construction sentence. Under the semantic analysis that I will propose below, this was again to be expected. A cleft construction is used to pick out and highlight a constituent, and as I claim that the *weg*-NP is not a semantic element in the meaning representation of the *weg*-construction, it cannot be clefted (see section (4.3)). If we leave out the reflexive *zich*, clefting is possible.

Nevertheless, hypothetically constructed clefted examples still sound at least less unnatural if NP and PP remain adjoined, in cleft as well as pseudo-cleft constructions:

(90) *?Het was een weg naar de finale die Bubka zich*
It was a way to the finals that Bubka REFL
*baande*
BAANDE

(91) **Het was een weg die Bubka zich naar de finale*
It was a way that Bubka REFL to the finals
*baande*
BAANDE

(92) *?Een weg naar de finale is wat Bubka zich*
A way to the finals is what Bubka REFL
*baande*
BAANDE

(93) **Een weg is wat Bubka zich naar de finale baande*
A way is what Bubka REFL to the finals BAANDE

Another argument in favour of the NP-adjunct analysis is the observation that the linear order of *weg*-NP and PP cannot be interrupted by adjuncts. In English, on the other hand, this is easily possible. A quick Google search for the string *his way slowly through* resulted in 71.100 hits[11] Hence, the English example sentences below are fine, while the Dutch sentences are not felicitous.

(94) Ali smiled as he made his way slowly through it.[12].

(95) He picked his way slowly through the darkened alley, trying to avoid the larger concentrations of refuse.[13]

(96) **Braid virus baant zich een weg langzaam door email.*
Braid virus *banen-3rdSG* itself a way slowly through email.
'Braid virus makes its way slowly through email.'

(97) **Twee bussen boren zich een weg langzaam naar het hart van Istanbul.*
Two busses drill themselves a way slowly to the hart of Istanbul.
'Two buses make their way slowly to the hart of Istanbul.'

The example sentences (96) and (97) are repetitions of the examples in (81) and (82), but I inserted a VP adjunct, *langzaam*

[11]March 2010. I am well aware of the complications that arise when one tries to exemplify linguistic intuitions using a search engine on the web. And particularly for a language like English, which is the lingua franca in many areas, questions like, e.g., the native language of the author of a certain text, are impossible to answer. Nevertheless, more than 35.000 occurrences are a strong indication that this is not an odd pattern in English.

[12]URL: http://www.touregypt.net/kids/story72.htm (10.07.2007)

[13]URL: http://www.goodwriters.net/wizardsbane.html (10.07.2007).

('slowly'), between the *weg*-NP and the PP adjunct, which leads to unacceptability. If NP and PP can stay adjacent and the adjunct is placed in front of the complex, the sentences are fine, again.

(98) *Braid virus baant zich langzaam een weg*
Braid virus *banen-3rdSG* itself slowly a way
*door email.*
through email.
'Braid virus makes its way slowly through email.'

(99) *Twee bussen boren zich langzaam een weg naar*
Two busses drill themselves slowly a way to
*het hart van Istanbul.*
the hart of Istanbul.
'Two buses make their way slowly to the hart of Istanbul.'

Another Google search for the patterns *baant zich een weg langzaam* ('BAANT REFL a way slowly') and *baant zich een weg snel* ('BAANT REFL a way fast') returned no hits. The string *baant zich een weg* returned 12.100 hits.

Although separation of the NP and the PP cannot be reached by inserting an adjunct of any kind, they can be separated by a verb form. In Poß (2005), I cited the following complex NP, which actually is an example of this:

(100) *een prachtige streek waarin zeven riviertjes*
a wonderful area where-in seven rivers-DIM
*zich een pad kronkelen naar de zee*
themselves a path wind to the sea
'a wonderful area where seven rivers wind to the sea'[14]

In this case, the lexical verb interrupts NP and PP. Nonetheless, this is not a counter-argument to the observation that NP and PP are inseperable, as this sentence is an example of extraposition of the usual kind. In Dutch, (parts of) lexical verbs can interrupt complex object constituents, as is demonstrated in the following examples:

[14] http://www.freewebs.com/maisjo/infooverdestreek.htm (01.02.2005).

(101) *jullie hadden nooit de eerste man gezien van Ma*
you had never the first husband seen of Mom
*Flodder*
Flodder
'you have never seen Mom Flodder's first husband'[15]

(102) *ik ben 26 en heb in sl een man ontmoet van 66*
I am 26 and have in sl a man met of 66
'I am 26, and in sl I met a 66 year old man'[16]

In both examples, the PP following the lexical verb is clearly a modification of the head of the object NP. Note that this interruption is only possible with object NPs, and that the interrupting element can only be a verb (or a verb particle). Interfering adjuncts lead to unacceptability, as we see if we slightly alter the above examples:

(103) **jullie hadden de eerste man nooit van Ma*
you had the first husband never of Mom
*Flodder gezien*
Flodder seen
'you have never seen Mom Flodder's first husband'

(104) **ik ben 26 en heb in sl een man laatst van 66*
I am 26 and have in sl a man recently of 66
*ontmoet*
met
'I am 26, and in sl I recently met a 66 year old man'

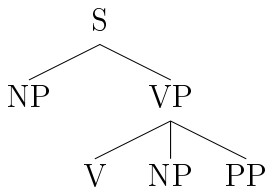
Therefore, the discontinuous *weg*-NP in (100) is not a counter-argument against the claim that the *weg*-NP is a complex NP that contains the PP modifier, but it follows from the more general principle of Dutch, that complex NPs can be interrupted by a lexical verb.

An explanation for why English allows adjuncts to separate the NP and the PP may come from the observation that with

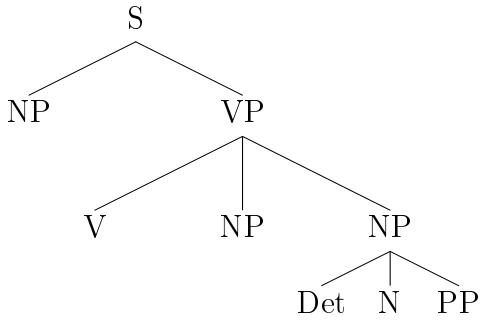
[15]URL: http://www.cvsjustin.aaijtwieverbal.htm (10-01-2007).
[16]URL: http://www.secretsecondlife.nl (10-01-2007).

few exceptions, Dutch as well as English allow for at most two VP-internal arguments.[17] While the English *way*-construction instantiates a monotransitive pattern with only *X's way* as an NP argument, there is still room for another obligatory argument. The Dutch *weg*-construction, instead, already has two internal argument positions occupied, *zich* and *een weg*, and therefore does not lean to an obligatory third argument. Even though the structures of the Dutch and the English construction might look similar, they are not, as the separability of the NP and the PP shows. English instantiates a pattern that looks like (105), while Dutch instantiates a pattern that looks like (106). I clearly do not want to contradict myself here by positing a phrase structure for a construction. This tree merely intends to exemplify the hierarchical structure of the latter NP.

(105) Structure English *way*-construction




(106) Structure Dutch *weg*-construction




[17] For the exceptions, think of verbs like bet, rent, etc.

### 4.1.2 The elements of the way-construction

Let us now sum up the formal parts that build this construction and take another look at Verhagen's schematic construction in (85), here repeated as (107):

(107)

| Sem: | creator | cr.-move, | for-self | cr.-way, | path |
|---|---|---|---|---|---|
| | \| | \|means | \| | \| | \| |
| Syn: | [SUBJ$_i$ | [V | [REFL$_i$ | [*een weg*] | OBL]]] |

The first element in this construction is the subject, which, as the "creator" and "traveller", has to be able to conduct some self-propelled motion. Therefore, the subject is usually animate, or can be interpreted as something having its own will, like a cancer cell (see example (108)) or, e.g., a river.

(108) *De kankercel heeft namelijk proteasen nodig [...]*
the cancer-cell has namely proteases necessary [...]
*om zich een pad door het lichaam te*
in-order-to REFL a path through the body to
*kunnen vreten*
can eat
'In fact, the cancer cel needs proteases in order to eat its way through the body'[18]

As I mentioned above, Dutch does not allow for sentence-initial indefinite NPs, hence the subject NP must either be definite, or the first argument slot must be occupied by *er*, which is comparable to the English existential *there*:

(109) *Er baant zich intussen overigens een*
there BAANT REFL in-the-meantime by-the-way a
*soort duizendpoot op mij voet een weg naar boven*
kind-of centipede on my foot a way to upstairs
'In the meantime by the way, there is a centipede on my foot making his way up'

[18] http://noorderlicht.vpro.nl/artikelen/10343958/ (05.07.2007)

The second component is the reflexive that functions as the indirect object. It is always co-indexed with the subject. What is interesting about this element is its strong preference for the weak form *zich*. Dutch has two reflexive pronoun forms, *zich* and *zichzelf*. Verhagen (2002:31) reports that he does not find any occurences of the strong reflexive in this construction in his search in the *Volkskrant* corpus, and only very rare hits in a Google search.[19]

(110) *Iedereen baant zichzelf steeds meer een eigen weg*
Everyone BAANT REFL always more an own way
*door de harde realiteit en iedereen heeft een eigen*
through the hard reality and everyone has an own
*wil*
will
'Increasingly, everyone makes his own way through the hard reality and everyone has their own will'[20]

This preference for weak reflexives can be explained straightforwardly, if we assume that the reflexive is not a semantic argument of this construction, but merely a formal position that needs to be filled without any lexical semantic contribution. Therefore, emphasizing this argument by means of using the strong form is not possible.

A comparable dichotomy is the difference between the real and the fake reflexive verbs. Real reflexives like *zich schamen* ('be ashamed') or *zich vervelen* ('be bored') cannot occur with a strong reflexive form (examples (111) and (112)). In these cases, the reflexives are in object position, but they are semantically

[19]On 10-10-07, the Google search returned 58 hits for *baant zichzelf* and 23.100 for *baant zich*. The string query is reliable to only return *way*-construction sentences, as the verb *baant* only occurs in this pattern.

[20]URL: http://weblog.fok.nl/Kort20verhaal/2 (10.10.2007). This sentence might look like a counter-example to my claim that the *weg*-noun cannot be modified, because here, it obviously is, while usually, we do not find any modification of the *weg*-element. But at a second glance, it is not the *weg* that is modified, but *eigen* behaves like a VP manner adverb and modifies the entire predication rather than merely the noun, as we know from sentences like "an occasional sailor walked by".

void. Fake reflexives behave differently, as the reflexives here are actual, semantically meaningful, arguments.[21] In these cases, the reflexive can adopt the strong form (examples (113) and (114)) For a deeper explanation of this phenomenon, see, e.g., Haeseryn *et al.* (1997), chapter 5.3.4.

(111) *Ik schaam me*
I shame me
'I am ashamed'

(112) **Ik schaam mezelf*
I shame myself
'I am ashamed'

(113) *Ik was me*
I wash me
'I wash myself'

(114) *Ik was mezelf*
I wash myself
'I wash myself'

The observation that the strong reflexive is very rare (and, as Verhagen reports, is judged infelicitous by many native speakers of Dutch), serves as additional evidence for the claim that it does not function as a semantic argument of the sentence.

Van Egmond (2006) claims that the reflexive is not a semantic argument of the verb, due to the observation above and the fact that the verbs are usually intransitive (or optionally intransitive) verbs. In my opinion, the status of not being a semantic argument of the verb follows from this observation. If an argument is not an argument of a verb, it will also not be a semantic argument of this verb. Rather, it could be seen as an argument position of the construction. Nevertheless, this only follows from the mismatch of numbers of argument roles, verb versus construction. The use of the weak reflexive as the default leads me to the claim that

[21] In the English literature, this categorization is sometimes exactly the other way around.

the reflexive is not a semantic argument at all, neither of the verb, nor of the construction. Below, I will discuss this in more detail.

The next element is the *weg*-NP, which I discussed earlier in connection with the hierarchical structure of the NP PP complex. Apart from the alternative analysis of the structure, the former analyses do not provide sufficient evidence for the assumption that only the lexical head *weg* can occur in this position. On the internet, there are plenty of natural sounding sentences with either *pad* ('path', see (115)) or *baan* ('pathway', see (116)). Sure, these are semantic equivalents of *weg*, but it appears that the restriction is on a semantic paradigm, the class of lexical elements that depict a sort of way, rather than on a single word.

(115) *De mannen moesten zich een pad banen, en de*
the men had-to REFL a path BANEN, and the
*bijl van den timmerman bewees goede diensten*
axe of the carpenter proved good services
'the men had to make themselves a path, and the carpenter's axe proved very useful'[22]

(116) *Meanderend slingert de rivier zich een baan*
Meandering winds the river REFL a pathway
*door het dal*
through the valley
'The river meanders its way through the valley'[23]

Note here that in (115) the directional PP is elliptic. Also, in (83), here repeated as (117), the *pad*-noun is modified, which normally is not possible.

[22] http://www.dbnl.org/tekst/fran018kant01_01/–fran018kant01_01_0008.html (06.07.2007)

[23] URL: http://www.olivierwillemsen.nl/oli4/column2.oli/258/ (10.10.2007)

(117) *De flits baant zich een gloeiend heet*
The lightning *banen-3rdSG* REFL a red hot
*pad door de lucht.*
path through the air.
'The lightning makes his red hot path through the air.'

It seems that the deviation from the more typical noun *weg* opens the door for a more literal interpretation, a less semantically bleached one. *Pad* can be modified, while *weg* cannot. *Pad* occasionally stands alone, i.e. without a directional PP, while *weg* cannot. Nevertheless, it is obvious that there is a default verb, *banen* (see Verhagen (2003:35), who found that in the Volkskrant corpus, 59 out of 92 examples of the *way*-construction instantiate the verb *banen*, whereas the rest of the examples distribute over 17 other verbs), and the default nominal head is *weg*. Still, both lexical items are not obligatory. It is not even the case that either *banen* or *weg* need to be present in the sentence, although they are the two highest frequent elements, which is illustrated in (118) and (119):

(118) *Dunk is wat trager en waggelt zich een*
Dunk is somewhat lazier and hesitiates REFL a
*pad door het leven*
path through the life
'Dunk is somewhat more lazy and spends his life hesitating'[24]

(119) *Met hakbijlen hakten de bosjagers zich een*
with axes chopped the wood-hunters REFL a
*pad door het woud*
path through the wood
'the hunters chopped their way through the forest with axes'[25]

[24] http://www.hetrietje.nl/contents/nl/d268_bacladi.html (06.07.2007)
[25] http://www.bertsgeschiedenissite.nl/boeren/mill7/bosjagers2.html (06.07.2007)

One observation we can make is that the less default the choice of lexical items is, the less idiosyncratic the meaning of the utterance becomes. Also, (118) is still a clear instance of the *weg*-construction, while (119) already leans strongly towards a more literal interpretation. It is easier to construe a situation where someone chops a path into a wood than a situation, where someone hesitates a path through a life. The degree of idiomaticity is not fixed in the case of the *weg*-construction, but seems to depend on the choice of lexical items.

But let us come back to the *weg*-NP element. The determiner of the *weg*-noun is always indefinite, *een*, at least in this sense of the construction. Verhagen (2003) defines three different family members of the *weg*-construction, namely a) making oneself a way, b) easing and blocking the way, and c) finding one's way. In my description so far I only treated the first sense, but the other two senses only differ slightly from this, the most frequent, construction. An example of subconstruction b) is the following sentence, taken from Verhagen (2003:40):

(120) *Sex baande voor hem ook de weg naar de*
sex 'banen'-PAST-SG for him also the way to the
*roem*
fame
'Sex also paved the way to fame for him'

In this type of *weg*-construction, instead of a reflexive there is a full NP or pronoun marked with *voor* ('for'), and the determiner of the *weg*-NP is definite. Verhagen (2003:42) argues that the *voor X* argument is not at all obligatory, and that it is more like a goal-like role, rather than a beneficiary, as he calls the *zich* in the a) version of the *weg*-construction. This is an important observation, and it provides evidence that one of the differences between making and easing ways is the argument status of this second argument. In the making-version, *zich* is not a semantic object, whereas in the easing-version, *voor PRO* is a regular argument that semantically contributes to the expression. Another piece of evidence is that the NP marked with *voor* is fully referential and not restricted to referring back to the subject, as the

following example, also from Verhagen (2003: 42), shows:

(121) *Ze blokkeerden de weg tot de kassa's voor*
They blocked the way to the cash-registers for
*de rest van de menigte*
the rest of the crowd
'They blocked the way to the cash-registers for the rest of the crowd'

This is only one of several differences between making oneself a way and easing or blocking a way, but for me, this is sufficient to exclude the blocking type from my analysis. Due to obvious similarities between the two types of sentences, Verhagen included the blocking-type of sentences in his description, but as I see it as a compositional instance of a transitive verb pattern with fully referential arguments, it falls outside of the scope of phenomena I want to deal with in this chapter.

The last type of sentences Verhagen discusses is not in the focus of my dealings, either. It includes sentences like (122):[26]

(122) *Veel kunst vindt via vlooienmarkten zijn weg naar de*
Much art finds via flea-markets his way to the
*kopers*
buyers
'A lot of art finds its way to the buyers via flea markets'

This pattern is formally and functionally very far from the central *weg*-construction.

Last, but not least, the paradigm of verbs that can occur in the *weg*-construction is restricted to activity verbs that can have an intransitive reading (see Egmond (2006) for a list of verbs that she found in this construction). Stative verbs are at least unusual (although it is imaginable that someone sleeps his way through school), and strictly transitive and ditransitive ones are infelicitous. The same holds for reflexive verbs. In a nutshell, all verbs that necessarily provide two or three argument slots do not enter the *weg*-construction. Semantically, that makes sense.

[26] In Verhagen (2003:48).

If the verb contributes an own slot for an object, this slot needs to be (meaningfully) filled. As previous studies claim, the two internal arguments are not semantic arguments of the verb, hence, there is a necessary mismatch if the verb demands a semantic argument position to be filled. There necessarily needs to be a mismatch between the argument slots of the verb and the argument positions of the construction, which cannot be fulfilled in the case of strictly transitive or strictly reflexive verbs.

In the section 4.3, I will go one step further and argue that the reflexive and the *weg*-NP are not at all semantic arguments of the construction, i.e., they don't contribute any meaning to this type of expression.

### 4.1.3 Semantic properties

Not only the morphosyntactic properties of the Dutch *weg*-construction are peculiar, the semantics of this type of sentence is remarkable, too. Verhagen (2003) describes the semantics of the Dutch *weg*-construction as being analogous to the English one, and therefore gives it the same meaning that Goldberg gives for the English *way*-construction. That is, in this view "the subject referent creates a (possibly metaphorcal) path and/or removes obstacles on it, and travels it" (Verhagen 2003:32). Goldberg (1995:199f.) explains that for the English *way*-construction, expressions like (123) express motion, even though the verb does not need to express motion. She contrasts these sentences with examples like (124), which can be continued as in (125), whereas instances of the *way*-construction cannot, as in (126).

(123) Frank dug his way out of prison

(124) Frank dug his escape route out of prison

(125) Frank dug his escape route out of prison, but he has not gone yet

(126) *Frank dug his way out of prison, but he has not gone yet

For Dutch, exactly the same generalization holds, and motion is without doubt an attribute that is crucial to the meaning of this type of expressions.

Verhagen, Goldberg and Van Egmond all argue for a semantics of the *way*-construction, be it English or Dutch, that involves a volitional agent who creates a path to a location and travels it. The acts of creating and travelling that path are indistinguishable, in the sense that they may form two sub-events, but they are not separable. I want to argue for a slightly different analysis here, and in order to do so, we will first have to take a look at another related type of expression, a construction which Egmond (2006) calls the Transition to Location construction.

## 4.2 The TLC

There is another construction in Dutch that exhibits similar properties to those of the *weg*-construction. Verhagen (2003:341) calls this construction the 'zich-verplaatsings-constructie' (the '*zich*-transition construction'), Van Egmond (2006) calls this the Transition to Location construction (TLC). Examples are given in (127) and (128).

(127) *Domingo zingt zich naar de top*
Domingo sings REFL to the top
'Domingo sings his way to the top'

(128) *De kankercel vreet zich door het lichaam*
the cancer-cell eats REFL through the body
'the cancer cell eats its way through the body'

Even though there is no obviously better English translation of this type of sentences than the *way*-construction, there are subtle semantic differences between the two Dutch constructions. However, the morphosyntactic difference between the TLC and the *weg*-construction is not so subtle: there is no *weg*-NP in the former.

The semantic differences between the Dutch *weg*-construction and the TLC are discussed in detail by Van Egmond (2006),

and she comes to the conclusion that the two constructions differ in the following respects, among others (see Van Egmond (2006:140f.)):

- The basic interpretation of the *weg*-construction is the incremental traversal of a path by means of (or while) V-ing, the basic interpretation of the TLC is the transition to a stative location by means of V-ing.
- Regarding the temporal relation between subevents, the action described by the verb is coextensive with the traversal of the path in the *weg*-construction, and temporally disjoint from the transition to the location in the TLC.
- Bounded events are interpreted as being iterated in the *weg*-construction, but as a single action in the TLC.
- The *weg*-construction is neutral with respect to telicity, but the TLC is necessarily telic.

Therefore, Van Egmond also assumes the semantics of the *weg*-construction to imply the element *weg*, which I will argue against below. The most important difference between the *weg*-construction and the TLC in favour of this, I consider to be the fact that bounded events are iterated in the former.

## 4.3 The *weg*-construction without *weg*

In this section, I will present a novel semantic interpretation of the Dutch *weg*-construction, which, in a nutshell, claims that both the *weg*-construction and the TLC have the meaning of 'transition to a location by means of V-ing', and that they merely differ with regard to aspect.

I argue that the *weg*-element does not play a role in the (lexical) semantic structure of the construction, and that the entire 'creating and traversing a path'-part can be skipped. In order to

do that, I will carry out two tests that provide arguments that *weg* is not accessible in the interpretation of *weg*-construction sentences, and therefore is not a semantic argument of the construction.

Nevertheless, that generalization does not hold for all instances of the *weg*-construction. In fact, it depends on the actual usage of the *weg*-construction whether some traces of the semantic *weg*-element are still visible. As I mentioned above, the construction is flexible in its degree of idiomaticity, and it depends on the lexical material, whether the sentence is completely idiomatic, or whether there are still traces of the literal meaning left. In sentences like (129), there is no literal path implied, whatsoever, whereas in sentences like (130), it is easily possible to imagine a concrete path.

(129) *Domingo zingt zich een weg naar de top*
Domingo sings REFL a way to the top
'Domingo sings his way to the top'

(130) *Hij hakt zich een weg uit de jungle*
he chops REFL a way out-of the jungle
'he chops his way out of the jungle'

Example (130) has a more compositional feel to it, exactly because a path can be the result of chopping (plants). (129), on the other hand, is completely idiomatic, since it is not possible to create a way by means of singing.

Nevertheless, we need to explain the question how the reflexive comes into the picture, if we opt to analyze example sentence (130) literally. A straightforward answer is that the 'self' is the beneficiary of the path-digging event. But note that Dutch does not normally have a productive ditransitive construction, so this interpretation should not be possible. At least, this holds for standard Dutch. German does have a productive ditransitive, and therefore the language can apply a fully compositional reading to those types of *weg*-construction sentences, and the same holds for a couple of Eastern Netherlandic dialects. The language border between Dutch and German forms a dialect continuum,

and the productive ditransitive is one grammatical feature that is available at both sides of the (regional) border.

Interestingly, the German equivalent to the *weg-* and *way*-construction looks like a superset of the two cognates.

(131) *Er bahnt sich seinen Weg durch die Menge*
He BAHNT REFL his way through the crowd
'he makes his way through the crowd'

The German verb *bahnen* behaves similar to the Dutch *banen*, and also only occurs in this construction. Furthermore, the German construction instantiates the reflexive that we know from the Dutch cognate as well as the possessive determiner that we find in the English construction. The reason for this phenomenon is a grammatical one: as I have shown above, expressions of the Dutch construction can be interpreted more or less literal, but at all times, it remains an instance of the *weg*-construction, since there is no productive ditransitive in Dutch, which could license a completely literal interpretation. In German, however, there is a fully productive ditransitive, which would lead to a mix-up between a *way*-construction interpretation and a ditransitive interpretation. The following examples will make that clear:

(132) *Er hackt sich einen Weg durch den Wald*
He chops REFL a way through the forest
'he chops a way through the forest for himself'

(133) *Er hackt sich seinen Weg durch den Wald*
He chops REFL his way through the forest
'he chops his way through the forest'

Sentence (132) is a normal ditransitive with a literal, ditransitive interpretation. There is no way that we could identify this sentence as an instance of the *way*-construction. Only in (133), it becomes unambiguous, since the markers *sich seinen* are a clear indication that this has to be interpreted as a *way*-construction expression. Of course, ambiguity in language is a natural phenomenon, but German can rely on this double marking, and so, a mix-up with the productive ditransitive can be avoided.

At this point, two different analyses are possible. Either, we assume that examples such as (129) are non-compositional, while examples like (130) are compositional to a certain extent, or we opt for an approach where both sentences are considered non-compositional, and treat them the same way in the analysis. I will argue for option two. I propose that all (synchronic) instances of the *weg*-construction are non-compositional, and that in all cases, the path-element should not be included in the semantic representation of the sentence, at least not as a concrete path. Nevertheless, I see the gradience of compositionality. I would like to claim that more compositional sentences like (130) are closer to the historically earlier examples of the *weg*-construction, which originally allowed for literal interpretations, while sentences like (130) are clear cases of the metaphorically extended, later examples. I base this claim on the historical sketch of the development of the *weg*-construction, as presented in Verhagen (2002).

In this paper, Verhagen provides empirical evidence that the *weg*-construction emerged out of a compositional pattern with the (independent) verb *banen*, which meant 'flatten' or 'level' in 17th century Dutch. He gives examples with reflexive and non-reflexive indirect objects, and definite and indefinite articles modifying the noun of the direct object, which very frequently was *weg*. Therefore, in this stage, there was a less constrained variation in a ditransitive pattern with the verb *banen*. What happened from this point on, he describes as follows:

"[S]peakers had been using the combination of words *weg banen* "to level a road, making it easier to travel" to convey messages of the type "to create a possibility to reach a goal", and then the "create" component of this message became a conventional meaning of *banen*. [...] The elements of the metaphorical interpretation are also metaphorically related to the elements of the literal interpretation: the possibilities are linked to *weg* 'way' and the action of creating to *banen* 'to level'. In the course of the development then, what was originally a contextually determined, extended interpretation of the conventional meaning "to level a path", became directly

associated with the formal combination *weg-banen* itself; it was no longer derived from the original conventional meaning [...]" (Verhagen (2002:424))

In other words, the compositional *banen*-sentences served as a model for metaphorical extensions, which did not have to rely on *banen* anymore, but still conveyed the same information, namely "create a possibility to reach a goal". This is a logical development, although I claim that nowadays, *weg*-construction sentences still contain a semantic path-element, as is suggested by Verhagen (2003) and Van Egmond (2006). Instead, the *weg*-noun is fully non-referential, and it does not appear in the semantic interpretation of the *weg*-construction. I base my claim on two independent tests that provide evidence for the assumption that the *weg*-element is non-referential.

### 4.3.1 Non-referentiality

My first argument is that the path-element is not part of the semantic representation of the *weg*-construction, due to the observation that the path-element in the syntax is non-referential in the sense that the NP *een weg* does not refer to a path or way. If it does have a semantic contribution, it is not the denotation of path or way. I conclude that if a noun has a non-empty reference, it needs to be given a place in the expression's lexical semantics, though if it has an empty reference and does not denote anything, it does not.

The first test I will carry out demonstrates the non-referentiality of the *weg*-element by showing that it cannot be picked up again in following discourse. A lexical element that is present in the semantic representation of an expression should be able to function as the antecedent of an anaphor. In case of the *weg*-construction, that is not possible. The path-noun cannot function as an antecedent. Therefore, the following sentence is (at least) highly unusual:

(134) **Zij zong zich een weg naar het podium, maar die*
she sung REFL a way to the podium, but it
*was zwaar*
was difficult
'She sung her way to the podium, but it (the way) was difficult'

This example is felt to be very odd by native speakers of Dutch, and in case that it is understood, it is usually judged as a language play. We can make sure that the pronoun *die* refers to *weg*, as both noun and pronoun are masculine. Would it refer back to *podium*, the pronoun would have to be the neuter form *dat*.

So, it is not possible to anaphorically refer to the *weg*-element alone, although it is perfectly easy to refer to the entire proposition, as is demonstrated in (135):

(135) *Zij zong zich een weg naar de top, maar dat was*
she sung REFL a way to the top, but it was
*zwaar*
difficult
'She sung her way to the top, but it was difficult'

In this example, I substituted *podium* with *top*, as the latter is also masculine. This way, it is clear that the pronoun *dat* does not refer to any of the nominal heads of the proceeding clause, but refers to the entire proposition expressed in that clause.

This observation contradicts the claim that the *weg*-construction opens two subevents, viz. creating and traversing a path, as—if there were two events that both contain a path argument—it should be possible to pick out this element and use it for anaphoric reference. The fact that this is not possible supports two points: the analysis does not contain two subevents, and also it does not contain a single (sub)event that involves a path element in the semantic representation.

The second argument in order to show that the *weg*-element is not part of the semantic representation is the observation that the noun cannot easily be modified. This, again, is a piece of evidence of the non-referentiality of *weg*. A sentence like (136),

where the *weg*-noun is modified by an adjective, is clearly infelicitous:

(136) **Hij zong zich een lange weg naar de top*
He sung REFL a long way to the top
'He sung his long way to the top'

Certain modifiers are possible, though they receive a floating interpretation:

(137) *Hij zong zich een snelle weg naar de top*
He sung REFL a fast way to the top
'He sung his way to the top fast'

In this example, *snel* modifies the entire proposition, not *weg* alone.

If *weg* were referential, there was no reason why it should not be possible to be modified. As Ghomeshi (1996) argues, "adjectives cannot take non-referential complements" (p.73). The *weg*-NP cannot function as the complement of an adjective, hence it cannot be referential. Modification by means of a relative clause is not possible, either, as is shown in (138):

(138) **Hij zong zich een weg die lang was naar de top*
He sung REFL a way that long was to the top
'He sung his way which was long to the top'

The two arguments I have presented for the non-referentiality of *weg* show that the *weg*-NP can neither be anaphorically referred to, nor can it be modified by an adjunct. I conclude from that that the path-noun in the Dutch *weg*-construction is non-referential. From that, a second claim follows, namely that the element 'path' should not be included in the semantic representation of the Dutch *weg*-construction.[27]

[27] As I only deal with Dutch here, I also only draw conclusions for the Dutch *weg*-construction. Nevertheless, it is not improbable that similar conclusions can be drawn for the English *way*-construction.

### 4.3.2 A matter of aspect

Now that I have argued that the path-element should not be included in the semantics of the *weg*-construction, the question remains what the semantic representation should then look like. In this section, I will propose an analysis that is based on one semantic representation for both, the *weg*-construction and the TLC, while the main difference between the two constructions is a matter of inner aspect (as defined in Arsenijević (2006)).

The basic interpretation for both constructions is very similar to Van Egmond's semantics of the TLC, namely

- transition to a location by means of V-ing

In what follows, I will defend this view.

Both example (139) and (140) mean that Bubka reached the finals by means of jumping:

(139) *Bubka sprong zich naar de finale*
Bubla jumped REFL into the finale
'Bubka jumped himself into the finals'

(140) *Bubka sprong zich een weg naar de finale*
Bubla jumped REFL a way into the finale
'Bubka jumped himself into the finals'

The difference between the two examples is that (139) exhibits perfective aspect, i.e., Bubka had at least one good jump and landed in the finals, while (140) exhibits iterative aspect, i.e., Bubka had a row of jumps, probably most or all of them good, and, as a result, landed in the finals. An iterative reading of (139) is probably not excluded but definitely not forced, while a perfective, non-iterative 'single-jump' reading of (140) is impossible. So, for verbs of achievements (in the sense of Vendler (1957)), where the action is instantaneous, this action is understood as being iterated. For verbs that fit into Vendler's accomplishment category, the action has durative aspect. This is why we do not get such a clear-cut dichotomy between the *weg*-construction and the TLC semantics with accomplishment verbs: the source of the

aspectual meaning contribution lies in the verb, already, so there is no durative/non-durative distinction between expressions of the two constructions. This is exemplified in (141) and (142):

(141) *Hij heeft zich een weg naar de top gelachen*
He has REFL a way to the top laughed
'He laughed his way to the top'

(142) *Hij heeft zich naar de top gelachen*
He has REFL to the top laughed
'He laughed his way to the top'

The second aspectual difference arises from the telicity of the two constructions. Van Egmond (2006) shows that the *weg*-construction can be both telic and atelic, while the TLC is necessarily telic. The standard test for telicity, the insertion of a durative and a non-durative adverbial (see Verkuyl (1989)), shows that the *weg*-construction allows for both types of adverbials, whereas the TLC only allows for the durative one:

(143) *Tarzan heeft zich dagenlang een weg door de jungle gehakt*
Tarzan has REFL days.long a way through the jungle slashed
'Tarzan slashed his way through the jungle for days' (Van Egmond 2006:71)

(144) *Tarzan heeft zich in twee dagen een weg door de jungle gehakt*
Tarzan has REFL in two days a way through the jungle slashed
'Tarzan slashed his way through the jungle in two days' (ibid.:72)

(145) **Marien heeft zich minutenlang in de finale*
Marien has REFL minutes.long in the final
*gezwommen*
swum
'Marien swam his way into the final for minutes' (ibid.:104)

(146) *Marien heeft zich in twee minuten in de finale*
Marien has REFL in two minutes in the final
*gezwommen*
swum
'Marien swam his way into the final in two minutes' (ibid.:104)

To sum up, it appears that we have the following picture: the basic interpretation of both *weg*-construction and TLC is 'transition to X by means of V-ing', but they differ with respect to aspectual properties. The *weg*-construction is underspecified for telicity and is necessarily iterative (if applicable), while the TLC is underspecified for durative or iterative aspect, but is necessarily telic.

(147)

| **Aspectual Properties** | | |
|---|---|---|
| | ***weg*-cx** | **TLC** |
| telic | o | x |
| durative/iterative | x | o |

I argued against an analysis which contains a *path*-element in the semantic interpretation. In previous analyses, this *path*-element used to be expressed by the *weg*-NP in the syntax. Nevertheless, it is possible to construe the *weg*-NP as semantically non-empty, if we consider it the bearer of the abstract meaning of 'extension'. The iterative as well as the durative aspectual component are both related to temporal extension, and it is not unusual to express abstract (temporal) concepts by means of concrete (spatial) vocabulary (see, e.g., Lakoff & Johnson (1980)

and Radden (2006)). This phenomenon is known as a conceptual metaphor (see Croft & Cruse (2004), Lakoff (1987)). Consequently, the *weg*-NP is not considered a semantically empty dummy, but the source of the aspectual properties of the *weg*-construction. This also explains why the nouns that can appear in the *weg*-NP must have a *weg*-related semantics. *Weg*, *pad*, and *baan* all describe a stretch in space, which metaphorically translates to a stretch in time.

The difference between the existing (constructional) analyses of the Dutch *weg*-construction and the analysis presented here is not subtle. While Verhagen's approach, just like Goldberg's or Jackendoff's for English, aims at an almost compositional treatment of the pattern, where every syntactic item is assigned a piece of meaning, I argued for a non-compositional approach. Not every part of the structure needs to have a contribution to the semantics, and even if it does, it does not need to be the literal interpretation. The latter insight is, of course, already prominent in the existing literature, but the first one is novel to my analysis. Not every item in the syntax of the *weg*-construction is mapped onto an item in the semantic structure.

I assume that—from a constructionist point of view—it would not even be necessary to explain the semantic (or functional) difference between two constructions by deriving it from the difference between the two patterns. Nevertheless, in the case of the *weg*-construction and the TLC, I propose that such a derivation is actually possible. Under my analysis, the semantic difference between the two constructions is an aspectual one, which I link to the *weg*-element. The *weg*-construction necessarily implies a stretch of time, be it in a durative or an iterative manner. The action expressed by the verb cannot be instantaneous, which is different for the TLC. The structural difference between the two constructions is the added *weg*-NP in the *weg*-construction, the semantic difference is the added stretch of time in the *weg*-construction. Conceptual metaphors, that express temporal notions with spatial lexical items is not unknown a process, and here, in this construction, this is exactly what happens.

## 4.4 Formal analysis

Now, let us turn to the formal analysis of the Dutch *weg*-construction. Under my analysis, the *weg*-construction and the TLC are lexical phenomena rather than phrasal patterns.

I use a derivational construction to build a *weg-v*-lexeme from a strictly intransitive verb lexeme:

(148) *weg-v* derivational construction:

$$
\textit{weg-v-cxt} \Rightarrow \left[\begin{array}{ll}
\text{MTR} & \left[\begin{array}{ll}
\textit{weg-v-lxm} & \\
\text{FORM} & \langle\ X\ \rangle \\
\text{ARG-ST} & \left\langle \begin{array}{l} \text{NP}_x\ \left[\textit{pro}\right], \\ \text{NP}_y\ \left[\text{SEM}\ \ \textit{way-rel}, \right] \end{array} \right\rangle \\
 & \qquad\qquad\qquad\qquad\quad \textit{goal-rel} \\
\text{SEM} & \left[\text{FRAMES} \left\langle \begin{array}{l} L_1, \\ \oplus\ \textit{move-fr}, \\ \oplus\ \textit{iter-fr} \end{array} \right\rangle\right] \\
\left[\ldots\right] &
\end{array}\right] \\
\text{DTRS} & \left[\begin{array}{ll}
\textit{siv-lxm} & \\
\text{FORM} & \langle\ X\ \rangle \\
\text{SEM} & \left[\text{FRAMES}\ \ L_1\right] \\
\left[\ldots\right] &
\end{array}\right]
\end{array}\right]
$$

For a derivational construction that derives *weg-v*-lexemes from *siv*-lexemes, the following information needs to be encoded: since it is a lexical class construction, all strictly intransitive verbs can enter the construction as daughters. The input (DTRS) is an intransitive verb, the output (MTR) is a *weg-v*-verb. This has the same form as the DTRS, but the semantics and the argument structure have changed. The *weg-v*-verb takes two internal arguments, the first being a pronoun, the second being an NP with the semantic equivalent of *way* plus, on top of that, *goal*-semantics (which I will explain in a minute). The semantics of

the *weg-v*-lexeme is the combination of the semantic frames of the daughter, plus a *move*-frame, plus an *iteration*-frame.

Note that one principle of SBCG is made use of twice in this lexical class construction: the principle of Locality (see Sag (2008)). In SBCG (as well as in HPSG), it is not possible to place constraints on items that are embedded lower in the structure than immediately accessible, selected items. Sag captures this idea in the hypothesis of Selectional Locality:

(149) **Selectional Locality:**
For purposes of category selection (subcategorization), (nonanaphoric) agreement, semantic role assignment, and case assignment, a lexical head has access only to those elements that it is in a grammatical relation with (subject of, complement of, etc.). In SBCG, this amounts to a restriction that the only non-local elements that can be selected are those whose grammatical information is encoded by constraints placed on elements of the ARG-ST list of the predicator that is doing the selecting at the higher level. Sag (2007b:44)

A second restriction on nonlocal processes is the hypothesis of Constructional Locality, which regulates in the same way the way a construction can only access immediate daughters:

(150) **Constructional Locality:**
'Constructions license mother-daughter configurations without reference to embedding or embedded contexts. Sag (2007b:45)

In the case of the *weg*-construction, Selectional Locality prohibits the verb to have direct access to the PP that modifies the *weg*-NP, since it only has access to the NP itself, not to a complement of it. Sag (2008) approaches a comparable problem with the XARG feature (for external argument), which localizes the nonlocal information by percolating it up to an accessible node. This way, the nonlocal information is encoded on every node that lies between the dependent elements.

For the *weg*-construction, the problem of Selectional Locality can be solved rather straightforwardly: since we cannot encode

the directional semantics of the embedded PP directly in the construction, I refer to the *weg*-NPs semantics instead. This NP needs to have a *way-rel* semantics to make sure that the noun is chosen from the group of *weg*-equivalents, and in needs to have *goal-rel* semantics to make sure that the noun is modified by a directional PP.

For flatter analyses, like the ones that Delilah yields, this step is unnecessary, since the given representation is flatter and can encode this dependency directly.[28]

The according Delilah template that licenses *weg*-construction verbs looks as follows:

(151) Delilah template *weg*-lexeme

```
template( weg-lxm, verb,
[ id:Top+ID,
synsem: [cat: vp, tense: untensed,
        predtype:nonerg,
        eventvar:EV, extth:Stheta~[Top+ID, A]],
sem: { { [ SemS&(ID+ID1)#A, SemPPNP&(ID3+ID4)#C,
        EStructure@some^E^and(quant(E, Quant),
        Main~[E], Etype~[E], entails1( E, FuncL),
        and( EStructure, entails(E, FuncR)))
        & (Top+ID)#EV],
        [],[]},
        Time@and( and( Stheta~[EV,A],
        some^EE^and(quant(EE,some),
        and(and(event(EE), move(EE)),
        theme_of(EE, A)),
        goal_of~[EE, C], entails1(EE, incr),
        cause(E, EE)),
        entails(EE, incr)) , attime(EV,Time)) },
head: [phon: _X,
```

[28] In a sense, the Delilah approach is comparable more to the analysis of idiomatic expressions in Sailer (2000) and later, for example, Soehn (2004). In Delilah, the only boundary is the sentence, and dependencies, agreement and case-assigning relations can be realized within the entire domain of a sentence.

```
        synsem:[vtype:intr, flex:infin,
               etype:Etype],
        sem: Main,
        concept:Main ],
arg(ID+ID1+10):[phon:_Subj,
        synsem:[theta:Stheta, person:Person,
               number:Number,
               obj:subject_of(Top+ID)],
               sem:SemS],
arg(ID+ID2+1):[phon:_Obj, head:[synsem:[cat:det]],
        synsem: [theta:theme_of, cat:np, person:3,
                funcl:FuncL,
                funcr:FuncR, number:sing,
                qmode:indef, dir:left(2), flag:0,
                case:obliq],
arg(ID2+_ID4+1):[head:[concept:road track  ],
        synsem: [number:sing, cat:n]],
arg(ID2+ID3+2):[phon:_PP,
        head:[concept:towards direction ],
        synsem:[theta:goal_of, flag:0,
               dir:left(1), cat:pp, case:obliq],
arg( ID3+ID4+1):[synsem:[cat:np],sem:SemPPNP],
                sem:_SemPP],
                sem:{_Stores0, _@Quant^_^_}],
arg(ID+_ID5+3):[phon:_Reflex,
        sem:_RefSem,
        synsem:[cat:np, dir:left(3), flag:0,
               person:Person, number:Number,
               subcat:pron, pron:refl, case:obliq,
               focus:nonfocus ] ]
               ]  ).
```

The important difference you see here is that in Delilah, the PP can be encoded directly (under arg(ID2+ID3+2)). As it is the default in Delilah, modifying PPs and relative clauses are usually parsed as a sister of the determiner instead of the noun. The argument of the sem value expresses that there is a moving

event towards a goal by means of V-ing. The head is a strictly intransitive verb (vtype:intr), the inflection is infinite, since this template gives a lexeme as its output. The reflixive pronoun is bound under arg(ID+_ID5+3), and arg (ID2+_ID4+1) binds the *weg*-NP, that is constrained to have the concept (semantic category) of 'road' or 'track'. Note that there is no trace left of this concept under the sem value.

Now compare the analysis of the *weg*-construction with the TLC. Again, the TLC is a lexical class construction rather than a phrasal construction, and again it derives a new *tlc-v*-lexeme from strictly intransitive verbs.

(152) The TLC

$$\textit{tlc-v-cxt} \Rightarrow \left[\begin{array}{ll} \text{MTR} & \left[\begin{array}{ll} \textit{tlc-v-lxm} & \\ \text{FORM} & \langle\ X\ \rangle \\ \text{ARG-ST} & \left\langle \begin{array}{l} \text{NP}\ \left[\textit{pro}\right], \\ \text{PP}\ \left[\text{SEM}\quad \textit{goal-rel}\right] \end{array} \right\rangle \\ \text{SEM} & \left[\text{FRAMES}\ \left\langle \begin{array}{l} L_1, \\ \oplus\ \textit{move-fr}, \\ \oplus\ \textit{telic-fr} \end{array} \right\rangle\right] \\ \left[\ldots\right] & \end{array}\right] \\ \text{DTRS} & \left[\begin{array}{ll} \textit{siv-lxm} & \\ \text{FORM} & \langle\ X\ \rangle \\ \text{SEM} & \left[\text{FRAMES}\quad L_1\right] \\ \left[\ldots\right] & \end{array}\right] \end{array}\right]$$

In this construction, the same information as in the *weg-v*-construction lexeme is encoded, apart from the lacking *weg*-NP which is not on the ARG-ST list, and the PP that *is* on the ARG-ST list. Under SEM, the same *move*-frame is opened, but where there was the *iteration*-frame in the *weg*-construction, there is a *telic*-frame in the TLC.

One answer that follows directly from the construction is why both *weg*-construction and TLC cannot feed passive. Since the verbs that can enter these constructions need to be intransitive (or, at least, be able to have an intransitive reading), passive versions of the two constructions are infelicitous due to the fact that intransitive verbs generally do not feed passive, apart from the impersonal passive with *er*, which, although we did not find any naturally occuring examples, does not sound infelicitous in the *weg*-construction.

The Delilah template for the TLC looks as follows:

(153) Delilah template tlc-lexeme

```
template( tlc-lxm, verb,
[ id:Top+ID,
synsem: [cat: vp, tense: untensed, predtype:nonerg,
        eventvar:EV, extth:Stheta~[Top+ID, A]],
sem: { { [SemS&(ID+ID1)#A, SemPPNP&(ID3+ID4)#C,
         EStructure@some^E^and(quant(E, some),
         Main~[E], Etype~[E], entails1( E, incr),
         and( EStructure, entails(E, incr)))
         & (Top+ID)#EV],
         [],[]},
        Time@and( and( Stheta~[EV,A],
        some^EE^and(quant(EE,some),
        and(and(event(EE), move(EE)),
        and(and(state(EE), be(EE)),
        theme_of(EE, A), location_of~[EE, C],
        entails1(EE, incr), cause(E,EE),
        entails(EE, incr))), attime(EV, Time))) },
head:   [phon: _X,
        synsem:[vtype:intr, flex:infin,
               etype:Etype],
        sem: Main,
        concept:Main ],
arg(ID+ID1+10):[phon:_Subj,
        synsem:[theta:Stheta, person:Person,
               number:Number,
```

```
                obj:subject_of(Top+ID)],
        sem:SemS],
arg(ID+ID3+2):[phon:_PP,
               head:[concept:towards direction ],
               synsem:[theta:goal_of, flag:0,
               dir:left(1), cat:pp, case:obliq],
arg(ID+_ID5+3):[phon:_Reflex,
                sem:_RefSem,
                synsem:[cat:np, dir:left(3), flag:0,
                        person:Person, number:Number,
                        subcat:pron, focus:nonfocus ] ]
                   ]  ).
```

Note here that the aspectual difference between the *weg*-construction and the TLC is expressed as follows: in the former, the sem value describes a moving event to a goal, while in the latter, the sem value describes a move-to and a be-at state. Also, in arg(ID+ID3+2), the PP is an argument of the verb directly.

## 4.5 Conclusion

In this chapter, I presented previous approaches and new ideas about the Dutch *weg*-construction. As the basis, I took the studies by Verhagen (2003) and Van Egmond (2006), which are both very detailed on the one hand and use empirical techniques on the other. I concentrated in this chapter on a more theoretical evaluation of the analysis, rather than on gathering data, as I did in the case studies on the NPN construction and the dative alternation, which will be the topic of the next chapter. And although I certainly agree with many of the results of previous studies of the *weg*-construction, I recast a number of points. Some of them are not very crucial, but some are of major importance for the theoretical analysis and the computational treatment of the *weg*-construction.

One major difference compared to former construction-based analyses of the *weg*-construction is the point that I do not view

this construction as a phrasal phenomenon, but I analyze the *weg*-construction as a lexical process, where there is one verb with its set of features derived from another verb. The *weg*-construction, be it in Dutch or in English, is not a phrasal, but a lexical construction. Of minor importance for the big picture is the claim that the directional PP is a modifier of the *weg*-noun, rather than a VP adjunct. I presented a series of arguments that support an analysis like that. A point of critique that might have some meta-theoretical importance is my disagreement with Verhagen's (2003) schematic formalization of the *weg*-construction. As I already discussed in chapter 2, I contradict the idea that phrase structure carries semantics, or that phrase structure is a fixed feature of (at least) argument structure constructions like the *weg*-construction. Verhagen's formalization presents the tree [SUBJ [V [REFL [*een weg*] OBL]]] as the syn side of the construction, but that is inconvenient in the sense that we do not always find this pattern. A formalization that provides information about the number and type of arguments without giving a fixed structure is to be preferred in this case.

Furthermore, I sketched an answer to the question why Dutch, which does not have a productive ditransitive, instantiates a ditransitive pattern for the *weg*-construction, while English does have a productive ditransitive, which is not used for the English *way*-construction.

With respect to the elements that constitute the *weg*-construction, my critique mainly concerns the inflexibility of the lexical head of the *weg*-NP. A quick Google search already provided a significant amount of example sentences with the lexical heads *pad* and *baan*, both semantic cognates of *weg*. Therefore, I argue, it is empirically more adequate to put a constraint on the semantic class of nouns that can enter this slot, rather than narrowing down the choice to one single lexical item.

When it comes to the semantics of the *weg*-construction, my claims differ more drastically from the previous accounts. Though all studies of the Dutch and the English *way*-construction assume that the reflexive and the *weg*-NP are not semantic arguments of the verb, I go one step further and argue

that neither the reflexive nor the *weg*-noun play any role at all in the semantic representation of the construction. Based on a comparison of the *weg*-construction and the Transition to Location construction, I claim that the major semantic difference between the two is the an aspectual one. Under my analysis, both constructions receive the basic semantic interpretation of 'transition to a location by V-ing', while the *weg*-construction is necessarily iterative and underspecified for telicity, and the TLC is necessarily telic and underspecified for iterative aspect. This semantic difference can be assigned to the *weg*-NP in the *weg*-construction, via the Time-Is-Space metaphor. The difference between the *weg*-construction and the TLC is an additional durative or iterative aspect in the former, which implies the action of the verb occupying a stretch of time. This stretch of time is encoded in the *weg*-NP, which describes a stretch of space.

In sum, I reject an analysis of the Dutch *weg*-construction that aims at a compositional treatment of the pattern. It is neither possible nor necessary to assign a certain part of the construction's semantics to each lexical item involved. Nevertheless, I claim that it is possible to trace the semantic difference between *weg*-construction and TLC back to the structural difference between the two, namely the additional *weg*-NP in the former.

# Chapter 5

# The Dative Alternation

The case study I will present in this chapter is going to be different from the ones before, in that it will adopt a cross-linguistic, or rather contrastive, perspective. The constructions under discussion are widely known as the 'dative alternation' (see Goldberg (1995), Bresnan & Nikitina (2007), Oehrle (1976), etc.), two alternating constructions that are used to express certain events with three participants. I will discuss the two constructions for Dutch as well as German and English, on an empirical basis. In order to be able to compare the three languages with regard to this alternation, I conducted an extensive corpus search for instances of the two constructions in a parallel corpus.

There are several theoretical questions that arise from an investigation like this. One major concern will be based on a hypothesis that was put forward by Levin und Rappaport Hovav. Levin & Rappaport Hovav (2005) argue that verbs of giving differ from verbs of, e.g., sending, in not having caused-motion semantics (Levin & Rappaport Hovav 2005:33).[1] As a consequence of this, they predict that languages with a morphological marking of the recipient do not display the prepositional alternative to the double object construction.

In this chapter, I will contrast the (apparent) dative alternations in Dutch, English, and German, trying to shed some more light on this issue. A comparison of these three languages seems

[1] For a discussion, see below.

particularly useful, as they are not only closely related—all three being Westgermanic languages—, but they also differ with respect to their case marking properties. German still exhibits the inflection of four cases, namely nominative, genitive, dative, and accusative, whereas English and Dutch only have a nominative (and a possessive, rather than grammatical, genitive)[2] form for nouns, and subject and object forms for pronouns. Nevertheless, English and Dutch differ with respect to the period in which the deflection process started or took place. English lost its case inflection the earliest, then came Dutch, and German is still in the very beginning of a deflection process (see, e.g., Köpcke (2003) and Köpcke (2005)), on a stage, where the genitive gets replaced by the dative or, alternatively, by prepositional complements, and where the case system increasingly becomes unstable.

The working hypothesis for this chapter, then, is that Dutch, English, and German differ vastly with respect to the distribution of the two alternants, and that these differences can be explained from the different case marking properties of the three languages. Furthermore, we will try and see whether we can draw conclusions about the grammatical status of the dative alternation.

## 5.1 The data

### 5.1.1 Dative alternations

Dutch, English, and German seem to instantiate the same set of alternating constructions, namely a double object construction (henceforth DO, examples in (154)), and a corresponding prepositional alternative, often called 'dative'[3] (henceforth PO,

[2]I make a distinction between possessive and grammatical genitive here, as German still has real genitive objects, i.e., verbs that select for a single genitive object as in *Sie gedachten der Toten* (They commemorated the(Gen) dead(Gen). Dutch and English only have possessive genitives, like in *Bob's car* or *a friend of Bob's*. On a sentence-complement level, there is no genitive in the two languages.

[3]I am hesitant to call the PO 'dative', as it is often done in the English literature (e.g., Goldberg (1995)). In my terminology, *dative* is reserved for

examples in (155)). The sentences in 1. are Dutch, the sentences in 2. are German:

(154) 1. *Jan geeft de vrouw de bal*
Jan gives the woman the ball
'Jan gives the woman the ball'

2. *Jan gibt der$_{DAT}$ Frau den$_{AKK}$ Ball*
Jan gives the woman the ball
'Jan gives the woman the ball'

3. John gives the woman the ball

(155) 1. *Jan geeft de bal aan de vrouw*
Jan gives the bal to the woman
'Jan gives the ball to the woman'

2. *Jan gibt den$_{AKK}$ Ball an die Frau*
Jan gives the ball to the woman
'Jan gives the ball to the woman'

3. John gives the ball to the woman

In the first set of examples, the recipient and the theme argument are both expressed as bare objects. In Dutch and English—due to lack of case marking—the two objects are not morphologically distinguishable, and they are referred to as first or indirect object (the recipient argument), and second or direct object (the theme).[4] German, however, does employ morpholog-

---

the notion of dative case, which is not retained in English. Especially in comparison with a language that does exhibit dative case, like German, calling the prepositional object or the construction that instantiates it, *dative* would be confusing. This confusion gets even worse as in German, the recipient in the DO is a dative, whereas the recipient in the PO is not. Hence, I will stick to the following convention: the prepositional counterpart of the double-object construction, I will call PO, and the term 'dative' will be reserved for the actual dative case. The terminology is merely a consequence of the differences between the languages in question, but to avoid confusion, I disregard the term 'dative' for the PO construction.

[4]The proper Dutch terms for indirect and direct object are *meewerkend voorwerp* ('participating object') and *lijdend voorwerp* ('suffering object', 'patient').

ical case marking, and therefore the two objects can be distinguished as a dative object (the recipient) and an accusative object (the theme). Form and function are terminologically intertwined, here. In the second set of examples, the recipient is expressed as a prepositional phrase. The order of recipient and theme is reversed in the default order of the PO.

With those alternations at hand, an obvious question is what the distribution of the two alternants is in each language, and how the three languages differ if we contrast them. This question will be addressed in the remainder of this chapter. Our focus, as in the two preceding chapters, is on the Dutch constructions, though, and an analysis will be provided only for Dutch.

### 5.1.2 The distribution of the alternating patterns

Much work has been done on the distribution of the double object and the prepositional construction for English and Dutch. Relevant work for Dutch has been carried out by Schermer-Vermeer (1991), Van der Beek (2004), and Colleman (2006), among others. The list of names of people who worked on the English dative alternation is long, and for an overview, I point the reader to Rappaport Hovav & Levin (2006) and Levin & Rappaport Hovav (2005).

In recent years, there also have been a number of empirically driven studies on the dative alternation. One such study of the alternation in English, provided by Bresnan *et al.* (2007) and reported in Bresnan & Nikitina (2007), found that in the 3 million words Switchboard corpus of English telephone conversation, 78.6% of all alternating dative constructions instantiate a double object pattern. 51% of all extracted instances were headed by the verb *give*, which raised the DO in 84.6% of the cases. So, English seems to be strongly skewed towards the DO pattern.

A comparable result is found by Gries & Stefanowitsch (2004), who found 607 significant instances of *give*-sentences in the ICE-

GB[5] (one million words of spoken British English). 461 (75.9%) sentences were ditransitives, 146 (24.1%) sentences instantiated the prepositional pattern. Out of 2954 ditransitive and dative sentences, 607 (20.6%) were headed by a form of *give*.

An empirical study of the dative alternation in Dutch has been carried out by Van der Beek (2004). She extracted DOs and POs from two corpora, the treebank part of the Spoken Dutch Corpus (1 million words of spontaneous speech), and the Alpino Treebank (150.000 words of newspaper text). In the CGN, she found a distribution of 78.5% for the DO pattern, and 21.5% for the PO. In the Alpino, she found 70.9% for the DO, and 29.1% for the PO. The results for Dutch and English surely signal a tendency, but they are not directly comparable. Neither the sizes of the corpora nor the genres are similar, and the methods of extracting the sets of candidates are different. So, in order to empirically contrast the dative alternation in the languages in question, I cannot rely on existing analyses, but I have to compile the dataset myself.

### 5.1.3 The method

A promising way to get an overview over the distribution of the two constructions in the three languages is an empirical analysis based on comparable corpora. I weighed the pros and cons of different available corpora, and came to the conclusion that the most suitable one for these purposes is the Europarl corpus (Koehn 2002), which was originally sampled as an evaluation corpus for machine translation. Europarl is a parallel corpus of speeches held in the European parliament, and it is available for 11 European languages. Each language is document- and sentence aligned to English, and each monolingual corpus consists of approximately 20 million words. Apart from the alignment, the corpora are not linguistically processed, so the data are available only as raw text.

Dealing with raw text makes the corpus analysis for a particular construction somewhat more complex, as the only possible

[5]URL: http://www.ucl.ac.uk/english-usage/projects/ice-gb/.

search pattern is based on strings, unless the data are prepared beforehand, e.g., by parsing. For that reason, I decided to restrict the count for this study to a set of three verbs (lemmas), namely *give* for English, *geben* for German, and *geven* for Dutch. The choice of these verbs is not arbitrary, as these verbs are often described as the prototypical verbs for the ditransitive construction. This is supported by the fact that those verbs are the highest frequent verbs that appear in the ditransitive construction, and that the semantics of the verb comes closest to the semantics of the construction, namely designating the transfer of a theme from an actor to a recipient.[6]

The actual process of gathering the data looked like this: First, I extracted all sentences instantiating any form of the lemmas *give*, *geben*, and *geven* automatically from the three monolingual subcorpora. In the next step, I manually sorted the data into three categories, namely 1) DO sentence, 2) PO sentence, and 3) trash. In order to be sorted into 1) or 2), the necessary condition to be met was the presence of three arguments with the semantic roles of agent, theme and recipient. The mapping of these roles to functional categories was not a decisive factor, which enabled me to also count relative clauses and other constructions. For passives sentences, only the presence of a theme and a recipient argument was necessary, as in all three languages, the agent is not instantiated in the default passive sentence (see also section (5.3.1)).

### 5.1.4 The results

The overall count of matches revealed the following picture. For the English dative alternation, this distribution was found:[7]

[6]For an elaboration of this idea of prototypical verbs for a certain sentence pattern and overlapping verbal and constructional semantics, see, e.g., Goldberg (1995).

[7]Note that for all verbs in this study, we only distinguish between forms, not between functions, i.e., person, tense, and voice. In further studies, it will be useful to distinguish between those functions, too, since it will cast even more light on the particular usage of the constructions.

(156) Distribution of the English constructions

| **English Constructions** | | |
|---|---|---|
| **Form** | **DO** | **PO** |
| *give* | 5349 (74.1%) | 1868 (25.9%) |
| *gives* | 887 (71.2%) | 359 (28.8%) |
| *gave* | 331 (59.3%) | 227 (40.7%) |
| *giving* | 941 (56.5%) | 723 (43.5%) |
| *given* | 3402 (63.2%) | 1978 (36.8%) |
| *TOTAL* | 10910 (67.1%) | 5155 (32.9%) |

For the Dutch dative alternation, the following distribution was found:

(157) Distribution of the Dutch constructions

| **Dutch Constructions** | | |
|---|---|---|
| **Form** | **DO** | **PO** |
| *geven* | 1689 (57.2%) | 1263 (42.8%) |
| *geef* | 118 (39.5%) | 181 (60.5%) |
| *geeft* | 457 (63.8%) | 259 (36.2%) |
| *gaf* | 32 (72.7%) | 12 (27.3%) |
| *gaven* | 5 (26.3%) | 14 (73.7%) |
| *gegeven* | 270 (30%) | 630 (70%) |
| *TOTAL* | 2534 (52.1%) | 2332 (47.9%) |

For the German data, the following distribution was found:

(158) Distribution of the German constructions

| **German Constructions** | | |
|---|---|---|
| **Form** | **DO** | **PO** |
| *geben* | 542 (100%) | 0 (0%) |
| *gebe* | 254 (98.9%) | 3 (1.1%) |
| *gibst* | 1 (100%) | 0 (0%) |
| *gibt* | 891 (99.4%) | 5 (0.6%) |
| *gaben* | 44 (100%) | 0 (0%) |
| *gabst* | 0 | 0 |
| *gab* | 121 (99.2%) | 1 (0.8%) |
| *gegeben* | 1082 (99%) | 11 (1%) |
| *TOTAL* | 2935 (99.4%) | 17 (0.6%) |

To summarize, the distribution of the two constructions in the three languages is presented in (159).

(159) Overall distribution

| **Dative Alternation** | | |
|---|---|---|
| | **DO** | **PO** |
| **English** | 67.1% | 32.9% |
| **Dutch** | 52.1% | 47.9% |
| **German** | 99.4% | 0.6% |

The numerical differences between the distribution of DO and PO in the three languages is striking. The distribution in German

is especially surprising.[8] English does have a clear preference for the DO construction, the PO construction is next to non-existent in German, and in Dutch the result is quite balanced.

Apart from the mere numbers of occurrences, the different word orders that are instantiated also vary across the three languages. For the double object construction, we find the following rough word orders:[9]

(160) DO patterns

| **Double Object** | | | |
|---|---|---|---|
| | **English** | **Dutch** | **German** |
| **SU IO DO** | X | X | X |
| **SU DO IO** | O | X | X |
| **DO SU IO** | X | X | X |
| **DO IO SU** | O | O | X |
| **IO SU DO** | O | X | X |
| **IO DO SU** | O | O | X |

For the prepositional construction on the other hand, we find the following orders:

[8]The results of a Chi Square test revealed the following significances: P = 0.0006 for the English distribution, P = 0.6745 for Dutch, and P = 0.0001 for German. This shows that the distribution of DO and PO differ significantly in English and German, while the difference is not significant in Dutch.

[9]In the scope of this research, it was not possible to find the time to evaluate all orders empirically. The Xs in the table only represent relative orders that do occur in the corpus, including theme extraposition, wh-sentences and embedded clauses. Counting and evaluating the numbers will be another fruitful topic for further research, particularly in the case of German. Random sampling revealed that the more frequent order for the double object construction in German is IO>DO, but a more careful examination might reveal more parallels between the German DO construction and the Dutch and English PO.

(161) PO patterns

| Prepositional construction | | | |
|---|---|---|---|
| | **English** | **Dutch** | **German** |
| **SU DO PP** | X | X | X |
| **SU PP DO** | O | X | X |
| **DO SU PP** | O | X | O |
| **DO PP SU** | O | O | O |
| **PP SU DO** | O | X | O |
| **PP DO SU** | O | O | O |

It is clear that the three languages differ with respect to their flexibility. In German, all possible word orders for the double object construction can be found, whereas English only allows for the standard word order SU > IO > DO and an order with a topicalized theme argument. Due to the small number of occurences, it is impossible to say much about the German PO. Dutch is less rigid than English, but less flexible than German. It allows for both possible orders with a subject in initial position, and scrambling one object over the subject is permitted as well. Two fronted objects could not be found, since none of the languages allows for more than one fronted element. Examples of the discussed patterns will be presented below.

After I set the stage by preparing the data we need to deal with, I want to turn to the theoretical considerations that play a role for an analysis of the dative alternation.

## 5.2 Theoretical considerations

### 5.2.1 Polysemy versus monosemy

One prominent discussion around the grammatical status of the dative alternation deals with the question of what the source for the alternation might be (for an overview, see Levin & Rappaport Hovav (2005), and work cited therein). Answers to that question generally fall into two categories: group 1 assumes both the verb and the two constructions to be semantically equivalent

and related to each other by derivational means (see, e.g., Baker (1988) and Larson (1988)). In Baker's account, for instance, the alternants need to be thematic paraphrases of each other, as follows from his Uniformity of Theta Assignment Hypothesis. This approach, however, does not provide an explanation for the full range of the known observations. The opposite is often claimed in polysemy approaches: that there is a semantic or pragmatic difference between the two alternants, which eventually leads to choosing one construction over the other. The observed differences are the strongest argument for a polysemy approach.

Among the group of polysemy approaches, opinions differ as to where the source of the polysemy has to be looked for. Either the verb itself can be seen as having two distinct meanings which, in turn, give rise to two different argument structures (as put forward by, among others, Hale & Keyser (1996), Krifka (1999), and Oehrle (1976)), or the verb itself is seen as monosemous, but it is capable of "fusing", to borrow a term from Goldberg (1995), with different argument structure constructions. Under such an analysis, one single verb can be combined with two different argument structure constructions, as defended by Goldberg (1995). In this case, the polysemy of the constructions is located in the different functions of the two constructions. Polysemy, no matter where it is placed in the grammar (or lexicon), is generally linked to two different semantic contents. In Goldberg's analysis for English, e.g., the double object construction has the semantics of caused possession, where X causes Y to have Z, whereas for the prepositional dative construction, the meaning is one of caused motion, where X causes Z to be at Y. In the lexical approaches, this is a semantic difference between the two verb senses, in Goldberg's constructional approach, this is reflected in the assumption that the prepositional dative is a subconstruction of the Caused-Motion Construction, namely the Transfer-Caused-Motion Construction.

For my analysis, I want to follow the idea first put forward by Levin & Rappaport Hovav (2005) and Rappaport Hovav & Levin (2006). They claim that the attempt to categorize the entirety of alternating verbs in an either polysemous or monose-

mous way necessarily fails, as different (groups of) verbs must be treated differently. They argue that verbs that "inherently involve causation of possession and select a possessional goal—that is, a recipient—[...] are monosemous" (Levin & Rappaport Hovav (2005:3)). Examples of this group are verbs of giving, like *give*, *hand*, *loan*, and *sell*, verbs of future having, like *allow*, *offer*, and *promise*, and verbs of type of communicated message, like *tell*, *ask*, and *teach*.

Polysemous dative verbs are verbs of sending, like *send*, *mail*, and *ship*, verbs of instantaneous causation of ballistic motion, like *kick*, *throw*, and *toss*, and verbs of causation of accompanied motion in a deictically specified direction, like *bring* and *take* ((Levin & Rappaport Hovav 2005:3), based on Gropen *et al.* (1989)).

They associate the first of the two groups, the monosemous verbs, with only one event schema, namely the caused possession schema, whereas the polysemous verbs can have either a caused possession schema or a change of location schema. Nevertheless, many of the monosemous verbs do alternate. A question that naturally comes up is why that is the case. Levin & Rappaport Hovav (2005) claim that, in contrast to the polysemous verbs, the monosemous *give*-type verbs are always associated with a caused possession event schema, even in the PO construction. In other words, the event schemata of *sent Mary the letter* and *sent the letter to Mary* are different, whereas for *gave Mary the ball* and *gave the ball to Mary*, they are the same. Polysemous verbs are associated with caused possession in the DO, and with change of location in the PO construction. The reason for the alternation with monosemous verbs is rooted in various other factors, like information structure and heaviness of a constituent (Rappaport Hovav & Levin 2006:31). Evidence from German and Dutch verbs (other than *geben* and *geven*) showing that, at least in these languages, some verbs do alternate and some do not, will be discussed by the end of this chapter.

Related to this problem is another one, namely the status of the two alternating constructions. The question here is whether DO and PO are semantically distinct, or whether they exhibit

the same truth conditions and only differ with respect to their information structural properties. Goldberg (1995) for instance comes to the conclusion that for English, the two constructions are semantically the same, but with different pragmatic values. The DO construction emphasizes the event and the theme, the PO construction puts the recipient in focus. Based on the Dutch data, I will argue something similar. Under my analysis, DO and PO are principally interchangable, and factors like focus and information structure properties play an important role in the choice of one alternative over the other.

### 5.2.2 Categorical versus gradient grammar

Another important issue deals with the question how rigid the constraints are that are imposed on the two constructions. In other words: if we have a constraint, how easily can it be violated? Nowadays, questions like this can be answered rather easily with the aid of large amounts of data. The accessibility of corpora and the internet assists the investigation of the limits of grammatical flexibility. In a series of papers, Bresnan and colleagues (see, e.g., Bresnan & Nikitina (2007), Bresnan *et al.* (2007)) have presented the results of a study which shows that the choice of one over the other construction, for alternating and reportedly non-alternating verbs, is gradient and less categorical than is often assumed. Indeed, there are tendencies, in some cases even very strong ones, and some factors, e.g. weight or pronominality, animateness, etc., heavily affect the choice of alternants. Consequently, Bresnan and colleagues argue for lexical as well as grammatical gradience.

Let me give an example for the lexical gradience of reportedly non-alternating verbs. There are—usually idiomatic—constructions that appear in only one of the two argument frames. Consider, e.g., the idiomatic expression of *giving someone a headache*. You do not normally find this idiom expressed in a PO pattern, which supports the semantic difference between the two constructions. If the DO encodes the meaning 'X causes Y to have Z' and the PO encodes the meaning 'X causes Z to

go to Y', you do not expect headaches to participate in the PO. X causes Y to have a headache, but there is clearly no transfer-meaning detectable, where X causes the headache to go to Y.

Nonetheless, Bresnan & Nikitina (2007) present a large variety of exactly those types of reportedly non-alternating verbs in the unexpected pattern. See, e.g., the examples in (162) and (163):

(162) From the heads, offal and the accumulation of fishy, slimy matter, a stench or smell is diffused over the ship that would give a headache to the most athletic constitution.

(163) She found it hard to look at the Sage's form for long. The spells that protected her identity also gave a headache to anyone trying to determine even her size, the constant bulging and rippling of her form gaze Sarah vertigo.[10]

Even though examples such as these may not be extremely frequent, they provide evidence for the conclusion that the expression *give a headache* is strongly skewed toward the double object construction, rather than categorically determined to only occur in this argument frame.

Bresnan & Nikitina (2007) distinguish lexical gradience from grammatical gradience. Not only are certain expressions more flexible than is usually claimed, grammatical constraints appear to be less rigid than it seems as well. One example for this grammatical gradience is the *NP Pron constraint. Although it is widely claimed that personal pronoun NP objects cannot follow a lexical NP object, which in turn predicts that personal pronouns as the theme argument can only occur in the PO construction, Bresnan and Nikitina cite a number of sentences that violate that constraint, like (164) and (165):

(164) Note: I don't give children peanut butter until they are 3 years old since it is recommended not to give children it to avoid possible allergies.

[10] Examples from Bresnan & Nikitina (2007:4).

(165) You should never give out your address or phone number online and you should never send someone them in the mail either.[11]

Apparently, the pronouns in these examples are chosen to avoid the ugly repetition of the noun phrase. But, even if the sentences look the way they do for stylistic reasons, they do show that the order NP Pron is not ruled out categorically.

To sum up, for my own analysis, I adopt the ideas presented by Levin & Rappaport Hovav (2005) and assume that *give*, *geven* and *geben* are monosemous verbs, while the two constructions, DO and PO instantiate two different event schemas. So far, the lexical approach is not very different from Goldberg's constructional approach. One difference, though, is that the lexical approach claims that, even in the PO construction, the event schema is not a change of location schema for monosemous verbs, whereas the constructional approach does not make any such claims.

My second basic assumption is that there is a lexical gradience that we have to keep in mind, since it is not a hard constraint that reportedly non-alternating verbs (or ELUs) can only appear in one frame, but it is rather a question of likelyhood. This insight was important for my own study in that respect that I decided to include collocations like *give a hand* in my data as well.

### 5.2.3 Factors that influence the alternation

Bresnan *et al.* (2007) show in their paper that reportedly non-alternating verbs can nevertheless alternate to a series of different factors that can influence the choice of an alternant. But the most prominent (or most often investigated) factor is the semantic difference between either the participating verb senses, or the participating constructions.

An extensive investigation into the semantic differences between DO and PO in Dutch has been carried out by Colleman (2006). He concludes that PO and DO construction are not syn-

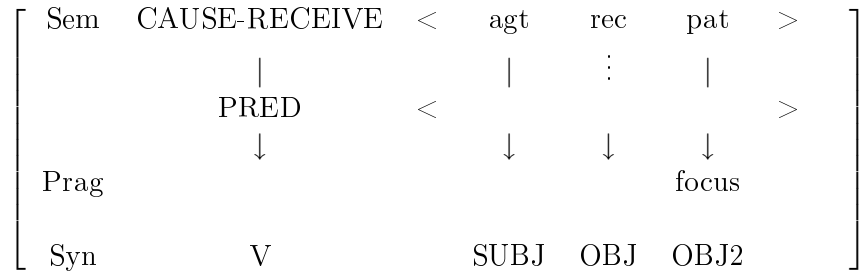
[11] Examples in Bresnan & Nikitina (2007:9).

onymous, as they cover different semantic grounds, that they transport different meanings (p. 609). He also rejects the view that discourse factors such as givenness, pronominality, definiteness, and weight, influence the choice of one alternative over the other. But first and foremost, he argues that DO and PO are two distinct modes of presenting the same state of affairs. The PO, Colleman claims, is a sort of 'extended monotransitive' construction, which backgrounds the recipient and puts emphasis on the interaction of agent and theme, whereas the DO explicitly foregrounds theme and recipient as affected by the agent's doings.

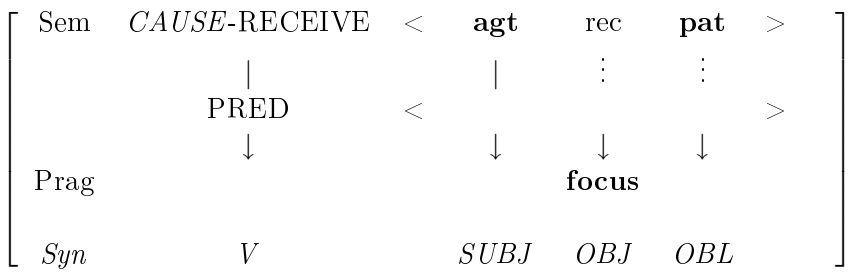
Colleman (2006) comes to the conclusion that the two constructions are semantically different, but that there is an overlap of the semantic fields that they cover. In this overlapping field, alternation can occur. Which of the two constructions is chosen in such a case is influenced by information structural reasons.

This conclusion differs strongly from the one drawn by Goldberg (1995), who bases her argumentation on findings by Erteschik-Shir (1979). Erteschik's claim is that in the DO construction, the recipient is not dominant and the theme is focused. The PO construction, on the other hand, puts focus on the recipient. This is the exact opposite of Colleman's analysis. Goldberg (1995:93) adopts Erteschik's idea and formalizes it as follows. The schema for the DO construction is repeated in (166), the schema for the PO construction is repeated in (167).

(166) Goldberg's Ditransitive Construction schema



(167) Goldberg's Transfer-Caused-Motion Construction schema[12]




Goldberg (1995) claims that DO and PO in English semantically cover the same ground, and the choice of one construction over the other is mainly depending on information structural reasons. Colleman (2006) compares Goldberg's analysis with the one by Bresnan & Nikitina (2003), which he evaluates as basically the same, only with a stronger claim about the influence of discourse-pragmatic factors in the latter.

The formal—rather than semantic—factors that seem to be influencing the dative alternation, have been described by Bresnan & Nikitina (2007). They found out that the *NP PRON constraint, which says that the sequence of a lexical NP followed by a pronoun (if both are arguments), influences the alternation, even though it is gradient. Furthermore, they found a soft effect of person, in the sense that first and second persons are skewed towards NPs, whereas third persons are skewed towards PPs.

Further effects were found for weight (in the sense of word length, see Thompson (1995), Wasow (1997), etc.) Also, (in)definiteness of the arguments played a role. In a nutshell,

"nominality, non-givenness, indefiniteness, inanimacy, and non-locality of persom in the recipient favor the non-Core (PP) realization, compared to the contrasting values of the recipient (pronominality, givenness, definiteness, animacy, and local-

[12]It is not entirely clear to me, why agt, pat and focus are put in boldface in (167), or why the lower line is in italics. On the other hand, it doesn't disturb readability, either.

ity of person), which favor the core(NP) realization" (Bresnan & Nikitina 2007:11).

## 5.3 Analysis of the corpus results

In the remainder of this chapter, I will see how we can make sense of the results of the corpus data I presented in 5.1, using the theoretical considerations I presented in 5.2. I will start with Dutch, since this is the focus, and I will turn to English and German, afterwards.

### 5.3.1 Dutch

The Dutch results of the corpus study are probably the most unexpected of the three. With a distribution of 52.08% for the DO and 47.92% for the PO construction, the data are much more balanced than one might have expected. Dutch is often grouped in one category together with English, as both have this particular alternation, and neither exhibit overt morphological case marking (see, e.g., Rappaport Hovav & Levin (2006)). Nevertheless, the numerical results of this corpus study reveal a drastic difference between English *give* and Dutch *geven*.

The results here also differ from those reported in Van der Beek (2004). In her study of the dative alternation, van der Beek found the following number of instances in the CGN and the Alpino treebank[13], for all alternating dative verbs.

(168) Results in van der Beek (2004)[14]

| | NP NP (unsh) | NP NP (sh) | NP PP | PP NP | TOT |
|---|---|---|---|---|---|
| CGN | 226 | 33 | 63 | 8 | 247 |
| Alpino | 122 | 7 | 43 | 10 | 182 |

[13]URL: http://www.let.rug.nl/vannoord/trees/.

[14]NP NP (unshift) stands for the IO DO order, NP NP (shift) for the DO IO order. Furthermore, in the original, the results from the CGN is not added up to 330, but 334. So, either the count or the addition was presented incorrectly.

This can of course be the source of the difference: not only is the genre entirely different from my data, but the behavior of one verb does not necessarily translate to the entirety of alternating verbs.

It goes without saying that the total number of sentences van der Beek covers, is remarkably smaller than the data from the Europarl. This is mainly due to two factors: the corpora in van der Beek's study are a mere 900 000 words (CGN) and 150 000 words (Alpino), respectively. Furthermore, she chose not to consider passives, object topicalizations, and split dative PPs with R-pronouns. All those were counted in this study, and I will try to show that, e.g., including passives actually reveals interesting facts.

Let's go back to the table in (157). In the Dutch data, it becomes rather obvious that looking at the separate word forms gives a slightly different picture than looking at the total numbers of the distribution of the lemma over the two alternating constructions. Overall, the two alternatives are instantiated in a rather balanced distribution. The form *geven*, which is either the infinitive or the form for 1st, 2nd and 3rd person plural present tense and therefore also covers the largest amount of examples, is slightly skewed towards favouring the DO construction. *Geeft*, the form for 3rd person singular present tense is skewed more clearly towards the DO construction, with almost two third of the examples. The picture changes when we look at *gegeven*, the past participle. I marked all sentences with *gegeven* for whether they are active or passive. And the results are surprising: 33 out of the 270 examples for the DO construction are passive sentences, which is a mere 12.2%. In the PO construction, the passive sentences make up 53.8% of the examples, namely 339 out of 630 sentences. We can conclude from this distribution that the verb *give* has a clear preference for the PO construction in passive sentences. If the assumption that the PO puts the recipient in focus is correct, it would follow from this that the recipient is more likely to be in focus in a passive sentence than the theme.

But even if we leave out the passive sentences, which clearly balance the results more toward the PO construction, the picture

looks different from the one given in van der Beek's data. The ratio without passives is 55.65% for the DO, and 44.35% for the PO construction. This result is still much more balanced than the results found in the Alpino and the CGN treebank, which we might blame on the abovementioned differences of the datasets, but mainly the extreme difference in size.

### Passives

The observation that passives behave differently from actives is interesting, as one might have expected a distribution of the passive sentences that reflects the distribution of active sentences. If we take a look at English, we find that it has two different passives (see, e.g., Huddleston & Pullum (2002)), one which is the passive corresponding to active DO sentences, as in (169), and one that is the passive corresponding to active PO sentences, as in (170):

(169) Grandma was given the cake

(170) The cake was given to grandma

In (169), the recipient *grandma* is the subject of the clause, in (170), the theme *the cake* functions as the subject. Using a pronoun instead of a grandma makes this clear:

(171) She was given a cake

(172) *Her was given a cake

In Dutch, there are two different passive forms as well, but none of them maps the recipient on the subject.

(173) *Mops werd een ei gegeven*
Mops was an egg given
'Mops was given an egg'

(174) *Aan Mops werd een ei gegeven*
to Mops was an egg given
'An egg was given to Mops'

We can be sure that *Mops* in (173) is not the subject of the sentence, because the pronominalization test fails with a nominative

pronoun. So, this is clear object fronting with an impersonal passive:

(175) **Hij werd een ei gegeven*
he was an egg given
'he was given an egg'

(176) *Hem werd een ei gegeven*
him was an egg given
'he was given an egg'

In German, passivizing the indirect or dative object is completely out, too.[15] Furthermore, the PO passives are, at the very least, just as questionable as PO actives.[16]

(177) *Mops wurde ein Ei gegeben*
Mops was an egg given
'Mops was given an egg'

(178) *Ihm wurde ein Ei gegeben*
him was an egg given
'he was given an egg'

(179) **Er wurde ein Ei gegeben*
he was an egg given
'she was given a cake'

(180) *?An Mops wurde ein Ei gegeben*
To Mops was an egg given
'Mops was given an egg'

[15]On German television, there even used to be an advertisement that utilized the ungrammaticality of passivized dative objects as the punchline, when Verona Feldbusch, who had a strong reputation for not being able to apply proper case inflection, starred in an advertisement for a telephone helpline with the sentence *Hier werden sie geholfen* ('Here are you$_{nom}$ helped'), a passive sentence with the recipient as subject.

[16]All three languages have alternative ways of expressing impersonal passives, which is described in Landsbergen (2006). English has the so-called get-passive, Dutch has a similar phenomenon with *krijgen* ('get'), and German uses the verb *kriegen* ('get'). These passives will not be considered here.

Furthermore, the verb agrees with the egg, not with the dative object:

(181) *Ihm wurden zwei Eier gegeben*
him were two eggs given
'he was given two eggs'

Wellander (1920) published a paper called *Over den datief als subject van een passieve zin* ('About the dative as the subject of a passive sentence'), where it is claimed that elements that originally used to be dative objects in Dutch active sentences, are 'wrongly' used as subjects of passive sentences. Even though his argumentation might at some points sound funny in the 21st century, his observations are important, and some questions he raises are very reasonable to ask. The examples Wellander covers are mainly monotransitive sentences which select for an object that he describes as a dative, which is rather uncommon nowadays. In German, however, where direct and indirect objects are case-marked as dative and accusative objects, verbs like *helfen* ('help') indeed take an overt dative object. In Dutch, due to lack of overt marking, the question of accusative or dative is not applicable. But Wellander also gives a few passive ditransitive sentences with *geven*, like the one in (182):[17]

(182) *zuigelingen worden geen morfine gegeven*
babies are no morphine given
'babies are not given morphine'

In this example, the disambiguation between subject and object of *zuigelingen* happens with the help of agreement features. If the first noun phrase was an object, the finite verb would agree with a singular, uninstantiated subject, and would have the form *wordt*. *Worden* is plural and can therefore only agree with the subject *zuigelingen*.

Wellander claims that the shift to the passivization of dative-like objects is a tendency that we find in many Germanic lan-

[17] For many native speakers of Dutch, the construction with the recipient of a ditransitive as the subject of a passive sentence sounds odd, to say the least.

guages. The furthest along the road, he argues, are English and Swedish, where this construction is completely standardized. German, he shows, does not allow for this shift, and Dutch seems to be somewhere in between.

Still, he looks mainly for extralinguistic factors that might influence or trigger this development, like the decreasing influence of languages like Latin on the Germanic languages, the lesser impact of written language and prescriptive grammars on spoken varieties, and the bigger democratic rights of the people. I would like to show that linguistic factors can facilitate this shift, too.

Wellander (1920:292) does mention the 'loss of feeling for different cases' in Dutch and English as one factor, but discards it as an explanation for the development. In his opinion, the cause for such a shift, from passivizing the one object to passivizing the other, should not be looked for in linguistic, but must be found in psychological reasons. In a way, he makes a distinction between the facilitating condition and the cause of a change, and the loss of case inflection is a precondition which initially makes a change possible, but the cause, he argues, is found in the fact that speakers prefer to make the 'psychological subject' the grammatical subject as well. The notion of 'psychological subject' goes back to von der Gabelentz (1891:371f.), who was the first to make a distinction between psychological and grammatical subjects and predicates. The psychological subject, in this view, is the element that the speaker actually 'thinks about', and the psychological predicate is whatever is thought about the psychological subject. These two terms can be compared to the more contemporary notions of *topic* and *comment*.[18] Wellander does not elaborate on the question of what exactly would make the recipient of a ditransitive active sentence the psychological subject, and also in van der Gabelentz, there are no clear underpinnings of a view like this. Nevertheless, the claim I want to make here goes into the same direction.

[18]The psychological subject is different from the logical subject in that it is not necessarily the actor of a dynamic verb. Logical subjects usually are the actors in a semantic frame.

For the monotransitive sentences that Wellander cites, an explanation based on purely language-internal grounds is possible. With the loss of morphological case marking, the overt distinction between dative and accusative, or indirect and direct object, is not possible anymore in English and Dutch. Monotransitive sentences with verbs that select for a direct object—the huge majority of verbs, that is—can be passivized, and the argument that was the object of the active sentence is expressed as the subject of the passive sentence. If there is no overt difference between the two groups of verbs, and there is no distinction felt by the speakers, a simple process of analogy can lead to treating all verbs the same. If you can passivize one type of monotransitive clauses, and there is a second type of similar looking clauses, why not passivize this type as well? Hence, speakers allow the monotransitive dative verbs to feed passive, too. Therefore, for this type of shift, the facilitation can also be the cause of the shift, given the pervasiveness of analogy. A generalization is formed over the phenomenon that clauses with one argument can be passivized, and this pattern is extended to another type of clauses.

For ditransitive sentences, the story is a little bit different. The mere loss of overt case marking facilitates but does not actually cause the shift from passivizing the direct to passivizing the indirect object, as has happened in English, and might become more standardized for Dutch in the future. In other words, the loss of the overt distinction between accusative and dative argument alone does not provide sufficient cause to switch from passivizing the one object to passivizing the other. There must be additional motivation for speakers to choose the indirect object as the candidate. It is not satisfying to assume that speakers may have the tendency to just pick the first object that they encounter after the verb as the candidate for passivization. There is little evidence that can be provided for this, and this argumentation has very little explanatory power. Furthermore, this answer would be based on the fact that the recipient is the first object in the default word order, and this fact would remain unexplained as well.

One factor draws on results by Levin (2006) about the status of first (or indirect) objects in English. Levin comes to the conclusion that, in general, English first objects are not comparable to the objects of monotransitive verbs, despite their passivizability and postverbal position, but they are comparable to dative NPs in languages that do mark dative case (Levin 2006:2). Further on, she argues that "[w]ith dative verbs, since recipients are typically human, the[y] are likely to be given, while themes are typically inanimates and, thus, less likely to be given." (Levin 2006:11). For information structure purposes, given information precedes new information (see, e.g., Lambrecht (1994), Vallduví (1992), and Ward *et al.* (2002)), hence the DO fulfils the need for a construction where the recipient linearly precedes the theme.

So how does all this connect to passives? If we claim that the recipient precedes the theme in active sentences for information structural properties, we can claim the same for passives as well. New information tends to be encoded toward the end of the sentence, so if the recipient is old information, it is likely to preceed the new information, namely in this case the direct object or theme, and the same would hold for passive sentences.

In the end, it makes no difference which of the two possible explanations mentioned is taken to account for the passivization of the indirect object. If the speaker broadens the generalization that the first object of an active sentence is the subject of the corresponding passive sentence, this object will be the the indirect one in the DO construction, due to information structural reasons. If the speaker wants to put the NP that is old information on the first sentence position, that will also be the indirect object, and also for information structural reasons.

These two factors, analogy on the one side and information structure on the other, answer the question of why there is a shift over time in English and Dutch, from passivizing the indirect object to passivizing the direct object. For the Dutch shift of monotransitive dative verbs, it can be said that this was just an analogical process that copied the examples of the vast majority of monotransitive verbs. Furthermore, this shift was facilitated, but not caused, by the loss of overt case marking. For the shift

of passivized objects of ditransitives it can be said that the shift from passivized direct object to passivized indirect object is facilitated by loss of case marking and caused by the information structural tendency to let given information precede new information.

### Word order

As we have seen above, one obvious difference between the English and the Dutch examples is the fact that in Dutch, both constructions appear in both possible relative orders of the object, while the shifted versions are not found in English. As van der Beek's results have shown, the "shifted" orders, i.e., SU DO IO and SU PP DO, are much less frequent than the "unshifted" orders. But still they do exist. Van der Beek (2004) models two factors that influence the choice of a pattern in Dutch, namely weight and pronominality.[19] There is indeed a tendency to align heavy constituents to the right and to shift the direct over the indirect object, if both are pronominal. But these are mere tendencies, and van der Beek follows Bresnan & Nikitina (2003) in assuming that these tendencies are constraints on probability[20] rather than categorical constraints.

In the Europarl data, I found the same result. Even though they are scarce, there are DO sentences with a heavy, unshifted direct object, just as there are PO sentences with a very heavy but still shifted PP, that precedes the DO.

[19] In contrast to the Bresnan's work, she did not find any effects of person or definiteness. I did not investigate these factors, but it will be interesting to see in the future, whether the larger dataset reveals any effects.

[20] Probability is understood here as the chance that a speaker choses one variant over the other. If there is an actual choice between two variants, both must necessarily be part of the grammar.

(183) *Ik hoop dat [aan de oproepen die u met name tot*
I hope that to the appeals that you with name to
*de bankwereld hebt gericht] gehoor zal worden*
the bank-world have directed hearing will be
*gegeven*
given
'I hope that the appeals that you particularly directed to the bank world, will be heard'

In (183), the recipient PP is *aan de oproepen die u met name tot de bankwereld hebt gericht*, and the direct object is *gehoor*. This is the combination of a very heavy shifted PP and a rather light NP. The question that arises is the following: if weight was a categorical factor, this sentence was improbable. If the order of the constituents were fixed in both DO and PO (or rather recipient and theme), this sentence would not have been felicitous. If the speaker wanted to mention the recipient first, why did she use this construction rather than the DO pattern? I will try to answer that question in the discussion below.

### 5.3.2 English

The set of English DO and PO sentences in our data is the biggest by far. An intuitive explanation might be that, different from Dutch and, even more clearly, German, English does have a large number of idiomatic expressions with the verb *give* that use a DO or PO pattern, like *give a damn* or *give a hand*. But when it comes to data, size does matter, and the bigger the set of matches, the more reliable the results.

English is clearly the least flexible of the three languages, when it comes to the rigidity of sentence patterns. As English is known to be a language with strict word order, this is not surprising. *Give* is strongly skewed towards the double object construction, with almost two third of the observed relevant sentences instantiating this pattern. The internal order of arguments is stable across the examples, with the subject in first position and the indirect object or recipient preceding the direct object or theme. There are particular constructions, like relative clause modifiers

and questions, where the direct object by default scrambles over the subject, like in (184):

(184) A comment and question to the Council's representative: is the answer he just gave us specific? (ep-00-04-12.al)

Scrambling of the direct object in simple sentences, for mere pragmatic reasons and without a grammatically obvious reason, is not found in the Europarl. Hence, it is valid to say that in the Europarl data, the double object construction in English never deviates from the standard word order.

The *to*-variant is in English just as inflexible as the double object variant. The standard word order is SU > DO > PP, and the data never deviates from this pattern. With a count of more than 16.000 sentences, this result speaks for itself. Apparently, the rigidness of the two default word orders SU DO PP and SU IO DO is not at all gradient. This result would fall into the category 'grammatical gradience' in the work by Bresnan *et al.* (2007). I did not, however, control for, e.g., NP Pron order. Therefore, this result does not contradict Bresnan *et al.* (2007), it only adds the factor that, for English, the internal order of arguments is not gradient, but categorical, at least in the Europarl data.

The distribution I found in the Europarl corpus differs slightly from the data that Bresnan *et al.* (2007) found in the Switchboard corpus of English Telephone conversation[21]. In their data, 78,6% of the set of alternating dative constructions are DOs. 51% of the DO or PO sentences are headed by the verb *give*, which patterns with the DO in 84.6% of the cases. This number is noticeably higher than the percentage of DOs I found in the Europarl data, but the difference can be explained from the nature of the data, as the two datasets differ in important respects. The Switchboard corpus covers 3 million words, Europarl contains ca. 20 million words per language. Furthermore, the Switchboard data are spontaneous spoken language, whereas the Europarl speeches, though eventually spoken, hardly qualify for speech, let alone spontaneous language. Last but not least, Switchboard

[21] http://www.ece.msstate.edu/research/isip/projects/switchboard/.

covers American English, and the Europarl is strongly skewed towards British English. With these factors in mind, it is not surprising that the results differ. Nonetheless, the tendency is the same in both corpora: *give* has a strong tendency to pattern with the double object construction.

### 5.3.3 German

I am not aware of an empirical study of German ditransitives or the dative alternation. That is not surprising, as—if we are looking at the distribution alone—there does not seem to be a systematic alternation, at all. In Europarl, more than 99% of the instances with the lemma *geben* are in a DO construction, and only 17 sentences overall exhibit a PO pattern. An easy way of explaining these examples would be just to blame the translators (in those cases where the sentences are translations from English). But I want to try and look at the PO examples more carefully in order to find principles other than mere mistakes.

One thing that seems to be regular, even with the obvious sparseness of the data, is that the NPs that are selected by the preposition *an* are usually very heavy. I will give two examples:

(185) *Wie bei anderen Gelegenheiten [...] stehen wir*
As with other opportunities [...] stand we
*vor einer rein technischen Prüfung*
in-front-of a purely technical examintion
*durch ein internes Organ des Parlaments, das*
through an internal organ the$_{GEN}$ Parliament, which
*eine Empfehlung an die Präsidentin, ein externes*
a recommendation to the president, an external
*Organ des Parlaments, gibt, das eine*
organ the$_{GEN}$ parliament$_{GEN}$, gives, which a
*Entscheidung treffen muss*
decision make must

'As in other situations, we are facing a purely technical examination, carried out by an internal organ of the Parliament, which will give a recommendation to the pres-

ident, an external organ of the Parliament, which must make a decision'

(186) *Meine Frage geht dahin, ob es möglich ist,*
My question goes there, whether it possible is,
*dass die Europäische Investitionsbank auch indirekte*
that the European Investment-bank also indirect
*Förderungen gibt, z.B. an Gruppen, die sich*
funding gives, e.g. to groups that themselves
*fuer Klein- und Mittelbetriebe einsetzen, [...]*
for small and medium-companies engage, [...]
'My question is whether the European Investment bank also gives indirect funding, e.g. to groups that support small and medium size companies, [...]'

In both examples, the recipient NPs are heavy NPs with a post-nominal modifier. In (185), the recipient is *die Präsidentin, ein externes Organ des Parlaments, das eine Entscheidung treffen muss. Das eine Entscheidung treffen muss* is a modification to *Organ*, and the entire constituent is an apposition to *Präsidentin*. The finite verb *gibt*, intervenes modifier and modified element of the apposition, which is not uncommon in German and which ensures interpretability if the material between subject and (lexical) verb becomes too long.

In (186), the recipient is *Gruppen, die sich für Klein- und Mittelbetriebe einsetzen*, a noun with a relative clause modifier. The heavy NP is scrambled to the right, which is indicated by *zum Beispiel* ('for example'). In addition to the fact that we are dealing with a scrambled heavy NP, another influencing factor for the PO pattern could be found in processing difficulties. For plural, case markings are syncretized in German. Therefore, it is not obviously visible that *Gruppen* ('groups') is a dative (in the DO variant), and so placing it out of its usual position in the sentence makes processing the NP more difficult. Scrambling it to the unusual position and making it a PP instead, guarantees an easy interpretation as the recipient.

In both cases, though, it is important that a realization of this proposition with a DO pattern would at least sound less

natural (to me, that is). A recipient NP of that size fits very well into a PO pattern. That does not mean, on the other hand, that modified heavy NPs do not occur in the DO sentences. We also find examples like the following (and, comparing the numbers for DO and PO examples in German, DOs clearly are much more frequent):

(187) *Als ich davon sprach, dass wir den Bürgern in*
When I there-from spoke, that we the people in
*den Kandidatsländern wie auch in unseren Ländern*
the candidate-countries as also in our countries
*Garantien bieten müssen, […]*
guarantees must offer, […]
'When I said that we have to offer guarantees to the people in the candidate countries as well as the people in our countries, […]'

(188) *Ich ziehe ein Europa vor, das all jenen eine*
I pull a Europe before, which all those an
*Integrationschance gibt, die sich zu seinen*
integration-chance gives, who themselves to its
*gemeinsamen Werten bekennen, ob sie*
common values commit, whether they
*sich nun einer Religion zugehörig fühlen oder*
themselves now a religion to-belonging feel or
*nicht, ob sie die Sprache ihrer Ahnen*
net, whether they the language of-their forefathers
*sprechen oder entschieden haben, sich von ihren*
speak or decided have, themselves from their
*Wurzeln zu lösen*
roots to part-from
'I prefer a Europe that gives a chance to integrate to all those, who commit to its common values, whether they feel like they belong to a religion or not, whether they speak the language of their forefathers or whether they have decided to part from their roots'

In (187), the recipient is *den Bürgern in den Kandidatsländern*

*wie auch in unseren Ländern*, a noun with two coordinated PPs as modifiers. In (188), the recipient is interrupted by *gibt*, again, as it is extremely long and has several dependencies, namely *all jenen, die sich zu seinen gemeinsamen Werten bekennen , ob sie sich nun einer Religion zugehörig fühlen oder nicht, ob sie die Sprache ihrer Ahnen sprechen oder entschieden haben, sich von ihren Wurzeln zu lösen.*

I assume that sentences with such long recipients can be expressed with a DO construction, exactly because it is possible to linearly interrupt one argument (or even constituent) with other material, like the finite verb, as this enhances the understandability of a proposition greatly. A factor that very likely plays a role in this, is the location of the NP. If the NP is in situ, like in (187), processing is easier, while a scrambled NP causes more difficulties and is more likely realized as a PO, thus being marked as the recipient overtly.[22]

(189) *[. . . ] kommt es hauptsächlich darauf an, dass*
comes it mainly R-PRO on, that itself
*sich die Europäische Union eine gute*
the European Union a good government
*Regierung gibt [. . . ]*
gives
'the most important thing is that the European Union gives itself a good government'

[22]This seems to be a very nice topic for a psycholinguistic experiment, e.g., a magnitude estimation setup. My intuition is that factor heaviness increases the acceptability vastly, just as the actual transfer of a physical object would. But as I cannot be sure that my personal intuitions are not skewed by high exposure to Dutch and English PO sentences, an experiment to test this hypothesis might yield more interesting and linguistically valid insights.

(190) *Der Mensch ist nicht nur darauf angewiesen,*
the human-being is not alone there-on dependent
*dass man ihm die Mittel zum Leben gibt, sondern*
that one him the means to live gives, but
*auch und vor allem die Gründe für das Leben*
also and for all the reasons for the life
'People not only depend on that they are given the means to live, but also, and most of all, a reason to live'

## 5.4 Discussion

The empirical comparison of the dative alternation in English, German, and Dutch reveals some significant structural and distributional differences between the three languages, at least for the target verbs *give*, *geben*, and *geven*. Bresnan & Nikitina (2007:15) point out that, even though it often is taken as the prototypical verb for the ditransitive, *give* behaves differently from other alternating verbs. They attribute this to the fact that *give* partakes in a variety of idiomatic expressions. I take this as the explanation for the comparably large amount of examples in the English part of the Europarl, compared to the Dutch and German examples. At the same time, Bresnan & Nikitina (2007) make a strong point, arguing that even in idiomatic expressions, *give* tends to alternate, and they illustrate that with a number of example sentences. Many of those sentences are similar to the German examples for PO sentences that I found in Europarl. Even though intuitively it looks like no alternation is possible, there can be pressing factors like understandability that lead to expressing the recipient as a prepositional phrase, after all. One case of this was sentence (186). Nevertheless, the observations I made in the corpus that only contains sentences with *give*, *geven* and *geben* lead me to conclude that the choice of this set of verbs still leads to interesting results, and I am convinced that an investigation of the entirety of alternating verbs will lead to comparable results. Of course, detailed analyses like Colleman (2006) for Dutch or Levin & Rappaport Hovav (2005) for En-

glish that include a large set of alternating verbs have concluded that different verbs behave differently, but still I predict that a larger study with a bigger set of alternating verbs will result in the same empirical tendencies that have been revealed here. The English *give* shows a preference for the DO construction, and obeys to grammatical constraints like, e.g., that no shifted orders are allowed. The German *geben* instantiates the DO construction by default, but there are pressing contextual factors that can force the speaker to use the PO construction. Dutch *geven* does not show a clear preference for one of the two constructions, and it does not seem to obey grammatical constraints on the internal argument order, as English does.

But now, let us move on to the part where we search for an explanation of the data. In the beginning of this chapter I claimed that case deflection can be a factor that influences the result of the contrastive analysis. I base this claim on an investigation of the rise of the *to*-dative by McFadden (2002).

According to McFadden, in Old English, only the ditransitive construction was available, but then with both word orders for the two objects. So, whatever information structure (or rather, order of thematic roles) had to be encoded, this was possible using a double object pattern. In a detailed corpus analysis, McFadden shows that the prepositional construction in English first arose in Early Middle English, coinciding with (or, more likely, triggered by) the loss of case marking morphology. His first attestations stem from 1150 CE, although he argues that the prepositional counterpart only became a fully viable option in the period between 1250 and 1350. He also provides evidence for the assumption that weight played an important role in the choice of constructions from very early on, as even in the beginning, heavy constituents preferably aligned to the right, which lead to IO DO for light recipients, and to DO PP for heavy recipients. He shows the influence of weight in his table 3, here repeated as (191). The N row gives the number of attestations, the weight row gives the average number of words in the constituent.

(191) Distribution of DO and PO in ME

| | Double Objects | | *to*-Datives | |
|---|---|---|---|---|
| | DO-IO | IO-DO | DO-IO | IO-DO |
| N | 57 | 109 | 231 | 55 |
| avg. DO weight | 1.86 | 3.18 | 1.95 | 4.45 |
| avg. IO weight | 3.65 | 2.1 | 2.58 | 2.73 |

McFadden's claim that present day German actually instantiates two distinct constructions, one which has an IO DO order and one that has an DO IO order, where the latter should be comparable to the prepositional dative constructions in English and Dutch, does not seem to be borne out in my data. The vast majority of sentences exhibit the IO DO order, and DO IO is only enforced by information structure properties. To me, this does not seem to be reason enough to establish a distinct construction. If we can arrive at the SU DO IO order by combining the ditransitive construction with information structural mechanisms that usually and generally apply to the German mittelfeld, positing a distinct grammatical structure seems redundant. If we look at information structure as constructions in their own right which combine with argument structure constructions, we get the following picture: German has one single double-object construction which can combine with information structure constructions, which then determine the internal order of the arguments.

Nevertheless, we learn from McFadden's study that the PO construction is a real alternative for the DO IO order, and that it came into existence for exactly this reason. Ambiguities arising from case deflection made an unambiguous alternative necessary, and the prepositional *to*-dative filled that void. In German, ambiguities are rare due to the case marking properties, which explains why there is almost no need for an alternating construction in German. In Dutch, on the other hand, the two constructions are still so obviously flexible with regard to choice of construction and word order within a chosen alternant, as they are still in the process of becoming more constrained. The speaker does not have to obey hard constrains on the choice of the construc-

tion, and once she chose either DO od PO, she can also choose whether the shifted or the unshifted order is the better option. As table (191) shows, McFadden found both constructions to appear shifted and unshifted in Middle English as well. I know that making claims about future language changes is not exactly in the realm of this work, but it is perfectly imaginable that Dutch will arrive at two fixed constructions SU IO DO and SU DO PP at one point as well.

### 5.4.1 Polysemy versus monosemy

In 5.2.1, I introduced the idea by Levin & Rappaport Hovav (2005) and Rappaport Hovav & Levin (2006) that some groups of alternating verbs are polysemous, while others are monosemous in nature. Rappaport Hovav & Levin (2006:32) predict that, if there is a connection between word order and the existence of a dative alternation in a language, then there should be a difference between languages like German and English. And this is borne out for the verbs *give* and *geben* as well as for the larger group of alternating verbs.

The almost entire lack of the PO construction in German is evidence that the verb *geben* does not alternate in German, and, following from that, cannot be polysemous either (at least not between cause of possession and change of location). In this respect, Rappaport Hovav and Levin's assumption that verbs of giving are monosemous is borne out for German. Extending this claim to Dutch and English, where there is an alternation, is in my opinion a valid step, since the verbs are semantically equivalent.

I do have a different perspective, though, on their German example (72) (cited here as (192)) in (Rappaport Hovav & Levin 2006:32), cited from (Hameyer 1979:235).

(192) **Ich gab einige Blumen zu diesem Mädchen*
I gave some flowers to this girl
'I gave some flowers to this girl'

The claim here is that *geben* does not occur with the allative

preposition *zu*. That is in fact true, but the example gets remarkably better with *an*, when it receives a strong directional or allative interpretation.

(193) *Ich gab einige Blumen an dieses Mädchen*
I gave some flowers to this girl
'I gave some flowers to this girl'

*Zu* might be used in certain allative contexts, e.g., with verbs of saying like *sagen* ('say'), but it is never used with verbs of giving or verbs of sending, independent of whether the recipient/goal is animate or locative.

Interestingly, *zu* does play a role with the verb *werfen* ('throw'). *Werfen* is inherently intransitive or monotransitive, and cannot surface with a recipient argument, only with a locative goal. The phrasal verb *zuwerfen* ('to-throw'), on the other hand, patterns with (what I claim is) the DO construction, with a stranded preposition, as in the following example:

(194) *Jan warf dem Mädchen den Ball zu*
Jan threw the girl the ball to
'Jan threw the girl the ball'

The stranded preposition is a regularity of prefixed or phrasal verbs in German. In analytic verb forms, the prefix is not split off, and then *zuwerfen* patterns exactly like *geben*. But the version with a well-behaved PP is grammatical, too:

(195) *Jan warf den Ball zu dem Mädchen*
Jan threw the ball to the girl
'Jan threw the ball to the girl'

This sentence is a clear case of a Caused-Motion construction, as *zu dem Mädchen* is interpreted only as a direction, and lacks any sense of possession. The scenario that the girl is looking the other way and the ball hits her in the head is just as likely as the scenario where the girl aptly catches the ball. Hence, there is an obvious meaning difference between the two examples (194) and (195).

The difference between (194) and (195) can be explained by assuming that two different verbs, *werfen* and *zuwerfen* are involved. Using an analytic verb form shows a morphosyntactic difference:

(196) *Jan hat dem Mädchen den Ball zugeworfen*
Jan has the girl the ball to-thrown
'Jan has thrown the ball to the girl'

(197) *Jan hat den Ball zu dem Mädchen geworfen*
Jan has the ball to the girl thrown
'Jan has thrown the ball to the girl'

To conclude this little excursion, the German verb *werfen* does not alternate, but only allows for a monotransitive version, where a locative PP can surface. For the ditransitive version, a different verb, namely *zuwerfen*, is used.

Let us now return to Rappaport Hovav and Levin's idea that not all alternating verbs can be treated uniformly. If the claim holds that—in contrast to verbs of giving—certain groups of verbs are polysemous, one would expect that German has means to express the different senses that a polysemous verb can encode. And indeed, this claim bears out. For verbs of sending, compare the German examples (198), (199), and (200):

(198) *Jan schickte Oma einen Brief*
Jan sent grandma a letter
'Jan sent grandma a letter'

(199) *Jan schickte einen Brief an Oma*
Jan sent a letter to grandma
'Jan sent a letter to grandma'

(200) *Jan schickte einen Brief nach London*
Jan sent a letter to London
'Jan sent a letter to London'

Example sentence (198) is an instantiation of the DO construction, with *Oma* as the dative and *einen Brief* as the accusative

object. A reverse order of the two objects sounds odd, to say the least:

(201) *?Jan schickte einen Brief Oma*
Jan sent a letter grandma
'Jan sent grandma a letter'

Example sentence (199), on the other hand, is an instantiation of the PO construction. Just as in (198), reversing the order of the objects makes the sentence odd, although in this case, contrastive focus helps making the sentence acceptable, as illustrated in (203):

(202) *?Jan schickte an Oma einen Brief*
Jan sent to grandma a letter
'Jan sent a letter to grandma'

(203) *Jan schickte an Oma einen Brief und an Opa ein Fax*
Jan sent to grandma a letter and to grandpa a fax
'Jan sent a letter to grandma and a fax to grandpa'

For English, it is often argued that the PO construction is a direct or metaphorically derived version of a more general Caused-Motion construction (see, e.g., Goldberg (1995)). Since it is argued by all the aforementioned authors that there is a semantic notion of movement along a path that seems to come with these sentences, and since the form is the same for both constructions, this seems a reasonable claim. Compare the examples given in Beavers (2005), where (204) is underspecified for possession and can obtain a locative as well as a recipient (the London office) reading, whereas (205) can only have a London-office reading.

(204) John sent a letter to London

(205) John sent London a letter

For German and Dutch, though, the argument based on formal features does not hold, as the prepositions instantiated in the caused-motion and in the PO construction are different, namely

*an* (German) and *aan* (Dutch) for recipients in the PO and *nach* (German) and *naar* (Dutch) for goals in the Caused-Motion construction. Example sentence (200) is a caused-motion variant, with a location as goal. The Dutch verb *sturen*, the equivalent to *send* and *schicken*, patterns the way the German verb does:

(206) *Jan stuurt oma een brief*
Jan sends grandma a letter
'Jan sends grandma a letter'

(207) *Jan stuurt een brief aan oma*
Jan sends a letter to grandma
'Jan sends a letter to grandma'

(208) *Jan stuurt een brief naar London*
Jan sends a letter to London
'Jan sends a letter to London'

As in German, the sentence pattern for caused-motion and PO constructions is the same, but the prepositions are different: *aan* for the PO with animated recipients, and *naar* for the Caused-Motion construction with spatial goals.[23]

Thus I conclude, for the verbs *schicken* and *sturen*, German and Dutch do have a Caused-Motion construction as well as both variants of the dative alternation. This supports the claim by Rappaport Hovav & Levin (2006) that a uniform categorization of all groups of alternating verbs in English might not be correct: certain verbs behave inherently differently from other verbs, even if they pattern the same way.

Goldberg's (1995) claim that the English PO construction is a subconstruction of the Caused-Motion construction, namely the Transfer Caused-Motion construction, does not generalize to German or Dutch. In those two languages, the PO is clearly not a subconstruction of the Caused-Motion construction. Not only

[23] *Jan stuurt een brief naar oma* is possible as well, but in my opinion, *oma* is construed more like a spatial goal rather than a recipient, then. The opposite holds for *?Jan stuurde een brief aan London*, which, if at all possible, gets a strong 'London-office' reading.

is the preposition different for PO and Caused-Motion construction, verbs can pattern with both constructions as well, which then leads to unambiguous sentences. I don't claim that the analysis is incorrect for English, too, but in the light of a contrastive analysis, it is questionable. Nevertheless, that the analysis of a phenomenon in one language should not necessary be influenced by findings in another language, but it can, to say the least, profit from a look over the fence. I wonder whether Goldberg had proposed an analysis of the English PO construction as a version of the Caused-Motion construction if she had been aware of the Dutch and German data. In case the answer might be 'yes', the follow-up question would be how insights into one language can be if the analysis cannot hold for another language with a comparable phenomenon.

### 5.4.2 Categorical versus gradient grammar

In section 5.2.2, I referred to work that provides evidence that certain constraints on the dative alternation in English are constraints on probability rather than grammaticality. I focused on the empirical distribution of the two patterns, but I also looked at the internal orders of arguments within the particular construction. The behavior of the various languages investigated turned out to be revealing with respect to the question whether grammatical constraints are categorical or gradient. Each language reveals a different picture.

For English, the word order constraint is categorical for each construction. Even in a set of more than 15.000 sentences, there is no example of the orders SU DO IO (unless both objects are pronouns) or SU PP DO. Therefore, it is valid to claim that the internal order is completely inflexible. Factors like weight and pronominality most certainly play a role in the choice of the construction, but there is a large number of examples with, e.g., heavy indirect objects and light PPs. The heaviness constraint is therefore gradient, in that it imposes a certain improbability on particular structures without ruling them out.

For Dutch, the answer to this question is a bit more compli-

cated. Certainly, the default order for the two constructions is SU IO DO and SU DO PP, as they form the majority in the set. Nevertheless, both shifted orders, SU DO IO and SU PP DO, appear as well. This is also what Van der Beek (2004) found in her (much smaller) corpora and showed in her Optimality Theory-analysis. It is obvious that the default word order constraint, different from the one in English, is a flexible one and can be influenced by other factors, in other words, in the presence of other factors, it is likely that the constraint is violated. The factors which were assumed to play the most important role in many previous studies, weight and pronominality, are gradient constraints as well, as there is a clear tendency to align heavy constituents to the right. But the existence of sentences with shifted heavy PPs must lead to the conclusion that there have to be other factors involved that influence the choice.

I propose that these additional reasons can be found in the information structural and semantic arguments that Colleman (2006) identified for Dutch, and Goldberg (1995) for English. If we follow Goldberg's analysis that the PO puts the recipient in focus, the speaker opts for the PO construction if she wants the proposition to be about the recipient in first place. This choice can co-occur with the discourse-pragmatic necessity to shift the PP before the DO, and then we can end up with a rather improbable sentence as (183), here repeated as (209):

(209) *Ik hoop dat aan de oproepen die u met name tot*
I hope that to the appeals that you with name to
*de bankwereld hebt gericht gehoor zal worden*
the bank-world have directed hearing will be
*gegeven*
given
'I hope that the appeals that you particularly directed to the bank world, will be heard'

## 5.5 Formal analysis

Just as I did in the case of the NPN and the *weg*-construction, I will finish the chapter on the dative alternation with a formal, computational analysis for the Dutch constructions. The prerequisites for the analysis are based on the sections above. Let me start with the verbs.

Polysemous verbs such as *zenden* are linked to two different lexeme classes which capture the semantic frames of caused possession on the one hand, and change of location on the other. (This analysis is based on the analysis that Sag (2007b:29f) gives for verbs of the '*spray/load*' alternation). In this view, the verbal entry itself is minimally specified, but must obey all the restrictions of the constructions that it is linked to. This approach offers an elegant way to solve the question of monosemous and polysemous ditransitive verbs. The two crucial lexeme classes are presented below:

(210) Caused-possession lexeme

$$\textit{c-p-lxm} \Rightarrow \begin{bmatrix} \text{ARG-ST} & \langle \text{NP}_x\ ,\ \text{NP}_y\ ,\ \text{NP}_z \rangle \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & s \\ \text{FRAMES} & \left\langle \begin{bmatrix} \textit{transact-1-fr} & \\ \text{SIT} & s \\ \text{ACTOR} & x \\ \text{RECIPIENT} & y \\ \text{THEME} & z \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

(211) Change-of-location/transfer lexeme

$$c\text{-}o\text{-}l\text{-}lxm \Rightarrow \begin{bmatrix} \text{ARG-ST} & \langle \text{NP}_x\ ,\ \text{NP}_z\ ,\ \text{PP}\big[to\big]_y \rangle \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & s \\ \text{FRAMES} & \left\langle \begin{bmatrix} \textit{transact-2-fr} & \\ \text{SIT} & s \\ \text{ACTOR} & x \\ \text{THEME} & z \\ \text{GOAL} & y \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

Polysemous verbs have one lexical entry that interacts with both of the lexeme classes, since the frames in both classes are the same, only with a different thematic-role assignment.[24] Both lexical entries use a transaction frame, but the thematic role assignments in both frames are different. Fot the *transact-1-fr*, which is used by caused-possession lexemes, the thematic roles of the three NP arguments are ACTOR, RECIPIENT, and THEME, whereas the thematic roles in the *transact-2-fr*, which is used by the change-of-location/transfer lexemes, assigns the roles ACTOR, THEME, and GOAL for the PP. The *transact-fr* must be a supertype of the *transact-1-fr* and the *transact-2-fr*, which enables the lexical entry to license expressions of both, caused-possession and change-of-location/transfer, semantic classes.

(212) Lexicon entry for *zenden*, DO

$$\begin{bmatrix} \text{FORM} & \langle \text{zenden} \rangle \\ \text{SEM} & \begin{bmatrix} \text{FRAMES} & \langle \big[\textit{transact-fr}\big] \rangle \end{bmatrix} \end{bmatrix}$$

More difficult to handle are verbs like *geven*, since they are supposed to be monosemous, but can appear in both frames. Linking them the same way as we linked *zenden* will allow

[24] In his analysis of the *spray/load* -alternation verbs, Sag assumes that both variants are semantically the same. In the approach here, they are not the same, but systematically related.

instances of *geven* to receive the interpretation of change-of-location/transfer, which we wanted to avoid. So what needs to be captured is the possibility of occuring in both syntactic configuration, but with a persistent role assignment, namely the *transact-1-fr* class. Hence, we capture the following information in this lexical entry.

(213) Lexicon entry for *geven*

$$\begin{bmatrix} \text{FORM} & \langle \text{geven} \rangle \\ \text{SEM} & \begin{bmatrix} \text{FRAMES} & \left\langle \left[ \textit{transact-1-fr} \right] \right\rangle \end{bmatrix} \end{bmatrix}$$

The lexical item *geven* can, unlike *zenden*, only combine with one type of *transact-fr*, namely *transact-1-fr*. Nevertheless, it can be instantiated in two alternating argument configurations. Hence, we need to add a third type of lexeme class, which combines the *transact-1-fr* with the argument configuration of the PO construction:

(214) Caused-possession-PP-lexeme

$$\textit{c-p-pp-lxm} \Rightarrow \begin{bmatrix} \text{ARG-ST} & \langle \text{NP}_x \text{ , NP}_z \text{ , PP}\left[ \textit{aan} \right]_y \rangle \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & s \\ \text{FRAMES} & \left\langle \begin{bmatrix} \textit{transact-1-fr} & \\ \text{SIT} & s \\ \text{ACTOR} & x \\ \text{RECIPIENT} & y \\ \text{THEME} & z \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

Of course, there are other possible solutions to this problem. We could have the lexical entry only combine with the appropriate *transact-1-fr* and add the information that it can have an alternative argument structure. But the presented analysis captures the fact that the two frames are systematically related in a stricter way. Adding an alternative argument structure gives room to all possible structures, but since it is systematically the

same structure for the entirety of these monosemous alternating verbs, this option seems too unconstrained. On top of that, we employ the same mechanism for the dative alternation as we do for the *spray/load*-alternation.

Cognitively, this is also a reasonable analysis, since it captures the insight that the speaker has the knowledge that *geven* can occur in both lexeme classes, but nevertheless has stable semantics throughout the different usages.

When it comes to the argument structure constructions that license DOs and POs, I'm hesitant to generalize over the entirety of data, when I only looked at one verb in detail. It is clear that there must be more than just one abstract DO and PO, which is one thing that the existing constructionist analyses of the dative alternation across languages agree on. The constructions are themselves polysemous. Barðdal (2007) gives at least 17 subconstructions of the Icelandic DO, Colleman (2006) presents a detailed semantic map of the Dutch construction, Goldberg (1995) posits six different senses of the English DO, and Kay (2005) posits three maximal subconstructions and lets most of the entailments follow from the semantics of the verb alone. My approach is closer to Kay's than it is to Goldberg's, in the sense that the semantic contribution of the verb (or the opening of frames) plays a more important role than in Goldberg's analysis. But since I did not look at more than one verb systematically, it would not make much sense to generalize over all possible verbs and draw conclusions about the exact nature of the argument construction.

In Delilah, the following lexical class construction templates are realized. For the DO construction, verbs are licensed by the following lexical class construction:

(215) Caused-possession verb

```
template( caused-possession lemma, verb,
[ id:Top+ID,
synsem: [cat:vp, tense:untensed,
        eventvar:EV, predtype:nonerg,
        extth:Stheta~[Top+ID, A]],
```

```
sem: { {[SemS&(ID+ID1)#A, SemO&(ID+ID2)#B,
        SemI&(ID+ID3)#C,
        EStructure@some^E^and(quant(E, some),
        Main~[E], Etype~[E], entails1( E, incr),
        and( EStructure, entails(E, incr)))&
        (Top+ID)#EV],[],[]},
        Time@and(and(and(actor~[EV,A],
        theme~[EV,B]), recipient~[EV,C]),
        attime(EV, Time)) },
head: [phon: _X,
      synsem:[vtype:bi_trans, flex:infin,
             etype:Etype],
      sem: Main,
      concept:Main ],
arg(ID+ID1+10):[phon:_Subj,
               synsem:[theta:Stheta,
                      obj:subject_of(Top+ID)],
               sem:SemS],
arg(ID+ID2+1):[phon:_Obj,
              synsem:[obj:dirobject_of(Top+ID),
                     theta:Otheta, cat:np,
                     dir:left(1),
                     flag:0, case:obliq],
                sem:SemO],
arg(ID+ID3+2):[phon:_IObj,
              synsem:[obj:indirobject_of(Top+ID),
                     theta:Itheta, flag:0, \
                     dir:left(2),
                     cat:np, case:obliq],
                sem:SemI]
 ]  ).
```

This template licenses verbs of the type 'bi_trans' to take three arguments, subject (A), direct object (B), and indirect object (C), and assigns the three arguments the roles of actor, theme, and recipient.

The change-of-location/transfer lexeme is realized as follows:

(216) Change-of-location/transfer verb

```
template( change-of-location-transfer lemma,
          verb,
[ id:Top+ID,
synsem: [cat:vp, tense:untensed,
        eventvar:EV, predtype:nonerg,
        extth:Stheta~[Top+ID, A]],
sem: {{[SemS&(ID+ID1)#A, SemO&(ID+ID2)#B,
       SemI&(ID+ID3)#C,
       EStructure@some^E^and(quant(E, some),
       Main~[E], Etype~[E], entails1( E, incr),
       and( EStructure, entails(E, incr)))&
       (Top+ID)#EV],[],[]},
       Time@and(and(and(actor~[EV,A],
       theme~[EV,B]),goal~[EV,C]),
       attime(EV, Time)) },
head: [phon: _X,
      synsem: [ vtype:bi_trans, flex:infin,
               etype:Etype],
      sem: Main ],
arg(ID+ID1+10):[phon:_Subj,
      synsem:[ theta:Stheta,
               obj:subject_of(Top+ID)],
      sem:SemS],
arg(ID+ID2+1):[phon:_Obj,
      synsem:[dir:left(1),
               obj:dirobject_of(Top+ID),
               flag:0, cat:np, case:obliq,
               theta:Otheta],
      sem:SemO],
arg(ID+ID3+2):[phon:_IObj,
      synsem:[obj:prepobject_of(Top+ID),
               flag:6, dir:left(2), cat:pp,
               theta:Itheta],
      sem:SemI]
 ]  ).
```

This template licenses verbs of the type 'bi_trans' to take three arguments, subject (A), direct object (B), and a prepositional object (C), and assigns the three arguments the roles of actor, theme, and goal.

For the caused-possession PP lexeme, Delilah has the following lexical class construction template:

(217) Caused-possession PP verb

```
template( caused-possession-PP-lemma, verb,
[ id:Top+ID,
synsem: [cat:vp, tense:untensed,
        eventvar:EV, predtype:nonerg,
        extth:Stheta~[Top+ID, A]],
sem: {{[SemS&(ID+ID1)#A, SemO&(ID+ID2)#B,
       SemI&(ID+ID3)#C,
       EStructure@some^E^and(quant(E, some),
       Main~[E], Etype~[E], entails1( E, incr),
       and( EStructure, entails(E, incr)))&
       (Top+ID)#EV],[],[]},
       Time@and(and(and(actor~[EV,A],
       theme~[EV,B]),recipient~[EV,C]),
       attime(EV, Time)) },
head: [phon: _X,
      synsem: [ vtype:bi_trans, flex:infin,
              etype:Etype],
      sem: Main ],
arg(ID+ID1+10):[phon:_Subj,
      synsem:[ theta:Stheta,
               obj:subject_of(Top+ID)],
      sem:SemS],
arg(ID+ID2+1):[phon:_Obj,
      synsem:[dir:left(1),
              obj:dirobject_of(Top+ID),
              flag:0, cat:np, case:obliq,
              theta:Otheta],
      sem:SemO],
arg(ID+ID3+2):[phon:_IObj,
```

```
synsem:[obj:prepobject_of(Top+ID),
        flag:6, dir:left(2), cat:pp,
        theta:Itheta],
sem:SemI]   ]  ).
```

This template licenses verbs that have the same argument structure as the change-of-location/transfer lexemes, but that assign a recipient role to the PP, rather than a goal.

## 5.6 Conclusion

In this chapter, I have undertaken a slightly different journey than in the two preceding ones, in carrying out an empirical, contrastive study in order to get better insight into the question of lexicon organization for Dutch. But new insights have been gained into a topic that is is obviously not new and that has been discussed extensively. The empirical comparison of the dative alternations in English, German, and Dutch provided evidence for (and against) some theoretical work that has been done before.

To begin with, I have tried to back up the claim made by Levin & Rappaport Hovav (2005) that in English, *give* in itself does not alternate, even though it can occur in both constructions of the dative alternation. Levin & Rappaport Hovav (2005) predict that a language with overt case marking will not allow for an alternation with the respective verb for *give*, and for German that is borne out, if we neglect the few counter-examples for the moment.

To the results of Bresnan & Nikitina (2007), this study has added a valuable insight, too. Even though English seems flexible enough to allow for certain kinds of improbable constructions (like, e.g., alternations with reportedly non-alternating verbs), and to exhibit gradience rather than categorical constraints on lexical as well as grammatical level, my data provide support for the claim that the internal order of arguments is completely inflexible in both constructions, at least for English. This is interesting, as it differentiates the English from the Dutch dative

alternation in a very strict way. This distinction, namely a categorical versus a gradient constraint, in combination with the historical explanation for the rise of the PO construction as the alternative to the IO DO order in Middle English, leads me to conclude that Dutch and English are on two different stages in the process of abandoning a DO IO word order.

The fact that German does not exhibit a clear alternation is understandable from this perspective, too. Where there is overt marking of objects, there is no (or not much) ambiguity, and hence, there is no need for an alternative pattern. That, and the fact that there are so few examples with a PO pattern, would have been reason enough to explain the scarce examples away as mere interferences from the translator's side. Nevertheless, a close look at the examples has revealed that all of them are instances of complex grammatical structures, most of the time with very heavy PPs. Although I do not want to base strong claims on this very small set of real examples, my interpretation is that German does have a PO construction (as opposed to: it does not and those are translator's mistakes), and that its function is to disambiguate between the two objects by overtly marking the recipient. This mechanism, however, seems to take place only in cases where the structure is so complex that the sentence would be close to uninterpretable otherwise. In that sense, the German PO construction can be seen as a kind of 'last resort' construction.

For questions of lexicon organization, the conclusion of this case study must be the following: all three languages only have one, monosemous, verb *give*, *geven*, and *geben*. Also, all three languages have a double object lexical class construction and a PO lexical class construction that can license the verb in question.

In this chapter, the emphasis has been more on the way lexical items are stored in the lexicon than on the constructions that eventually license expressions. In a way, this is also the more intriguing question, since neither the DO nor the PO construction exhibit odd syntactic or semantic restrictions for Dutch. Both constructions are free with respect to the internal order of the arguments, and the constructional semantics are less remarkable

than, e.g., the semantics of the *weg*-construction. Moreover, I do not want to go so far as to make claims about the entirety of DO and PO constructs, as not all of them behave the same way. Nevertheless, the results of this study suggest important follow-up work. Particularly the empirical part contributes an important piece to the ongoing puzzle of the dative alternation, since for the first time, there is reliable contrastive, empirical data.

Also, the formal discussion sheds some more light on the important role that verbs play in this proposed model. Since verbs, as all lexical types, are organized hierarchically, and concrete verbs belong to more abstract lexeme classes, we can utilize this organization in order to account for alternations. Sag (2007b) presented an approach for the supposedly semantically identical *spray/load*-alternants, which I expanded to account for the semantically related, though not identical, dative alternants. The resulting structures for lexical items and lexical classes capture the relations between DO and PO systematically.

# Chapter 6

# Conclusion and Outlook

This dissertation presented research on three different constructions in Dutch. Although it was planned to be a mainly theoretical thesis, it eventually turned out to be more data driven than was expected. The reason is quite simple: in order to gain new insights about phenomena, it seems natural to take a closer look at them.

But I did not abandon theoretical issues completely. The framework that seemed most suitable to investigate the constructions I chose is that of Construction Grammar, since this family of approaches outspokenly deals with all kinds of uncommon grammatical phenomena. And not only for a grammarian it is desirable to be able to describe the core constructions of a language as well as the more peripheral patterns. Also for computational linguistics, particularly grammar analyzing systems, being able to deal with the quirky things is a crucial factor to reach wide coverage. Therefore, taking a closer look at Construction Grammar was a natural thing, since this approach aims at a generalized treatment of all phenomena, as idiosyncratic as they might be.

The problem that arose was the observation that there seems to be a lot of unclarity about the actual architecture of a Construction Grammar system, since there has not been much research on formal Construction Grammar, yet. There are two groups that do extremely interesting work on embodied systems,

namely the Embodied Construction Grammar (Bergen & Chang (2005)), and Fluid Construction Grammar (Steels (2004)). Both research lines yield valuable results, the former on working out the relationship between the semantic aspects of constructions and the embodiment of mental representations, the latter on simulations of the emergence and development of language in embodied agents.

One formal system for Dutch that aims at wide coverage and uses certain principles that are pillars of Construction Grammar is the Alpino Parser (Bouma *et al.* (2001)). Alpino is an HPSG grammar, but a variety of rules and principles were discarded in favour of a type hierarchy. The second system for Dutch that aims explicitly at the implementation of Extended Lexical Units is Delilah (Cremers (2004)), a semantic parser/generator. Delilah is still far from being a wide-coverage application, and since it is a purely knowledge-based system, it will not succeed in expanding its coverage significantly, but eventually, Delilah was designed to allow for detailed analytics rather than wide coverage (this could be compared to, e.g., the SUSANNE treebank,[1] which is extremely small, but has a great depth of analysis). Delilah fruitfully served as the test playground during this research project.

In a nutshell, there is a lot of work being done on the field of computational Construction Grammar, although there remains to be a lack of shared assumptions and techniques. This might mainly be due to the problem that generally, research projects in the field of computational linguistics often work towards a practical goal, and in many cases, linguistic analyses are subservient to the higher goal. However, in the future the different flavours of computational Construction Grammars will likely become more streamlined.

The same holds for the research on linguistic theory. The theoretical field of Construction Grammar still seems a bit scattered, so that it is often target of the criticism, it was not much more than a pool for linguists working on unusual phenomena. But it does not need to be like this. There are a number of tenets in constructionist frameworks that have been formulated clearly

[1] See Sampson (1995).

(see for instance Goldberg (2003)), and researchers should have these tenets in mind, if they formulate their analyses.

There is at least one formal Construction Grammar approach that is being worked on lately, and that is Sign-Based Construction Grammar. Being a successor of Head-Driven Phrase Structure Grammar that tries to implement constructionist ideas, SBCG is interesting for researchers of HPSG as well as the less formal Construction Grammar approaches. And we can see movement towards SBCG from both directions (for examples of construction-style HPSG, see for instance Bilbie & Laurens (2009) and Chaves (2009), for examples of SBCG-style Construction Grammar, see Michaelis (To Appear) and Barðdal & Eythórsson (To Appear 2010), and generally the work presented in Boas & Sag (To Appear 2010)).

One point of discussion that seems to divide the field of construction grammarians is grounded in the question whether a phrasal or a lexical analysis is preferred, when both are applicable. I argue in favor of lexical analyses for cases like idioms of the *spill the beans*-type, but also for what traditionally is considered a construction, namely the ditransitive. Also for the Dutch version of the *way*-construction, I argue for a lexical approach. *Weg*-construction verbs that license expressions like *hij graaft zich een weg uit het gevangenis* ('he digs his way out of prison') are derived from intransitive verbs by means of a derivational construction.

In the more practical part, I hope to have presented some new insights for all the three constructions I analyzed. For the NCoN and NPN constructions, I argued for a lexicon-based analysis with specific lexical entries for the expressions of the type NCoN, and for a set of phrasal constructions for expressions of the type NPN. But as a corpus study revealed, the members of the NPN group show remarkable differences in their combinatorics as well as their semantics. Therefore, I posited six different subtypes of NPNs, all of which equipped with their own particular set of constraints.

The *weg*-construction, I analyzed as a lexical class construction, which deviates from the recent constructionist analyses

which tend to be phrasal. Another novel point about the analysis is that it is shown that in Dutch, it is not appropriate to categorize the *weg*-construction as a subtype of the Caused-Motion construction, as it has been done for English by Goldberg (1995). Another novelty about the analysis in this dissertation is that the semantic pole of the construction has completely been reconsidered. In the tradition of *weg*-construction analyses, the usual description of the meaning of the construction was something in the realm of "creating and travelling a path by means of V". I provide arguments which reveal that there is sufficient reason to doubt that the *way*-element is part of the semantic structure of the Dutch construction. In a direct comparison with a related pattern, the Transition-to-Location construction, I argue that the difference between the two constructions is not a missing versus a present *way*-element, but rather an aspectual issue.

For the third construction I chose, I carried out a major corpus study. For German, English, and Dutch, I extracted all instances of the verb *give* and its cognates from the Europarl Corpus, in order to see what the numeric distribution of double-object and prepositional construction is. This work was not only time consuming, it also lead to surprising results. It appears that the distribution in German is at an expected 99:1 ratio, English shows a solid 2/3:1/3 ratio, and Dutch, which came quite unexpected, has a 50:50 distribution. I conclude from this balanced result that semantics cannot be a dominant factor that drives the choice of one construction over the other. In comparison with English and German, I argue that the lack of morphological case marking facilitates a more constrained use of the two alternants, and that this is a process that takes place over a long stretch in time. Since Dutch has lost its morphological case marking later than English did, the patterns are still more flexible than they are in English. Furthermore, I analyze the two alternants as being lexical-class constructions, which are licensed by according lexical entries.

There are a lot of possible further topics around that analysis that I can think of. First of all, to get a clearer picture of the factors that drive the alternation in Dutch will be to analyze the corpus data with respect to their word order, heaviness, pronom-

inality, and other factors that are known to be relevant for this phenomenon. The first step is made by compiling the corpora, now it is a question of more fine-grained annotation and further research to be able to come to more detailed conclusions. The data are of course available from the author.

The distinction between lexical and phrasal constructions does not need to lead to so much differences in the assumptions that underlie a grammatical theory. Although this is a highly debated topic, it seems that a lot of insights point towards the existence of both. Nevertheless, it still needs to be determined, which processes are lexical in nature, and which are phrasal. SBCG is a proper framework that assumes a lexical approach to begin with, but also allows for phrasal constructions where they are appropriate.

There is one more issue that I want to raise at the end of the discussion. Chomskyan generative approaches are often criticized from more functional parties for their anglocentricity. A quote by Goddard on Universal Pragmatics, e.g., says the following:

"[...] it is increasingly evident that these avowedly universalist models are [a]nglocentric, in the sense that they adopt some aspect of Anglo norms or practices as a baseline or template, and then attempt to generalize or adjust this to suit all other cultural settings." (Goddard 2006)

In the case of the Dutch *weg*-construction I have shown that an analysis of the Dutch data that is based on the analysis of the English data is not appropriate. Also, the distributions of the English, Dutch and German dative alternations are so extremely different that it is hard to believe that comparable factors influence the choice of one alternant over the other. Although it is a difficult thing to do, linguists should pursue their investigations independently of results that have been proposed for other languages, since one language system always poses very different constraints on structures than another one does.

Eventually, we will not succeed in building a model that re-

sembles cognitive reality. First of all, although massive research is done in linguistics as well as related fields like psychology and neurophysiology, even with a huge apparatus of experimental setups, we will only ever be able to test whether a hypothesis we have formed in advance turns out to be true or false. Whether the hypothesis is true will remain unanswered. The same holds for computational models of language. If a system is consistent and working, all we have shown is that we have succeeded in writing a consistent model of our assumptions. That does not mean that linguists should stop searching for consistent language models. But whether the result is an equivalent of how language works, is easy to answer. It is not.

# Appendix

## NPNs in the CGN

**Universal quantification:**
alinea per alinea
bladje per bladje
bladzij voor bladzij
centimeter per centimeter
centimeter voor centimeter
graadje voor graadje
kaart voor kaart
kamer per kamer
knijper voor knijper
kolom voor kolom
komma voor komma
laagje voor laagje
letter voor letter
noot voor noot
oester per oester
punt voor punt
regio per regio
stapje per stapje
stapke voor stapke
steen voor steen
stuk voor stuk
vezel voor vezel

**Temporal succession:**
angina op angina
avond aan avond
avond na avond
broek na broek
corner op corner
dag aan dag
dag na dag
dag na dag na dag
dag op dag
frustratie op frustratie
generatie na generatie
keer op keer
koorts op koorts
maand na maand
onderwerp voor onderwerp
pagina na pagina
preventie op preventie
project voor project
samenkomst na samenkomst
slag na slag
stap voor stap
stoot na stoot
TV over TV
vetlaag na vetlaag
week aan week
week na week
zegening na zegening

van dag tot dag
van minuut tot minuut
van moment(je) tot moment(je)
van roes naar roes
van tijd tot tijd

**Spatial succession:**
van café naar café
van camping tot camping
van club naar club
van deur tot deur
van dorp tot dorp
van jeugdherberg naar jeugdherberg
van kamer naar kamer
van kamer tot kamer
van kerk naar kerk
van kroeg naar kroeg
van laag tot laag
van land tot land
van minnehof tot minnehof
van schilderij tot schilderij
van school naar school
van schrijn tot schrijn
van sector tot sector
van stad tot stad
van streek tot streek
van tent naar tent
van tree tot tree
van waterval naar waterval
van zaal tot zaal

**Juxtaposition:**
arm in arm
camping naast camping
hand in hand
laag over laag
mond aan mond (reclame)
neus aan neus
rug aan rug
schouder aan schouder
zij aan zij

**Spatial extension:**
van deur tot deur
van dijk tot dijk
van oever tot oever
(van top tot teen)
(van vloer tot plafond)
van winkel tot winkel

**Transition:**
van gemeenschap tot gemeenschap
van(generatie op generatie)
van hand tot hand
van leerling naar leraar
van man tot man
van mond tot mond
van oor tot oor
(van vader op zoon)
(van verkoper naar koper)
van vrouw tot vrouw
(van zoon op kleinzoon)

# Bibliography


Altenberg, Bengt. 1998. On the phraseology of spoken English: The evidence of recurrent word-combinations. In *Phraseology. Theory, Analysis, and Applications*, ed. by A.P. Cowie, 101–22. Oxford:Clarendon Press.

Arsenijević, Boban. 2006. *Inner Aspect and Telicity. The decompositional and the quantificational nature of eventualities at the syntax-semantics interface.*. Leiden University dissertation.

Baker, Mark C. 1988. *Incorporation: A Theory of Grammatical Function Changing*. Chicago:University Press.

Barðdal, Johanna. 2007. The semantic and lexical range of the ditransitive construction in the history of (north) Germanic. *Functions of Language* 14, 9–30.

Barðdal, Jóhanna, & Thórhallur Eythórsson. To Appear 2010. Reconstructing syntax: Construction grammar and the comparative method. In *Sign-Based Construction Grammar*, ed. by Hans C. Boas & Ivan A. Sag. Stanford:CSLI Publications.

Beavers, John. 2005. Towards a semantic analysis of argument/oblique alternations in HPSG. In *Proceedings of the 12th International Conference on Head-Driven Phrase Structure Grammar*, ed. by Stefan Müller, 28–48.

Beck, Sigrid, & Arnim von Stechow. 2006. Dog after dog revisited. In *Proceedings of the Sinn und Bedeutung 10*, ed. by

Christian Ebert & Cornelia Endriss, volume 44 of *ZAS Papers in Linguistics*, 43–54.

Bergen, Benjamin K., & Nancy Chang. 2005. Embodied Construction Grammar in simulation-based language understanding. In *Construction Grammar(s): Cognitive and Cross-Language Dimensions (Constructional Approaches to Language 3)*, ed. by J.-O. Östman & M. Fried, 147–190. Amsterdam:John Benjamins.

Bilbie, Gabriela, & Frederic Laurens. 2009. A construction-based analysis of verbless relative adjuncts in French and Romanian. In *Proceedings of the 16th International Conference on Head-Driven Phrase Structure Grammar, University of Göttingen, Germany*, ed. by Stefan Müller, 5–25. Stanford:CSLI Publications.

Boas, Hans, & Ivan A. Sag (eds.). To Appear 2010. *Sign-Based Construction Grammar*. Stanford:CSLI Publications.

Bouma, Gosse, Robert Malouf, & Gertjan van Noord. 2001. Alpino: Wide coverage computational analysis of Dutch. In *Computational Linguistics in the Netherlands CLIN 2000*, 45–59.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina, & Harald Baayen. 2007. Predicting the dative alternation. In *Cognitive Foundations of Interpretation*, ed. by Gerlof Bouma, Irene Kraemer, & Joost Zwarts, 69–94. Amsterdam:Royal Netherlands Academy of Science.

——, & Tatiana Nikitina. 2003. The Gradience of the Dative Alternation. Manuscript, Stanford University.

——, & ——. 2007. The gradience of the dative alternation. In *Reality Exploration and Discovery: Pattern Interaction in Language and Life*, ed. by Linda Uyechi & Lian Hee Wee. Stanford:CSLI Publications.

Bybee, Joan. 2005. The impact of use on representation: grammar is usage and usage is grammar. Presidential address LSA, Oakland, January 8, 2005.

Casenhiser, Devin, & Adele E. Goldberg. 2005. Fast mapping of a phrasal form and meaning. *Developmental Science* 8, 500–508.

Chaves, Rui P. 2009. Construction-based cumulation and adjunct extraction. In *Proceedings of the 16th International Conference on Head-Driven Phrase Structure Grammar, University of Göttingen, Germany*, ed. by Stefan Müller, 47–67. Stanford:CSLI Publications.

Colleman, Timothy. 2006. *De Nederlandse datiefalternantie: een constructioneel en corpusgebaseerd onderzoek*. Ghent University dissertation.

Cremers, Crit. 2004. ('n) betekenis berekend. *Nederlandse Taalkunde (7)* 4, 375–395.

——, & Hilke Reckman. 2008. Exploiting logical forms. In *Computational Linguistics in the Netherlands 2007*, ed. by P.A. Coppen S. Verberne, H. van Halteren, 5–20. Utrecht:LOT.

Croft, William. 2001. *Radical Construction Grammar. Syntactic theory in typological perspective.* Oxford:University Press.

——. 2004. Logical and typological arguments for Radical Construction Grammar. In *Construction Grammar(s): Cognitive and cross-language dimensions (Constructional Approaches to Language 3)*, ed. by J.-O. Östman & M. Fried, 273–314. Amsterdam:John Bejamins.

——, & D. Alan Cruse. 2004. *Cognitive Lingusitics*. Cambridge:University Press.

Culicover, Peter W. 1999. *Syntactic Nuts: Hard Cases, Syntactic Theory, and Language Acquisition*. Oxford:University Press.

——, & Ray Jackendoff. 2005. *Simpler Syntax*. Oxford:University Press.

de Swart, Henriëtte, Yoad Winter, & Joost Zwarts. 2005. Bare predicate nominals in Dutch. In *Proceedings of Sinn und Bedeutung 9*, 446–460.

——, & Joost Zwarts. n.d. Less form – more meaning: why bare singular nouns are special. submitted.

Egmond, Marie-Elaine Van. 2006. Two way-constructions in Dutch: motion along a path and transition to a location. Master's thesis, University of Canterbury.

Erman, Britt, & Beatrice Warren. 2000. The idiom principle and the open choice principle. *Text, an interdisciplinary journal for the study of discourse* 20, 29–62.

Erteschik-Shir, Nomi. 1979. Discourse constraints on dative movement. *Syntax and Semantics 12: Discourse and Syntax* 441–467.

Fillmore, Charles J., Paul Kay, & Mary C. O'Connor. 1988. Regularity and idiomaticity in grammatical constructions: the case of *let alone*. *Language* 64, 501–39.

Flickinger, Daniel. 1987. *Lexical Rules in the Hierarchical Lexicon*. Stanford University dissertation.

Ghomeshi, Jila. 1996. *Projection and Inflection: A Study of Persian Phrase Structure*. University of Toronto dissertation.

Ginzburg, Jonathan, & Ivan A. Sag. 2000. *Interrogative Investigations: the Form, Meaning and Use of English Interrogatives*. Number 123 in CSLI Lecture Notes. Stanford:CSLI Publications.

Goddard, Cliff. 2006. Ethnopragmatics: a new paradigm. In *Ehtnopragmatics. Understanding Discourse in Cultural Context*, ed. by Cliff Goddard, 1–30. Berlin:Mouton de Gruyter.

Goldberg, Adele E. 1995. *Constructions. A Construction Grammar Approach to Argument Structure.* Chicago:University Press.

—— 1996. Making one's way through the data. In *Grammatical Constructions*, ed. by Masayoshi Shibatani & Sandra A. Thompson, 29–54. Oxford:University press.

—— 2003. Constructions: a new theoretical approach to language. *Trends in Cognitive Sciences* 7, 219–224.

—— 2006a. *Constructions at Work. The Nature of Generalization in Language.* Oxford:University Press.

—— 2006b. Learning linguistic patterns. In *Categories in Use. Psychology of Learning and Motivation*, ed. by A. Markman & B. Ross, volume 47, 33–65. Academic Press/Elsevier.

——, Devin Casenhiser, & Nitya Sethuraman. 2004. The role of prediction in construction learning. *Journal of Child Language* 2, 407–426.

——, & Ray Jackendoff. 2004. The English resultative as a family of constructions. *Language* 80, 532–568.

Gries, Stefan Th., & Anatol Stefanowitsch. 2004. Extending collostructional analysis. a corpus-based perspective on alternations. *International Journal of Corpus Linguistics* 9, 97–129.

Gropen, Jess, Steven Pinker, Michelle Hollander, & Richard Goldberg. 1989. The learnability and acquisition of the dative alternation in English. *Language* 65, 203–257.

Haeseryn, Walter, K. Romijn, G. Geerts, J. de Rooij, & M.C. van den Toorn. 1997. *Algemene Nederlandse*

*Spraakkunst*. Groningen/Deurne:Martinus Nijhoff uitgevers/ Wolters Plantyn, tweede, geheel herziene druk edition.

Hale, Ken, & Jay Keyser. 1996. On the Double Object Construction. manuscript, MIT.

Hameyer, K. 1979. English and German constraints on dative movement. *Word* 30, 235–244.

Huddleston, Rodney, & Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge:University Press.

Jackendoff, Ray. 1990. *Semantic Structures*. Cambridge, MA:MIT Press.

——. 2008. Construction after construction and its theoretical challenges. *Language* 84, 8–28.

——. n.d. Construction after construction. unpublished manuscript.

Kay, Paul. 1995. What is construction grammar. Manuscript, U.C. Berkeley.

——. 2002. An informal sketch of a formal architecture for construction grammar. *Grammar* 5, 1–19.

——. 2005. Argument structure constructions and the argument-adjunct distinction. In *Grammatical Constructions: Back to the Roots*, ed. by M. Fried & H. Boas, 71–98. Amsterdam:Benjamins.

——, & Charles J. Fillmore. 1999. Grammatical constructions and linguistic generalizations: the what's x doing y? construction. *Language* 1–33.

——, & Ivan A. Sag. 2009. How hard a problem would this be to solve? In *Proceedings of the 16th International Conference on Head-Driven Phrase Structure Grammar, University of Göttingen*, 171–191. Stanford:CSLI Publications.

Koehn, Philipp. 2002. Europarl: A multilingual corpus for evaluation of machine translation. Draft.

Köpcke, Klaus-Michael. 2003. Grammatische Komplexität und die Beherrschung der Kasusmorphologie durch Grundschulkinder. *Didaktik Deutsch* 55–68.

——. 2005. "die Prinzessin küsst den Prinz" – Fehler oder gelebter Sprachwandel? *Didaktik Deutsch* 67–83.

Krifka, Manfred. 1999. Manner in the dative alternation. In *WCCFL*, volume 18, 260–271.

Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*. Chicago:University Press.

——, & Mark Johnson. 1980. *Metaphors we live by*. Chicago:University Press.

Lambrecht, Knut. 1994. *Information Structure and Sentence Form*. London:CUP.

Landsbergen, Frank. 2006. Krijgen, kriegen en get: een vergelijkend onderzoek naar betekenisverandering en grammaticalisatie. In *Nederlands tussen Duits en Engels*, ed. by T. van der Wouden M. Hüning, U. Vogl & A. Verhagen, 259–272. Leiden:SNL.

Langacker, Ronald W. 1987. *Foundations of cognitive grammar: Theoretical Prerequisites*. Stanford:University Press. Vol 1.

—— 1991. *Foundations of cognitive grammar: Descriptive Applications*. Stanford:University Press. Vol 2.

—— 2003. Construction grammars. cognitive, radical and less so. Plenary Paper ICLC 8, Logroño, Spain.

Larson, Richard K. 1988. On the double object construction. *Linguistic Inquiry* 335–391.

Levin, Beth. 2006. First objects and datives: Two of a kind? Handout, Parasession on Theoretical Approaches to Argument Structure, 32nd Annual Meeting of the Berkeley Linguistics Society, University of California, Berkeley.

——, & Malka Rappaport Hovav. 1995. *Unaccusativity: at the syntax-lexical semantics interface*, volume 26 of *Linguistics Inquiry Monographs*. Cambridge, MA:MIT Press.

——, & Malka Rappaport Hovav. 2005. All dative verbs are not created equal. manuscript, Stanford University and The Hebrew University of Jerusalem.

Marantz, Alec. 1992. The way-construction and the semantics of direct arguments in English: a reply to Jackendoff. In *Syntax and the Lexicon*, ed. by Tim Stowell & Eric Wehrli, 179–188. San Diego:Academic Press.

McFadden, Thomas. 2002. The rise of the to-dative in Middle English. In *Syntactic effects of morphological change*, ed. by D.W. Lightfoot. Oxford:University Press.

Meurers, Walt Detmar. 1997. Using lexical principles in HPSG to generalize over valence properties. In *Proceedings of the Third Conference on Formal Grammar. Aix-en-Provence, France*, 137–146.

Michaelis, Laura A. 2009. Sign-Based Construction Grammar. In *The Oxford Handbook of Linguistic Analysis*, ed. by B. Heine & H. Narrog, 155–176. Oxford:Oxford University Press.

—— To Appear. Complementation by construction. In *The Proceedings of the 32. Annual Meeting of the Berkeley Linguistics Society: Parasession on Theoretical Approaches to Argument Structure*. Berkeley:BLS, Inc.

Müller, Stefan. 2006a. Phrasal or lexical constructions. *Language* 82, 850–883.

——. 2006b. Resultativkonstruktionen, Partikelverben und syntaktische vs. lexikonbasierte Konstruktionen. In *Konstruktionsgrammatik. Von der Anwendung zur Theorie*, ed. by Kerstin Fischer & Anatol Stefanowitsch, number 40 in Stauffenburg Linguistik, 177–202. Tübingen:Stauffenburg Verlag.

——. 2007. Phrasal or lexical constructions? some comments on underspecification of constituent order, compositionality, and control. In *Proceedings of the HPSG-2007 Conference, Stanford University*, ed. by Stefan Müller, 373–393.

——. To Appear 2010. Persian complex predicates and the limits of inheritance-based analyses. *Journal of Linguistics* 46.

Oehrle, Richard. 1976. *The Grammatical Status of the English Dative Alternation*. MIT dissertation.

Oostdijk, Nelleke. 2000. The Spoken Dutch Corpus. Overview and first evaluation. In *Proceedings of the second International Conference on Language Resources and Evaluation*, ed. by M. Gavralidou, G. Carayannis, S. Markantonatou, S. Piperidis, & G. Stainhaouer, 887–893. Paris:ELRA.

Paardekooper, Petrus C. 1956. Een schat van een kind. *De Nieuwe Taalgids* 49, 93–99.

Pollard, Carl, & Ivan A. Sag. 1994. *Head-driven Phrase Structure Grammar*. Chicago:University Press.

Poss, Michaela. 2005. Towards an implementation of constructions. In *Proceedings of the Sixth Annual High Desert Linguistics Society Conference*, 71–82.

——, & Ton van der Wouden. 2005. Extended lexical units in dutch. In *Computational Linguistics in the Netherlands 2004: Selected papers from the fifteenth CLIN meeting*, ed. by Hilke Reckman Ton van der Wouden, Michaela Poß & Crit Cremers, 187–202. Utrecht: LOT.

Postma, Gertjan. 1996. *Zero Semantics. A Study of the Syntactic Conception of Quantificational Meaning*. The Hague:Holland Institute of Generative Linguistics.

Radden, Günter. 2006. The metaphor time as space across languages. In *Metonymy—Metaphor Collage*, ed. by Elżbieta Gńska & Günter Radden, 99–120. Warsaw:University Press.

Rappaport Hovav, Malka, & Beth Levin. 2006. The English dative alternation: The case for verb sensitivity. manuscript, the Hebrew University of Jerusalem and Stanford University.

Reckman, Hilke. 2009. *Flat but not Shallow: Towards Flatter Representations in Deep Semantic Parsing for Precise and Feasible Inferencing*. Utrecht:LOT Publications.

Sag, Ivan A. 2007a. Remarks on locality. In *Proceedings of the HPSG07 Conference, Stanford University*, ed. by Stefan Müller, 394–414. Stanford:CSLI Publications.

—— 2007b. Sign-Based Construction Grammar: An Informal Synopsis. manuscript, Stanford University.

—— 2008. Feature geometry and predictions of locality. In *Proceedings of the Workshop on Features*, ed. by Greville Corbett & Anna Kibort. Oxford:University Press.

—— to appear 2010. Sign-Based Construction Grammar: An Informal Synopsis. In *Sign-Based Construction Grammar*, ed. by H. Boas & I.A. Sag. Stanford:CSLI Publications.

——, Timothy Baldwin, Francis Bond, Ann Copestake, & Dan Flickinger. 2001. Multiword expressions: A pain in the neck for NLP. LinGO Working Paper No. 2001-03 (CSLI Linguistic Grammars Online (LinGO) Lab, Stanford University); also in Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics (CICLING 2002), Mexico City, Mexico, pp. 1-15.

——, Thomas Wasow, & Emily Bender. 2003. *Syntactic Theory: A formal introduction*. Stanford:CSLI, 2 edition.

Sailer, Manfred. 2000. *Combinatorial Semantics and Idiomatic Expressions in Head-Driven Phrase Structure Grammar*. Eberhard-Karls-Universität Tübingen dissertation.

Sampson, Geoffrey. 1995. *English for the Computer. The SUSANNE Corpus and Analytic Scheme*. Clarendon Press.

Schermer-Vermeer, Everdina C. 1991. *Substantiële versus formele taalbeschrijving. Het indirect object in het Nederlands*. Universiteit van Amsterdam dissertation.

Soehn, Jan-Philipp. 2004. License to COLL. In *Proceedings of the HPSG-2004 Conference, Center for Computational Linguistics, Katholieke Universiteit Leuven*, ed. by Stefan Müller, 261–273. Stanford:CSLI Publications.

Sprenger, Simone A. 2003. *Fixed expressions and the production of idioms*. Nijmegen dissertation.

Steels, Luc. 2004. Constructivist development of grounded construction grammars. In *Proceedings Annual Meeting of Association for Computational Linguistics*, ed. by W. Daelemans, 9–16.

Thompson, Sandra. 1995. The iconicity of the "dative shift" in English: Considerations from information flow in discourse. In *Syntactic Iconicity and Linguistic Freezes*, ed. by Marge E. Landsberg, 155–175. Berlin:Mouton de Gruyter.

Tomasello, Michael. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard:University Press.

——. 2006. Acquiring linguistic constructions. In *Handbook of Child Psychology*, ed. by D. Kuhn & R. Siegler. New York:Wiley.

Vallduví, Enric. 1992. *The Informational Component*. New York:Garland Press.

Van der Beek, Leonoor. 2004. Argument order alternations in Dutch. In *Proceedings of the LFG04 Conference*, ed. by Miriam Butt & Tracy Holloway King, 39–59.

van der Wouden, Ton. 1997. *Negative Contexts: Collocation, Polarity and Multiple Negation*. London and New York:Routledge.

——. 2001. Collocational behaviour in non content words. In *COLLOCATION: Computational Extraction, Analysis and Exploitation. Proceedings of a Workshop during the 39th Annual Meeting of the Association for Computational Linguistics and the 10th Conference of the European Chapter, Toulouse, France, July 7th*, ed. by Béatrice Daille & Geoffrey Williams, 16–23. CNRS – Institut de Recherche en Informatique de Toulouse, and Université de Sciences Sociales.

——. 2005. On the phraseology of stop words. In *Proceedings of Phraseology 2005: The many faces of Phraseology*, ed. by F. Meunier C. Cosme, C. Gouverneur & M. Paquot, 417–420.

——. 2007. On the phraseology of stop words. *Leiden Papers in Linguistics* 4, 56–67.

——. to appear 2010. *Dutch as a Construction Language. Grammatical aspects of spoken Dutch*. Utrecht:LOT.

Vendler, Zeno. 1957. Verbs and times. *The Philosophical Review* 66, 143–160.

Verhagen, Arie. 2002. From parts to whole and back again. *Cognitive Linguistics* 13, 403–439.

——. 2003. The Dutch *way*. In *Usage-Based Approaches to Dutch. Lexicon, Grammar and Discourse*, ed. by Arie Verhagen & Jeroen van de Weijer, 27–57. Utrecht: Utrecht:LOT.

——. 2007. English constructions from a Dutch perspective. where are the differences? In *Structural-Functional Studies in English grammar. In honour of Lachlan Mackenzie*, ed. by Mike Hannay & Gerard J. Steen, 257–274. Amsterdam:John Benjamins.

Verkuyl, Henk J. 1989. Aspectual classes and aspectual composition. *Linguistics and Philosophy* 12, 39–94.

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

Ward, Gregory, Betty J. Birner, & Rodney Huddleston. 2002. Information packaging. In *The Cambridge Grammar of the English language*, ed. by R. Huddleston & G. Pullum, 1363–1447. Cambridge:University Press.

Wasow, Thomas. 1997. Remarks on grammatical weight. *Language Variation and Change* 81–105.

Wellander, Erik. 1920. Over den datief als subject van een passieve constructie. *De Nieuwe Taalgids* 14, 291–299.

WNT. 1864-1998. *Woordenboek der Nederlandsche Taal.* 's Gravenhage:Martinus Nijhoff. Digital version (1995): Het Woordenboek der Nederlandsche Taal op CD-ROM. Rotterdam: AND.

# Samenvatting in het Nederlands

Natuurlijke taal gedraagt zich helaas niet altijd zo netjes als grammatici zouden willen. Overal maken onregelmatigheden het leven van theoretische en computationele taalkundigen zwaar. Dit proefschrift levert een bijdrage aan het onderzoek van de zogenaamde "Extended Lexical Units", uitgebreide lexicale eenheden, aan de hand van drie case studies, en de resultaten van de theoretische uiteenzetting zijn in de semantische ontleedmachine "Delilah" geïmporteerd.

De theoretische basis van het onderzoek is een formele versie van de grote groep van constructiegrammatica's, namelijk de "Sign-Based Construction Grammar", de tekengebaseerde constructiegrammatica. Als een derivaat van "Head-Driven Phrase Structure Grammar", de hoofdgestuurde constituentenstructuurgrammatica, brengt SBCG een formaliteit met zich mee die deze aanpak bijzonder geschikt maakt voor een computerimplementatie. Als een versie van constructiegrammatica houdt SBCG rekening met inzichten uit de cognitive disciplines die van belang zijn voor het bouwen van een model van natuurlijke taal.

Hoofdstuk 2, ELUs en grammatica, gaat in op een aantal vragen omtrent het verschijnsel "constructie". De thema's die ik behandel zijn onder andere wat een constructie nou eigenlijk is, hoe semantiek en syntaxis hand in hand gaan, en of constructies lexicaal of frasaal van aard zijn. De meerderheid van constructionele aanpakken geeft een frasale analyse voor zo bekende verschijnselen als de *weg*-constructie of de datiefalternantie, maar ik argumenteer voor een lexicale aanpak waarbij de constructie

opereert op werkwoordsniveau en niet op zinsniveau. En toch verschilt de hier gekozen aanpak duidelijk van bv. HPSG door de mogelijkheid onderscheid te maken tussen frasale en lexicale constructies.

In hoofdstuk 3, NCoN en NPN, kom ik ook meteen tot een voorbeeld van frasale constructies. De twee besproken verschijnselen hebben de vorm naamwoord-conjunctie-naamwoord (zoals in *kind noch kraai*) en naamwoord-voorzetsel-naamwoord (zoals in *bumper aan bumper*), waarbij in beide gevallen het naamwoord enkelvoud, telbaar en zonder lidwoord moet zijn. Zij verschillen in hun flexibiliteit: de NCoNs zijn allemaal vaste uitdrukkingen, terwijl de NPN-klasse productief is en ad hoc nieuwe uitdrukkingen kan vormen. Omdat ik met behulp van een corpusstudie verschillende soorten NPNs geclassificeerd heb die met betrekking tot hun betekenis van elkaar verschillen, presenteer ik een netwerk van lexicale-klasseconstructies die nieuwe uitdrukkingen licenseren. Voor de NCoNs ligt het een beetje anders. Hier beschrijf ik een netwerk van types die allemaal niet meer productief zijn maar een set van vaste lexicale eenheden licenseren, uitdrukkingen die net zoals eenvoudige woorden in het lexicon zijn opgeslagen. De formele en semantische overeenkomsten van de verschillende uitdrukkingen zijn geformuleerd in een overkoepelend supertype, en abstracte lexicale klasse, die alle gemeenschappelijke kenmerken van zijn leden beschrijft, waardoor deze informatie niet verloren gaat. Tenslotte vergelijk ik mijn analyse met een eerdere Zero Semantics analyse van dezelfde verschijnselen. Hierbij kom ik tot de conclusie dat een constructionele aanpak de grammaticale feiten beter kan weergeven.

Hoofdstuk 4 houdt zich bezig met een constructie die vaak is besproken in de bestaande literatuur, namelijk de *weg*-constructie. Deze constructie manifesteert zich in zinnen zoals *hij wurmt zich een weg naar buiten*. De hier beschreven analyse verschilt op een aantal punten van de eerdere aanpakken: ten eerste geef ik argumenten voor een frasale structuur die de voorzetselconstituent niet een oblique argument, maar een modificerend element van de *weg*-NP maakt. Verder, en dat is zeker

het meest afwijkende punt, verlaat ik de oude weg waar elk element in de constituentenstructuur een corresponderend element in de semantische structuur van de constructie moet krijgen toegewezen. Ik presenteer een analyse van de *weg*-constructie waarbij er geen *weg*-element in de semantische representatie aanwezig is. Om deze argumentatie te onderbouwen, vergelijk ik de *weg*-constructie met de overgang-naar-locatie constructie (zoals in *hij springt zich de finale in*). Die verschilt alleen maar van de *weg*-constructie door het ontbreken van de *weg*-NP aan de formele kant, en door een ander werkwoordelijk aspect aan de semantische kant. Allebei de constructies krijgen dezelfde basisbetekenis, namelijk "overgang naar locatie door werkwoord-en", waarbij de *weg*-constructie is ondergespecificeerd voor teliciteit en noodzakelijk iteratief. De TLC constructie is ondergespecificeerd voor iterativiteit en noodzakelijk telisch.

Hoofdstuk 5, de datiefalternantie, verschilt van zijn voorgangers door het contrastieve perspectief dat wordt ingenomen. Bovendien voer ik een kwantitatieve corpusstudie door. Het verschijnsel dat er behandeld wordt is de zogenaamde datiefalternantie, de relatieve uitwisselbaarheid van zinnen als *Jan geeft oma een bal* en *Jan geeft een bal aan oma*. In het Europarl corpus zoek ik alle zinnen met *geven*, Duits *geben* en Engels *give*, die voorbeeld van óf de dubbelobjectsconstructie óf de prepositionele constructie zijn. En het resultaat is nogal verbazingwekkend: terwijl de verdeling tussen de twee constructies in het Nederlands bijna 50/50 is, komt in het Engels de dubbelobjectsconstructie in exact tweederde van de voorbeeldzinnen voor, de prepositionele constructie in eenderde. Het Duits vertoont de prepositie-variant zo goed als nooit, wat te maken heeft met het feit dat het Duits nog zichtbaar onderscheid maakt tussen de verschillende naamvallen, terwijl het Engels en het Nederlands dat niet meer (op zo'n uitgebreide manier) doen.

Dit onderscheid neem ik ook als de basis voor een analyse van de verschillen tussen de Nederlandse, Duitse en Engelse datiefalternantie. Voor het Engels is er onderzoek gedaan naar het ontstaan van der prepositionele constructie als alternatief van de dubbelobjectsconstructie, en het blijkt dat de eerste on-

der bepaalde voorwaarden de tweede is gaan vervangen, toen het direct en indirect object niet meer door hun naamvalsuitgangen te onderscheiden vielen. Hetzelfde lijkt in het Nederlands te zijn gebeurd. Voor het Duits is het niet uitgesloten dat deze vervanging ook plaats gaat vinden, zodra het naamvalssysteem nog meer bedreigd wordt dan nu al het geval is.

Ik concludeer dat alle drie de talen een monoseem werkwoord *geven*, *geben* en *give* in hun repertoire hebben, die met twee lexicale-klasse constructies kunnen combineren om de gewenste argumentstructuur toegewezen te krijgen.

In de gepresenteerde case studies staat altijd één vraag centraal: hoe ziet de organisatie van het lexicon eruit? Dit is van grammaticaal en van computationeel belang. Ik geef voorbeelden voor zowel lexicale constructies, waarbij de benodigde informatie aan een lexicaal element of een lexicaal proces is opgehangen, als ook voor frasale constructies, waarbij een abstracte constructie de kombinatoriek van bepaalde elementen stuurt. Uiteindelijk pleit ik dus niet voor een strict lexicale aanpak, maar ik argumenteer wel voor een zuinige omgang met frasale constructies.

# Curriculum Vitae

Michaela Poß was born in Duisburg, Germany, on the 19th of July, 1976. After secondary school, she moved to the Netherlands in 1995 to study Psychology. She returned to Germany a year later and enrolled to study General Linguistics, Dutch and English language and literature in Münster, where she graduated with an MA in 2002. From 2002 to 2004 she worked as a Research Assistant at the English Department (chair of Prof. Wolf Paprotté), before she once again moved to the Netherlands.

In May 2004 she started her AiO position in Ton van der Wouden's NWO project "Dutch as a Construction Language", and the research she carried out in the following four years resulted in this dissertation. In 2005 and 2006 she was happy to spend four terms at Stanford University, where her collaboration with her promotor, Prof. Ivan A. Sag, started.

In 2008 she took up a position at the Institute for Dutch Lexicology (INL) in Leiden, which she terminated when her son Jonathan was born in 2009. At present, she spends her maternity leave in Düsseldorf, Germany, with her son, her dog, and her partner.