

Subject or Object?

Syntactic disambiguation in Norwegian, stochastic
Optimality Theory and application in an automatic system

Lilja Øvrelid

Thesis submitted in partial fulfillment
of the requirements for
the degree of Cand. Philol.
Department of Linguistics
University of Oslo, Norway

November 3rd 2003

Acknowledgements

As the work on my thesis is drawing towards its end, I am happy for the opportunity to thank some of the people who have assisted me along the way and made the whole process such an interesting and enjoyable venture.

Firstly, I want to thank my two advisors. My principle advisor, Prof. Janne Bondi Johannessen, has been a true pleasure to work with. She has been involved in the whole process, from choice of topic to concluding remarks, and has always given me pointed criticism, whilst at the same time installing in me a belief in the project and my capabilities therein. Thanks also for giving me a glance into academic life by inviting me to two lovely seminars, and in general making me feel welcome in settings that would otherwise have seemed quite intimidating from a mere student's perspective.

Prof. Helge Lødrup has been my secondary advisor and an invaluable source of knowledge and theoretical insight. He taught my first course in syntax and had me interested from the very start. His enthusiasm for the field is truly inspiring. It is my pleasure to thank him for all the time he has spent on me during the course of this thesis and for always having an open door.

Many thanks to Kristin, Anders, Åsne and Lars at Tekstlaboratoriet for being so supportive and always making me feel welcome.

My fellow students at Language, Logic and Information also deserve thanks. In particular, I want to thank my good friends Iselin, Fredrik and Bjørnar for all the lunches and coffee breaks along the way. The social life here has made the difficult writing process seem a lot less impossible and I will miss you all.

My friends outside of university - Liza, Iselin, Kari and all the rest of you, thanks for all the good times!

A special thanks also to Anne Karin Paaske and Nils Øvrelid for making the last few months endurable and for all the good food and comfort.

Finally, and on a more personal note, I want to thank Stian Nygaard for endless support and patience in everything.

Lilja Øvrelid
November 3rd, 2003

Contents

1	Introduction	1
1.1	Focus of the thesis	1
1.2	Conceptual restrictions and clarifications	2
1.3	Overview of the thesis	3
2	Syntactic functions and disambiguation in Norwegian	5
2.1	The subject and object functions	5
2.1.1	Coding Properties	6
2.1.2	Behavioral properties	7
2.1.3	Prototypical subject properties	11
2.2	Disambiguation	11
2.2.1	Case	11
2.2.2	Word Order	13
2.3	Conclusion	16
3	Theoretical Framework	19
3.1	Typological Linguistics - markedness and hierarchies	20
3.1.1	Markedness criteria	20
3.1.2	Prominence hierarchies	20
3.2	Topic, focus and related notions	25
3.2.1	Pragmatic Roles	25
3.2.2	Heaviness and newness	27
3.2.3	Related notions	27
3.3	Optimality Theory (OT)	28
3.3.1	The emergence of Optimality Theory	28
3.3.2	Ranked constraints	29
3.3.3	An input-output device	30
3.4	Aissen's theory of markedness through harmonic alignment	32
3.4.1	DOM	32
3.4.2	Formal Framework	33
3.4.3	The realization of DOM	35
3.4.4	Conclusion	36
3.5	Stochastic OT	37

3.5.1	Modeling variation in language	37
3.5.2	The Gradual Learning Algorithm	39
3.5.3	The GLA in use	39
3.6	“Reverse-animacy” verbs	40
3.6.1	Introduction	40
3.6.2	Thematic roles	41
3.6.3	An aspectual dimension	44
3.6.4	Dowty’s theory of thematic roles	45
3.6.5	Causative verbs	47
3.6.6	Summing up	48
3.6.7	Practical consequences	48
3.7	Word Order Freezing	49
3.7.1	Lee: Markedness Reduction in Word Order	49
3.7.2	Morphological syncretism and word order freezing	52
3.7.3	Animacy-related freezing effects	53
3.7.4	Relevance of freezing phenomena	54
3.8	Ambiguity: Bidirectional optimization and the problem of in- effability	55
4	Prerequisites for a data analysis	57
4.1	Earlier corpus studies - the case of Swedish	58
4.2	Sampling and annotating the data	60
4.2.1	What to include in the sampling process	60
4.2.2	What to exclude in the sampling process	65
4.2.3	What to count?	67
4.3	The database	68
4.3.1	From corpus to database	69
4.3.2	Extracting the results	71
4.4	Non-literal language	73
4.4.1	Metonymies	74
4.4.2	Metaphors	75
4.4.3	Practical consequences	76
5	Animacy, definiteness and word order variation	79
5.1	Sampling results	79
5.1.1	Animacy	79
5.1.2	Definiteness	85
5.1.3	Object topicalization	88
5.1.4	Equally placed arguments	90
5.2	Summary, discussion and practical consequences	93
5.2.1	Summary and discussion	93
5.2.2	Practical consequences	95
5.3	New challenges and their theoretical implications	98

6	Modeling variation: a stochastic OT analysis	101
6.1	Input and candidates	102
6.2	Constraints	103
6.2.1	Animacy constraints	103
6.2.2	Word order constraints	105
6.2.3	Definiteness	109
6.3	Stochastic OT analysis	112
6.3.1	Training towards convergence	112
6.3.2	Results	115
6.4	Topics for further research	118
6.4.1	An alternative approach through bidirectional OT	119
6.4.2	Coverage	124
6.5	Conclusion	125
7	Constraint Grammar application	127
7.1	The Oslo-Bergen tagger and the Constraint Grammar (CG) formalism	127
7.1.1	The formalities of CG Rules	129
7.2	Practical implementation explored	130
7.2.1	Practical outline	131
7.3	Results	139
7.4	Problem areas for further work	141
7.4.1	Metonymy revisited	142
7.4.2	Other areas	144
7.5	Conclusion	145
8	Conclusion	147
8.1	Summary	147
8.2	Topics for further work	148
8.2.1	Animacy and word order revisited	148
8.2.2	Specificity	149
8.2.3	Other topics	149
A	OT Grammar and Distributions	151
A.1	OT Grammar	151
A.2	Pair distributions	153
B	Constraint Grammar application	155
B.1	Constraint Grammar rules	155
B.2	Constraint Grammar sets	159

Chapter 1

Introduction

- (1) a. Jenta skrev brevet
girl-DEF wrote letter-DEF
'The girl wrote the letter'
- b. Brevet skrev jenta
letter-DEF wrote girl-DEF
'The letter, the girl wrote'

The minimal pair in (1) above illustrates the essential phenomenon that triggered the work of this thesis. These sentences are very similar from a linear point of view, but represent quite different word order patterns. The sentence in (1a) is a transitive sentence with a SVO¹ word order, whereas in (1b), we see a transitive sentence where the object has topicalized, giving us an OVS word order. Any native speaker of Norwegian would rule out the second structurally possible readings from (1a) and (1b) above, namely the semantically odd one, that the letter wrote the girl. The question of which factors that contribute to the disambiguation of the syntactic functions of subject and object in sentences like the above, has led to some interesting and hitherto unnoticed generalizations relating to formal and semantic properties of these functions, as well as their positioning within the sentence in Norwegian.

1.1 Focus of the thesis

The focus of this thesis is on properties of the subject and object functions in Norwegian transitive constructions. The work presented here has originated from a problem of automatic disambiguation of syntactic functions in Norwegian, and a need to discover facets of word order and argument realization apart from solely structural criteria. Work on a morphosyntactic tagger for

¹SVO stands for Subject Verb Object, and OVS, Object Verb Subject.

Norwegian² has raised some interesting questions which served to trigger the work presented in the following.

In typological literature, prominence hierarchies expressing properties such as relative animacy or definiteness have been posited to play an important role in various linguistic phenomena (Croft, 1990; Comrie, 1989). Lately, a renewed interest in typological generalizations can be found within the Optimality Theoretic community. Aissen (2003) points to the role that prominence hierarchies of definiteness and animacy might have in the choice of subject and object, and in particular the form that these might take. Through formal operations on prominence hierarchies, relations of markedness are expressed that pertain to the realization of the syntactic functions of subject and object. In particular, a so-called markedness reversal is observed, expressing the observation that what is marked for a subject is unmarked for an object and vice versa.

A key question in the following thus becomes whether these prominence hierarchies might also be operative in Norwegian as conditioning factors on argument realization and positioning, and to what extent? In many languages, subjects and objects must bear explicit morphological marking if they are semantically or pragmatically marked as subject or object. This however is not the case in Norwegian, which only marks pronominal objects overtly for case, but not consistently. Even so, might it be that “the effects of hierarchy alignment which are categorical in some languages are expressed as usage preferences in others” (Bresnan, 2002), for instance in Norwegian? Do these hierarchical properties also condition word order variation?

We will in the following restrain ourselves to looking at the influence of *animacy* and *definiteness* on the realization and positioning of the syntactic functions of subject and object within the sentence. We will see that these properties both contribute towards disambiguation, albeit to a varying degree. Animacy, in particular, restrains the possibility for word order variation, and we will see examples of so-called freezing effects on word order at a level of performance. Stochastic Optimality Theory provides us with the formal apparatus for modeling our observations within a structured and linguistically well-founded theory. Also, we will see how our findings may be implemented, and contribute towards an improved level of disambiguation of syntactic functions within an automatic system.

1.2 Conceptual restrictions and clarifications

Before we proceed, some restrictions and clarifications regarding the content of this thesis and the use of various terms in the following should be stated.

The work presented in this thesis pertains exclusively to Norwegian, even

²The Oslo-Bergen tagger has been developed at Tekstlaboratoriet at the University of Oslo, Norway.

though examples are drawn from several other languages, including English, to exemplify different phenomena. Judgments of grammaticality are based on my own intuitions as a native speaker of Norwegian.

We are in the following, unless otherwise specified, dealing with transitive sentences, i.e. sentences containing a transitive verb with two arguments - a subject and an object. We are thus not dealing explicitly with indirect or prepositional objects, and when employing the term *object* in the following, we are referring to the direct object. Also, when dealing with transitive sentences, we are, unless otherwise stated, dealing with a transitive sentence with a simple verb phrase, as in (1) above, the reasons for which will be made clear in the next chapter.

As explained above, central to the work described in the following are notions of ambiguity and disambiguation. These obviously pertain to structural, not lexical ambiguity. A sentence is structurally ambiguous when one and the same surface sentence may have different structural interpretations, i.e. parses. The fact that an object may topicalize, as in (1b) above, constitutes the source of this structural ambiguity. Our starting point for this work was that of automatic disambiguation, and we want to ascertain other properties of subjects and objects, besides the strictly structural, in order to improve disambiguation. Our data material will, for instance, contain only those sentences that are structurally ambiguous from an automatic point of view. However, most of these are readily disambiguated by any native speaker, and it is precisely this implicit knowledge that is in focus. Because of this, all structurally possible readings are not always represented graphically, even though they constitute the underlying motivation for the analysis.

Wherever possible, examples containing a topicalized object in Norwegian will be translated with a topicalized construction in English. As we shall see in chapter 3.2, however, there is not always a one-to-one correspondence between the two languages in this respect.³

1.3 Overview of the thesis

In chapter 2, we take a closer look at the problem of disambiguation of syntactic functions in Norwegian. We examine some criteria for subjecthood in Norwegian and contrast these with criteria pertaining to the object function. Furthermore, an account of some complicating factors for disambiguation is provided by looking at morphological case and word order in Norwegian.

³The problem of translating Norwegian sentences containing a topicalized object with a corresponding structure in English, resides in the distinction between contrastive and continuous topics. Norwegian may express both types of topics with a topicalized structure, whereas English only does so for the former (Engdahl, 1997). See chapter 3.2 for more on this distinction.

Chapter 3 provides an overview of the theoretical framework that constitutes the backbone of this thesis. We start out by examining the notion of markedness and its relation to prominence hierarchies, as emphasized in typological linguistics. The discourse-related notions ‘topic’ and ‘focus’ are then given a short account. Optimality Theory (OT) constitutes the main theoretical approach in this thesis. We start out by examining the main tenets of traditional OT, followed by a thorough presentation of the formal framework developed in Aissen (2003, 1999) and then an account of stochastic OT is presented. Ensuing, we take a closer look at some verbs that constitute a challenge to our theoretical predictions, the “reverse-animacy” verbs. Finally, we examine the phenomenon of word order freezing, and, in particular, one formal account of this phenomenon, as presented in Lee (2002b). In closing, bidirectional OT is given a short presentation, and contrasted with traditional OT.

A corpus study of one thousand transitive sentences has provided the data material central to this thesis. In chapter 4, we establish some prerequisites for the analysis of these data. First of all, a similar corpus study of Swedish is described. We go on to examine the criteria set up to delimit the data set, and provide a short description of the organization of the database containing our data. Following this, we move on to examine a challenge for the annotation of the data material, namely that of non-literal language, a phenomenon which deserves a thorough treatment.

In chapter 5 we examine and discuss the results from the data analysis. The implications of our findings, both for the practical side of the thesis, as well as the theoretical, are then discussed further.

By making use of the theoretical framework of stochastic OT, we are, in chapter 6, able to model our findings from the data analysis, and thus obtain a clearer picture of the influences of animacy and definiteness on the subject and object in Norwegian. In particular, this model reflects the data material and the real frequencies observed there. Also, the stochastic OT modeling reflects our theoretical framework from chapter 3.

As mentioned above, the point of departure for the following work, was based in a problem of automatic disambiguation of sentences like the ones exemplified in (1) above. In chapter 7, we will venture into a practical outline and see how our findings from the previous chapters may be implemented through disambiguation rules written in the Constraint Grammar formalism. We will also review some results from a test run of the automatic system following the addition of our rules.

Chapter 2

Syntactic functions and disambiguation in Norwegian

Norwegian has a fairly rigid word order, and marks syntactic functions in large part through structural positioning, as opposed to languages with a freer word order, where morphological case or agreement is a crucial disambiguator in this respect. In Norwegian, pretty much all constituents may topicalize, however, when the object in a simple transitive construction topicalizes, disambiguation of the syntactic functions of subject and object is at risk. In linear surface structure there are no structural indicators as to the syntactic functions of the two arguments in such a construction. In the following we will start out by looking at some criteria for the positing of the subject and object functions in Norwegian. We will then examine further the problem of disambiguation of syntactic functions in Norwegian and some complicating factors for this.

2.1 The subject and object functions

Various syntactic theories differ in how they view the status of syntactic functions. Chomskyan syntax claims that syntactic functions are strictly structural and can be defined by structural means alone, e.g. a subject is an element in the specifier-position of the IP, whereas an object might be the sister of V. On this view, grammatical relations are simply structural constellations which can be derived from a structural representation (a tree). In a competing theory however, such as Lexical Functional Grammar (LFG) (see for example Bresnan (2001)), grammatical relations are primary notions expressed at a separate level from that of the tree representation.

In the following section, however, a more theory-independent, functional account of the syntactic functions will be presented. Criteria for the use of the subject function as well as characteristic properties will be contrasted with those of the object, in order to differentiate clearly between these. The

presented criteria are language-specific for Norwegian, even though some of them are quite common cross-linguistic subject-criteria as well.

It is common practice to differentiate between two types of criteria for syntactic functions; criteria that pertain to the function itself or so-called *coding* properties and criteria that pertain to the role of the function in certain grammatical/syntactic phenomena, so-called *behavioral* properties (Keenan, 1976; Faarlund et al., 1997). These criteria in isolation are not seen as providing the necessary and sufficient conditions for discriminating the subject from, for instance, the object. Rather they provide, in combination, a clearer picture of the properties and behaviour of a subject, thus defining it more clearly.

2.1.1 Coding Properties

Form

Firstly, there are criteria which delimit the domain of possible subjects - only nominal constituents may serve as the subject of a sentence:¹

- (1) a. Noun phrase:

Fredsforhandlingene i Midtøsten har fullstendig
Peace-negotiations-DEF in Middle-East-DEF have completely
stagnert
stagnated

‘The peace negotiations in the Middle East have stagnated completely’

- b. Pronominal phrase:²

Han har vært toleransens og verdighetens
He-NOM has been tolerance-DEF-GEN and dignity-DEF-GEN
talsmann
spokesman

‘He has been the spokesman of tolerance and dignity’

- c. Non-finite clause:

Å lese klassikere er noe en må komme inn i
To read classics is something one must get in to

‘Reading classics is something you have to get into’

- d. Finite clause:

¹All of the examples in (1) are taken from the Oslo Corpus, a Norwegian corpus of approximately 18 million words, available at <http://www.tekstlab.uio.no/norsk/bokmaal/>.

²The pronominal subject is always in nominative case.

At du har eget hus, blir nok også tillagt
 That you have your-own house will-be enough also given
 vekt
 weight

‘The fact that you also own your own house will surely be weighted’

The formal criteria above pertain to objects as well. In addition, a pronominal object is often marked morphologically with accusative case. Nominative case, however, is not a criterion from which one may automatically assume subjecthood, as we shall see later on in this chapter.

Since a subject is required in all declarative Norwegian sentences, there is also the possibility of an expletive subject:

- (2) Expletive subject:

Det er fokus på hår i høst
 It is focus on hair in autumn

‘The focus is on hair this autumn’

Position

The subject of a declarative sentence is either in first position, i.e. it is the first constituent, or it is the constituent which immediately follows the finite verb. Structurally speaking, the subject is often said to be base generated in the specifier position of IP, and the object as sister to V (see for example Carnie (2001)). As we have seen, the object may also occupy the first position, or the specifier position of CP, when it topicalizes. Topicalization is not an option for weak pronouns. These may, however, occupy a position outside of VP through the process known as object shift (Sells, 2001). Section 2.2.2 below will provide us with a more detailed picture of word order in Norwegian, thus for now the above will suffice.

2.1.2 Behavioral properties

Behavioral properties are mainly syntactic properties, which contribute towards discriminating the subject from the object of a sentence.

Passive

A well-known test for subjecthood is the passive construction, or rather the process from active to passive, where the active subject is suppressed:

- (3) a. *Jenta skrev brevet*
 girl-DEF wrote letter-DEF
 ‘The girl wrote the letter’

- b. Brevet ble skrevet (av jenta)
 letter-DEF was written (by girl-DEF)
 ‘The letter was written (by the girl)’

The object, on the other hand, becomes the subject of the passive sentence, or it may remain object in the impersonal passive construction with an expletive subject. The impersonal passive however, requires an indefinite object:

- (4) Det ble skrevet et brev
 It was written a letter
 ‘A letter was written’

Imperative

If an argument is deleted in an imperative sentence, it is always the subject:³

- (5) a. Jenta skrev brevet
 girl-DEF wrote letter-DEF
 ‘The girl wrote the letter’
 b. Skriv brevet!
 write letter-DEF
 ‘Write the letter!’
 c. *Du skriv!
 you write!

This is not the case with direct objects, as shown in (5c) above.

Deletion in sentence-conjunctions

When two sentences with identical subjects are conjoined, the subject in the second conjunct sentence may be deleted:

- (6) a. Jenta skrev brevet og sendte det med en gang
 Girl-DEF wrote letter-DEF and sent it with one time
 ‘The girl wrote the letter and sent it right away’
 b. *Jenta skrev brevet og (hun) sendte med en gang
 Girl-DEF wrote letter-DEF and (she) sent with one time

³Imperatives may occur without subject-deletion. In this case the subject follows the finite verb:

- (1) Skriv du det brevet!
 write you that letter!
 ‘You write that letter!’

This property is not usually one that holds for objects, as we can see from (6b) above. There are, however, some exceptions:

- (7) Jenta tok vedkubbene og satte vedkubbene i skjulet
girl-DEF took firewood-DEF and put firewood-DEF in shed-DEF
'The girl took the firewood and put the firewood in the shed'
- (8) Jenta tok vedkubbene og satte i skjulet
girl-DEF took firewood-DEF and put in shed-DEF
'The girl took the firewood and put (it) in the shed'

Common to these is that the subject, which must be the same for both the conjoined sentences, often is deleted as well in the second conjunct, like in (8) above. Also, the verb in the first conjunct is commonly the verb *ta* 'take', which occurs in these types of constructions as more of an aspectual marker, thus in reality describing only one verbal action (Faarlund et al., 1997, p. 716).

Control and raising

Only the subject of a non-finite subclause may be controlled by an argument in the matrix clause or raised out of a non-finite clause to assume either subject or object position in the matrix clause. In the case of control constructions, as in (9) below, the matrix argument and the controlled subject are co-referent, much like a noun and a pronoun. This is the reason why the subject of the non-finite clause has been proposed to be a null and phonologically silent pronoun, represented as PRO⁴:

- (9) Jenta_i liker å PRO_i skrive brev
girl-DEF_i likes to PRO_i write letters
'The girl enjoys writing letters'

In raising constructions, however, the subject of the non-finite clause is raised, thus leaving a trace in subject position:

- (10) Jenta_i pleide å t_i skrive brev
girl-DEF_i used to t_i write letters
'The girl used to write letters'

Raising verbs, as we know, do not subcategorize for an external role, but rather receives this from the verb of the subclause.

Obligatory relative marker

When a matrix argument is subject of a modifying relative clause, the relative marker *som* is obligatory, as shown in (11) below. This is not the case when

⁴See for example Carnie (2001).

a matrix argument is the object of a modifying relative clause, then the relative marker is optional, as we can see in (12) below:

- (11) a. Jenta som liker å spise epler skrev brevet
 girl-DEF who likes to eat apples wrote letter-DEF
 ‘The girl who likes eating apples wrote the letter’
 b. *Jenta liker å spise epler skrev brevet
 Girl-DEF likes to eat apples wrote letter-DEF
- (12) a. De brevene som jenta skrev så vi aldri
 Those letters-DEF that girl-DEF wrote saw we never
 ‘Those letters that the girl wrote, we never saw’
 b. De brevene jenta skrev så vi aldri
 those letters-DEF girl-DEF wrote saw we never
 ‘Those letters the girl wrote, we never saw’

Quantifier movement

In this criterion, a quantifier may be positioned directly after the main verb, even though the subject it has scope over is positioned in front of the verb. The equivalent is not, however, possible for the object:

- (13) a. *Begge* jentene skriver brev
 Both girls-DEF write letters
 ‘Both girls write letters’
 b. Jentene skriver *begge* brev
 girls-DEF write both letters
 ‘The girls both write letters’
- (14) a. Jentene skrev *begge* brevene
 girls-DEF wrote both letters-DEF
 ‘The girls wrote both letters’
 b. *Jentene *begge* skrev brevene
 girls-DEF both wrote letters-DEF

Tag questions

The echoing pronoun of so-called tag questions always refers to the subject, and never the object:

- (15) Jenta_i skrev brevet_j, gjorde hun_i/*det_j ikke?
 girl-DEF wrote letter-DEF, did she/*it not?

2.1.3 Prototypical subject properties

Apart from the formal and structural criteria reviewed above, subjects and objects have both semantic and pragmatic properties that more or less serve to identify them cross-linguistically. There seems, for instance, to be a universal tendency towards equating the subject with the agent and the topic of the sentence. Rather than being an absolute definition, however, this is more of a prototypical property, which subjects more or less will adhere to:

The kind of definition of subjects towards which we will be working is the following: the prototype of subject represents the intersection of agent and topic, i.e. the clearest instances of subjects, cross-linguistically, are agents which are also topics. (Comrie, 1989, p. 107)

As we shall see, there are several other properties that are also prototypically properties of the subject. As these are the topics of the following chapter, as well as the ensuing analyses, we will not pursue this aspect any further here.

2.2 Disambiguation

The fact that the two arguments in a transitive sentence may differ in structural position is a cause of distress when it comes to automatic disambiguation of syntactic functions. An automatic computer program will not necessarily have access to the semantic restrictions available to a speaker of a language, which easily facilitates disambiguation in most cases.

In what follows we will take a closer look at some complicating factors with regards to disambiguation of syntactic functions in Norwegian. The focus will be on transitive sentences throughout.

2.2.1 Case

As mentioned above, case is a crucial factor in the disambiguation of syntactic functions in many languages. In particular, it allows for a large degree of word order variation. An often remarked fact is a correlation between the degree of morphological marking in a language and the degree of word order variation (Lee, 2002a). Norwegian must, in this respect, be said to have little of both, i.e. due to a lack of morphological marking, word order becomes quite fixed.

Only personal pronouns are marked for morphological case in Norwegian. One might argue that in a transitive construction containing a topicalized object, the subject will usually be pronominal and thus help to disambiguate the sentence. This however, does not necessarily solve the problem. Norwegian pronouns are not consistently unambiguous when it comes to case.

Firstly, several of the Norwegian pronouns in nominative form syncretize with the accusative form, as can be seen from the pronominal paradigm, as given in table 2.1 below. The 3rd person singular nominative pronoun *han* ‘he’ and the 3rd person plural pronouns may syncretize with the accusative, whereas 3rd person singular *det* ‘it’, *den* ‘it’ and 2nd person plural *dere* ‘you’ always do so. Secondly, Johannessen (1998) points to the fact that

Person/Number	Nominative	Accusative
1. sing.	jeg	meg
2. sing.	du	deg
3. sing. masc.	han	han/ham
3. sing. fem.	hun/ho	henne
3. sing.neut (inanimate)	det	det
3. sing.masc./fem. (inanimate)	den	den
1. pl.	vi	oss
2. pl.	dere	dere
3. pl.	de	dem

Table 2.1: The Norwegian pronominal paradigm

a nominative pronoun alone very rarely occurs as object of a verb. However, as head of a phrase, a pronoun will be stressed and when stressed it is strongly preferred in nominative form regardless of syntactic function. So, when followed by for instance a relative clause or a prepositional phrase, the pronoun will be in its nominative form even when functioning as an object. This obviously leads to quite a bit of ambiguity, since pronouns are often followed by relative clauses or prepositional phrases.⁵

- (16) Dette gjelder i tillegg *de* som håndterer etc.
 this concerns in addition they-NOM who handle etc.
 ‘This also concerns those who handle’

- (17) *De* som fortsatt tror at idyllen kan bevares [...] [*...*]
 they-NOM who think that the idyll can maintain-PASS [*...*]
 tar alvorlig feil
 take seriously wrong
 ‘Those who still believe that the idyll can be maintained [*...*] are seriously mistaken’

In (16) we see a nominative form of the 3rd person plural pronoun *de* ‘they’ followed by a relative clause, functioning as an object, whereas the same type of construction plays the role of subject in (17). When a nominative pronoun that is head of a phrase may function as an object, it may, in theory, very well topicalize as well. From the viewpoint of automatic analysis then, these must be considered to be possibly ambiguous.

⁵The examples in (16) and (17) are taken from the Oslo Corpus.

This generalization does not, however, extend to all nominative pronouns. Local, i.e. 1. and 2. person, pronouns may not function as objects with the same ease as the others, even if they are modified by another clause:

- (18) a. * Dette gjelder *jeg* som skrev brevet
 this concerns I-NOM who wrote letter-DEF
- b. * Dette gjelder *du* som skrev brevet
 this concerns you-NOM who wrote letter-DEF
- c. ?? Dette gjelder *vi* som skrev brevet
 this concerns we-NOM who wrote letter-DEF

This mirrors a morphological split that is present in many Scandinavian dialects (Dahl, 1997), where only 1. and 2. person pronouns must bear morphological marking when functioning as objects. This split, which pertains to simple pronouns, i.e. pronouns that are not modified by another clause, seems, however to be present for 1. and 2. person pronouns also when they are modified. The above examples are also semantically odd, because local persons, that are readily identified by the immediate context, are not usually in need of restrictive modification. Rather, the examples above in (18) would take on some sort of causal interpretation, rather than a restrictive one, had they been in accusative case.

There is some sense in which the plural example in (18c) above is more acceptable than the singular forms. It is possible that the mentioned split is influenced by number, as well.

Faarlund et al. (1997) state that nominative pronouns occur as objects almost exclusively when they are 3. person plural pronouns. It seems clear, however, that the singular 3. person pronouns also may occur quite freely as objects when modified (Johannessen, 1998):

- (19) Dette gjelder *han/hun* som skrev brevet
 this concerns he-NOM/she-NOM who wrote letter-DEF
- ‘this concerns he/she who wrote the letter’

2.2.2 Word Order

The basic word order of a language is “typically identified with the order that occurs in stylistically neutral, independent, indicative clauses [...], it is the ordering of constituents in prototypical transitive clauses”. (Siewierska, 1988, p. 8). In this respect, Norwegian must be said to be a SVO language.

Norwegian clause structure and V2

Like German and the other Scandinavian languages, Norwegian is also a V2-language, i.e. the verb is always the second constituent in declarative

main clauses.⁶ The V2 property explains why the structural ambiguity in the sentence pair from chapter 1 pair is present:

- (20) a. Jenta skrev brevet
girl-DEF wrote letter-DEF
(i) ‘The girl wrote the letter’
(ii) ‘The girl, the letter wrote’
b. Brevet skrev jenta
letter-DEF wrote girl-DEF
(i) ‘The letter, the girl wrote’
(ii) The letter wrote the girl

Most types of constituents may be sentence-initial, the subject being the most common:⁷

- (21) a. Sentence-initial Subject:
Jeg liker dette
I like this
‘I like this’
b. Sentence-initial Object:
Dette liker jeg
This like I
‘This, I like’
c. Sentence-initial Adverbial:
Her bor jeg
Here live I
‘Here, I live’

Through the process of topicalization constituents other than the subject may become sentence-initial, thus enabling the marked word order OVS, as in (21b) above.

In transformational theories, topicalization is viewed as a movement to the “topic” position, SpecCP. The traditional generative view of clause structure in V2-languages is that all sentences are CPs, and that V2 is preserved

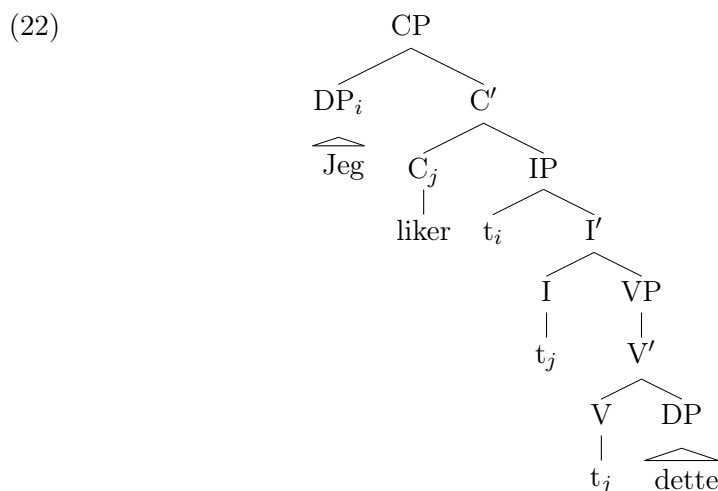
⁶English, on the other hand, is SVO, but is not a V2 language:

- (1) I like this.
(2) This, I like.
(3) * This like I.

⁷These examples are not in need of disambiguation, as the nominative pronoun *jeg* ‘I’ is an unambiguous subject. Rather they are meant simply to illustrate the word order patterns without any chance of ambiguity.

through a demand that C must be lexicalized (Nordgård and Åfarli, 1990, p. 71)⁸. Subordinate clauses, on the other hand, do not exhibit V2 order, due to their filling the C-position with a complementizer.

The finite verb in Norwegian declarative main sentences thus moves from I to C, thereby assuring that only one constituent may precede it. No constituent is thought to be base-generated in SpecCP, but rather, a constituent will move here and is thus defined as the topic of the sentence. This describes the structural topic of the sentence, but not necessarily a pragmatic topic, as we will see in chapter 3.2. In a majority of cases, this constituent will be the grammatical subject of the sentence, thus moving from SpecIP to SpecCP, as in (22) below.

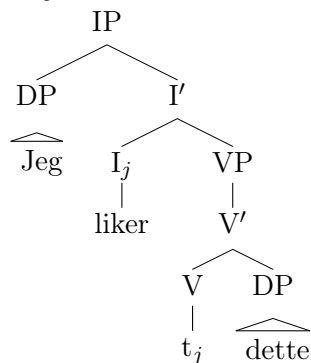


An alternative analysis of V2 clause structure (Sells, 2001)⁹, has as of late gained some ground as well. Here, the minimal matrix sentence is an IP, as analyses for English and other non-V2 languages have claimed, and not a CP, cf. example (23) below. The idea is that the subject is fixed in SpecIP, and only topicalization of non-subject constituents incur a CP structure, as in (24). This thus incorporates a notion of economy in syntax, as emphasized in theories such as LFG. In LFG, only constituents marked with the discourse functions TOPIC or FOCUS are found in SpecCP, and the subject is the default topic. Therefore, there is no reason why the subject should move when it is sentence-initial. On this account then we in fact operate with two possible structures for declarative main sentences:

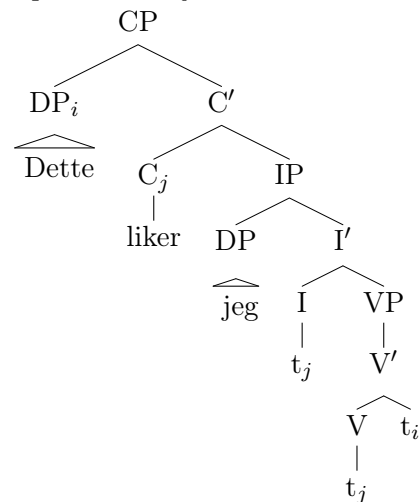
⁸This demand is justified by claiming that the C functions as head of the sentence. A typical head-property is thus to determine the properties of the phrase as a whole.

⁹Sell's analysis is directed primarily at Swedish, but makes use of examples from Norwegian as well.

(23) Subject-initial sentence



(24) Topicalized object



Complex VPs

With regards to our initial problem of disambiguation, it is important to note that word order may also act as a disambiguating factor in some cases. When the transitive sentence contains a complex verb phrase, the structural positioning of the arguments indicate their syntactic function, thus excluding the second reading present in (20) above:

- (25) a. *Jenta* har skrevet *brevet*
 girl-DEF has written letter-DEF
 (i) 'The girl has written the letter'
 (ii) * 'The girl, the letter has written'
- b. *Brevet* har *jenta* skrevet
 letter-DEF has girl-DEF written
 (i) 'The letter, the girl has written'
 (ii) * 'The letter has written the girl'

The reason for this is found in the property of Norwegian being a V2-language. Only one constituent may precede the finite verb:

- (26) **Brevet jenta* har skrevet
 letter-DEF girl-DEF has written

2.3 Conclusion

The criteria for the subject and object functions set out above, provide us with an account of the formal and structural properties associated with the arguments of a transitive sentence. The subject/object criteria are not,

however, disambiguating from the point of view of this thesis, and are useless from an automatic point of view, as they rely on grammaticality judgments.

The phenomena of interest in the following chapters are thus transitive sentences where there are no structural or morphological factors present that disambiguate the syntactic functions of the arguments. What this means is that we are, in essence, looking at sentences with a simple verb phrase and arguments that are not morphologically marked as objects, i.e. bearing accusative case. As we saw above, nominative case is not always sufficient for the inference of subjecthood.

Chapter 3

Theoretical Framework

It might be useful to clarify the theoretical background the thesis will be drawing on. Thus, in the following I will review the relevant theories for the later chapters and analyses.

We will start out by examining the concept of *markedness* and the role it has played in typological linguistics. In particular, we will see how a concept of relative markedness has led to the positing of feature hierarchies, and examine some of these more closely. Following that, we will take a look at the discourse-oriented notion of topic, and see how this type of information interacts with other features of a sentence and its arguments.

Optimality Theory (OT) will be a central theory in this thesis, underlying much of the work reviewed here and also forming the theoretical backbone of the analysis performed in chapter 6. We will therefore review some of its main tenets below, as well as examine the work of Judith Aissen on the subject of Differential Object Marking, which combines in a fascinating manner insights from typological linguistics with a formal OT framework. A recent addition to OT, stochastic OT, has opened up for a modeling of variable phenomena in OT using stochastic measures based on relative frequencies. This will prove to be crucial for our analysis, and we will therefore present this theory and contrast it to traditional OT.

A challenge may be found in sentences containing verbs that may take arguments that deviate from our theoretical predictions. We will examine these verbs closer and find that they are a rather diverse group. Ensuing, we will take a look at a particular phenomenon related to word order variation, namely that of word order freezing. Finally, another extension to OT, namely that of bidirectional OT will receive a short overview.

3.1 Typological Linguistics - markedness and hierarchies

The notion of ‘markedness’ has been of great importance in the typological field of linguistics, underlying much of the cross-linguistic comparative work performed there. The notion is based on “asymmetrical or unequal grammatical properties of otherwise equal linguistic elements”(Croft, 1990).

3.1.1 Markedness criteria

Based on Joseph Greenberg’s work on markedness and his thirteen criteria of markedness, Croft (1990) outlines some basic criteria for the relative markedness of a linguistic structure:¹

Structure An unmarked structure will contain fewer than or an equal amount of morphemes as a marked structure.

Behavior An unmarked structure will be more versatile:

- It might have a greater number of morphological distinctions,²
- and/or it might figure in a greater number of linguistic contexts.

Frequency An unmarked structure will be more frequent than a marked structure in textual discourse. This criterion obviously relates to the second part of the behavioral criterion, a structure which figures in a greater number of contexts will also be more frequent.

Croft (2003) introduces a notion of *typological markedness*, and emphasizes that this notion differs from language-particular markedness and its language-particular formal expression. Rather, typological markedness is a “universal property of a conceptual category, not a language-particular property of a language-particular grammatical category” (Croft, 2003, p.88). This is an important point, because languages will necessarily differ in their formal expression of the same typological markedness relations.

3.1.2 Prominence hierarchies

The early versions of theories of markedness, especially of the Prague School, proposed a view of the markedness phenomenon as a binary opposition between a marked and unmarked form. Either an element was marked, or it was not. However, this binary notion often lead to so-called “markedness

¹‘Relative’ due to the fact that an element is not marked in isolation, but always relative to another element of the same grammatical dimension.

²The distinction between singular (unmarked) and plural (marked) English 3. person pronouns might exemplify this. The singular inflection has three different forms (*he, she, it*), whereas the plural has only one (*they*) (Croft, 1990).

paradoxes”. An example of such a paradox can be found e.g. if one first establishes a markedness relation between the singular and the plural form of a noun, and then, in attempting to do the same in a language containing the *dual* as well, one might be forced to conclude that the plural is the unmarked form compared to the dual. Thus the plural is at the same time marked and unmarked (Croft, 1990).

A strict binary distinction of markedness proved insufficient in a number of cases, thus a softer version was introduced through scales or hierarchies. These hierarchies express the relative prominence of a structure, and incorporate the relativity of markedness into the theory. The notion of prominence has been linked to several properties such as most likely as topic, agent, most available referent etc. Among the hierarchies established in typological literature are those of syntactic functions, animacy and definiteness.³ Prominence and markedness are connected concepts, but may not be equated. This point is illustrated by the person hierarchy, for instance, where the third person is considered to be the unmarked form, whereas the local persons certainly are the most prominent. What we will see, however, is that the features expressed through prominence hierarchies play an important role in several phenomena, and in doing so, express general markedness patterns. As we shall see later on, features placed high on one hierarchy tend to attract other prominent or high-placed features; subjects, for instance, will tend to be animate and definite.

Syntactic Functions

In typological linguistics syntactic functions are attested through cross-linguistic comparison, thus examining how they are expressed in various languages and relating these facts to markedness criteria. Croft (1990) states the hierarchy of syntactic functions as follows:

- (1) Subject > Object > Oblique

As we have already mentioned, syntactic functions may be expressed differently in various languages, e.g. through case, verb agreement and word order. In the case of word order, the prominence relation expressed by the hierarchy is also one of *iconicity*, where basic word order in the sentence tends to follow the hierarchy of syntactic functions (Croft, 1990). In this respect then, prominence on the scale of syntactic functions seems to incur prominence or precedence in the syntactic tree.

³Croft (1990) also mentions a hierarchy of number, but this will not be reviewed here. Another common prominence hierarchy is that of thematic roles. An example of such a hierarchy is provided in section 3.6 below.

Animacy

Comrie (1989) differentiates clearly between the dimensions of semantic or pragmatic roles, that relate a given noun phrase to its verb or the discourse as a whole, and properties that are inherent properties of a noun, such as definiteness and animacy. These do however interact, as inherent properties of a noun are sure to influence the types of semantic or pragmatic roles it will receive. As we will see, it also influences the expression of syntactic functions. Animacy is one such inherent property, an “extra-linguistic conceptual property” (Comrie, 1989), which creates distinctions found in many languages. Comrie (1989) states the hierarchy as follows:⁴

- (2) Human > Animate > Inanimate

Evidence for this hierarchy comes again from cross-linguistic examination of the realization of animacy in different languages, and especially in how animacy motivates morphological “splits” in various ways. An often cited example of this is found in inverse languages, such as Navajo, which grammaticalizes the tendency in languages to place the subject of a transitive clause higher in animacy than the object. Here, the inverse, more marked form, expressed by the verbal affix *bi* is employed, when the subject is lower in animacy than the object. Dahl and Fraurud (1996, p. 50) provide the following example:

- (3) 'ashkii at'éé yiníí
the-boy the-girl he-is-looking-at-her
'the boy is looking at the girl'

The above example shows the unmarked SOV word order, where the verb has the prefix *yi*. However, when the subject is inanimate and the object animate, a passive-like word-order is obligatory and the verb is inverse (Dahl and Fraurud, 1996), and marked with the verbal affix *bi*:

- (4) 'ashkii k'asdáá tó biisxí
the-boy almost water it-killed-him
'the boy was almost killed by water' = 'the boy nearly drowned'

This shows us one way of expressing the typological markedness of an inanimate subject and an animate object, where there is a coalescence between a formal, morphological marking and a conceptual, typological markedness. This type of situation is one of *iconicity*, often contrasted with *economy* of expression (Croft, 2003, p. 101). Languages will to differing degrees lean towards the one or the other with regards to expression of typological markedness.

⁴Comrie (1989) calls the middle category in the hierarchy *Animal*, whereas Aissen uses the term *Animate*. We will follow Aissen on this.

We also find the phenomenon of Differential Object Marking (DOM) to varying degrees in different languages, where objects high in animacy tend to have case marking, thus rendering them overtly marked. We will take a closer look at the phenomenon of DOM below, when reviewing Aissen's theory on the matter.

Definiteness

Definiteness as used in this context is a semantic property which can be marked morphologically in a language in different ways. Croft (1990) and Comrie (1989) both emphasize the importance of definiteness in various phenomena, and thereby expressing markedness relations. The following prominence hierarchy is presented (Croft, 1990):⁵

- (5) Definite > Referential Indefinite > Non-referential Indefinite

Central to a notion definiteness is the property of *identifiability* (Lyons, 1999): the speaker assumes that the hearer is familiar with the referent, or, based on the situation of the utterance, the previous discourse or general background knowledge, the hearer is able to work out the referent of the noun. Another often mentioned characteristic of definiteness, is that it involves an implication of *uniqueness*, i.e. that the referent is in some sense unique in a certain context. An example taken from Lyons (1999) illustrates this:

- (6) I've just been to a wedding. *The bride* wore blue.

Clearly, this is not a matter of the hearer identifying the referent of the definite noun phrase, but rather acknowledging that there is usually only one bride at a wedding, a fact which paves the way for a definite form. Uniqueness as a criterion may be stretched to cover mass and plural nouns under the term *inclusiveness* (Lyons, 1999), which means that a definite refers to the whole set denoted by the noun, not a subset of it. Due to this property, a sentence making use of a definite plural noun will sound odd or even contradictive when the ensuing discourse implies that only a subset is relevant:

- (7) ?I've read the books you gave me, but only a few of them

As for indefinites, they are often used to introduce a new discourse referent, thus are not assumed to be familiar or identifiable to the reader. With re-

⁵As we shall see later on, Aissen makes use of a more fine-grained scale for definiteness, incorporating information regarding type of DP, as well as a notion of specificity, which is certainly related to Croft's notion of referential indefinite:

Definiteness scale : Personal Pronoun > Proper Noun > Definite NP > Indefinite Specific NP > Indefinite Non-Specific NP (Aissen, 2003, p. 437)

gards to the property of uniqueness, indefinites may go both ways, depending on the context.

Interaction between the above hierarchies

As already mentioned, the prominence hierarchies presented above are often related by participation in different phenomena. A generalization that will be predominant in the following, is that arguments bearing high-ranking properties on one scale tend to be realized together with high-ranking properties on another, and the reverse for low-ranking elements. Also, if this generalization does not hold, a tendency towards explicit formal marking is prevalent.

The important role of animacy with regards to syntactic functions has been noted several places. Firstly, the distinction between ergative and accusative case marking depends heavily on the animacy of the referents of arguments. Some split-ergative languages employ both ergative and accusative case-marking, depending on the animacy of the arguments. The absolutive-ergative distinction treats the subjects of intransitive clauses together with the objects of transitive clauses as the unmarked arguments, whereas the nominative-accusative distinction treats the subjects of both intransitive and transitive clauses uniformly, whereas the object is in opposition. Languages where this split is present, thus have the possibility of overtly marking both subjects and objects of transitive clauses. This is where animacy comes in, governing the marking of arguments, so that a semantically marked member is also marked overtly by morphological case. Thus, an inanimate subject of a transitive clause is marked with ergative case, whereas an animate object is marked with accusative case. Animacy is not, however, the only decisive factor in split-ergative languages⁶. The features represented in the animacy hierarchy often interact with other inherent features of an argument. Croft (1990) in fact incorporates information from two other scales into his hierarchy of animacy, namely the scales for person and DP-type.⁷ Comrie (1989) also relates the hierarchies of animacy and syntactic functions, stating that:⁸

the most natural kind of transitive construction is one where A is high in animacy and definiteness, and the P is lower in animacy and definiteness; and any deviation from this pattern leads to a more marked construction (Comrie, 1989, p. 128)

⁶The category of person is also central to the morphological split in many split-ergative languages, like Dyirbal (Aissen, 1999).

⁷Croft's hierarchy of animacy:

- (1) first, second person pronouns > third-person pronouns > proper names > human common nouns > nonhuman animate common nouns > inanimate common nouns

⁸Comrie uses S for the subject of an intransitive sentence, A for the subject of a transitive sentence and P for the corresponding transitive object.

In the section on Judith Aissen’s theory of markedness (cf. section 3.4 below) we will see an alternative and perhaps formally more explicit method of combining hierarchies in order to express these markedness generalizations. We will also see that the definiteness hierarchy seems to be inescapably related to the animacy hierarchy, especially when it comes to case marking of objects.

3.2 Topic, focus and related notions

First a clarification regarding levels of linguistic interpretation. The notions of topic and focus can be viewed from both a syntactic and a more discourse-oriented angle. From a syntactic angle, the topic of a sentence is simply a structural position, SpecCP. It is the position filled by topicalization or wh-movement. Obviously, these structural criteria are not universal, but portray the situation in Norwegian and many other languages. The pragmatic side of these notions is certainly more complex, and has been the subject of an extensive amount of literature. Some of these ideas will be reviewed below. This is not to say that these levels are completely independent of each other and do not interact. It is often the case that it is the topic of discourse that is in fact placed in the topic-position. However, it can also be quite useful to keep the two levels distinct.

3.2.1 Pragmatic Roles

The notions of topic and focus are often related to the information-flow in sentences:

... the different ways in which essentially the same information, or the same semantic content, can be structured differently to reflect the flow of given and new information. (Comrie, 1989, p. 62)

A common distinction is that between, on the one hand, the *focus* of a sentence, denoting the new information, and on the other hand the *topic*, denoting given information, or “what the sentence is about” (Comrie, 1989, p. 64). Both of these may be differentiated from *ground* material, representing background knowledge or information the speaker assumes to be inferable from the context. Pragmatic roles or functions are, in essence, relational, in that they denote a relationship between a verb and its arguments, rather than being an inherent property of noun phrases, like animacy and definiteness.

Norwegian, in contrast to English or the Romance languages has two different types of fronted topics, *contrastive* and *continuous* topics (Engdahl, 1997)⁹.

⁹Engdahl (1997) deals with relative clause extractions in Scandinavian and her findings are therefore relevant for Norwegian.

Contrastive topics A contrastive topic typically picks up on a previously introduced entity and contrasts it with another entity. An example using a question-answer method illustrates this point (the contrastive element is boldfaced):¹⁰

- (8) a. Hva syns du om guttene?
 what think you about boys-DEF?
 ‘What do you think of the boys?’
 b. **Ole** liker jeg, men **Per** kan jeg ikke fordra
 Ole like I, but Per can I not stand
 ‘Ole I like, but Per I cannot stand’

The contrastive topics *Ole* and *Per* are part of the previously introduced set of boys, thus therefore topics. As we can see, contrastive topics may be translated with a topicalized structure in English as well. Both English and the Romance languages have methods for fronting contrastive topics (Engdahl, 1997).

A similar, but distinct, construction is the contrastive focus construction (Engdahl, 1997):

- (9) a. Hva liker du å drikke?
 what like you to drink
 ‘What do you prefer to drink?’
 b. **Kaffe** liker jeg, men ikke te
 Coffee like I, but not tea
 ‘Coffee, I like, but not tea’

Here the contrasted element is not related to any given discourse element, as it is the answer to the ‘what’ in the above wh-question. These two constructions seem difficult to differentiate in isolation, i.e. independent of a question-answer dialogue or a larger context.

Continuous topic

- (10) a. I går kjøpte jeg en jakke
 yester day bought I a jacket
 ‘Yesterday, I bought a jacket’
 b. Den skal jeg ha på meg i kveld
 it will I wear on me tonight
 ‘I am wearing it tonight’

The topic *den* ‘it’ in (10b) is in no way in contrast to the antecedent *jakke* ‘jacket’, but rather creates cohesion by referring back to the already introduced referent. In contrast to the contrastive topics, continuous topics do not

¹⁰The examples in (8) - (11) are adapted from Engdahl (1997).

have a fronted counterpart in English or the Romance languages (Engdahl, 1997).

Stress is the major differentiating factor between contrastive and continuous topics. Engdahl (1997) provides the following example, in contrast to (10) above:

- (11) a. I går kjøpte jeg en jakke
yester day bought I a jacket
'I bought a jacket yesterday'
- b. **Den** skal jeg ha på meg i kveld, ikke dongerijakken
it will I wear on me to night, not jeansjacket-DEF
'That one, I am wearing tonight, not the jeans jacket'

3.2.2 Heaviness and newness

As we have already seen, topics tend to denote given information. It follows from this then, that given information will usually be positioned earlier in the sentence than new information. Another factor, which has also been noted to figure in information structuring in the sentence, is the relative heaviness of a constituent.

Arnold et al. (2000) report of a study where the correlation between newness and heaviness is explored through a quantitative corpus-study and an elicitation experiment. The study focuses on the effects of newness and heaviness when it comes to word order variation in speech, in this case for English heavy NP-shift and dative alternation. They report that new information tends to be placed later in the sentence than old or given information, and that heavy constructions, i.e. long, complex constructions, are usually also placed towards the end of a sentence:

...items that are new to the discourse tend to be complex, and items that are given tend to be simple. (Arnold et al., 2000, p. 34)

3.2.3 Related notions

The idea that the notion of topic, especially, is closely linked to other aspects of DPs has been noted several times in the literature. Givón (1976) attributes a DPs *topicality* to the likelihood for its becoming topic of a sentence. It follows then, that DPs differ in their topicality, in accordance with other properties. The hierarchies he mentions are in close correspondence to some of the ones discussed above (given in parenthesis) (Givón, 1976):

- Human > Non-Human (Animacy)

- Definite > Indefinite (Definiteness)¹¹
- More Involved Participant > Less Involved Participant (Thematic Roles)
- 1st Person > 2nd Person > 3rd Person

He does not explicitly mention a hierarchy of syntactic functions, but links the subject function to the discourse role of topic and the semantic role of Agent, pointing to a “consistent and highly universal pairing of the discourse function “topic” and the semantic function “agent” into one highly universal *aggregate* function “subject”” (Givón, 1976, p. 152, author’s emphasis). He points to the fact that there is also a strong link between a high ranking of definiteness on the one hand, and subjecthood on the other: “subjects tend to be overwhelmingly definite and referential even in languages which tolerate, at the “competence” level, indefinite subjects”. Conversely then, Givón (1976) observes in objects a tendency towards indefiniteness and also inanimacy, and points to this as one reason for why they are more rare as topics. Once again, then, we encounter the idea that high-ranking elements on one prominence hierarchy tend to attract prominent elements of other prominence hierarchies.

3.3 Optimality Theory (OT)

In the following we will take a closer look at Optimality Theory (OT) in its most general form. OT is a central framework in this thesis, and constitutes the backbone of stochastic OT, as presented later on. We will see how the framework of OT allows us to represent the generalizations from typological linguistics regarding markedness relations within a formally unified approach.

3.3.1 The emergence of Optimality Theory

OT has developed from generative linguistics, but differs from it in several fundamental aspects. Both of these theories seek to capture the universality of certain language phenomena, as well as language-particularities within one and the same model.

According to the Principles-and-Parameters approach of generative grammar a language either has a certain property or it does not. Universality is captured through a set of principles assumed to be part of Universal Grammar (UG), whereas variation between different languages is modeled by the binary settings of so-called parameters. OT also assumes universality, through a universal set of constraints. A crucial difference between the two approaches, however, resides in the fact that the constraints of OT

¹¹Givón (1976) states explicitly that he views definiteness as an encoding of new vs. given information.

are *violable*. Whereas the P&P approach is based on a view of universality as absolute, OT approaches the issue of universality through the notion of markedness, which is, as we have seen above, a relative notion.

3.3.2 Ranked constraints

OT is not a linguistic theory of representations, such as e.g. generative grammar or LFG, but rather a theory based on interactions of *constraints*. Kager (1999) defines a constraint as a “structural requirement that may be either satisfied or violated by an output form” (Kager, 1999, p. 9). OT typically operates with two main types of constraints, markedness constraints and faithfulness constraints.

Markedness vs. faithfulness

A theory of markedness, as seen above, focuses on an opposition between marked and unmarked linguistic structures, where the unmarked is the preferred form cross-linguistically and the marked is one which is avoided if possible. The function of the marked structure is to mark a contrast. OT is in many ways an implementation of markedness, a notion which is pivotal in an OT grammar. Markedness is expressed through constraints, stating what is marked and unmarked in languages, e.g. ‘Topics come first’. Faithfulness constraints serve to preserve coherence between the input and the output of an OT-grammar, e.g. stating that some feature of the input should be preserved in the output. Markedness and faithfulness constraints are thus in conflict, mirroring a tension present in language between avoiding contrast (markedness) and preserving it (faithfulness) (Kager, 1999).

Ranking of universal, violable constraints

All constraints in an OT-grammar have the properties of universality and violability in common. They are thought to be part of Universal Grammar, however, the constraints are not necessarily active in every language and they are not absolute principles, but may be violated by a language.

An important aspect of the constraints in OT is the fact that they are ranked with respect to one another. The rankings of constraints create language-particularity, languages may differ with respect to which constraints they rank high or low, but the constraints themselves are thought to be universal. This explains why some constraints may not be active in a language, as they may be ranked so low as to not really participate in the grammar at all.

3.3.3 An input-output device

An OT-grammar is, in essence, an input-output device, i.e. for a specific input it generates the optimal, most harmonic output. The notion of *optimality* is a central one in OT. One speaks of an optimal output, as being an output incurring the fewest, least serious violations of constraints. Importantly, an optimal output will not be perfect, i.e. have zero violations, due to the idea that constraints are in conflict, and satisfying one constraint will usually cause the violation of another.

The form of the input to a grammar will typically differ from linguist to linguist. Usually, however it will consist of a minimal representation of a sentence, for instance represented by the verb and its argument structure and usually, some semantic properties of the arguments. The input may take the form of representational devices from other theoretical frameworks, e.g. underspecified LFG f-structures, as OT is not a representational framework.

Possible output candidates are thought to be generated by GEN(erator), a function which generates all possible candidates from a given input, or underlying form (Kager, 1999, p. 19):

$$(12) \text{ GEN(input)} \Rightarrow \{\text{cand}_1, \text{cand}_2, \dots, \text{cand}_n\}$$

An important notion which pertains to GEN is that of “the richness of the base”, i.e. that no constraints restrict underlying forms. All linguistic generalizations are made with regards to the level of output. GEN can freely generate “all logically possible” output candidates, as long as these represent licensed structures from the different linguistic levels, i.e. taking into account prosodic structures or X-bar theory.

The second main component of the input-output device of an OT-grammar is the EVAL(uation) component. Here, the different candidates, as generated by GEN, are evaluated with regards to the constraints, and an optimal output is arrived at through repeated evaluation¹² (Kager, 1999, p. 19):

$$(13) \text{ EVAL}(\{\text{cand}_1, \text{cand}_2, \dots, \text{cand}_n\}) \Rightarrow \text{output}$$

An OT evaluation is usually illustrated graphically by a so-called *tableau*. An example tableau, taken from Boersma (1999), is provided below, where the input is an underlying phonological form and the output an overt phonological form:

(14) An OT tableau

pat	NoCODA	PARSE
pa		*
pat	*!	

¹²The repeated evaluation goes like this: first all candidates are evaluated with regards to the highest ranked constraint. The ones that do not violate it, proceed to the next round, where these are evaluated with regards to second-highest constraint etc. (Kager, 1999, p. 22)

We see that a candidate’s violation(s) of a certain constraint is marked with an asterisk, where the crucial violation (the violation which ultimately excludes one candidate) is marked with an exclamation mark. The optimal candidate is indicated with a ‘hand’ symbol, and shaded cells represent violations that are not relevant to optimization. Thus, in the tableau above two constraints are ranked - NOCODA, a markedness constraint stating that a syllable should not have a coda, and PARSE, a standard faithfulness constraint, which states that all input must be parsed, i.e. be present in the output.¹³ For the input *pat* then, two candidates are generated, *pa* and *pat*, where the first candidate becomes the optimal output after evaluation. The reason for this is that the competing candidate, *pat*, violates a higher-ranked constraint (NOCODA) than the winning candidate *pa*.

As mentioned above, constraints in an OT grammar are ranked in a hierarchy of dominance, related through *strict domination* (Kager, 1999, p. 22):

Strict domination: Violation of higher ranked constraints cannot be compensated for by satisfaction of lower-ranked constraints.

What becomes evident here then, is that the seriousness of a violation, i.e. the ranking of the violated constraint(s), is crucial for the outcome of an evaluation. An illustration by way of example tableaux will clarify this central point:

(15) Strict domination 1

	C ₁	C ₂
☞ candidate1		**
candidate2	*!	

(16) Strict domination 2

	C ₁	C ₂	C ₃
☞ candidate1		*	*
candidate2	*!		

What we see illustrated in the above tableaux are two implications of the definition of strict dominance that might not be immediately clear. In (15) we see that candidate 1 is the optimal candidate, even though it violates one constraint, C₂, twice, whereas candidate 2 violates C₁ only once. The reason for this then, is that C₁ is ranked higher than C₂, and a violation of a high-ranking constraint is more severe than several violations of a lower ranked constraint. Another point, as illustrated by tableau (16), is that violations of several different constraints do not compensate for a single violation of a higher ranked constraint, so there is no adding of different violations.

¹³It is important to remember that OT constraints do not represent absolutes, i.e. a syllable with a coda may emerge as the optimal candidate even if this constraint is present in the grammar. In a larger grammar, other higher-ranked constraints may exclude opposing candidates.

3.4 Aissen’s theory of markedness through harmonic alignment

In her 1999 article *Subject Choice and Markedness in Optimality Theory* Judith Aissen presented an analysis of linguistic factors influencing subject choice in a number of languages, which made use of well established hierarchies from typological linguistics, and most importantly, ordered combinations of these, in order to express markedness relations. She showed how Optimality Theory with its rankings and constraints could provide a formalized, uniform account of these relations. In the following we will look more closely at her newest article, *Differential Object Marking: Iconicity vs. Economy* (Aissen, 2003), which also explores the possible insights that combinations of typological hierarchies might provide, in regards to the form a syntactic function might take, in this case that of the direct object. The reason why we focus in particular on this article is that it deals with exactly the dimensions of prominence that we are interested in, namely those of animacy and definiteness. Aissen’s work has influenced a wave of literature in syntactic Optimality Theory since, and constitutes the formal and theoretical starting point for the work of this thesis. We will therefore take the time and space to review Aissen (2003) in some detail.

3.4.1 DOM

In many case-marking languages there seems to be a close correlation between certain types of semantic or pragmatic information and overt case marking on objects. This is the phenomenon Aissen calls Differential Object Marking (DOM):

The higher in prominence a direct object, the more likely it is to be overtly case marked. (Aissen, 2003, p. 436)

Prominence is expressed through well-known prominence hierarchies or scales from typology, namely a definiteness scale, an animacy scale and a scale for syntactic functions:

Animacy scale: Human > Animate > Inanimate

Definiteness scale: Personal Pronoun > Proper Noun > Definite NP > Indefinite NP¹⁴

Syntactic functions scale: Subject > Object

¹⁴Aissen’s original scale splits the ‘Indefinite NP’ element into ‘Specific Indefinite NP’ which is ranked above ‘Non-specific Indefinite NP’. The notion of specificity will be discussed later on (cf. chapter 5). We will not, however, operate with this distinction in the following.

Aissen observes a so-called *markedness reversal* between the syntactic functions of subject and object which is such that what is unmarked for a subject is marked for an object and vice versa, i.e. subjects tend to be animate and definite whereas objects tend to be inanimate and indefinite. DOM is encountered, then, when objects deviate from this norm.

What Aissen is aiming for then is a theory that generalizes over DOM making use of typological hierarchies, while at the same time accounting for language-particular variation. Optimality Theory combined with the typological generalizations contained in the prominence hierarchies provides the formal framework, as well as capturing the universality through constraints.

3.4.2 Formal Framework

Harmonic alignment

Aissen (1999) makes use of *harmonic alignment*, a technique imported from phonological OT, which provides a method for expressing the markedness of various combinations of prominence hierarchies. She aligns the prominence hierarchy of syntactic functions with the hierarchies of animacy and definiteness to yield subhierarchies of constraints which are put actively to use in the analysis. Harmonic alignment aligns the dominant elements of a scale with the dominant elements of another and the lower ranked elements of one scale with the lower ranked of another, expressing the idea that prominence on one scale will attract prominence on another. Markedness constraints are derived by reversing the output scales from the alignment and adding the ‘avoid’-marker, ‘*’, to them:

Suppose given a binary dimension D1 with a scale $X > Y$ on its elements $\{X, Y\}$, and another dimension D2 with a scale $a > b \dots > z$ on its elements. The harmonic alignment of D1 and D2 is the pair of harmony scales:

H_x : $X/a > X/b > \dots > X/z$

H_y : $Y/z > \dots > Y/b > Y/a$

The constraint alignment is the pair of constraint hierarchies:¹⁵

C_x : $*X/z \gg \dots \gg *X/b \gg *X/a$

C_y : $*Y/a \gg *Y/b \gg \dots \gg *Y/z$

(Prince and Smolensky 1993, as quoted in Aissen (2003, p. 441))

Aissen aligns the scale of syntactic functions, Subject $>$ Object, with the scales of animacy and definiteness in turn, in order to achieve the relevant constraints.

Harmonic alignment of the syntactic function scale and the animacy scale gives us:

¹⁵The ‘ \gg ’ operator marks the fact that we are now dealing with constraints which adhere to a specific internal ranking, i.e. we are dealing with a *subhierarchy*.

- (17) a. Su/Hum > Su/Anim > Su/Inan
 b. Oj/Inan > Oj/Anim > Oj/Hum

The above alignment reads something like “a human subject is less marked than an animate subject etc.”. These new scales are then reversed in order to yield the appropriate markedness-constraints:¹⁶

- (18) a. *SU/INAN ≫ *SU/ANIM ≫ *SU/HUM
 b. *OJ/HUM ≫ *OJ/ANIM ≫ *OJ/INAN

The constraint subhierarchy above reads as follows: “avoiding an inanimate subject is more important than avoiding an animate subject etc.”, or “an inanimate subject is more marked than an animate subject” etc.

The definiteness¹⁷ hierarchy is also aligned with the syntactic functions hierarchy:

- (19) a. Su/Pro > Su/PN > Su/Def > Su/Indef
 b. Oj/Indef > Oj/Def > Oj/PN > Oj/Pro

This provides us with the following constraints:

- (20) a. *SU/INDEF ≫ *SU/DEF ≫ *SU/PN ≫ *SU/PRO
 b. *OJ/PRO ≫ *OJ/PN ≫ *OJ/DEF ≫ *OJ/INDEF

These constraints are thus markedness constraints, expressing universal markedness patterns. Also, they are ranked internally with respect to each other in a subhierarchy. We see then, that through a simple, yet fully explicit procedure, we may in fact derive constraint subhierarchies from the prominence scale from typological linguistics. The resulting constraint subhierarchies of syntactic functions and animacy and definiteness thus predict that the most marked construction, or the construction most to be avoided, would be one which had in it either an inanimate subject and/or a human object or an indefinite subject and/or a pronominal object, or even a combination of the two.

Additional constraints

Returning again to the phenomenon of DOM, we see that the present constraints will penalize any marked object. It is not the case, however, that

¹⁶Here, and in the following, we will represent Optimality Theoretic constraints in small capitals.

¹⁷Aissen considers it important to define definiteness irrespective of a language’s particular morpho-syntax, thus opting for a discourse-oriented definition:

the relevant scale has to do with the ‘extent to which the value assigned to the discourse referent introduced by the noun phrase is fixed’

Aissen (2003, p. 444) This coincides quite well with the criteria of identifiability and uniqueness, as presented in Lyons (1999).

these marked object constructions are non-existent. These are precisely the constructions which give rise to DOM. What is missing from the analysis, then, is an expression of the relation between morphological complexity and markedness. The constraint $*\emptyset_c$ (“star zero”) which penalizes null morphological expression of case will provide exactly that. In order to introduce this constraint into the constraint subhierarchies derived by harmonic alignment, while at the same time preserving the individual rankings, Aissen makes use of the technique of *local conjunction*:

The local conjunction of C_1 with subhierarchy $[C_2 \gg C_3 \gg \dots \gg C_n]$ yields the subhierarchy $[C_1 \ \& \ C_2 \gg C_1 \ \& \ C_3 \gg \dots C_1 \ \& \ C_n \gg]$. (Aissen, 1999, p. 698)

The local conjunction of $*\emptyset_c$ with the subhierarchies for syntactic function/animacy and syntactic function/definiteness thus express the idea that marked associations should be marked morphologically, e.g. $*OJ/HUM \ \& \ *\emptyset_c \gg *OJ/ANIM \ \& \ *\emptyset_c \gg \dots$, i.e. it is more important to avoid a null-marked human object than a null-marked animate object etc.. In order to violate a locally conjoined constraint, then, a candidate must violate both conjuncts at least once.

The analysis as it stands, however, would be too radical in that it would rule out all objects which are not marked morphologically for case. A final constraint is introduced to improve this, namely $*STRUC_C$, which is a version of a more general constraint penalizing structure, i.e. enforcing economy of expression. In this version it penalizes morphological case. The function of this constraint will be to delimit the penalization of null morphological case. The constraint will have different rankings in various languages, depending on the realization of DOM in these languages. I.e. if only object pronouns are morphologically marked in a language, it would be ranked as follows:

(21) $*OJ/PRO \ \& \ *\emptyset_c \gg *STRUC_C \gg *OJ/PN \ \& \ *\emptyset_c \gg \text{etc.}$

This constraint ranking would read that “it is more important to avoid a pronominal object that has null marking for case than it is to avoid morphological structure, but it is more important to avoid morphological structure than it is to avoid a proper name object with null marking for case etc.” This would result in a penalization of pronouns that do not have case marking, whereas proper names (and all lower ranked categories) would not be penalized for null marking.

3.4.3 The realization of DOM

Aissen goes on to provide a number of examples of DOM’s realization in different languages, together with the formal OT-analysis of DOM in these languages. This illustrates the interplay between iconicity and economy when it comes to the various realizations of DOM in different languages. On the

one hand, some languages lean strongly towards economy of expression, thus case marking very few objects, while others display a tendency for iconicity thus case marking all but very few objects.

Two types of DOM

Aissen differentiates between two main types of DOM, one-dimensional DOM and two-dimensional DOM. One-dimensional DOM describes a situation where either animacy or definiteness in isolation determine DOM in a language. The differences between these languages are modeled through differing placement of $*\text{STRUC}_c$ with respect to the other constraints in the two hierarchies. Through the interpolation of the $*\text{STRUC}_c$ at different places in the subhierarchies, she is able to illustrate the typological generalization, that if a language marks one element on the scales, then it will also mark the higher ranked elements. The languages thus differ in so-called “cut-off” points.

A language where both hierarchies of animacy and definiteness in combination determine DOM exhibits two-dimensional DOM. Aissen merges the two relevant hierarchies by taking the cross product of them, thus deriving a tree containing all possible combinations of the two, while maintaining the internal rankings. Thus, the root node of the tree is inhabited by ‘Human Pronoun’, the most marked object possible, which in turn dominates ‘Human Name’ and ‘Animate Pronoun’ etc. At the bottom of the tree we find ‘Inanimate Non-specific Indefinite’, as the least marked object possible. The language-particular variation when it comes to two-dimensional DOM is also modeled as being relative to the placement of the $*\text{STRUC}_c$ -constraint within the tree structure described above. This divides the tree into three different zones, which might or might not be present in a language: i) the nodes that dominate $*\text{STRUC}_c$, i.e. with obligatory case marking, ii) The nodes that neither dominate nor are dominated by $*\text{STRUC}_c$, but which are surrounded by a ‘floating $*\text{STRUC}_c$ ’, i.e. optional case marking, and iii) the nodes that are dominated by $*\text{STRUC}_c$, i.e. null case marking.

3.4.4 Conclusion

Aissen has very convincingly demonstrated that Optimality Theory offers the possibility to account for what seems to be a universal principle, DOM, whilst at the same time allowing for extensive language-particular variation. We have seen that languages to a varying degree lean towards iconicity of expression on the one hand, and economy of expression on the other.

In Aissen (2003) the main focus has been on the syntactic function of object. In Aissen (1999), the focus is on the subject and its realization in languages where the active/passive distinction is categorically conditioned

by markedness relations conditioned by the person hierarchy¹⁸. However, in employing the syntactic functions hierarchy (Subject > Object) in her analysis of DOM, she also makes a claim about the subject, thus expressing the mentioned ‘markedness reversal’. One would expect to find a mirror of DOM, i.e. if any subject should be marked, that would most likely be an inanimate, indefinite subject. So-called DSM-languages have also been attested,¹⁹ and some languages, like the split-ergative languages, must be said to have the possibility for both DOM and DSM.

As we have seen however, Norwegian is not a language that to any large degree makes use of case to mark syntactic functions. It does however mark pronominal objects to a certain extent, thus displaying a marginal DOM. As mentioned above, however, there seems to be a split in marking that adheres to person when it comes to accusative marking of pronouns, wherein the local persons, i.e. the most prominent persons, must be case-marked when functioning as objects.

Even so, the main ideas regarding properties of transitive arguments, typological markedness relations and the formal machinery outlined above, will prove essential in the following chapters.

3.5 Stochastic OT

3.5.1 Modeling variation in language

An aspect of traditional OT, as described in section 3.3 above, is its inability to model variation in a principled manner. In OT terms, variation occurs in a situation where one and the same input may result in different outputs at different evaluations. Traditional OT will always provide the same optimal candidate for the same input, given identical grammars. Within traditional OT variation has been modeled using so-called floating constraints. Here, the constraints that compete regarding the variable output are unranked with respect to each other. This, however, results in a situation where the possible outputs vary with completely equal probability, a scenario which is usually an excessive idealization of the phenomenon in focus. Below, we will examine an alternative OT model, which makes use of a stochastic component in order to reflect actual frequencies of variation in a language in the ranking of OT constraints.

One of the key differences between traditional OT and stochastic OT resides in the fact that constraints in stochastic OT are ranked along a *linear* scale, i.e. constraints are not only ranked with respect to each other, but they are also ranked a specific length apart, and have a real-number value.

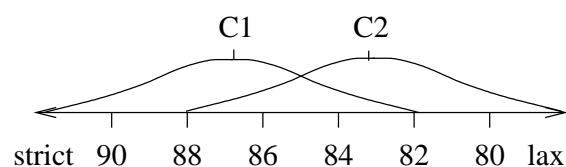
¹⁸The person hierarchy is the binary hierarchy: Local (1st and 2nd) > 3rd

¹⁹The Papuan language Fore (Donohue, 1999) is a typical DSM language, where the DSM is conditioned by animacy. Here inanimate subjects always receive morphological case.

The distance between constraints is important, a short distance between two constraints indicate that their rankings are less fixed with respect to each other.

In order to allow for variation within a language and specific constructions, each constraint in a stochastic OT grammar is associated with two different values - a *ranking value* and a *selection point*. The ranking value constitutes a constraint's more permanent value²⁰ through several evaluations, whereas the selection point is the value chosen for a constraint at one single evaluation. At the time when the constraints are to be evaluated, the specific placement of each constraint on the scale (its ranking value) is disrupted somewhat by adding a random value of *evaluation noise* to each constraint, thus the constraints simulate covering a range of points on the scale, rather than just one. The noise added to each constraint is the same for all constraints and represents a probability distribution for the selection of a specific constraint. The selection point at one evaluation is chosen from within the range created by the adding of noise to the ranking value. The resulting range follows a normal (Gaussian) distribution, i.e. a range which peaks in the middle and goes downward from there, with the ranking value as its mean, thus being the most probable selection point and probability sinking as the curve diminishes. At this point in the evaluation there are two possibilities, either the ranges of some constraints overlap, indicating that there will be some variation between the two, as in (22) below, or they do not overlap and go back to their original ranking value.

(22) Overlapping ranking distributions (Boersma and Hayes, 2001)



The point is that constraints whose ranges overlap will display a certain amount of variation because the selection point, i.e. the value chosen for a constraint at one evaluation, may be picked from anywhere within a range according to its probability. So, in the figure in (22) above, most of the evaluations will rank C1 above C2, however, since their ranges do overlap, a low value on C1 and a high value on C2 will result in a reversal of the rankings of the two constraints. This will be a rare case because the curve of the range also says something about the probability of a certain point being chosen. However, the fact that the ranges do in fact overlap indicates that variation will occur.

²⁰The ranking value may however be changed, through constraint promotion or demotion, as we shall see below.

It is certainly not the case that all constraints overlap. The further two constraints are apart, the less is the probability for their rankings ever to be reversed. Constraints which are far apart thus contribute to obligatory rankings, mirroring categorical aspects of a language.

3.5.2 The Gradual Learning Algorithm

The Gradual Learning Algorithm (GLA), as described in Boersma and Hayes (2001), “tries to locate an empirically appropriate ranking value for each constraint” (Boersma and Hayes, 2001, p. 51). This approach then is a learning approach, where, based on a sufficient amount of data and a number of evaluations, the algorithm supplies a ranking of constraints which takes into account observed variation in a language. The GLA thus describes an implementation of stochastic OT.

In the initial state of the algorithm, every constraint involved is given the same value, which may be chosen arbitrarily as long as it is the same for all. The algorithm is then fed a learning datum. The learning datum is part of the data set fed to the algorithm, which consists of all relevant input forms with frequencies for the possible output forms. The algorithm then attempts to generate from the underlying input form based on the grammar and the constraints involved. The learning datum and the generated form are then compared. If the generated form matches the learning datum, no adjustments are made. However, as the algorithm is error-driven, the grammar is adjusted if there is a mismatch between the learning datum and the generated form, so that it will be capable of generating the learning datum correctly in the next evaluation. Adjustments to the grammar are made in two steps:

1. So-called *mark cancellation* is performed, i.e. all the violations which the two forms have in common are canceled, as these obviously did not influence the evaluation either way.
2. The constraints that penalized the learning datum in the evaluation are demoted (their ranking values decreased), whereas the constraints that penalized the generated form are promoted (their values increased).

After this, more data are added, and the algorithm repeats the above steps a number of times. With enough data the rankings of the constraints will mirror variation and actual user tendencies in a language very closely. This is so, because more frequent forms will influence the ranking values to a greater extent than less frequent forms.

3.5.3 The GLA in use

Boersma and Hayes (2001) report of several experiments where they have employed the GLA on phonological data, with good results. However, the

algorithm²¹ has lately also received attention from the syntactic community (Dingare, 2001; Lee, 2002c). Dingare (2001) employs the algorithm in order to examine the effects that hierarchies of definiteness and person might have on passivization in English. The choice between the active and passive voice is not categorically determined in English, thus stochastic OT, with its variation-oriented approach, provides some interesting generalizations. Dingare (2001) makes use of Judith Aissen’s technique of harmonic alignment, as reviewed above, which provides the necessary constraints by aligning hierarchies of syntactic functions with those of definiteness and person. Dingare’s training data are obtained from two corpora, one of written and one of spoken material, and the effects of definiteness and person on frequencies of passivization are studied independent of each other. She concludes that hierarchies of definiteness and person influence the choice between the active and passive voice, and uses stochastic OT to represent this conditioned variation.

The idea that the same patterns of markedness figure in languages to a varying degree certainly makes Optimality Theory enhanced with stochastic methods an interesting approach. Norwegian does not categorically restrict which elements may topicalize, however,

... even if the grammatical distinction is not categorical, it supports the markedness pattern as long as the tendency is in the right “direction”. Markedness, including hierarchies, is a matter of relative degree. (Croft, 1990, p. 111)

3.6 “Reverse-animacy” verbs

3.6.1 Introduction

The main generalization from the typological literature on prominence hierarchies and their role in argument realization, provides us with the generalization that subjects will in a majority of cases be animate and objects inanimate. This, however, is not always the case. Consider the following examples:

- (23) a. Boka *interesserer* jentene
 book-DEF interests girls-DEF
 ‘The book interests the girls’
- b. Saken *gjelder* jentene
 case-DEF concerns girls-DEF
 ‘The case concerns the girls’

²¹The implementation of the GLA discussed below is that of the Praat software (Boersma, 1999), available from <http://www.praat.org>.

- c. Bomben *drepte* jentene
 bomb-DEF killed girls-DEF
 ‘The bomb killed the girls’

All of the above sentences contain an inanimate subject and an animate object. In (23a) we see an example of a psychological verb, *interessere* ‘interest’ or a psych-verb, as it is often called in the literature. We will refer to the reverse-animacy psych-verbs as *amuse*-verbs, following Levin (1993). The verb in (23b) is not part of a larger group already identified, but differs from the two other verbs, as we shall see later on. For lack of a better name, we will call these verbs for *concern*-verbs, due to the fact that many of these translate into English *concern*. Finally, the verb in (23c) is a classical causative verb.

The title of the present section alludes to verbs which exhibit a ‘reversed’ relationship between their core arguments with regards to animacy. This is, as we shall see, closely related to other properties of the arguments of some of these verbs, such as their thematic role. On several levels then, the subject of these verbs will tend to be more object-like than the object itself. A relationship of this type between the arguments, will allow for a reversed relationship between the arguments also with regards to animacy. This group of verbs thus creates problems for our theory based on the alignment of high-ranking elements of one markedness hierarchy with the high-ranking elements of others. The fact that the group is quite limited however does in fact support our theory. In the majority of cases subjects will have high-ranking properties.

We will use the term *reverse-animacy* verbs to refer to verbs that often take an inanimate subject and an animate object. This does not mean that this is the only possible argument realization for them. Neither is this term meant to be a classification of a group of verbs that have any internal semantic resemblance, or that are in some sense a ‘natural’ group of verbs. Rather, the term is meant to cover simply those verbs that have in common the fact that they may occur with an inanimate subject and an animate object. In the following we will try to come up with a closer characteristic of these verbs, which form a rather diverse group. The following treatment is not, however, in any way exhaustive. Rather, it will provide us with a picture of a rather diverse group of verbs, which is by no means closed.

3.6.2 Thematic roles

The possibly reverse relationship between the arguments of these verbs are indeed an idiosyncratic property of the verb itself, and its argument structure. The restrictions the verb places on its arguments thus restrict the domain of possible subjects and objects. Closely related to a verb’s selectional restrictions we find the notion of thematic roles.

Comrie (1989) regards the thematic dimension as a continuum, and markedness hierarchies for thematic roles have been proposed several places in the literature. A classic hierarchy for thematic roles might be given as follows (Bresnan, 2001, p. 307):²²

- (24) Agent > Benefactive > Experiencer > Instrument > Theme > Location

Following up on our earlier assumptions regarding the alignment of hierarchies then, it seems quite clear why the subject, being the highest ranked element on the hierarchy of grammatical relations, usually has the properties of being agent, animate, definite and topic, all of which are properties ranked high on their respective scales.

Lødrup (2000) lists a few of the *amuse* and the *concern* verbs, and point to the fact that these verbs do not passivize in Norwegian.²³

- (25) a. *Jentene interesseres av boka
 girls-DEF interest-PASS of book-DEF
 b. *Jentene gjeldes av saken
 girls-DEF concern-PASS of case-DEF

The reason for this resistance towards passivization resides in the fact that these verbs all have a Theme subject (Lødrup, 2000). Lødrup (2000) differentiates much in the same way as we have done here, between two groups of verbs with a Theme subject. Some examples are provided below (Lødrup, 2000):

- (26) a. *interessere* ‘interest’, *forundre* ‘puzzle’, *behage* ‘please’
 b. *gjelde* ‘concern’, *vedrøre* ‘concern’ *gavne* ‘benefit’

The first group of verbs are psychological verbs, a subset of our *amuse*-verbs, which typically take a Theme subject and an Experiencer object, whereas the second group of verbs, a subset of our *concern*-verbs, take a Theme subject and a Benefactive object (Lødrup, 2000). This reversed relationship between the arguments on the thematic scale, opens for a reverse relationship with regards to animacy, as we have seen.

We may not, however, equate our group of *amuse* verbs and *concern* verbs completely with the two groups of verbs from Lødrup (2000).

Firstly, we will wish to include several psych-verbs on our list of reverse-animacy verbs that do not conform to the criteria of non-passivization. Levin (1993) account for the *amuse*-verbs as follows:

²²Aissen (1999) also makes use of a hierarchy of thematic roles, albeit a simple one: Agent > Patient.

²³Norwegian has both a morphological passive, an s- ending, and a periphrastic passive ‘auxiliary *bli* ‘become’/ *er* ‘is’ + participle’. There is a noted aspectual difference between these two, as is common for languages with several passive forms.

The members of this subclass of the psych-verbs describe the bringing about of a change in psychological or emotional state. They are transitive verbs whose object is the Experiencer of the emotion and whose subject is the cause of the change in the psychological state. (Levin, 1993)

Levin does not explicitly state that the thematic role of the subject for these verbs is in fact Theme. This is due to the fact that her grouping is not as fine-grained with regard to thematic roles and as she herself states, that “some of these verbs [...] allow the subject/stimulus argument to receive an agentive interpretation.” (Levin, 1993) This classification is, however, sufficient for our present purpose, as all the *amuse* verbs of Levin (1993) may take an inanimate subject and an animate object. The psych-verb *more* ‘amuse’, for instance, certainly may have an agentive use, with an animate subject, it may also, however, have an inanimate subject:

- (27) a. Klovnene morer barna ved å danse klumsete
 Clown-DEF amuses children-DEF by to dance clumsily
 ‘The clown amuses the children by dancing clumsily’
- b. Klovnens klumsete dans morer barna
 Clown-DEF-GEN clumsy dance amuses children-DEF
 ‘The clown’s clumsy dance amuses the children’

Due to the agentive aspect of several of the psych verbs included in the *amuse* class, we find that many of these passivize well.

Secondly, our group of *concern*-verbs are not exhaustively classified by a lack of passivization either. Consider, for instance the verb *omfatte* ‘include’ and *innbefatte* ‘include’. These may very well passivize, but seem no different from the other *concern*-verbs. Lødrup (2000) mentions one member of his group from (26b) which does not passivize, namely *romme* ‘contain’. These three verbs might be said to have a spatial quality in common, a factor which differentiates them from the other *concern*-verbs, and might be a reason for their passivization. Either way, they are all reverse-animacy verbs.

The *amuse*-verbs thus have an Experiencer object and a subject which is “the cause of change in psychological state” (Levin, 1993, p. 191). The *amuse* verbs thus have a causative dimension in common with the regular causative verbs, like *drepe* ‘kill’ in (23c) above. The *concern* verbs, however, are not causative in any sense. In the following we will take a closer look at some of the proposals in the literature, especially regarding psych-verbs, which have received a fair amount of attention, in the hope of gaining some more insight into the nature of the reverse-animacy verbs. We will see that causation is a central property to some of these verbs, an insight which recurs within several different frameworks.

3.6.3 An aspectual dimension

Grimshaw (1990) emphasizes the importance of argument structure with regards to syntactic behaviour. She proposes a theory of prominence, wherein the argument structure reflects the lexical semantics of the verb, and in particular, the relations of prominence that exist between them. The external argument, she claims is the most prominent argument. She expresses these prominence relations in her chosen representation of argument structure (Grimshaw, 1990, p.4):

(28) *announce*(Agent(Goal(Theme)))

The above example clearly shows a hierarchical prominence relation, where the more embedded a thematic role is, the less prominent it is. Thus, Agent is the most prominent in the argument structure in (28) above, followed by Goal and lastly Theme.

Far from being marginal, Grimshaw claims that, within her theory, the reverse animacy of the psych-verbs (or *frighten*-verbs as she names them²⁴) is to be expected. She uses this group of verbs which display a Theme subject and an Experiencer object as an example when laying out the proposal for her theory. She points to the asymmetry between the thematic relations of the verbs' arguments and their actual grammatical realization. The Experiencer role is more prominent than Theme, but is still realized as the object, and not as an external argument. In order to account for this, Grimshaw introduces an aspectual dimension into the theory. She compares the *frighten* verb to the verb *fear*, which links to the same thematic roles, but realizes them differently:

- (29) a. *Maria*_{Exp} *fears* *dogs*_{Th}
b. *Dogs*_{Th} *frighten* *Maria*_{Exp}

What she claims then is that it is in fact the aspectual dimension, represented by event structures, which determines the realization of arguments. It is the fact that *frighten* is a causative verb, and that the Theme argument happens to be Cause, which in fact determines the mapping to subject status. The Cause argument will always be the most prominent.

However, many verbs are not causative, yet they still manage to realize a subject. Grimshaw associates the notion of an aspectual dimension with a representation in terms of event-structures. These break events down into subevents, where arguments involved in the first subevent always will be more prominent than arguments involved in the second. Cause is always associated with the first subevent, as it initiates an event.

²⁴The *frighten*-verbs are a subset of our *amuse*-verbs, because, as we have seen, the latter may also have agentive readings, as in (27a) above.

3.6.4 Dowty's theory of thematic roles

David Dowty's seminal 1991 article offers a different view on thematic roles than the theories proposed earlier by among others, Fillmore and Jackendoff. Dowty points to the lack of disagreement as to what thematic roles actually are and also how they should be delimited. As part of a solution he puts forward a theory of thematic roles in order to better account for the relationship between argument selection and thematic roles. As we will see, causation, as in Grimshaw (1990), becomes an important factor on this account as well.

Dowty (1991) introduces his novel approach to thematic roles through the notion of *Proto-roles*. These are roles which do not attempt to display clear-cut boundaries, but rather are clusters of role-features. Arguments, by this hypothesis, are members of a role-type to differing degrees. Only two roles, which are complete opposites of each other, are needed - Proto-Agent and Proto-Patient.²⁵ Dowty (1991) lists the following properties as proto-typical of the P-Agent and P-Patient (Dowty, 1991, p. 572):

1. Contributing properties for the Agent Proto-Role
 - volitional involvement in the event or state
 - sentience²⁶ (and/or perception)
 - causing an event or change of state in another participant
 - movement (relative to the position of another participant)
 - (exists independently of the event named by the verb)
2. Contributing properties for the Patient Proto-Role
 - undergoes change of state
 - incremental theme
 - causally affected by another participant
 - stationary relative to movement of another participant
 - (does not exist independently of the event, or not at all)

All of the criteria mentioned above are thought to be independent, i.e. a candidate may satisfy just one of these and still be accepted as Proto-Agent/Patient. Usually, however, a Proto-Agent/Patient will have more than one of the above characteristics.

In order to explain the distribution of roles to arguments of verbs, Dowty introduces an *Argument Selection Principle*, which states that the arguments with more P-Agent properties, relative to the other argument(s), is realized

²⁵Abbreviated as P-Agent and P-Patient by Dowty (1991).

²⁶Dowty stresses the importance that this property involves sentience of the event, not just sentience in general.

as the subject of the sentence, whereas the argument with more P-Patient properties becomes object. If they have an equal amount of properties from a Proto-role, either argument may be realized as either core function. When it comes to ditransitive verbs, Dowty finds that the non-subject argument with the most Proto-Agent properties is realized as indirect object, whereas the other argument becomes direct object.

Dowty's approach to thematic roles departs from the generative approach in that on his account, roles are characterized by their *nondiscreteness*. What this means is that Proto-roles do not classify arguments

- exhaustively, all arguments do not have to have a role
- uniquely, arguments may have the same role
- discretely, arguments may be equally fit for a role

Only subcategorized arguments of the verb (i.e. not adjuncts) are classified by the Proto-roles.

Psych-verbs

Dowty continues to explicate his new account of thematic roles, by looking at psych-verbs. He examines the difference between so-called *fear* and *frighten* verbs (cf. example (29) above), and provides it with a rather interesting explanation. Dowty notes that the two arguments of these verbs, have a “weak, but equal claim to subjecthood” (Dowty, 1991, p. 579), because they both only have one P-Agent property. The Experiencer argument has *sentience*, and the Stimulus argument (Theme) has *cause of change*. Interestingly, Dowty points to the fact that the Stimulus arguments do not have to be sentient of the fact that they are feared by/frightening someone.

Dowty cites Croft in pointing to the fact that *fear*-verbs are always stative, whereas *frighten*-verbs may be either stative or inchoative²⁷. Thus, the inchoative reading adds another P-Patient property to the Experiencer argument, namely change of state:

Hence, though the two arguments are still equal in Agent properties, they are unequal in that one is a ‘better’ Patient, so it must be the direct object according to the selection principle. (Dowty, 1991, p. 580)

This is an interesting claim, once again emphasizing the property of causation with regards to the psych-verbs.

²⁷An inchoative verb brings the focus of attention to the beginning of a new state or a change of state (Saeed, 1997, p. 110):

- (1) The ice melted

3.6.5 Causative verbs

Often cited examples of causative verbs are verbs like *break* and *kill*. These conform to a lexical representation like the following, taken from Levin and Hovav (1995, p. 83):

- (30) *break*: [[x DO-SOMETHING] CAUSE [y BECOME BROKEN]]

What this illustrates is that some action or event performed by x, causes a predicate, in this case BROKEN, to be true of y.

As we have seen above, causative verbs are possible reverse-animacy verbs, thus may take an inanimate subject and an animate object. Levin and Hovav (1995) emphasize an event-based representation of the lexical meanings of verbs. For a causative verb, it is the event that causes something and not necessarily the entity/agent expressed by the subject. This opens for the use of an inanimate subject as the entity that sets the causing event in motion. Often this will be a *natural force*, as in (31) below:

- (31) Sola varmet alle som satt der
sun-DEF warmed everyone who sat there
'The sun warmed everyone who sat there'

Levin and Hovav (1995) differentiate between internally caused verbs, like *laugh* and *shudder*, and externally caused verbs, like *break* and *kill* in an attempt to explicate further the distinction between unergative and unaccusative verbs. Whereas an internally caused verb not necessarily has a transitive use, externally caused verbs do, exactly because the cause of change is external to the entity in question. Internally caused verbs place strict restrictions on their arguments, externally caused verbs are not that strict, and "unlike most internally caused verbs, most externally caused verbs do not impose restrictions on their external cause argument, taking agents, natural forces and instruments as the external cause" (Levin and Hovav, 1995, p. 94). The set of externally caused verbs, does, however, seem to be larger than the set of traditional causative verbs, including verbs like *write* and *build*.

Whether or not the object in a causative construction with an inanimate subject may in addition be animate depends somewhat on the idiosyncratic, lexical properties of the causative verb in question and the restrictions it poses on its arguments, as in (32a) below. Also, some causative verbs require an animate, agentive subject and may not take an inanimate subject, as in (32b) below.

- (32) a. *Steinen brakk jenta
rock-DEF broke girl-DEF
b. *Bomben myrdet jenta
bomb-DEF murdered girl-DEF

3.6.6 Summing up

It seems difficult then to arrive at some properties that characterize the totality of all reverse-animacy verbs. Even so, we may sum up some of the most central properties associated with these verbs.

As we have seen, causation is a property common to both the *amuse* verbs and the causative verbs. A focus on an event-based representation of these verbs, as seen in both Grimshaw (1990) and Levin and Hovav (1995), provides an explanation for the possibility of a reversed animacy relationship between the arguments of these verbs, since an event may be initiated both by an animate and an inanimate ‘causer’.

Many of the *amuse* verbs also have something in common with the *concern* verbs, namely their resistance towards passivization, as we saw above. This has been explained through a generalization that sentences with Theme subjects may not passivize. This, however, cannot be maintained as a sufficient criterion for this group, as there are obvious members that certainly may passivize. In terms of Dowty’s theory of Proto-roles, however, these verbs do not seem to have that much in common. The subject of the *amuse*-verbs have in common the fact that they are causes and that their objects are causally affected, whereas none of these properties seem apparent in the case of *concern*-verbs. The main point is that there does not seem to be any one unifying property available to account for all the reverse-animacy verbs, apart from their possibility for taking an inanimate subject and an animate object.

3.6.7 Practical consequences

As observed above, the reverse-animacy verbs represent a problem, as they do not fit neatly into the relations predicted by our theory of markedness and alignment of hierarchical information. However, as the groups of verbs that may partake in a reverse-animacy construction are rather well-defined, it might not be such a difficult task after all to account for these. In support of the tendency expressed by the theory of hierarchies is the fact that the reverse-animacy group of verbs is a restricted one, and the general tendency is thus in the right direction. In chapter 5, we will see that this observation is supported by our data material, where only a very small percentage of the sentences in question are reverse-animacy sentences. We will return to further practical remarks on the reverse-animacy verbs in chapter 5, where we review the results from the data analysis, as well as in chapter 7, where we outline a Constraint Grammar implementation of our findings.

3.7 Word Order Freezing

Norwegian mainly marks syntactic function through a fairly rigid word order. As we have seen, however, some variation is permitted by way of topicalization. In languages where word order is to a greater extent free, syntactic functions are often marked by case. However, so-called *freezing effects* on word order have been observed, mainly in two slightly different, but related circumstances:

1. When properties of the arguments are semantically or functionally marked (Lee, 2002b; Morimoto, 2000)
2. When the arguments are indistinguishable with regards to some disambiguating dimension like case or animacy (Bloom, 1999; Lee, 2002a; Morimoto, 2000)

In the following we will examine both types of freezing phenomena. We will look in detail at the OT analysis performed in Lee (2002b) to account for word order freezing in Hindi, as well as providing a richer picture by looking at word order freezing in other languages, such as Russian (Bloom, 1999) and Haida (Morimoto, 2000).

3.7.1 Lee: Markedness Reduction in Word Order

Lee (2002b) examines an assumably free word order language, Hindi, which exhibits word order freezing in certain contexts. She attributes this to a tendency towards avoidance of *the worst of the worst*, i.e. maximally marked subject-object configurations in marked structural positions:

free word order becomes fixed when the unmarked association among grammatical functions, semantic roles, case and positions in phrase structure does not match the relative prominence relations of these dimensions. (Lee, 2002b, p. 2)

In order to model this phenomenon, she develops an OT account which can derive the word orders needed, as well as providing an explanation through markedness hierarchies interacting with more language-particular phenomena.

The data in short

The unmarked or canonical word order in Hindi is SOV, however the three elements of S, O, and V may be scrambled to reveal all possible permutations. These different constellations are used in order to express differences in information structure and other shifts of semantic effects²⁸.

²⁸Lee (2002b) is not specific regarding what these semantic effects are, but she does mention that they can provide a *definiteness effect*, a term which has been employed to describe restrictions on the definiteness of the element occupying object position (Mikkelsen, 1999).

However, in the case of so-called non-volitional transitives, a verb class which takes arguments lacking the prototypical agent/patient properties of transitive constructions, the word order freezes in the canonical SOV order. These verbs take one argument which is obligatorily sentient (the Experiencer) and another which may or may not be sentient (the Theme). The obligatorily sentient argument always bears dative case. A property of these verbs is that either of the two arguments of the verb may be realized as subject, whereby the other becomes object. When the subject is linked to the Theme role and the object to the non-volitional Experiencer, however only the unmarked word order may be employed. Lee (2002b) provides the following example of this, which is meant to illustrate a standard Hindi subject test which shows that the Theme argument *Nina* is in fact subject, by showing that the pronoun cannot be bound to it:

- (33) Niinaa Anuu-ko uskii bastii-mē
 Nina-NOM Anu-DAT PRON-GEN neighborhood-LOC
 dik^hii
 appear-PERF
 ‘Anu_i saw Nina_j in her_{i/*j} neighbourhood’ ($S_{th} O_{exp} V, *O_{th} S_{exp} V$)

With all other permutations of the arguments, however, this reading is no longer available, and the Experiencer must be the grammatical subject:

- (34) Anuu-ko Niinaa uskii bastii-mē
 Anu-DAT Nina-NOM PRON-GEN neighborhood-LOC
 dik^hii
 appear-PERF
 ‘Anu_i saw Nina_j in her_{j/*i} neighbourhood’ ($S_{exp} O_{th} V$)

Thus, in order to avoid the worst of the worst, the unmarked word order is employed for the marked arguments.

OT formalization

Lee (2002b) derives free word order, and ultimately also the word order freezing, through an OT-LFG framework, where the input for evaluation in an OT grammar is a LFG functional structure (f-structure) which may be underspecified. On this approach, GEN is assumed to be a function which, from the f-structure input, generates candidates by specifying their f-structures and corresponding c-structures. It is important, as mentioned earlier, that the input be recoverable from the output. This property of the grammar is ensured by the relation of *subsumption*²⁹, which holds between the input f-structure and the outputs.

²⁹Informally this means that the input f-structure is more general than the output, thus the output is obtained from the input through unification. This is a common relation between two feature structures in unification-based grammars, like LFG and HPSG.

Through so-called *alignment* constraints, Lee (2002b) derives free word order in Hindi. These constraints express correspondence between syntactic functions in f-structure and structural positions in c-structure, as well as thematic roles in argument structure (a-structure) and their preferred positions in c-structure, e.g. that the subject should be aligned left in the sentence (SUBJ-L), and so should the the topic (TOP-L) and the Proto-Agent (PA-L). Lee (2002b) also introduces a few strictly structural constraints which derive right-branching trees, since Hindi is a right-headed language.³⁰ What these constraints in combination do, is to derive unmarked word order when the arguments do not differ in informational status, whilst allowing for a marked word order when the elements do differ.

Word order freezing in Hindi Lee (2002b) makes use of the, by now well-known, technique of harmonic alignment (cf. section 3.4 above), aligning the scale of syntactic functions (SUBJ \gg Non-SUBJ) with the scale for semantic role (P(roto)A(gent)_{vol} \gg P(roto)A(gent)_{-vol} \gg P(roto)P(atient)). She differentiates between a volitional and a non-volitional Proto-Agent, as volition is a particularly important Proto-Agent property in Hindi.³¹ The following constraints are derived through harmonic alignment (Lee, 2002b, p. 30):

- (35) a. C1: *SUBJ/PP \gg *SUBJ/PA_{-vol} \gg *SUBJ/PA_{vol}
 b. C2: *Non-SUBJ/PA_{vol} \gg *Non-SUBJ/PA_{-vol} \gg *Non-SUBJ/PP

In order to express a markedness in linking between, on the one hand, syntactic function and thematic role, and on the other, structural position, Lee (2002b) makes use of another well-known technique, namely that of local conjunction (cf. section 3.4 above). As we remember, in Hindi it is not possible to combine a marked linking between a non-volitional transitive verb and its arguments and a marked word order. A local conjunction between the structural align-constraint SUBJ-L and the constraints in (35a) gives us the following subhierarchy (Lee, 2002b, p. 33):

- (36) Conjoining SUBJ-L with C1:
 C3: *SUBJ/PP & SUBJ-L \gg *SUBJ/PA_{-vol} & SUBJ-L \gg *SUBJ/PA_{vol} & SUBJ-L

Lee (2002b) states that in order for a conjoined constraint to be violated, the candidate must violate both the conjunctive constraints at least once. This

³⁰We will not go further into exemplifying these as it does not seem necessary for the present discussion.

³¹Lee (2002b) notes that two Agent properties, in particular, seem to be ranked above the others as to importance with regards to subject selection, namely volition and causation. In English and Romance languages causation is more important than volition, as she states. We have already seen an example of the importance of causation in subject selection in section 3.6 above, regarding the reverse-animacy verbs.

expresses the central idea that a maximally marked subject (Proto-Patient) is the worst candidate if it, in addition, is not in an unmarked structural position.³²

3.7.2 Morphological syncretism and word order freezing

Russian is a language that exhibits free word order in most declarative sentences. However, all of the possible word orders, apart from the basic SVO order, carry additional information on discourse status, such as topic and focus, encoded through word order. Morphological case usually contributes to the fixing of syntactic function, however, this is not always the case. In Russian, all nouns syncretize their ending in accusative either with the genitive ending or the nominative. The largest class of nouns, counting all masculine, animate nouns and feminine nouns ending in a vowel, syncretize with the genitive ending, thus clearly marking syntactic function through case. However, this is not the situation with regards to the other group of nouns, counting the masculine inanimate nouns and the other feminine nouns. These nouns syncretize their endings in the accusative with the nominative, thus making them indiscernible as to syntactic function based on morphology alone. These are the cases where Russian exhibits word order freezing. In the example below, taken from Bloom (1999, p. 20), we see that a sentence with two nouns whose case morphology syncretizes their endings in nominative and accusative, freeze in canonical word order:

- (37) a. Mat' ljubit doč
 mother-NOM-ACC loves daughter-NOM-ACC
 'The mother loves the daughter'
- b. Doč ljubit mat'
 daughter-NOM-ACC loves mother-NOM-ACC
 'The daughter loves the mother'
- c. *Doč ljubit mat'
 daughter-NOM-ACC loves mother-NOM-ACC
 'The mother loves the daughter'

Lee (2002b) reports of similar freezing effects in Hindi, when nominal arguments are indiscernible with regards to case. Russian and Hindi are both fairly free word order languages that mark syntactic function in majority through morphological case. It is thus not a surprising fact that word order freezes when there is a severe risk of ambiguity.

³²Importantly, Lee remarks that it is not the conjunction as a whole which is part of UG, but rather its conjuncts and the &-operator.

3.7.3 Animacy-related freezing effects

From the viewpoint of this thesis we are interested primarily in the effects of animacy and definiteness on variation in word order. Freezing effects conditioned by animacy have been observed in several languages.

Haida, an indigenous language of British Columbia, exhibit freezing effects that are sensitive to the relative animacy of the arguments (Morimoto, 2000). Haida is an SOV language, but permits non-canonical OSV when the subject is higher in animacy than the object. When the subject-object relation is maximally marked, however, i.e. when the subject is inanimate and the object is animate, word order freezes to canonical SOV (Morimoto, 2000, p. 8)³³:

- (38) a. 7adaáhl-ǵuu t'sagt'sag-gee gyuúdan-ee 7is-tlagaay-gan
 yesterday-FOC wagon-DEF horse-DEF CA-hurt-PAST
 'The wagon hurt the horse yesterday'
- b. *7adaáhl-ǵuu gyuúdan-ee t'sagt'sag-gee 7is-tlagaay-gan
 yesterday-FOC horse-DEF wagon-DEF CA-hurt-PAST
 'The wagon hurt the horse yesterday'

Also, when the subject and object are of equal animacy, word order may freeze (Morimoto, 2000, p. 9):

- (39) a. ʔaay gyuúdan-ee k̄ing-gan
 dog-DEF horse-DEF see-PAST
 'The dog saw the horse'
- b. gyuúdan-ee ʔaay k̄ing-gan
 horse-DEF dog-DEF see-PAST
 '*The dog saw the horse'
 'The horse saw the dog'

An interesting point, however, is that freezing in cases of equal animacy is not completely categorical. If there is no chance of ambiguity, due to for instance a property of the verb, word order does not freeze, e.g. in sentences where the main verb clearly marks one of the arguments as the only possible agent. An example from Haida is provided with a verb corresponding to the English *peck*, an activity which only may be performed by an agent with a beak (Morimoto, 2000, p. 9):

- (40) a. 7adaáhl-ǵuu skaw-ee ʔaay skayj-an
 yesterday-FOC chicken-DEF dog-DEF peck-PAST
 'Yesterday the chicken pecked the dog'

³³Morimoto's data on Haida are taken from Enrico (1986).

- b. 7adaáhl-ʔuu ʔaay skaw-ee skayj-an
 yesterday-FOC dog-DEF chicken-DEF peck-PAST
 (i) ‘Yesterday the chicken pecked the dog’
 (ii) #‘Yesterday the dog pecked the chicken’

3.7.4 Relevance of freezing phenomena

Lee (2002b) proposes that “the account developed here can be naturally extended to languages in which the freedom of word order is sensitive to other dimensions of prominence”. Norwegian has a fairly rigid word order and a poor morphological case system, thus largely marking syntactic function through word order. Our claim here then, is that the interplay between word order variation and other properties of the arguments is crucial to the expression of syntactic function in Norwegian.

For the purpose of this thesis, the above accounts of freezing effects are interesting for several reasons. Norwegian is not a free word order language and does not exhibit any categorical freezing effects. However, from a probabilistic viewpoint there might be tendencies in the same direction.

One interesting aspect of Lee’s account is found in the status of verbs with maximally marked argument expression. We saw that, in Hindi, the non-volitional transitives with a Theme subject and an Experiencer object caused word order freezing. These verbs are certainly reminiscent of some of our *amuse*-verbs psychological verbs (cf. section 3.6 above). In Norwegian, word order can vary between SVO and OVS (topicalized object), as we have seen several examples of. Might it be then that psych-verbs exhibit a strong tendency towards an unmarked word order, thus a markedness reduction and avoidance of the worst of the worst? For instance, one might examine this by looking at a psych-verb, such as *interessere* ‘interest’. This is a verb which subcategorizes for rather marked arguments, a Theme subject and an Experiencer object. These may very well be realized as inanimate and animate, respectively, thus representing a maximally marked configuration with regards to animacy, as we have seen in section 3.6 above.

- (41) Filmen interesserer Karin
 movie-DEF interests Karin
 ‘The movie interests Karin’

Creating a structurally marked version of the above sentence by topicalizing the object, gives us the following sentence:

- (42) ??Karin interesserer filmen
 Karin interests movie-DEF
 ‘Karin, the film interests’

The above sentence is not strictly ungrammatical, however, it certainly borders on it. To stretch the parallel even further, might it be the case that

there exists a tendency to avoid a marked word order when the subject and object are marked, for instance in the case of the reverse-animacy verbs? These are interesting questions, which from the perspective of this thesis certainly seem relevant.

Another point, which is illustrated by the freezing found in Russian (Bloom, 1999), Hindi (Lee, 2002a) and Haida (Morimoto, 2000), is freezing that occurs when the arguments are, in some sense, too similar, either by morphological syncretism or equality in animacy. This might also prove to be a possible parallel to Norwegian, at least on statistical grounds.

3.8 Ambiguity: Bidirectional optimization and the problem of ineffability

Traditional OT, as well as the stochastic OT described above, is unidirectional. This means that it depicts a directed one-way process from input, via different candidates, to an output. The evaluation process leading up to a certain output thus represents a speaker producing a spoken output. In a recent extension of OT, however, the hearer has been included in the process, mainly in the modeling of discourse-oriented phenomena. A bidirectional OT approach would thus operate with both *production* and *comprehension*:

Production and comprehension functions defined (Smolensky 1996: 725)

$$f_{prod}(/i/) = \text{H-maxs} \in \text{UGen} \mid /i/ = \text{Input}(s)$$

$$f_{comp}([o]) = \text{H-maxs} \in \text{UGen} \mid [o] = \text{OvertForm}(s)$$

H-max = maximum Harmony; UGen = the universal set of all possible structural descriptions generated by the OT generator of candidates, Gen for all universally possible inputs /i/.

(Lee, 2002a)

What happens then is that the production function, f_{prod} , as we know it from traditional OT, takes an input or ‘meaning’ and returns the maximally harmonic structure from the set of structures (candidates) with the underlying form /i/. This maximally harmonic form is obviously our output, the candidate that has performed the best with regards to the ranked constraints. At this point, novelty comes into the picture through the comprehension function f_{comp} . Operating with the same OT grammar as the production function, the comprehension function takes an overt form, [o], i.e. output from production and goes the other way, so to speak. It returns the maximally harmonic candidate among the structures that are underlying forms for [o], i.e. the candidate that has performed the best with regards to the ranked constraints of the same grammar.

It is an important tenet in OT, that the input should be recoverable from the output. The so-called problem of *ineffability* relates to a “situation where there is no acceptable output for a given input” (Morimoto, 2000). In bidirectional terms this means that the original meaning, or input, is not recoverable from the output, i.e. that the comprehension function returns an output different from the input presented to the production function in the first place. This is what happens in cases of ambiguity (Morimoto, 2000).

As described above, Morimoto (2000) looks at animacy related freezing in the Haida language. Perhaps even more interesting, she looks at topicalization in Swedish, where the animacy of the arguments are equal. We will return in some detail to this at the end of chapter 6, where bidirectional OT will provide us with an extension of our analysis.

Chapter 4

Prerequisites for a data analysis

The theoretical framework of the previous chapter provides us with some interesting predictions regarding properties of transitive sentences, their arguments and, to a certain degree, word order. In particular, we have seen that subjects tend to be animate and definite, and objects inanimate and indefinite. We have also looked at some cases where these predictions do not follow through and the consequences that this might have for word order, in particular. In order to test these predictions, we need to obtain a clearer picture of the situation for Norwegian. In particular, we want to obtain some frequency distributions regarding different constructions for use in the analysis later on. In this chapter, then, we will set up some prerequisites for the data analysis. The results obtained in the data analysis will be reviewed and discussed further in chapter 5.

There is no corpus annotated for both animacy and definiteness available for Norwegian. In order to obtain data material for the analysis of this thesis, it was therefore necessary to extract a reasonable portion of data and annotate this manually. In the following we will first take a look at a similar corpus study conducted for Swedish, and review the results obtained there. After this we will review the guidelines set up for the sampling process, thus defining the boundaries of our data material. The extracted data was collected and imported to a database, a format which facilitated annotation and the following extraction of results. A short overview of the database is provided in section 4.3. In the last section, we will look at non-literal language use - metaphors and metonymies. These pose a challenge for annotation, and are thus important to obtain an understanding of. We will provide a short overview of non-literal language, and go on to explain how these will be dealt with in the following.

4.1 Earlier corpus studies - the case of Swedish

Dahl and Fraurud (1996) have performed a corpus study in order to examine the effect of animacy on the syntactic distribution of grammatical relations in Swedish. The corpus is composed of a variety of non-fictional, written texts and counts approximately 85,000 words. This is obviously a rather small and unbalanced corpus, however, it does have the advantage of being annotated for animacy.¹

Dahl and Fraurud (1996) observe that many of the phenomena grammaticalizing the relationship between animacy and syntactic function in other languages, are paralleled by statistical tendencies in Swedish. The positions favored by human DPs are without a doubt the transitive subject (56.5% are human) and the indirect object (83.1% are human). Also, there is a sharp contrast between the subjects of intransitive clauses and those of transitive clauses. Intransitive subjects do not display this strong tendency towards attracting human arguments.

In their study, Dahl and Fraurud (1996) single out the transitive clauses and conduct a frequency count in these with regards to the relationship between degree of animacy and syntactic functions. Their results are presented in table 4.1 below, as these are interesting from the viewpoint of this thesis. The most interesting result perhaps, is the vanishingly small amount

Subject	Direct Object	N	%
Person	Non-person	1484	47.7
Non-person	Non-person	1268	40.8
Person	Person	276	8.9
Non-person	Person	81	2.6
Total		3109	100.0

Table 4.1: Distribution of 3109 transitive clauses according to animacy of subject and direct object (Dahl and Fraurud, 1996, p. 53)

of clauses where the subject is Non-person and the object is Person (2.6%). Dahl and Fraurud (1996) conclude that:

more than 97% of all transitive sentences obey the constraint that the subject should not be lower than the object in animacy. Thus, this constraint, which is grammaticalized in a language such as Navajo, could be said to be approximated statistically in Swedish texts. (Dahl and Fraurud, 1996, p. 53)

¹Dahl and Fraurud (1996) remark that the corpus is annotated with the distinction Person/Non-Person, where Person includes all humans, as well as a few “individuated animals”, and Non-Person includes all other arguments.

Possible explanations

Dahl and Fraurud (1996) also attempt to explain their findings in the corpus. What might help explain the close link between a high rank of animacy and subjecthood? One possible direction is in the domain of thematic roles, as alluded to by the authors. Arguments are commonly associated with a set of prototypical roles, as indicated by the selectional restrictions of verbs. Agency and animacy for instance are inextricably connected and the Agent role is, if possible, linked to the relation of subject, a fact which has a few implications. It is certainly a point that there are several relevant dimensions with regards to the syntactic function of subject, through which agency and animacy are connected.

Another characteristic often associated with the subject is topicality (cf. chapter 3.2), and that humans are more fit to be topics than other referents. Dahl and Fraurud (1996) wish to focus on a notion of *point of view* instead,² pointing to the fact that discourse tends to be narrated from a human point of view, and that “we tend to think of the world as being organized around animate beings which perceive and act upon their inanimate environment” (Dahl and Fraurud, 1996, p. 60).

Conclusively, Dahl and Fraurud (1996) mention the problem that the animacy hierarchy does not represent a clear-cut dichotomy. There are borderline cases, and problematic usages. They provide a list of these types of exceptions with examples, which includes the following (Dahl and Fraurud, 1996, p. 62):

- metaphorical use, e.g. *Nature is generous*
- metonymical extension, e.g. *Norway is rejoicing after the Olympic victory*
- collective nouns, e.g. *The family are happy*
- DPs referring to ‘non-personal agents’ such as institutions, companies, associations etc.

These problems represent a pointer to problem areas we might encounter in our own corpus study. Dahl and Fraurud (1996) do not propose a solution towards the treatment of these areas. The above problems will be dealt with in the section on non-literal usage below, where we will attempt a unified treatment of these phenomena within our analysis. Also, our study includes topics not dealt with in Dahl and Fraurud (1996), such as the reverse-animacy verbs (cf. chapter 3.6).

²The close relationship between animacy and point of view is exemplified in Dahl and Fraurud (1996) by the Algonquian language Fox and the grammatical category of *obviation*. It distinguishes within the 3rd person between *proximate* category, which places the referent as a central participant, and *obviative* which is less central.

4.2 Sampling and annotating the data

The data material for the analysis is sampled from the Oslo Corpus, a corpus of Norwegian texts of approximately 18,5 million words.³ The corpus has been automatically annotated using the Oslo-Bergen tagger, a morphosyntactic tagger. As we remember, this is the tagger that our work here is, in part, aimed at improving, by providing insights than can be put to use in the automatic disambiguation of syntactic functions. We shall return in detail to this aspect of the thesis in chapter 7, where we will test our theoretical predictions in practice.

In order to get an adequate picture of the task at hand, it will be necessary to sample a reasonable size of data from the corpus, annotate it for the additional information needed, namely a semantic dimension of animacy and definiteness following the hierarchies in Aissen (2003), and then being able to run counts of frequencies on this material. Before sampling the data from the Oslo Corpus it is clear that, without a proper scheme for extraction and annotation, the data might not prove very useful in the end after all. First and foremost, a consistent view of what to include in the data set and what to exclude is crucial. Furthermore, a clear picture of what we are interested in counting in the sampled data obviously needs to be defined before annotating it.

4.2.1 What to include in the sampling process

The construction of interest in the following is a sentence containing a transitive verb along with its two core arguments, the subject and the object. These arguments should not be disambiguated by case⁴. Another important point is that there must be present a possibility for the interpretation of the sentence as topicalized. It should therefore exhibit V2 word order, i.e. in most cases it should be a declarative, main sentence. Another point is that the verb phrase should be a simple one, consisting of a single finite verb, and no auxiliaries. As we remember, complex VPs are not ambiguous with regards to word order.

Main clauses

Declarative sentences and interrogative sentences with interrogative pronoun are the two types of main clauses which exhibit V2 word order, i.e. the verb is the second constituent from the left (cf. chapter 2.2). These two sentence types differ in the fact that the declarative sentence may place almost any type of constituent in the topic position (SpecCP), whereas the interrogative

³The corpus consists of texts of three main genres: fiction (1.7 million words), newspapers/magazines (9.6 million words) and factual prose (7.1 million words).

⁴Cf. section on pronouns below.

sentence must fill this position with an interrogative pronoun. Even so, the interrogative pronoun may refer to (among others) both subject or object, thus displaying many of the properties of a topicalized declarative sentence.

A subject or a topicalized object in a regular declarative sentence is usually a discourse topic, i.e. it denotes given information. In contrast, interrogative pronouns, whether topicalized or not, are certainly not topics, but rather foci, i.e. they represent new information in the sentence. As mentioned in chapter 3.2, we may also find a contrastive focus (Engdahl, 1997) in SpecCP, however, without a larger context available, these are quite indistinguishable from contrastive topics. The difference in discourse functions between interrogative sentences and normal declarative sentences may contribute towards obscuring the picture somewhat, and we will therefore focus on declarative sentences only, i.e. excluding interrogative sentences from the sample.

Subordinate clauses

Only one type of subordinate clause in Norwegian allows for topicalization, namely those initiated by the subjunction *at* ‘that’. Declarative main clauses demand a V2 word order, i.e. the verb comes second, whereas most subordinate clauses do not. Subordinate clauses beginning with *at* ‘that’ however may follow the V2 pattern, and thus may also topicalize the object⁵ (Faarlund et al., 1997, p. 866):

- (1) Hun vet at den mannen kan hun få kjær
she knows that that man-DEF can she have dear
‘She knows that that man can become dear to her’

This type of subordinate clauses may also not follow the V2-pattern:

- (2) Hun sa at jenta ikke skrev brevet
she said that girl-DEF not wrote letter-DEF
‘She said that the girl did not write the letter’

Form of the arguments

In the following we will have a look at some formal properties of the subjects and objects included in the sample, and also how these will be annotated. In particular, we will examine a few cases which will need further discussion and justification for choice of annotation.

⁵The normal assumption in subordinate clauses with V2 word order is that these sentences contain two C-projections (Nordgård and Åfarli, 1990, p. 81)

Nominal The form of the subject and object in the sampled transitive constructions will have to be nominal, i.e. pronouns⁶, proper nouns, or other DPs containing common nouns.⁷ These will also be annotated accordingly; pronouns will be annotated as pronouns, proper nouns as such, and with regards to DPs containing common nouns these will be differentiated along the lower dimension of definiteness: definite noun or indefinite noun.⁸

Pronouns As mentioned above, the arguments of the sampled transitive sentences should not be disambiguated by their morphological case. As we saw in chapter 2, this will exclude all personal pronouns marked for accusative case, as well as 1. and 2. person pronouns marked for nominative case. As we remember, the 3. person singular pronoun *han* ‘he’ the inanimate pronouns *det/den* ‘it’ and the 2. person plural pronoun *dere* ‘you’ all syncretize with the accusative, so these are inherently ambiguous. With regards to the remaining nominative 3. person pronouns, *hun* ‘she’ and *de* ‘they’, these are possibly ambiguous when modified by a prepositional clause or a relative clause. As we remember, when functioning as head of a phrase these are often in nominative form, even when they function as object. With regards to pronouns then, we will include transitive sentences containing the following:

- (3) a. Morphologically ambiguous pronouns - 3.pers.sing. *han* ‘he’, *den/det* ‘it’ and 2.pers.pl. *dere* ‘you’
- b. Pronouns that are ambiguous when modified - 3.pers.sing. *hun* ‘she’ and 3.pers.pl. *de*

For more complex nominal arguments we adhere to the DP-hypothesis⁹, which makes the determiner head of the DP, thus the carrier of properties concerning the phrase as a whole, in particular, definiteness. Animacy is not a difficult property to assign to DPs, definiteness, however, is a bit more complex. We will be in need of clear guidelines when annotating for this property. There are several syntactic environments which may be used as tests for definiteness, as they posit certain restrictions on the definiteness of a constituent. Lyons (1999) gives several examples from English, and a suitable candidate for Norwegian might be the impersonal active or presentational construction,¹⁰ which contains an expletive subject and demands of its object

⁶Not all pronouns however, cf. section on pronouns and reflexive objects below.

⁷Complement clauses may also cause ambiguity with regards to syntactic function, as these may function as both subject and object, and thus also topicalize. These are not, however, classifiable along the dimensions of animacy and definiteness, and will therefore not be in focus in the following.

⁸As mentioned above, we will not differentiate with regards to specificity within the category of indefinites, even though Aissen’s definiteness scale does.

⁹See for example Carnie (2001) for more on the DP-hypothesis.

¹⁰More on these types of constructions and their relationship to definiteness in chapter 5.1.2.

that it must be indefinite:

- (4) Det kommer ei jente
it comes a girl
'A girl is coming'
- (5) * Det kommer jenta
it comes girl-DEF

Determiners A group of nominal arguments which will be included in the sample constructions are demonstratives. These are determiners which may occur alone, as in (6b) below:¹¹

- (6) a. *Dette alternativet* betyr sparte investeringer
this alternative means saved investments
'This alternative means saved investments'
- b. *Dette* betyr sparte investeringer
this means saved investments
'This means saved investments'

According to the DP-hypothesis then, the fact that a demonstrative may occur alone does not really make a difference, as it is the determiner (in this case a demonstrative) which is the head, and thus determines the definiteness of the whole DP. Determiners are considered to be definite due to a clear property of *identifiability* (Lyons, 1999), the hearer is assumed to be able to work out the reference based on the context. Demonstratives have a *deictic* property, which locates the referent in space and time. These determiners differ from the definite article (or, in Norwegian the definite affix) in that the speaker intends the referent to be directly accessible to the hearer without further inference (Lyons, 1999, p. 21). When occurring alone they may often simply refer to some part of the previous discourse, thus being perfect continuous topics.

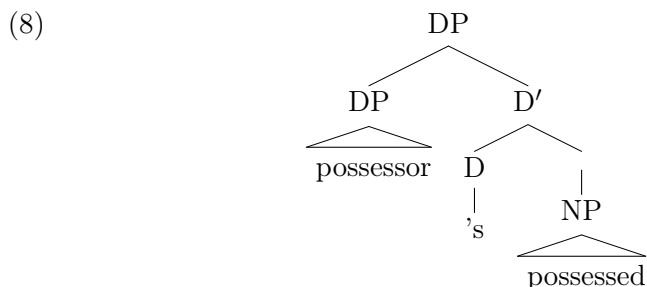
Our diagnostic for definiteness, the presentational construction, also indicates that demonstratives are definite, as they are ungrammatical in this context:

- (7) a. * Det kommer denne jenta
it comes this girl-DEF
- b. * Det kommer denne
it comes this

When occurring alone as an argument in a transitive sentence, the determiners are pronominal (Faarlund et al., 1997) and will be annotated accordingly. When the determiner has a nominal complement, however, the DP will be treated as a definite DP.

¹¹The example in (6a) is taken from the Oslo Corpus.

Possessives Following Lyons (1999), we include under this heading possessive determiners like *min* ‘my/mine’ and *sin* ‘their’, as well as genitives with an -s ending. Under the DP hypothesis, possessives are considered to be determiners, and are given the following analysis (Carnie, 2001, p. 144):



For the possessive determiner, the specifier position would be empty and it would be head of the DP, like a regular determiner.

Possessives are definite, even though the head noun is often indefinite morphologically in Norwegian. One way of seeing this, which works equally well for English and Norwegian, is that they are both easily paraphrased into a definite form:

- (9) a. Det er jentas bil
 it is girl-DEF-GEN car
 ‘It is the girl’s car’
 b. Det er bilen til jenta
 it is car-DEF to girl-DEF
 ‘It is the car belonging to the girl’

Possessives are not allowed in the presentational construction:

- (10) *Det kommer Kari’s mann
 it comes Kari’s man

Quantifiers Quantifiers are usually classified as determiners, denoting amount or quantity. They share with the demonstratives the property of occurring with or without their nominal arguments.

Alle, *hver*, *enhver* ‘everyone/thing’ and *begge* ‘both’ are all definite determiners, as they refer to the totality of a group. This satisfies the criterion of *inclusiveness* (Lyons, 1999), which requires that a definite refers to the whole of a set, not a subset. *Alle* ‘everyone/thing’ refers collectively, i.e. to the group as one unit, whereas *hver*, *enhver* ‘everyone/thing’ refers distributively, i.e. referring to the whole group by “pointing” to each member of the group. *Begge* ‘both’ is only used when referring to a group of two. It goes for all of these that may not occur as object in a presentational construction:

- (11) Det kommer *hver/*enhver/*alle/*begge mann/mennene
 it comes *every/*every/*all/*both man/men

Among the typical indefinite quantifiers we find *noen* ‘some’ and *ingen* ‘no’. In contrast with the definite quantifiers described above, these are quite grammatical as objects in a presentational construction:

- (12) Det kommer *noen/ingen* menn
it comes some/no men
‘Some/no men are coming’

Cardinality Cardinal DPs denote a number or an amount, e.g. *tre* ‘three’ or *mange* ‘many’. Seeing that these may be preceded by a definite determiner, thus rendering the whole DP definite, Lyons (1999) claims that indefiniteness is simply the absence of a definite marker. In this respect, it is a default value; indefinite DPs do not have to be preceded by an indefinite determiner to be interpreted as indefinite, as is the case for mass and plural nouns, both in English and Norwegian:

- (13) Vi har kjøpt melk
we have bought milk
‘We have bought milk’

The cardinal DPs may very well be the argument of a presentational construction, as we see in (14) below:

- (14) Det kommer *tre/mange* menn
it comes three/many men
‘Three/many men are coming’

4.2.2 What to exclude in the sampling process

Just as important as being clear as to what to include in the sampling process, is to be clear as to what to exclude. We will therefore take closer look at what our sample will not contain in the following.

Indirect Objects

We will only include transitive constructions with two arguments, a subject and a direct object. This rules out indirect objects¹².

Copular constructions

Even though a copular construction may include two nominal arguments, there is not really a subject-object like relation between the two, but rather a relation of ‘subject’ and ‘subject predicative’. The tagger also makes this distinction, thus these constructions are automatically excluded when searching for co-occurring subjects and objects in the corpus.

¹²We will however include so-called two-place unaccusatives when they are not in double-object constructions. More on this in chapter 5.1.1 below.

Subordinate clauses

As mentioned above, no subordinate clauses other than the *at*-clauses may in fact follow the V2 pattern. This means that topicalization of object is impossible in these clauses.

Topics other than subject or object of matrix clause

In Norwegian, pretty much any constituent may topicalize. However, only sentences where the subject or the object of the matrix verb is the initial constituent will be included. The reason for this, is that we are interested in the relationship between the subject and object of one and the same verb.

Preposition stranding A topicalized constituent may be topicalized from a prepositional phrase, and these will in the following be excluded as they are not ‘true’ arguments:¹³

- (15) *Det_i* tapte de over 300.000 kroner på *t_i*
it lost they over 300.000 kroners on
‘They lost over 300.000 kroners on it’

Phrasal verbs A similar problem is that of so-called phrasal verbs or collocate constructions involving a verb and a preposition:¹⁴

- (16) *Historien_i* fant han på *t_i*
story-DEF found he on
‘The story, he made up’

Here, we see the phrasal verb *finne på* ‘make up’, and its topicalized object *historien* ‘the story’.

Topicalization from subordinate clauses Again, constituents which are topicalized from a position other than the subject or object positions of the main clause will be excluded. When it comes to constituents which are topicalized from a non-finite embedded clause, the subject and the topicalized element are certainly not arguments of the same verb, and are thus not within the scope of this analysis:¹⁵

- (17) *Treningen like ved Heathrow-flyplassen_i* regnet han ikke med
training-DEF just at Heathrow-airport-DEF count he not on
å rekke *t_i*
to make

¹³The example in (15) is taken from the Oslo Corpus.

¹⁴Faarlund et al. (1997) pair the preposition in these cases with the object, instead of the verb, calling them prepositional objects (Faarlund et al., 1997, p. 697).

¹⁵The example in (17) is taken from the Oslo Corpus.

‘He did not count on making the training close to Heathrow Airport’

This will also cover topicalization from raising and control constructions, i.e. from the non-finite clause.

Reflexive objects

Reflexive objects, i.e. a reflexive pronoun, will be excluded, as they are not objects in the sense used here, but rather co-referent with the subject:¹⁶

- (18) De vasket *seg* i innsjøen den kvelden
They washed themselves in lake-DEF that evening-DEF
‘Themselves, they washed in the lake that evening’

Another point is that reflexive pronouns are not, in fact, possibly ambiguous, they may not function as subjects or topicalize as objects.¹⁷

Indefinite pronouns

The indefinite pronoun *man* ‘one’ will not be included in the sample, as it does not posit a problem with regards to ambiguity. It is always a subject and never an object:

- (19) * Vi så man i går
We saw one yesterday

4.2.3 What to count?

Before annotating the data, an explicit operationalization of what to count is needed. The data was extracted by sampling randomly a *quota* of data, i.e. “a set number of the construction searched for selected at random from the corpus as a whole” (Wolters, 2002). In order to achieve a fair amount of data pertaining to the problem of ambiguous tagging of subjects and objects, only sentences where both the subject and the object were ambiguously tagged were sampled.¹⁸ The size of the data sample was set at one thousand transitive sentences, in accordance with the above criteria.

The following questions are the ones I want answered from the data:

¹⁶The example in (18) is taken from the Oslo Corpus.

¹⁷This is a truth with modifications. Simple reflexive objects may not topicalize, whereas the complex reflexive pronoun *seg selv* ‘oneself/-selves’ may:

- (1) *Seg selv* vasket de i innsjøen den kvelden
Themselves washed they in lake-DEF that evening-DEF
‘They washed themselves in the lake that evening’

¹⁸Regarding sentences where only one of the two arguments are tagged ambiguously one might conclude the function of the other argument, as we wish for the tagger to adhere to a uniqueness principle, i.e. each of the main functions may only occur once.

- The number of transitive sentences containing topicalized objects in the sample.
- The semantic relationship between the subject and object in the transitive sentence with regards to the dimensions of:
 - Animacy: the number of subjects higher than the object in animacy or equal as the object in animacy, and vice versa.
 - Definiteness: the number of subjects higher than the object in definiteness or equal as the object in definiteness, and vice versa.

In order to find an answer to these questions then, the sample data will be annotated for the following:

- Whether the sentence in question has a topicalized object or not.
- Subjects and objects are annotated for animacy and definiteness. The values for animacy and definiteness follow the elements in the prominence hierarchies of Aissen (2003), i.e. human, animate or inanimate for animacy and pronoun, proper noun, definite noun or indefinite noun for definiteness.
- Verbs are annotated using a NorKompleks¹⁹ code.

We will return to the specifics of the annotation scheme in the following section on the database.

4.3 The database

In order to obtain empirical material for analyses, I have collected and annotated a database of one thousand transitive sentences adhering to the criteria of the above section. In the following we will take a closer look at the search method used, the general organization of the database, as well as methods for result extraction from the database.

¹⁹NorKompleks is a lexicon of verb forms, specifying their argument structure: thematic roles of arguments, types of arguments, e.g. DP, PP, etc. It has been developed at NTNU in Trondheim, and may be downloaded at <http://mime.hf.ntnu.no/hf/prosjekter/spraktek/prosjekter/nkl>. We will return to the language resources employed in the application, as described in chapter 7.

4.3.1 From corpus to database

Search methods

The Oslo corpus has a web-interface which makes it quite easily searchable. One may use regular expressions²⁰ in order to perform as focused and specialized a search as possible. The following regular expression gives us a transitive sentence with two arguments which are ambiguously tagged. These are nominal arguments - nouns, proper nouns and pronouns:

```
([tagg=".*@obj.*" and tagg=".*@subj.*" and tagg=".*subst.*"] |
[tagg=".*@obj.*" and tagg=".*@subj.*" and tagg=".*pron.*"]) []*
[tagg=".*verb.*" and tagg=".*@fv.*"] []*
([tagg=".*@obj.*" and tagg=".*@subj.*" and tagg=".*subst.*"] |
[tagg=".*@obj.*" and tagg=".*@subj.*" and tagg=".*pron.*"])
within s
```

This reads as follows: search for a word tagged ambiguously as subject (@subj) and object (@obj), as well as being tagged as a noun (subst), OR (|) a word tagged ambiguously as subject (@subj) and object (@obj), as well as being tagged as a pronoun (pron), followed by anything ([]*), followed by a word tagged as a verb, bearing the syntactic tag of finite verb (@fv), followed by anything again, and then another occurrence of a word tagged ambiguously as subject and object, as well as being tagged as a noun, OR (|) a word tagged ambiguously as subject and object, as well as being tagged as a pronoun. Finally all of this within the boundaries of one and the same sentence. The sampled constructions were drawn at random from the 18,5 mill. word corpus.

The regular expression above gives us sentences which contain a transitive verb, and a nominal subject and object, which are both tagged ambiguously.²¹

- (20) Knut Wickstrøm solgte huset på Seiersbjerget
Knut Wickstrøm sold house-DEF on Seiersbjerget
'Knut Wickstrøm sold the house on Seiersbjerget'

The regular expression does not give us sentences with a clausal argument²², as in example (21) below or a topicalized constituent other than the subject

²⁰Regular expressions are strings of symbols used to match specific patterns in text.

²¹The example in (20) is taken from our data sample of the Oslo Corpus.

²²This is not to say that a transitive sentence with a clausal argument may not be ambiguous with regards to syntactic function:

- (1) a. At huset blir solgt bekymrer jenta
that house-DEF becomes sell-PASS.PART worries girl-DEF
'That the house will be sold worries the girl'
- b. At huset blir solgt vet jenta
that house-DEF becomes sell-PASS.PART knows girl-DEF

or object, as in (22):

- (21) Knut tror at huset blir solgt til slutt
Knut believes that house-DEF becomes sell-PASS.PART to end
'Knut believes that the house will be sold eventually'
- (22) Neste måned selger Knut huset
Next month sells Knut house-the
'Next month, Knut is selling the house'

Neither of the above sentences are ambiguously tagged.

The database

My search of the Oslo Corpus resulted in a text file of transitive sentences, the content of which was then imported to a database file. Here the sentences were manually annotated with regards to certain properties.

The database contains the following columns:

1. **Sentence** - a text field containing the sentence in question, where '<' and '>' mark the boundaries set by the tagger for the subject-verb-object sequence. The sentence constitutes the only piece of information which is taken directly from the corpus. The rest of the fields contains my annotations of the respective sentences.
2. **TopObj?** - a Boolean field, where the value may only be one out of two (yes/no), indicating whether or not the sentence in question has a topicalized object.
3. **DefSubj, DefObj** - text fields giving the definiteness of the subject and object along Aissen's scale of definiteness (cf. chapter 3.4). The scale for definiteness is repeated below, where the tags used in the database are given in parentheses:

'The girl knows that the house will be sold'

In the above examples we see that a complement clause initiated by the complementizer *at* 'that' may function as a subject in (1a) and a topicalized object in (1b). Here, however, the complement clause is in first position. The same ambiguity does not seem as present when the complement clause is placed after the verb:

- (2) a. Jenta vet at huset blir solgt
girl-DEF knows that house-DEF becomes sell-PASS.PART
'The girl knows that the house will be sold'
- b. ?? Jenta bekymrer at huset blir solgt
girl-DEF worries that house becomes sell-PASS.PART

As we see then, a topicalized version of (1a) in (2b) seems highly unlikely, and has not been taken into account when constructing the Oslo-Bergen tagger.

Personal Pronoun (pro) > Proper Noun (pn) > Definite NP (def) >
Indefinite NP (indef)

4. **AnimSubj, AnimObj** - text fields giving the animacy of the subject and object along Aissen's scale of animacy (cf. chapter 3.4). The scale for animacy is repeated below, where my tags, again, are in parentheses:

Human (hum) > Animate (anim) > Inanimate (inan)

5. **Verb** - a text field stating the transitive verb of the sentence. Due to restrictions on the sample, as described in the above section, there is only one verb in focus, the matrix verb, and no auxiliaries.
6. **NorKompleks** - a text field giving the NorKompleks code of this verb. The code has been taken from the NorKompleks list of verbs which are coded for argument structure.
7. **Comment** - a text field for comments regarding non-literal²³ language use, like metonymy and metaphor.

An example of a few entries are given in table 4.2 below.²⁴

4.3.2 Extracting the results

SQL-queries

In order to extract the results I have been interested in from the database, I have written a number of queries in SQL (Structured Query Language) and directed these towards the database file of transitive constructions. Queries select records from one or more tables in a database so they can be viewed, analyzed, and sorted. SQL easily facilitates conjunctive or disjunctive queries, i.e. queries which posit several subconditions with a relationship of either conjunction or disjunction holding between them. This proved very useful in the case of the sampled data, where we are looking for e.g. a sentence with an inanimate subject and an object which is animate²⁵ or human. The query for this is provided below:

```
SELECT [sample].[AnimSubj],[sample].[AnimObj],  
       Count(*) AS CountAnim  
FROM sample  
WHERE ([sample].[AnimSubj]="inan" And  
       ([sample].[AnimObj]="anim" Or
```

²³More on non-literal language in section 4.4 below.

²⁴*etc.* indicates that the sentence continues. In the real database table, however, the fields are not subject to such strict requirements of space and contain the full sentence. The real database does not contain translations of the sentences, but they are provided in the example table 4.2 for the benefit of the reader.

²⁵The animate property, as we remember, pertain to all living things apart from humans.

Sentence	TopObj?	DefSubj	AnimSubj	DefObj	AnimObj	Verb	NorKompleks	Comment
<En person brukte ulykken> etc. 'A person used the accident'	No	indef	hum	def	inan	bruke	tr1,tr3	
<Høyre krever direkte utbetaling> etc. 'The right wing party demands direct payment'	No	pn	hum	indef	inan	kreve	tr1,tr2	Metonymy
<Det gjorde han> som vanlig bra etc. 'He did it well as usual'	Yes	pro	hum	pro	inan	gjøre	tr1	
<Mobutu eier en villa> etc. 'Mobutu owns a villa'	No	pn	hum	indef	inan	eie	tr5	
<Tauet understreker det robuste preget> 'The rope underlines the robust look'	No	def	inan	def	inan	understreke	tr1,tr2	Metaphor
Pål <Lydersen spilte hele kampen> etc. 'Pål Lydersen played the whole game'	No	pn	hum	def	inan	spille	tr1	
<Loven omfatter ikke folkeskolelærerne> etc. 'The law doesn't include the schoolteachers'	No	def	inan	def	hum	omfatte	tr5	

Table 4.2: Example entries from the database table

```
[sample].[AnimObj]="hum"))
GROUP BY [sample].[AnimSubj],[sample].[AnimObj];
```

This is a SELECT query, which selects entries according to conditions provided in the WHERE-clause. In this case we also want to count the number of occurrences satisfying the WHERE-clause. The table in question is called *sample* (as specified in the FROM-clause). The WHERE-clause states that the sentences selected should have the following properties: the subject's value for animacy should be *inan* and the object's value for animacy should be either *anim* or *hum*. This would give us the number of sentences in the whole of the sample where the object is higher in animacy than the subject.

Queries may also be directed towards the results of other queries. This has proved very useful in the extraction of information from the sample database. An example of this is a query which simply counts the number of sentences containing a topicalized object in the database.²⁶ The result of this query was called *topicalized*, and other queries pertaining only to topicalized constructions could be directed to the *topicalized* table, instead of complicating the call by building this into a subquery. For instance, I wanted to know how frequently an inanimate pronoun occurred as a topicalized object, i.e. I wanted to count all the topicalized constructions where the

²⁶The query for this is given below:

```
SELECT sample.[TopObj?],Count(sample.[TopObj?]) AS [CountOfTopObj?]
FROM sample
WHERE sample.[TopObj?]=True
```

The query simply states that all fields in the TopObj? column that bear the value of *true* are to be counted in CountOfTopObj?.

object was an inanimate pronoun, and made use of the following query:

```
SELECT topicalized.DefObj, topicalized.AnimObj, topicalized.Sentence,
       Count(*) AS Countpro
FROM topicalized
WHERE topicalized.DefObj="pro" And topicalized.AnimObj="inan"
GROUP BY topicalized.DefObj, topicalized.AnimObj,
         topicalized.Sentence;
```

We will examine this result, as well as the other results from the database analyses, further in the next chapter.

4.4 Non-literal language

The title alludes to non-canonical usage of arguments in a transitive sentence, opening for relations between the arguments that do not agree with the hierarchies posited in chapter 3. This is not to say that this represents usage that is ungrammatical or improper in any way. When the relationship between the subject and object does not match the hierarchical predictions, it is crucial to obtain a systematic understanding of these non-conforming cases, as to better be able to account for them. Through work on the data of transitive constructions, as outlined in the above sections, it has become apparent that not all constructions conform neatly to the theoretical predictions of chapter 3. We have already looked at some cases of deviance from the theoretical predictions in form of the reverse-animacy verbs. These are, however, different from non-literal language use, as their deviance follows from properties of the verb. Non-literal language is not as easily classified, and a key question, as we shall see, pertains to how these types of sentences shall be annotated.

Metonymy and metaphor seem to be the most common examples of non-literal language. Lakoff and Johnson (1980) differentiate clearly between the two, stating that they, in fact, involve two quite different processes. Metaphorical usage is employed as a means of understanding, “conceiving of one thing in terms of another” (Lakoff and Johnson, 1980, p. 10), whereas the main function of metonymy is to be referential, “it allows us to use one entity to stand for another” (Lakoff and Johnson, 1980, p. 36). Metonyms are usually thought to be easier to decipher, as they involve a clearer relationship to the referent. Fass (1988) embellishes on this, stating that “the core of metonymy is a semantic relationship, [...] whereas the core of a metaphor is a relevant analogy” (Fass, 1988, p. 178).

In the following we will take a look at both phenomena, supplied with examples from the literature and our data of transitive constructions, as well as examining the practical consequences of metonymies and metaphors from the viewpoint of this thesis and its topic.

4.4.1 Metonymies

(23) The ham sandwich is waiting for his check

The classic example above employs metonymy in that it uses “one entity to refer to another that is related to it” (Lakoff and Johnson, 1980), in this case a customer in a restaurant and a ham sandwich (the meal the customer has ordered). The distinction drawn between metonymy and metaphor as being a distinction between *understanding* and *referentiality*, is not necessarily as simple as that. Lakoff and Johnson (1980) also consider metonymy to figure in understanding, and not only as a fixing of reference. The fact that a speaker chooses a specific metonymic concept, conveys information regarding the focus of the speaker. In the above example, the waiter uttering the sentence chooses to focus only on the order of the customer, as this is what he/she conceives of as important in the relevant setting.

A central point for understanding non-literal usage, and in this case metonymy, is that it is not necessarily a random or arbitrary phenomenon. Lakoff and Johnson (1980) would argue that it is, in fact, mostly systematic, and they list several patterns of metonymic usage to illustrate,²⁷ e.g.:

- Controller for controlled

(24) Nixon bombed Hanoi.

- Institution for people responsible

(25) The Senate thinks abortion is immoral.

- The place for the institution

(26) Hollywood isn't what it used to be.

As Fass (1988) points out, metonymy often involves a violation of the semantic selectional restrictions of the verb. The reason for this is that in metonymic constructions “the actual argument of a predicate is not the literal argument, but is instead implicit and related to the literal argument through an implicit binary relation.” (Stallard, 1993, p. 87). In (25), for instance, *think* usually selects an animate, thus usually cognate, subject. The actual subject here is the institution of the Senate, whereas the implicit argument obviously is the members of the Senate, who certainly are animate and cognate.

Stallard (1993) proposes a test for the view that there is in fact an implicit argument linked to the actual argument through a semantic relation, the “indirect reference view”. This is a simple test of anaphoric agreement:

²⁷The examples in (24) - (26) are taken from Lakoff and Johnson (1980).

- (27) [The ham sandwich]_i is waiting for his check.
He_i is getting a little impatient.

The anaphoric pronoun *he* can felicitously be used to refer back to *the ham sandwich*, thus showing that there is an implicit reference of a human customer implied. Stallard (1993) compares the above example with the following:

- (28) Nixon_i bombed Hanoi.
They*_i sang all the way back to Saigon.

Here, *Nixon* is a metonym of the type “Controller for the controlled”, and is used instead of mentioning the exact group of people, who, under Nixon’s government, bombed Hanoi. The test of indirect reference, however, fails. This fact motivates a distinction, Stallard (1993) claims, between two types of metonymy, *referential* and *predicative* metonymy. Referential metonymy, exemplified by (23) and (27) above, denotes the type of metonymy where the indirect reference holds, i.e. there is an intended referent different from, but related to, the literal referent. Predicative metonymy, however, describes the case where, as in (28), the literal and the intended referents are identical, and rather than a reference shifting of the DP, we are dealing with a “coerced predicate”, a stretching, so to speak, of the predicate’s usage:

- (29) Nixon_i bombed Hanoi.
He_i wanted to force the Communists to negotiate.

Since there is no indirect reference, the appropriate anaphora is the one agreeing with *Nixon*, namely *he*.

4.4.2 Metaphors

As is the case for metonymies, metaphors are also quite commonly used in language. Lakoff and Johnson (1980) argue that metaphors are also systematic and list numerous patterns of metaphoric usage, e.g. (Lakoff and Johnson, 1980):

- Argument is war

- (30) I *demolished* his argument

- Time is money

- (31) You’re *wasting* my time

Lakoff and Johnson (1980) argue convincingly that these patterns exist. They are not, however, as easily determinable as the metonyms and the metonymic patterns.

There is one type of metaphor which constitutes a problem from the viewpoint of this thesis, namely that of the metaphor of *personification*. Lakoff and Johnson (1980) use the term personification to cover a wide range of metaphors, which all have in common that they employ a physical object which is given human characteristics. The motivation for this is that “it allows us to comprehend a wide variety of experiences with nonhuman entities in terms of human motivations, characteristics and activities” (Lakoff and Johnson, 1980, p. 33). In particular, such a non-human entity may be used as subject for a verb which usually would require an animate subject, e.g. (Lakoff and Johnson, 1980):

(32) Inflation is eating up our profits.

4.4.3 Practical consequences

As mentioned above, these non-literal usages cause irregularities in the data which are important to be aware of. However, as all the authors cited above point out, most types of metonymies at least are in fact quite regular and systematic. They have even attempted to account for these systematicities by proposing patterns of metonymy.

Metonymy, in particular, is very systematic and perhaps the most common non-literal phenomenon observed in the data material. A few patterns in particular abound:²⁸

(33) The place for the institution

- a. USA fremsatte et endringsforslag
USA proposed an amendment
'USA proposed an amendment'

(34) Institution for people responsible

- a. Bedrøftet har mange års erfaring
company-DEF has many years experience
'The company has many years of experience'

These two patterns, in particular, are frequent in the data, and posit a problem as they usually make use of inanimate subjects.

In general, these systematic usages of metonymy may in principle be accounted for in the lexicon. In addition, one obviously would need a named entity recognizer capable of discerning countries and organizations in particular.

Markert and Nissim (2002) propose a scheme for annotating corpora for metonymy, and argue that this is not a perfectly simple task, as metonymies are open-ended. The scheme they propose makes use of several annotators

²⁸All Norwegian examples in the following section are taken from the data sample of the Oslo Corpus.

and several stages of annotation in order to obtain something close to a gold standard. They do, however, make use of metonymic patterns from the literature, such as the ones reviewed above.

Personification also occurs in the data:

- (35) a. Komedien holder hodet over vannet
 comedy-DEF holds head-DEF above water-DEF
 ‘The comedy manages to keep its head above water’
- b. Denne fanen sier noe om hvor engasjert
 this banner-DEF says something about how engaged
 barna har vært i arbeidet
 children-DEF have been in work-DEF
 ‘This banner says something about how engaged the children
 have been in their work’

In contrast to metonymy, this process is not systematic and regular, but rather, creative and difficult to pin down. A possible approach might be to claim that this type of creative metaphoric usage relies heavily on unmarked, canonical word order in order to mark grammatical relations.

The annotation of non-literal language

Dealing with non-literal language when it comes to annotation for animacy is problematic. As mentioned above, both metaphors and metonymies violate the selectional restrictions of the verb for which they serve as arguments.

In the case of metonymies, however, I have chosen to annotate according to the indirect reference of the metonymic concept, that is, making use of the test of anaphoric agreement, as outlined by Stallard (1993) in (27) - (28) above, in order to determine the animacy of the indirect reference. Here, predicative and referential metonymy will differ. It is only in the cases of referential metonymy that an indirect referent is postulated, thus revealing an underlying referent which in most cases is animate, even if the actual argument is not. It will be annotated accordingly:

- (36) USA_[ANIM,PN] fremsatte et endringsforslag_[INAN,INDEF]
 USA proposed an amendment
 ‘USA proposed an amendment’

The above example represents a regular pattern, as we saw in (33) above. However, creative metonymy, like the example with the ham-sandwich in (23), does occur:

- (37) Villaen på Stabekk åpnet gjerne sine dører
 villa-DEF at Stabekk opened gladly its doors
 ‘The villa at Stabekk gladly opened its doors’

This type of metonymy is harder to spot, and easier to confuse with metaphoric use, especially that of personification. However, the main difference remains; most types of metonymic use are concerned with reference, in the case of (37), an indirect reference to the inhabitants of the villa.

Metaphors, which do not have an indirect referent like metonymies, will be annotated according to their actual argument:

- (38) Komedien_[INAN,DEF] holder hodet_[INAN,DEF] over vannet
 comedy-DEF holds head-DEF above water-DEF
 ‘The comedy manages to keep its head above water’

Typical metonymic subjects like organizations, governments etc. are not always used metonymically. They may also be used in the original collective sense, and will be annotated accordingly:

- (39) Horten Videregående skole_[INAN,PN] representerer et
 Horten high school_[INAN,PN] represents an
 gjennomsnitt_[INAN,INDEF] av de allmennfaglige videregående
 average_[INAN,INDEF] among the allmennfaglige high
 skolene
 schools
 ‘Horten high school_[INAN,PN] represents an average_[INAN,INDEF] among
 the allmennfaglige high schools’

Chapter 5

Animacy, definiteness and word order variation

In this chapter we shall take a closer look at the results from the data analysis. We shall see that there are some interesting generalizations regarding the influence of animacy and definiteness on the realization of syntactic functions in Norwegian, and also, with regards to their relative positioning within the transitive sentence. Below, the results from the data analysis will be presented and discussed. We will also consider the implications of our findings with regards to our theoretical predictions, as well as the practical side of our thesis, towards the disambiguation of the subject and object.

5.1 Sampling results

The sample, as we remember, counts a thousand transitive sentences fulfilling the criteria set up in the previous chapter and annotated in a database.

5.1.1 Animacy

The results for animacy are shown in table 5.1 below.¹

When assembling the sample I was in particular interested in examining the relationship between the subject and object of a transitive construction with regards to the dimensions of animacy and definiteness. Dahl and Fraurud (1996) conclude that a constraint expressing a reluctance for the subject to be lower in animacy than the object is prevalent in Swedish as a strong statistical tendency. Our data sample certainly adheres to this tendency as well:

¹Due to the fact that only 5 animate, i.e. animal, not human, subjects or objects were found in the sample, these are joined together with the results for human subjects and objects, giving us the composite category Hum/Anim.

Subject	Direct Object	N	%
Hum/Anim	Hum/Anim	74	7.4
Hum/Anim	Inan	615	61.5
Inan	Hum/Anim	24	2.4
Inan	Inan	287	28.7
Total		1000	100.0

Table 5.1: Distribution of 1000 transitive sentences according to the animacy of subject and object

- Subject higher than or equal to object in animacy (i.e. the combinations Hum/Anim - Hum/Anim, Hum/Anim - Inan and Inan - Inan): 97.6%
- Object higher than subject in animacy (i.e. the combination Inan - Hum/Anim): 2.4%

Metonymic Subjects

As chapter 4.4 above made clear, non-literal language, and metonymy in particular, is quite common in natural language. With regards to annotation I have chosen to annotate metonyms according to their indirect reference, which in the majority of cases is human, despite an inanimate actual argument. Dahl and Fraurud (1996) are unfortunately not explicit as to their choices of annotation of metonyms. Therefore it is not at all clear how they arrive at their fairly good results. As we remember, Dahl and Fraurud (1996) arrive at the conclusion that more than 97% of the transitive sentences in their material have a subject which is higher than the object in animacy. In comparison with the analyses done in this thesis, it seems highly unlikely that they would achieve such high counts for animacy without treating the indirect referents of metonymies as human.

The arguments found to be metonymic are annotated as such, giving us an idea as to the frequency of this type of language use:

- Number of metonymic subjects: 156, 15.6%

The number of animate subjects in total is 68.9%, i.e. the first two rows in table 5.1 above. However, had we not annotated metonymic subjects according to their indirect reference, the result would have been lower, 53.2%. Conversely, the number of objects higher in animacy than the subjects would probably also have risen.

The deviant 2.4%

As stated above, the number of sentences adhering to the principle that the subject should be higher or equal to the object in animacy is 97.6%. This

leaves us with 24 sentences which deviate from this norm. An interesting question then, pertains to whether these sentences have any common characteristics, as opposed to the ones that follow the norm.

The predictions are that these cases will either be ones where a reverse-animacy verb (cf. chapter 3.6) figures as the main verb, giving us sentences with reverse animacy relationships between the arguments, or sentences exhibiting non-literal language, e.g. making use of metaphorical language to reverse the subject-object relationship (cf. chapter 4.4). These two scenarios turned out to cover all the 24 sentences in question, as we see from table 5.2 below.

Reason for deviance	N	%
Reverse-animacy verb	22	91.7%
Non-literal	2	8.3%
Total	24	100.0%

Table 5.2: Properties of the deviant 24 sentences with object higher than subject in animacy

Deviant cases: reverse-animacy verbs Earlier, we have looked in some detail at the reverse-animacy verbs. These are transitive verbs which may take an inanimate subject and an animate object. We differentiated three main groups of verbs which belong to this meta-group of verbs:²

1. *amuse*-verbs, a subgroup of psych-verbs, e.g. *interessere* ‘interest’
2. *concern*-verbs, e.g. *gjelde* ‘concern’
3. causative verbs, e.g. *skade* ‘hurt’

Reverse-animacy verbs obviously pose a difficulty to our theory of scalar alignment between the arguments.

Some examples of sentences containing reverse-animacy verbs from the deviant 2.4% of the sample include the following:

²The first two groups of verbs contained within the reverse-animacy verbs, the *amuse*-verbs and the *concern*-verbs, bear the codes *trans5*, *trans6* and *trans7* from the NorKompleks lexicon. These are all codes that involve an argument structure with a Theme subject. Contrary to traditional theory on thematic roles, the codes *trans5* and *trans6*, in fact distribute the Theme role to both arguments.

The *amuse*-verbs are coded as *trans7*, however, most of these have an agentive reading as well and are then coded as *trans10*. Often a psych-verb is coded as both *trans7* and *trans10*. In four of the cases examined here (i.e. sentences with a reverse relationship between subject and object), however, the verbs are coded only as *trans10*. This is obviously wrong, as they are not agentive in this use. Even so, due to the fact that these were correctly coded as psych-verbs, I have included them in the reverse-animacy group of deviant cases. The *concern*-verbs are coded as *trans5* or *trans6*, according to whether they passivize or not. The causative verbs are not uniformly coded in NorKompleks. We will return to this matter in chapter 7.

- (1) a. Denne samtalen plager Oline i lang tid
 this conversation-DEF bothers Oline in long time
 ‘This conversation bothers Oline for a long time’
- b. Loven omfatter ikke folkeskolelærerne
 law-DEF includes not schoolteachers-DEF
 ‘The law does not include schoolteachers’
- c. Terrorbomben som eksploderte drepte hundrevis av
 terror-bomb-DEF that exploded killed hundreds of
 mennesker
 people
 ‘The terror bomb that exploded killed hundreds of people’

The above examples show an example from all the three identified groups contained within the reverse-animacy verbs: an *amuse*-verb, *plage* ‘bother’ in (1a), a *concern*-verb, *omfatte* ‘include’ in (1b) and a causative verb, *drepe* ‘kill’ in (1c).

There are three sentences in this part of the sample that are not as easily classified as the others. This is partly due to the fact that they were unsatisfactorily coded in the NorKompleks lexicon, which was used as a basic resource for classification when examining the data:

- (2) Amerikansk økonomi belaster ved innledningen til
 American economy burdens at start-DEF of
 primærvalgkampen ikke Bill Clinton
 primary-election-DEF not Bill Clinton
 ‘American economy does not bother Bill Clinton at the start of the primary elections’
- (3) Den nye ordningen rammer brukerne
 the new arrangement strikes users-DEF
 ‘The new arrangement strikes the users’
- (4) Dette rammet ikke FN-styrkene
 this striked not UN-troops-DEF
 ‘This did not strike the UN troops’

The above examples contain two verbs, *belaste*, ‘burden’ and *ramme*, ‘strike’, both coded as trans1 in the NorKompleks lexicon, i.e. prototypical transitive verbs, taking an agentive subject and a Theme object. However, as the examples make abundantly clear, these are not used in a prototypical agentive manner here. A short search in the Oslo corpus for these two verbs, made clear that they in fact have a lot in common. They were both passive in the majority of cases in the corpus:³

³The examples in (5) and (6) are taken from the Oslo Corpus.

- (5) Avdrag på lån *belastes* den person som står som eier
 downpayments on loan burdens that person who stands as owner
 ‘Mortgage downpayments are burdened the person who stands as owner’
- (6) Kvinner *rammes* ofte av beinskjørhet
 women struck often by osteoporosis
 ‘Women are often struck by osteoporosis’

Obviously, passive cases are not that interesting in the view of this thesis. This piece of information, does, however, serve to point us in the right direction with regards to the proper classification of these verbs. Lødrup (2000) uses thematic roles in order to explain exceptions to the Norwegian passive, and states that verbs with a Theme subject do not passivize. This generalization does not, however, seem to hold up with regards to the two verbs mentioned above. These take theme subjects and certainly passivize. As we remember from chapter 3.6, however, non-passivization is not an absolute criterion for the *amuse*-verbs or the *concern*-verbs. Even though our two initial verbs are used frequently in the passive, they may also be used actively, as in examples (2), (3) and (4) above. However, the verbs are used agentively extremely rarely (only once for *ramme*, ‘strike’ in the whole of the Oslo corpus). A reverse-animacy use however, is common among the active cases, as in examples (2), (3) and (4) above. These facts lead us to suggest an extension of the coding of these two verbs and inclusion of them into the reverse-animacy group of the deviant part of the sample. As they lack the property of causation, they must be said to resemble most the *concern*-verbs.

An interesting group of verbs, which are not represented in our classification of possibly reverse-animacy verbs, but which deserves mention are *two-place unaccusatives*. Unaccusative verbs are typically one-place verbs which lack an external argument. They are in a sense inherently passive, but without morphological marking for passive, and cannot passivize. Although unaccusatives are usually one-place, there is a small group of two-place unaccusatives, e.g. *tilfalle* ‘fall to’⁴. We find two examples of these in the data material, both with the verb *tilfalle* ‘fall-to’. As mentioned above, the one-place unaccusatives only subcategorize for an internal argument, so this argument must either move to assume subject-position or an expletive subject must be inserted in order to realize a subject. As for the two-place unaccusatives, these subcategorize for two internal arguments, much like di-

⁴It is difficult to come up with an adequate translation to English for the two-place unaccusatives as they usually realize their indirect object as an oblique:

- (1) En ulykke hendte jenta
 an accident happened-to girl-DEF
 ‘An accident happened to the girl’

transitive verbs, and also hand out the roles of Benefactive and Theme to their arguments (Lødrup, 1995). Much like ditransitive verbs, the two-place unaccusatives may occur in double-object constructions, however with an expletive subject, as in (7b). The direct object may also assume subject position, as in (7a). As we see in (7c), two-place unaccusatives may not passivize.

- (7) a. En arv tilfalt jenta
 an inheritance fall-to girl-DEF
 ‘An inheritance “fell” to the girl’
- b. Det tilfalt jenta en arv
 it fell-to girl-DEF an inheritance
 ‘An inheritance “fell” to the girl’
- c. *Jenta ble tilfalt en arv
 girl-DEF was fell-to-PASS.PART an inheritance

The sentence in (7a) above, looks very much like our other reverse-animacy constructions, and will therefore have to be accounted for. The fact that this is a rather limited group of verbs, however, makes them easy to handle. Even though the apparent object in (7a) above, is in fact an indirect object, and the subject the underlying object, we will include these within our meta-group of reverse-animacy verbs. Part of the reason for this, is obviously that we are dealing with an automatic, linear analysis which does not take into account underlying levels of linguistic structure.

In general then, as we saw in chapter 3.6, languages have a quite wide range of verbs which may allow for a reverse-animacy use, especially when we take into account the causative verbs. It is therefore quite striking that we find so few of these in our material. This indicates that our assumptions regarding the markedness relations expressed by combinations of the prominence hierarchies, as accounted for in chapter 3, to a large extent condition the realization of arguments in Norwegian.

Deviant cases: metaphoric language The rest of the group of deviant sentences, which amount to only three sentences, are metaphoric:

- (8) a. Flaks reddet Ine (8) fra døden
 luck saved Ine (8) from death
 ‘Luck saved Ine (8) from dying’
- b. Fjorårets messe trakk 10 000 mennesker
 Last-year’s fair drew 10 000 people
 ‘Last year’s fair drew 10 000 people’

The verbs employed here are all prototypical transitive verbs, taking an agentive subject and a theme/patient object. Their non-prototypicality with regards to arguments in these cases can be attributed to metaphoric use.

5.1.2 Definiteness

The results for definiteness are far less conclusive than those for animacy. The fact that we are dealing with a more finely grained scale is by all likelihood partly responsible for the more “scattered” distribution of definiteness and grammatical relations. The results for definiteness are presented in table 5.3 below.

The picture becomes clearer when looking at the hierarchical relationship

Subject	Direct Object	N	%
Pro	Pro	21	2.1
Pro	PN	5	0.5
Pro	Def	56	5.6
Pro	Indef	48	4.8
PN	Pro	50	5.0
PN	PN	35	3.5
PN	Def	178	17.8
PN	Indef	108	10.8
Def	Pro	39	3.9
Def	PN	18	1.8
Def	Def	196	19.6
Def	Indef	134	13.4
Indef	Pro	10	1.0
Indef	PN	4	0.4
Indef	Def	56	5.6
Indef	Indef	42	4.2
Total		1000	100.0

Table 5.3: Distribution of 1000 transitive sentences according to the definiteness of subject and direct object

between the subject and object with regards to definiteness. There does seem to be a tendency towards avoiding a less definite subject than object:

- Subject higher than or equal to object in definiteness: 823 sentences - 82.3%
- Object higher than subject in definiteness: 177 sentences - 17.7%

However, this is to be expected, as the subject is, after all, the default topic, as the unmarked word order is SVO. As the topic tends to be given information, thus often definite in form, this tendency is not at all surprising.

Indefinite subjects

From the harmonic alignment of the hierarchies of definiteness and syntactic functions (cf. chapter 3.4), it follows that indefinite subjects constitute the maximally marked subject with regards to the dimension of definiteness. Indefinite subjects are ungrammatical in several languages and subject to strict interpretational restrictions in others (Aissen, 2003, p. 445).

They are thought to be associated with special types of readings and are thought to be indirectly responsible for the impersonal active or *presentational* construction (Mikkelsen (2002), Sveen (1996)).⁵

- (9) Det oppsto brudd mellom stoffet og
it occurred break-INDEF between substance-DEF and
tankveggen
tank-wall-DEF

‘A break occurred between the substance and the wall of the tank’

This construction then, contains an expletive subject and an object which is the logical subject of the sentence. The object position may only be occupied by an indefinite argument, a so-called definiteness effect. It is thought to arise through a strategy for avoiding an indefinite subject in initial subject position, thus inserting an expletive subject instead (Sveen, 1996; Mikkelsen, 2002).

Sveen (1996) for Norwegian and Mikkelsen (2002) for Danish, both state that indefinite arguments in subject position incur an interpretational effect which is not present in the presentational construction. An example pair taken from Sveen (1996, p. 144) illustrates this:

- (10) Noen studenter spiser på Frederikke hver dag
some students eat at Frederikke every day
‘Some students eat at Frederikke every day’
- (11) Det spiser noen studenter på Frederikke hver dag
it eats some students at Frederikke every day
‘Some students eat at Frederikke every day’

The personal active in (10) and the impersonal active or presentational construction in (11) incur different interpretations of their argument. (10) indicates that a specific group of students consume their food at Frederikke every day, whereas (11) does not impose this restriction of *specificity* on its argument. Another commonly mentioned property of indefinite subjects is *genericity*, a reading which is simply not available in the presentational construction, as some examples taken from Sveen (1996, p.143) below should make clear:

⁵The example in (9) is taken from the Oslo Corpus.

- (12) En student sover i parken
 a student sleeps in park-DEF
 ‘A student sleeps in the park’
- (13) Det sover en student i parken
 it sleeps a student in park-DEF
 ‘A student sleeps in the park’

Where the personal version in (12) certainly allows a generic reading, the presentational (13) does not. In (12) the property of sleeping in parks may be attributed to students in general, whereas (13) simply states that some student is sleeping in a park.

In addition to genericity and specificity, indefinite subjects may also be interpreted as *partitive* (Mikkelsen, 2002). Here, the indefinite subject is seen to be a subset of a previously mentioned set, as in (14b) below, taken from Sveen (1996, p. 145):

- (14) a. Per har *fem venner*
 Per has five friends
 ‘Per has five friends’
- b. *Tre* bor i Oslo og *to* i Bergen
 three live in Oslo and two in Bergen
 ‘Three (of them) live in Oslo and two in Bergen’

These properties of indefinites leads to a distinction between *strong* and *weak* indefinites, where the strong indefinites are indefinites which are specific, generic or partitive, whereas the weak indefinites are the standard referential indefinites that correspond fairly well to an existentially quantified expression where the quantifier has narrow scope (Mikkelsen, 2002). Strong indefinites have more in common with definites than the weak indefinites, thus Mikkelsen posits a hierarchy⁶, where the strong indefinites are placed in between definites and the weak indefinites. Part of the same distinction is reflected in Aissen’s split between Specific and Non-Specific Indefinites (Aissen, 2003).

Both Mikkelsen and Sveen deal in essence with intransitive verbs. This is an important difference, as transitive sentences do not have the possibility for inserting an expletive in order to avoid an indefinite subject. Even so, Sveen (1996) claims that the avoidance of indefinite subjects tends to extend to transitive sentences as well. Many languages categorically ban indefinite subjects that are weak in the sense of Mikkelsen (2002).

Our data sample includes 113 sentences with indefinite subjects. Quite a few of these are clearly either specific, as in (15) below, or generic, as in (16) below:

⁶Definite > Strong Indefinite > Weak Indefinite (Mikkelsen, 2002, p. 14)

- (15) En nabo oppdaget innbruddstyven
 a neighbour discovered burglar-DEF
 ‘A neighbour discovered the burglar
 ,’
- (16) En grend uten skole mister sitt sosiale senter
 a town without school loses its social center
 ‘A town without a school loses its social center’

Both specificity and partitivity are difficult to ascertain, as they often rely heavily on context. This is the reason why this distinction within the indefinite DPs has not been pursued further in the present context. A possible reason for the difficulty of ascertaining specificity, in particular, in indefinite subjects is the fact that they occupy subject position, and might thus be coloured somewhat by the fact that the subject is default topic. This will tend to give indefinite subjects a strong reading, as opposed to a weak one.

Another interesting feature of the sentences with indefinite subjects in the sample, is that they are very seldomly found as subjects in sentences with a topicalized object. A reason for this is that an initial subject will be default topic, thus, perhaps, more easily interpreted as a strong indefinite. With a topicalized object, however, this reading is probably less available, so one strategy for avoiding a weak indefinite subject might be to position it sentence-initially. Weak, indefinite subjects after all constitute the maximally marked configuration along the dimension of definiteness. Only three sentences of all the sentences with an indefinite subject has a topicalized object. Also, all of these subjects are modified restrictively, so as to enforce a strong reading, as in (17):

- (17) Det sier en oppgitt nestleder av Fløen vel,
 it says a resigned deputy-leader of Fløen welfare-organization,
 Aida Hansen, til Bergens Tidende
 Aida Hansen, to Bergens Tidende
 ‘So says a resigned deputy leader of Fløen welfare organization, Aida Hansen, to Bergens Tidende’

5.1.3 Object topicalization

The number of sentences fronting a topicalized object in the sample amounted to 97 sentences, or 9.7% of the sampled sentences. However, with only the immediate context of the sentence available, sentences containing topicalized objects which are disambiguated mainly by a larger context may have been overlooked. The sentences with topicalized objects found in the sample might represent a prototypical, and therefore, easily recognizable, type of topicalization. In other cases, perhaps, will the larger linguistic context contribute towards disambiguation. In cases like these, the context might

for instance disambiguate a sentence where the arguments are equal with regards to the animacy or definiteness scales.

However, as I have marked for topicalization only the cases which are fairly clear, the results are conservative, rather than over-generating. The following results will have to be viewed in light of this.

Animacy and definiteness

We are interested in the influence that the relative animacy or definiteness of the arguments in a transitive construction might have on word order variation in the sentence. We will therefore in the following take a closer look at the relationship between the subject and object along these dimensions in the sentences containing a topicalized object.

We will first examine the results for animacy in the sentences containing a topicalized object:

- Subject animacy higher than or equal to that of the topicalized object: 100%, wherein (Subject-Object)
 - Hum/Anim - Hum/Anim: 2 sentences - 2,0%
 - Hum/Anim - Inan: 89 sentences - 91,8%
 - Inan - Inan: 6 sentences - 6,2%

There was not a single sentence with a topicalized object where the object was higher in animacy than the subject.

The corresponding results for definiteness in the sentences containing a topicalized object are presented below:

- Subject higher than or equal to topicalized object in definiteness: 48 sentences - 49.5%
- Topicalized object higher than subject in definiteness: 49 sentences - 50.5%

Once again, the picture for definiteness is much more fragmented. We see that a little over half of these sentences have an object which is placed higher in definiteness than the corresponding subject. This is certainly a high number, however, when one reflects over this, it is not that surprising. Obviously, an object which is topicalized will be an entity which displays a high degree of topicality. A common characteristic of such an entity would be that it represents *given* information, i.e. information that is in some sense known to the speaker and hearer. It is not surprising then, that this object which is topic, will be placed high in definiteness, whereas the corresponding subject may often be lower. One reason for the significant rates of high-placed objects in this case is that the inanimate pronoun *det*, 'it', as well as demonstratives, such as *dette* 'this' are very commonly used object topic, referring back in the discourse:

- (18) *Det* lovet Leif
 It promised Leif
 ‘Leif promised it’
- (19) *Dette* sier en skuffet Miss Norway
 this says a disappointed Miss Norway
 ‘This is what a disappointed Miss Norway says’

Det ‘it’ is in many ways a problematic pronoun, as it serves several functions in Norwegian. As mentioned above it is a 3. person singular pronoun which refers to inanimates or abstract concepts. It also serves as an expletive subject of, among others, presentational constructions and impersonal passives. It seems, however, that confusion between presentational constructions and a topicalized *det*-object when it comes to automatic disambiguation might not be as severe as one might think. First of all the verb in presentational constructions will usually be intransitive, whereas the verb of the topicalized version will certainly be transitive.

Turning to the examples in (18)-(19) above, we see that it should be fairly easy to distinguish these from presentational constructions. This does not however indicate that all other instances of initial *det* ‘it’ are topicalized objects. The more difficult case is to be found in distinguishing transitive sentences with topicalized *det* ‘it’ object from transitive sentences with a pronominal subject *det* ‘it’. Examples of the pronominal version are provided below:

- (20) *Det* oppsummerte den norske gårdsdagen
 it summed-up the Norwegian yesterday
 ‘That summed up yesterday’s events in Norway’
- (21) *Det* gjelder blant annet reglene om vedlikehold
 it concerns among others rules-DEF about maintenance
 ‘It concerns, among other things, the rules on maintenance’

The pronominal use of *det* ‘it’ refers to an inanimate entity, or a part of earlier discourse. In comparison with the topicalized *det* ‘it’, as in examples (18) and (19) above, these cases display a quite different relationship between the subject and object. The topicalized cases in the data show a clear asymmetry between the arguments in animacy, i.e. the subject is almost always higher in animacy than the object, whereas the cases with pronominal *det* ‘it’ subjects display an equal relationship between the arguments, they are both inanimate, except for one case which has a reverse-animacy verb and thus a human object.

5.1.4 Equally placed arguments

Quite a few of the arguments in the data are equally placed on one or both of the scales of animacy and definiteness. This constitutes a middle ground with

regards to the markedness predictions expressed through harmonic alignment of the scales of syntactic functions with the scales for definiteness and animacy respectively. In these sentences, one of the arguments represent a marked constellation whereas the other argument represents the more unmarked possibility. For instance, in sentences where both arguments are inanimate, the subject represents the marked combination, whereas the object has unmarked properties with regards to animacy. In example (22) below, we see two arguments bearing the same value for animacy, namely inanimate, whereas in (23) the arguments are equally placed on the scale for definiteness:

- (22) Meldingen utløste dette årets første runde med
 announcement-DEF triggered this year's first round of
 gjetninger
 guesses
 'The announcement triggered this year's first round of guesses'
- (23) Pasientene aksepterer stort sett de nye reglene
 Patients-DEF accept largely seen the new rules-DEF
 'The patients generally accept the new rules'

A reasonable prediction might perhaps be that in cases where the arguments are equally placed on one scale, they will not be so on the other, so that the other scale contributes towards disambiguation. Let us examine this prediction closer.

A quite high number of sentences has arguments placed equally on one of the scales of animacy or definiteness (see table 5.4 below).⁷ Of these

Scale	No. of sentences	% of whole sample
Animacy	361	36.1
Definiteness	294	29.4

Table 5.4: Number of sentences with equally placed placed on the scale of animacy or definiteness

sentences there are 136 sentences or 13.6% of the original sample of 1000 sentences, that contain arguments placed equally on both scales. These thus represent the intersection of the two sets represented in table 5.4 above. Two examples of these are provided below:

- (24) Eva-Britt besøker Hamilton i Moskva
 Eva-Britt visits Hamilton in Moscow
 'Eva-Britt visits Hamilton in Moscow'

⁷Due to the fact that a sentence may have equally placed arguments on both scales, the columns in table 5.4 will contain duplicates. The 136 sentences that have arguments placed equally on both scales are thus represented in both rows.

- (25) Trafikkulykkene representerer de største kostnadene
 traffic-accidents-DEF represent the largest costs-DEF
 ‘The traffic accidents represent the largest costs’

In (24) we find two arguments which are both proper nouns and human, whereas, in (25) the arguments are both definite and inanimate.

The rest of the sentences of table 5.4 have arguments placed equally on one scale but not on the other, as in examples (22) and (23) above. These count 383 sentences, or 38,3% of the original sample.⁸ It might be interesting, then, to examine the hierarchical relationship between the arguments in these sentences, where they are equally placed on one scale, but not on the other. Is it the case that their placements on the scale where they differ may help disambiguate?

This type of sentences may be roughly divided into two main groups:

1. Sentences that are equal in definiteness, but differ in animacy
2. Sentences that are equal in animacy, but differ in definiteness

Sentences of type 1 above are, not surprisingly, disambiguated by animacy in a majority of cases. What this means is that their arguments differ in animacy, and can be disambiguated based on this, as in example (26) below:

- (26) Departementet forbereder disse sakene
 ministry-DEF prepares these cases-DEF
 ‘The ministry prepares these cases’

This example shows a typical case, where the arguments have the same value for definiteness, namely definite, but have different values for animacy; the subject is human⁹ and the object is inanimate. In the sentences of type 1, this accounts for 95,6% of the cases. The rest are accounted for in section 5.1.1 above on the deviant cases. This indicates that in sentences where the arguments are equally placed on the scale for definiteness, the argument’s value for animacy are central to disambiguation. Or rather, animacy is a strong disambiguating factor in general, and also in sentences where the arguments are of equal definiteness.

When a sentence is of type 2, i.e. containing arguments that are equally placed in animacy, but not in definiteness, however, the picture is more complex. There are two subgroups then, sentences where the arguments are both animate (human), as in (27) below, and sentences where both arguments are inanimate, as in (28) below:

⁸Due to the fact that the 136 sentences with arguments that are equally placed on both scales are duplicated in the above figures, this number is obtained by subtracting 136 from the sum in table 5.4 twice, i.e. $361 + 294 - 136 - 136 = 383$.

⁹This is a form of metonymy and is coded for the indirect referent, cf. chapter 4.4.

- (27) Gro møtte PLO-lederen
 Gro met PLO-leader-DEF
 ‘Gro met the leader of PLO’
- (28) Alle vognene på Marina har slike hjul
 all wagons-DEF on Marina have such wheels
 ‘All the wagons on Marina have wheels like that’

Within both of these subgroups there is certainly a tendency towards a subject placed higher than the object on their differing scale of definiteness, so as to help disambiguate. In the subgroup where both arguments are human (animate), 69.6% have subjects that are higher in definiteness than the objects, whereas in the subgroup where both arguments are inanimate, the figure rises to 83,2%.

With regards to word order the sentences of type 2, seem to resist topicalization at a higher frequency than the sample as a whole. An interesting feature of the few sentences that are in fact topicalized (8 all together), is that they seem to a large extent to be disambiguated by definiteness, i.e. the subject is higher in definiteness than the object. The example sentence in (29) below illustrates this, it contains a topicalized, indefinite object and a pronominal subject:

- (29) Innpåslitne menn holder hun på avstand
 pushy men keeps she at distance
 ‘Pushy men, she keeps at a distance’

This generalization holds for all but two of these topicalized sentences, with arguments that are equally placed in animacy, as in (30) below:

- (30) Det kan ikke den hyggelige kirkelunsjen heller
 it can not the pleasant church-luncheon-DEF either
 ‘Neither can the pleasant church luncheon’

Both of these typically front the topicalized pronoun *det* ‘it’, the common object topic, as we have seen above.

5.2 Summary, discussion and practical consequences

In the above sections, the results from the data analysis have been presented and reviewed. In the following, we will sum up these results, as well as outlining some possible practical consequences of these results.

5.2.1 Summary and discussion

Throughout the analysis, we have been interested in the relationship between the arguments of a transitive construction along the scales of animacy and

definiteness. The status of these properties as a conditioning factor on word order variation has also been central.

Section 1 was attributed to animacy. We saw that Norwegian does seem, to a very large extent, to adhere to a constraint which might be stated as follows:

Constraint on subject The subject must never be lower in animacy than the object

It is important to remember, however, that this is not a categorical constraint in Norwegian. Earlier we have seen that in inverse languages, like Navajo, this constraint is categorical, or rather, a sentence where the object is higher than the subject in animacy must be formally marked (by the inverse). In Norwegian, however, no such marking is obligatory, rather, we have observed a strong statistical tendency in this direction. We looked in detail at the deviant cases which do not follow the subject-constraint above, and found that they in general could be classified as belonging to one of two groups:

1. Sentences with reverse-animacy verbs
2. Sentences with metaphoric language

In section 2 we looked at definiteness. In particular we saw that a corresponding subject-constraint for definiteness does not seem as strongly operative in Norwegian as the one for animacy described above.

Another main goal of the data analysis was to survey the frequency of object topicalization in order to obtain a clear picture of the challenge it poses to automatic disambiguation. Also, we wanted to examine the relationship between the arguments in topicalized constructions parallel to the above analysis. We saw that 9.7% of the sentences in the sample had a topicalized object. A complicating factor however, is the degree of context-dependency of object-topicalization, and we were forced to conclude that some topicalizations may have been missed due to the limited context available in the sample.¹⁰ It might be possible, however, to argue that these types of constructions are likely to be avoided, as they easily cause misinterpretations. However, data to support this is not easy to obtain. We will leave it at that for the present, and rather focus on the constructions of which we are certain are topicalized.

We looked at the animacy and definiteness relations between the arguments in the sentences containing a topicalized object. With regards to animacy, the constraint that the object should never surpass the subject in animacy was observed in all the sentences, i.e. there was no topicalized sentence with an object higher than the subject in animacy. In the case of definiteness, the picture was even less conclusive than for the sample taken as

¹⁰The linguistic context of the transitive constructions was only the immediate sentence.

whole. This was in part due to the frequent topicalization of the 3rd person pronoun *det* ‘it’ and the determiner *dette* ‘this’¹¹, as well as the property of givenness which often surfaces in definite form.

Finally, we examined cases where the arguments are equally placed on one or both of the scales. Again, we saw that the scale of animacy is central to disambiguation. When the arguments of a sentence in the sample were equally placed in definiteness, 95,6% of these could be disambiguated solely by animacy. When taking into account the above reflections on reverse-animacy verbs and metaphoric use, all of them could be accounted for.

The sentences equal in animacy could, to a certain extent, be disambiguated based on their differing properties in definiteness. Importantly, the few sentences with equally animate arguments and a topicalized object, seemed to depend on definiteness for disambiguation to a much greater extent than the rest of the group of topicalized sentences.

5.2.2 Practical consequences

The practical side of this analysis, focuses on the possibility for disambiguation of syntactic functions based on properties of their arguments along the dimensions of animacy and definiteness. In addition there is the dimension of word order, which is closely related to syntactic functions in Norwegian. The overall majority of sentences, as we have seen above, have the canonical SVO word order.

When it comes to disambiguating between the subject and object we have focused on semantic properties of these functions through the hierarchies of animacy and definiteness. Animacy certainly seems to be the most conclusive factor in disambiguation. Based on this then, three main scenarios can be envisioned in a disambiguation process where we have a transitive sentence with a linear surface structure like $[\alpha \text{ Verb } \beta]$:

Disambiguating asymmetry $\alpha = \text{hum-anim}, \beta = \text{inan}$

Topicalization/Deviant asymmetry $\alpha = \text{inan}, \beta = \text{hum-anim}$

Equally placed $\alpha = \text{hum-anim/inan}, \beta = \text{hum-anim/inan}$

In the first case, that of disambiguating asymmetry, we find a first argument that is human or animate and a second which is inanimate. This is a case of a prototypical transitive sentence, where we can conclude that we are dealing with an SVO word order. The reason for this is that none of the sentences containing a topicalized object in the sample have a human or animate object and an inanimate subject. This however could have been a possibility in a case of disambiguating asymmetry where we had a topicalized object in a

¹¹As we remember, a determiner which occurs alone is pronominal and is thus annotated as such (cf. chapter 4.2).

reverse-animacy construction. In chapter 3.7, we examined Lee (2002b)'s theory of markedness reduction in Hindi, and asked tentatively whether a similar point might be made for Norwegian, i.e. will there in cases where the subjects and objects are marked with regards to the hierarchies in question be an avoidance of a marked (topicalized) word order? This is certainly the case for the sentences in the deviant group of 2.4%, where we find that none of these contain a topicalized object. This is not to say that a topicalized version of these verbs would be strictly ungrammatical. As we remember, we are here dealing with two main types of sentences deviating from the norm - reverse-animacy use and metaphoric use. The boundaries of grammaticality seem fuzzy with regards to topicalization in these sentences, which certainly are marked. This is exemplified below with the non-topicalized and the topicalized version of the same sentence, taken from the deviant group of the sample - a metaphoric use in (31), an *amuse*-verb in (32), a *concern*-verb in (33) and a causative use in (34) :

- (31) a. Flaks *reddet* Ine (8) fra døden
luck saved Ine (8) from death-DEF
‘Luck saved Ine from dying’
- b. ?? Ine (8) *reddet* flaks fra døden
Ine (8) saved luck from death-DEF
‘Ine (8), luck saved from dying’
- (32) a. Spørsmålet *plager* Espen
question-DEF bothers Espen
‘The question bothers Espen’
- b. ?? Espen *plager* spørsmålet
Espen bothers question-DEF
‘Espen, the question bothers’
- (33) a. Loven *omfatter* ikke folkeskolelærerne
law-DEF includes not schoolteachers-DEF
‘The law doesn’t include the schoolteachers’
- b. ?? Folkeskolelærerne *omfatter* ikke loven
schoolteachers-DEF includes not law-DEF
‘Schoolteachers, the law doesn’t include’
- (34) a. Terrorbomben som eksploderte *drepte* hundrevis av
terror-bomb-DEF that exploded killed hundreds of
mennesker
people
‘The terror bomb that exploded killed hundreds of people’

- b. ?? Hundrevis av mennesker *drepte* terrorbomben som
 hundreds of people killed terror-bomb-DEF that
 eksploderte
 exploded
 ‘Hundreds of people, the terror bomb that exploded killed’

Not surprisingly, the metaphoric use in (31) borders on ungrammaticality. This is not surprising, as the main tool for metaphoric use is a violation of the subcategorization properties of the verb in question, usually a quite prototypical, agentive transitive verb. When this is combined with a marked word order, something is obviously wrong. Our claim here would be that metaphoric use relies heavily on canonical word order in order to function properly.

The reverse-animacy verbs must be said to be more acceptable with a topicalized object, although they are in need of a lot of contextual help, like stress, in order to be fully acceptable. However, a topicalized version of one of the reverse-animacy verbs, like the *amuse*-verb *plage* ‘bother’, becomes much more acceptable when used with an accusative pronoun as topicalized object, thus marking its grammatical function morphologically:

- (35) Meg plager det veldig
 me bothers it very
 ‘Me, it bothers very much’

The fact that none of the deviant sentences are topicalized in the data set tells us that this might not be such a probable case. In comparison, uses of the reverse-animacy verbs with arguments that are not reverse, i.e. agentive uses or uses where the arguments are equally placed in animacy, are topicalized in a number of sentences in the sample.

In the case of the Topicalization/Deviant asymmetry scenario, where we have an inanimate first element and a human or animate second element, we either have a case of a topicalized object construction or a reverse-animacy construction. Here we will have to look closer at the verb. If the verb is a reverse-animacy verb, we may conclude a SVO word order. However, if the verb is a regular transitive verb, we have in fact two conflicting situations - either a topicalized object as in (36) below or a personifying metaphoric use as in (37) below:

- (36) Seks kroner pr. kilo betaler Fjell
 six crowns per kilo pays Fjell
 ‘Six crowns per kilo, Fjell pays’
- (37) Flaks reddet Ine (8) fra døden
 luck saved Ine (8) from death
 ‘Luck saved Ine (8) from dying’

Both of the verbs in the above examples are trans1-verbs, i.e. prototypical agent-theme transitive verbs. Being able to separate these two, then, is an extremely difficult task and would require a lexicon with a much richer semantic specification for verbs in particular, as well as specification for animacy etc. for nouns. This is well beyond the scope of this thesis. However, this type of metaphoric use (i.e. with a deliberate reversal of argument properties) is not common in the data set at all. Only 2 sentences, i.e. 0,2% in the sample have this property. The conclusion for the cases which are not reverse-animacy then, would be to tag these according to an OVS word order.

This leaves us with the possibility for equally placed arguments, i.e. arguments that are equally placed on one or both of the scales. We have seen above that the cases where the two arguments are equally placed only in definiteness, animacy, once again disambiguates almost the whole group. These cases thus become cases of disambiguating asymmetry as defined above.

Our real concern in this third group of possible constellations, becomes sentences where the arguments are equally placed in animacy. These cases constitute 36.1% of the data sample or 361 sentences, which is a fairly high number. Of these 74 have arguments that are both human, and 287 have arguments that are both inanimate. The rate of topicalization is very low for these types of sentences 2.7% for the human group and 2% for the inanimate. Again, this is a result that must be viewed with caution, as it is precisely in this group of sentences where the context would be a decisive factor for disambiguation. However, as we choose to overlook this out of necessity, it tells us that in the whole of the group of sentences with equally placed arguments in animacy, only 2.35% of these have a topicalized object, a much lower number than for the rest of the sample. This might be taken as an indication that a constraint blocking marked word order is more operative in these sentences which certainly are prone to ambiguity. We have also seen that in the few sentences with arguments of equal animacy, where the object has been topicalized, these are often disambiguated by the relative definiteness of their arguments.

5.3 New challenges and their theoretical implications

Norwegian is in general a language which marks syntactic functions through structural positioning, a fact which relates to its lack of morphological case. Another often noted fact regarding Norwegian is that it may in principle topicalize any nominal constituent. These generalizations are unchallenged at a level of complete grammaticality, or *competence*. The b) sentences in (32) - (34) above, where we find topicalized versions of reverse-animacy sentences from the data material, are all grammatical at a competence level.

So are sentences with equally animate arguments and a topicalized object which are not disambiguated by definiteness, as in (30) above.

In chapter 3 describing our theoretical framework, examples are drawn from distant languages like Navajo, Hindi and Haida, exemplifying languages where the relationship between certain properties of the arguments in transitive sentences and their realization and positioning as syntactic functions, are categorical or grammaticalized at the competence level. Our main question then, has been whether some of the same conditioning factors, in this case, animacy and definiteness, are operative at a statistical level in Norwegian, i.e. as strong usage preferences at a level of *performance*. These are aspects of language that may be explained through so-called *soft* constraints, in contrast to the *hard* constraints of categorical grammars.

What we have seen above then, is that Norwegian, whilst in a majority of cases marking syntactic function through structural position, also relies heavily on animacy, in particular, for disambiguation of the syntactic functions of subject and object. In short, our findings that constitute new knowledge on Norwegian may be summarized within three main headings:

Properties of syntactic functions Subjects are almost exclusively higher or equally placed in animacy than their corresponding objects.

Word order freezing A freezing effect on word order in reverse-animacy constellations is present in Norwegian.

Ambiguity and word order Sentences with equally animate arguments resist object topicalization. If they do topicalize, however, they must be disambiguated in some other way - most often by definiteness.¹²

We have seen that, as common to many languages, animacy figures as an important conditioning factor for syntactic functions, their realization and positioning within the sentence. Whereas it certainly is true that most constituents may topicalize, we see that they are not associated with the same probability. Sentences with a reverse-animacy argument constellation, for instance, will topicalize at a vanishingly low rate. This, however, is not arbitrary or random, but rather, founded in principles common to a range of different languages. A strict level of competence will not capture this aspect of similarity, and thus explains too little and too much at the same time:

Categorical linguistic theories claim too much. They place a hard categorical boundary of grammaticality where really there is a fuzzy edge, determined by many conflicting constraints [...]

¹²Another way of disambiguating these types of sentences is obviously by lexical means. As Morimoto (2000) noted for Haida, transitive sentences freeze to canonical word order when arguments are of equal animacy, unless lexical properties of the verb makes it clear which arguments are linked to which syntactic functions (cf. chapter 3.7.3).

Categorical linguistic theories explain too little. They say nothing at all about the soft constraints which explain how people choose to say things. (Manning, 2003)

In the next chapter we will see how our findings regarding the influence of animacy and definiteness on the realization of syntactic functions and on word order in the transitive sentence may be modeled using stochastic Optimality Theory, a theory which allows for probabilistic and variable data, whilst at the same time founding this variation in Optimality Theoretic constraints assumed to be universal.

Chapter 6

Modeling variation: a stochastic OT analysis

Our point of departure for the work presented here, as mentioned several times, has been that of ascertaining whether properties along the dimensions of animacy and definiteness condition the realization of syntactic functions and their placement within the sentence in Norwegian. To be more specific, we want to investigate whether the variation between different syntactic interpretations of a transitive sentence might be governed by universal principles expressing markedness relations.

In the following, a stochastic OT (StOT, cf. chapter 3.5) analysis will be presented that reflects the theoretical framework of chapter 3, as well as accounting for the findings presented in chapter 5. A stochastic model has been chosen in order to reflect the variation observed in the data, representing the fact that we are not dealing strictly with a scale of grammaticality, but rather of probability. Firstly, we will examine our chosen input for OT evaluation, as well as the output candidates. Secondly, we will introduce the constraints central to our stochastic OT grammar and motivate further the choice of a stochastic, rather than an ordinal OT approach by way of examples. Following, we will take a closer look at the specific implementation of stochastic OT employed here, and present and discuss the results obtained from stochastic evaluation. Finally, we will focus on some topics for further research, and review in some detail a possible extension of our analysis through stochastic bidirectional OT.

6.1 Input and candidates

The input to GEN¹, then, will consist of a rough argument structure, specified for the arguments' value for animacy, definiteness and topicality. An example input is given below:

- (1) V(Anim/Pro/Top, Inan/Def)

The arguments are in random order, i.e. do not posit any restrictions on the output word order, which is to be generated by constraint interaction. We also limit ourselves to treatment of the types of sentences represented in the data material, i.e. declarative transitive sentences which are possibly ambiguous, where neither the case of their arguments nor the word order in the sentence determines syntactic function.

As OT is not a representational theory as such, input might just as well have been represented as an underspecified f-structure, as typical of the OT-LFG direction represented here by Lee (2002b,a). The idea is that GEN generates more specific² f-structures with corresponding c-structures as candidates. An underspecified input f-structure representing identical information as the more informal input above, is represented in (2) below:³

$$(2) \left[\begin{array}{l} \text{PRED} \quad \text{“} \\ \text{GF}_1 \quad \left[\begin{array}{ll} \text{ANIMACY} & \text{anim} \\ \text{DEF} & \text{pro} \end{array} \right] \\ \text{GF}_2 \quad \left[\begin{array}{ll} \text{ANIMACY} & \text{inan} \\ \text{DEF} & \text{def} \end{array} \right] \\ \text{TOPIC} \end{array} \right]$$

Following a common assumption in the OT-literature, we limit our analysis to as little structure as possible. In contrary to the traditional analysis of V2-languages,⁴ we will not assume that all Norwegian declarative sentences are CPs, but rather that a CP is only needed when something other than the subject is topicalized, like in Sells (2001), for Swedish (cf. chapter 2.2). This relies on the common notion of the subject as default topic, and we thus adopt an IP-analysis of canonical SVO declaratives.

¹As we remember from chapter 3.3, possible output candidates are thought to be generated by GEN(erator), a function which generates all possible candidates from a given input.

²As we remember from Lee (2002b), the candidate f-structure is more specific, in the sense that the input f-structure subsumes the candidate f-structures.

³GF stands for Grammatical Function. The below f-structure is thus underspecified with regards to the grammatical (syntactic) function of its arguments.

⁴See for example Nordgård and Åfarli (1990) for an analysis of Norwegian.

From the input, then, GEN in turn generates four possible output candidates, representing possible surface structure:⁵

Candidate 1: $[IP \text{ Subject}_{Anim/Pro/Top} [I' \text{ V } [VP \text{ Object}_{Inan/Def}]]]$

Candidate 2: $[IP \text{ Subject}_{Inan/Def} [I' \text{ V } [VP \text{ Object}_{Anim/Pro/Top}]]]$

Candidate 3: $[CP \text{ Object}_{Inan/Def} [C' \text{ V } [IP \text{ Subject}_{Anim/Pro/Top}]]]$

Candidate 4: $[CP \text{ Object}_{Anim/Pro/Top} [C' \text{ V } [IP \text{ Subject}_{Inan/Def}]]]$

In the evaluation process from input to output, we want in essence to arrive at two different but inter-related properties of the output structure: 1) to ascertain from the properties of the arguments in the input the subject- or objecthood of these, and 2) to ascertain the resultant word order in the output structure, based on properties of the arguments and specific word order constraints.

6.2 Constraints

As we remember, OT constraints should be well-founded linguistically. The idea behind StOT is that constraints that are categorical in some languages are expressed as statistical tendencies in others. This means that constraints posited for one language should also be candidates for universal constraints.

6.2.1 Animacy constraints

In chapter 5, we saw that there is a strong tendency in Norwegian that the subject should not be lower in animacy than the corresponding object. This constraint was observed by 97.6% of the sentences in the sample. As we have seen earlier, this is not an isolated tendency, but one that is categorical in several languages, e.g. the inverse language of Navajo⁶. In the languages Tzotzil and Chamorro, transitive sentences with a 3rd person inanimate subject and a 3rd person animate object are excluded (Aissen, 2003, p. 443).⁷

⁵Here we abstract away from the possibility of object shift, where the object is in fact in a position forward to the left. This is a phenomenon where a weak pronominal object may precede negation and other adverbials, usually assumed to mark the left edge of VP:

- (1) Vi likte *han* ikke
 we-NOM liked he-NOM not
 ‘We did not like him’

⁶As we remember, in Navajo, an inverse version of a transitive sentence is required when the object is higher than the subject in animacy (cf. chapter 3.1.2).

⁷Here, the dimension of person is obviously also a conditioning factor. However, if assuming a person hierarchy, such as the one in Aissen (1999), Local > 3rd, it follows from harmonic alignment that a subject which is 3rd person is marked, and a subject that

There are also languages, such as Jakaltek and Halkomelem, where inanimate subjects in transitive sentences are banned altogether, irrespective of the objects rank in animacy (Aissen, 2003, p. 443).

The tendency described above is a generalization in line with the properties expressed by harmonic alignment of the scales of syntactic function and animacy (cf. chapter 3.4), which gives us the following subhierarchies:⁸

- (3) a. *SU/INAN \gg *SU/ANIM
An inanimate subject is more marked (and should to a larger extent be avoided) than a human/animate subject.
- b. *OBJ/ANIM \gg *OBJ/INAN
A human/animate object is more marked (and should to a larger extent be avoided) than an inanimate object.

The most marked construction according to the above constraints, then, is a sentence with an inanimate subject and an animate object. This is in line with our data analysis. These maximally marked constructions are, however, not ungrammatical in Norwegian, they are simply quite rare. Another property of these (as we saw in chapter 5) is that they topicalize at a vanishingly low rate.⁹

As we remember from chapter 5, three subject-object configurations may be envisioned with regards to animacy:

1. The subject is higher in animacy than the object (i.e. animate subject and inanimate object)
2. The arguments are of equal animacy
3. The subject is lower in animacy than the object (i.e. inanimate subject and animate object)

In order to express these three scenarios and their relative markedness by constraints we introduce the technique of local conjunction (cf. chapter 3.4). This helps us express the idea that a construction marked in two or more aspects is more marked than a construction marked in only one of these. As the local conjunction is always ranked above its conjuncts, this property is preserved in a constraint ranking. We thus conjoin the two subhierarchies in

is inanimate is also marked. A subject which is thus marked in both these respects, i.e. 3rd person and inanimate, must be said to be cumulatively more marked than one marked only along one of these dimensions.

⁸As no significant differences between human and non-human animates have been observed in the data, the categories Human and Animate are collapsed into one category: Anim.

⁹Our data material does not include a single topicalized sentence with an inanimate subject and an animate object. This does not however, mean that these are ungrammatical. Even though they sound odd to say the least, in a proper context, with appropriate stress they could probably pass as grammatical.

(3) above (Donohue, 1999; Morimoto, 2000) and obtain the product partial order below:

1. *SU/INAN & *OBJ/ANIM
- 2a. *SU/INAN & *OBJ/INAN 2b. *SU/ANIM & *OBJ/ANIM
3. *SU/ANIM & *OBJ/INAN

The above figure thereby also indicates the relative markedness of the resulting conjunctions. The top constraint, 1., is the most marked, as it results from conjunction of the two highest ranked constraints of the subhierarchies, i.e. the two configurations most to be avoided. The two conjunctions in 2. are unranked with respect to each other, resulting from conjunction between one highest ranked constraint and a lowest ranked constraint. Finally, the least marked configuration is the one found in 3., where we have the configuration resulting from conjunction of the two lowest ranked constraints of the subhierarchies. We generalize the above picture into three constraints (Donohue, 1999, p. 20):

Result from local conjunction	Constraint
*SU/INAN & *OBJ/ANIM	*S<O(ANIM)
*SU/INAN & *OBJ/INAN, *SU/ANIM & *OBJ/ANIM	*S=O(ANIM)
*SU/ANIM & *OBJ/INAN	*S>O(ANIM)

Following from the local conjunction and the markedness relations obtained between the resulting constraints, the above constraints are ranked internally in a subhierarchy:¹⁰

$$(4) \quad *S<O(ANIM) \gg *S=O(ANIM) \gg *S>O(ANIM)$$

The subhierarchy above may be paraphrased as follows: “A construction with a subject that is lower than its object in animacy is more marked (thus more to be avoided) than a construction where the arguments are of equal animacy etc.”

6.2.2 Word order constraints

We have looked at the interplay between properties of the arguments in a transitive sentence and the preferred word order in the sentence. As we have seen, Norwegian allows for both the canonical SVO word order, and the more marked OVS word order, where the object is topicalized. Word order is often expressed in OT through so-called alignment constraints (Lee,

¹⁰This ranking is also supported by our findings in the data. An exact numerical ranking of these constraints will be presented in the section on the stochastic evaluation.

2002b), which align the element in question with a position in the clause (left- or rightmost). These constraints express a generalization akin to harmonic alignment, namely that the structural position of constituents mirror their rank on the hierarchy of grammatical functions. The subject, which is highly ranked on the hierarchy of grammatical functions, is also placed structurally high, or far to the left, thus representing a type of iconicity with regards to word order. The following constraint, based on a similar constraint in Choi (1999)¹¹, expresses these generalizations:

- (5) $CANON_{GF}$ - Grammatical functions remain in their canonical positions

An important point here, is that the above constraint on word order will not penalize a sentence where something other than the grammatical core functions is topicalized. This is in line with what we want, as we do not wish a sentence like (6) below, with a topicalized adverbial, to violate $CANON_{GF}$:

- (6) I dag skrev jenta brevet
 To day wrote girl-DEF letter-DEF
 ‘Today, the girl wrote the letter’

Even though such sentences are not examined here, we want our constraints to permit an extension of coverage. The constraint $CANON_{GF}$ will, however, penalize a sentences with a topicalized object.

We saw that certain freezing effects seem to be present in the data, at least to a statistical measure. As we remember (cf. chapter 3.7), this is a tendency that is not unique for Norwegian. Haida, the indigenous language of British Columbia, reviewed in Morimoto (2000), for example, exhibits a categorical word order freezing in sentences where the subject is lower in animacy than the object, and also when they are of equal animacy.

In our data analysis for Norwegian, we saw that sentences where the animacy of the arguments is reversed in comparison to the prototypical transitive sentence, i.e. where the subject is inanimate and the object is animate, there seems to be a strong tendency towards an unmarked word order, expressing a markedness reduction or avoidance of *the worst of the worst* (marked subject and object and marked word order) as in Lee (2002b). In order to express this generalization, we once again make use of the technique of local conjunction. We wish to express the freezing effect observed when the

¹¹Choi gives the following definition of the constraint (Choi, 1999, p. 40):

- (1) $CANON_{gf}$
- a. SUBJ should be structurally more prominent than (e.g. c-command) non-SUBJ functions.
 - b. Non-SUBJ functions align reversely with the c-structure according to the functional hierarchy.

subject is inanimate and the object is animate, thus expressing that a sentence marked in the above respect is cumulatively more marked if it also has a marked word order. We thus locally conjoin the constraint $*S<O(\text{ANIM})$ with the constraint enforcing canonical word order, CANON_{GF} and obtain the following constraint:

$$(7) \quad *S<O(\text{ANIM}) \ \& \ \text{CANON}_{GF}$$

A local conjunction, then, is violated only when both the local conjuncts are violated. In this case, a sentence where the subject is inanimate, the object is animate and the word order is marked, i.e. OVS, will violate the constraint as a whole.

We also wish to allow for topicalization, thus needing a constraint that will align the topic in the sentence appropriately. As the data we are dealing with is not marked for discourse topicality, we will adopt a structural definition of a topic. This means that whatever is situated in the first position of the sentence, be it the subject or the object, will be regarded as marked for topic. The fact that we are not dealing with interrogative sentences, which topicalize a focused object, makes this a plausible approximation also with regards to discourse status.¹² We make use of the following alignment constraint (Lee, 2002b, p. 23):

$$(8) \quad \text{TOP-L} - \text{topic aligns left in the sentence, i.e. IP for subject topics and CP for object topics}$$

A consequence of this notion of topic is that the above constraint becomes categorical, i.e. no candidate violating this constraint will ever be a winning candidate. This commits to only allowing for one topic in a sentence, and demanding that whatever is marked for topic is either in subject position, SpecIP, (as subject is default topic) or is topicalized to SpecCP in the case of object topics.

At this point we may hypothesize a ranking¹³ of the above constraints:

$$(9) \quad \text{TOP-L}, \ *S<O(\text{ANIM}) \ \& \ \text{CANON}_{GF} \ \gg \ *S<O(\text{ANIM}) \ \gg \ \text{CANON}_{GF} \ \gg \ *S=O(\text{ANIM})$$

¹²As we remember, however, also foci other than interrogative pronouns may topicalize. These are the so-called *contrastive foci* (Engdahl, 1997). However, as they are quite indistinguishable from contrastive topics without the assistance of a larger context, we will abstract away from this distinction in the following.

¹³The individual constraints from harmonic alignment of the hierarchies of animacy and syntactic function will be excluded from the rankings here and in the following. The reason for this is that they will necessarily be ranked lower than the constraints resulting from their local conjunction. The two high-ranked constraints in these subhierarchies will not play a decisive role, as the conjoined constraints will deal with the same phenomena. The lowest ranked constraints will in reality be ranked very low, and do not play a role in any of the evaluations we are to discuss. This is a consequence of the way harmonic alignment is done formally, that one will have low-ranking constraints that are ranked so low as to be virtually non-effective. For the same reason, $*S>O(\text{ANIM})$ will also be left out in the following analysis.

The above ranking is based on frequencies in the data material, as well as principles governing local conjunction and subhierarchies. TOP-L and *S<O(Anim) & CANON_{GF} are never violated in the data material, and are thus ranked highest of all the constraints. This ensures that a candidate violating one of these two constraints will never be an optimal candidate, i.e. these constraints are categorical. Following the two categorical constraints is the constraint *S<O(ANIM), penalizing constructions with inanimate subjects and animate objects. This is ranked above *S=O(ANIM), because constructions with inanimate subjects and animate objects are less frequent than constructions with arguments of equal animacy.

As mentioned above, the hypothesized ranking in (9) above, is based on the relative frequencies in the data material, and is thus *data-driven*. However, it is also a categorical ranking, rather than a stochastic one. This means that from one input there will always be one and the same optimal output, i.e. there is no room for variation. An example tableau may help to clarify this point:¹⁴

(10)

	V(Anim, Inan/Top)	TOP-L, *S<O(ANIM) & CANON _{GF}	*S<O(ANIM)	CANON _{GF}	*S=O(ANIM)
a.	[<i>IP</i> <i>S_{Anim}</i> [<i>I'</i> <i>V</i> [<i>VP</i> <i>O_{Inan/Top}</i>]]]	*!			
b.	[<i>IP</i> <i>S_{Inan/Top}</i> [<i>I'</i> <i>V</i> [<i>VP</i> <i>O_{Anim}</i>]]]		*!		
☞ c.	[<i>CP</i> <i>O_{Inan/Top}</i> [<i>C'</i> <i>V</i> [<i>IP</i> <i>S_{Anim}</i>]]]			*	
d.	[<i>CP</i> <i>O_{Anim}</i> [<i>C'</i> <i>V</i> [<i>IP</i> <i>S_{Inan/Top}</i>]]]	*!*	*	*	

For the input of an animate, non-topic argument and a topic inanimate argument, we see from the above tableau in (10) that the optimal candidate is candidate c., i.e. one where the topic, inanimate argument is realized as a topicalized object and the animate, non-topic as the subject. However, as we have seen in the data analysis, candidate b., with an inanimate subject and an animate object, should certainly also constitute a possibility, albeit

¹⁴At this point, definiteness is not represented in the input, as no constraints targeting definiteness have been introduced yet.

a less frequent one. This is where stochastic OT comes in, as we shall see shortly.

6.2.3 Definiteness

It seems fair to say that the dimension of word order is closely tied to a certain asymmetry in animacy. With regards to arguments that are equal in animacy, this asymmetry is not present.

We concluded the last chapter with some interesting points regarding constructions with equally animate arguments in Norwegian:

- They have a much lower rate of topicalized objects (2,35%) than the data material as a whole.
- The sentences that do contain a topicalized object are, in a majority of cases, disambiguated by definiteness, i.e. the subject is higher than the object in definiteness.

This gives us an indication that definiteness might be a decisive factor when it comes to the disambiguation of this type of sentences. We may attempt something like the following generalization: sentences with arguments that are equally placed in animacy tend to resist word order variation, but if they do vary, the arguments differ in definiteness to the extent that they help disambiguate the sentence.¹⁵ This is, as we remember, not the norm for the data as a whole. In the whole group of sentences with topicalized objects, we found that as many as 50% of these had a subject that was lower than the object in definiteness. As we have seen above, this is not that surprising, as definiteness is often linked to topicality and givenness. Precisely because of this, it is striking that this does not seem to be the case to such a large extent for the group of object-topicalized sentences with equally animate arguments.

The first point above states that the sentences with equally placed arguments tend to resist topicalization to a significantly larger extent than the rest of the sentences in the data material. We have seen that a possible ambiguity may cause word order freezing in other languages (Lee, 2002a; Bloom, 1999; Morimoto, 2000) (cf. chapter 3.7). In particular, we can mention the language Haida again (Morimoto, 2000), where word order categorically freezes in sentences with arguments of equal animacy, unless lexical properties of the verb clearly indicate the function of each argument.

¹⁵This is a truth with modifications. The majority of sentences where the object topicalizes and the arguments are equal in animacy are disambiguated by definiteness. There is however, a problem with scarce data, as our whole data material only includes 8 of these types of topicalized sentences. As we remember, two of these were not disambiguated by definiteness, but fronted a topicalized inanimate pronoun, *det* 'it', a common object topic.

Following from the tentative ranking of the above constraints in (9) above, topicalized objects will be excluded completely as outputs for candidates where the arguments are of equal animacy:

(11)

V(Anim/Top, Anim)		TOP-L, *S<O(ANIM) & CANON _{GF}	*S<O(ANIM)	CANON _{GF}	*S=O(ANIM)
☞	a. [IP S _{Anim/Top} [I' V [VP O _{Anim}]]]				*
	b. [IP S _{Anim} [I' V [VP O _{Anim/Top}]]]	*!			*
	c. [CP O _{Anim} [C' V [IP S _{Anim/Top}]]]	*!		*	*
	d. [CP O _{Anim/Top} [C' V [IP S _{Anim}]]]			*!	*

We see from the above tableau in (11) that the topicalized version, candidate d., will never be a winning candidate, because it violates the word order constraint CANON_{GF}, and the actual winning candidate does not.

How then, are we to explain the fact that in sentences with equally placed arguments the objects do in fact topicalize, albeit at a much lower frequency? We have seen that there seems to be a connection to the relative definiteness of the arguments in these topicalized cases. The tendency is that the subject is high in definiteness and the object is lower. This is reminiscent of the generalization resulting from harmonic alignment of the hierarchies of definiteness and syntactic function (cf. chapter 3.4), as repeated below:

- (12) a. *SU/INDEF ≫ *SU/DEF ≫ *SU/PN ≫ *SU/PRO
 b. *OBJ/PRO ≫ *OBJ/PN ≫ *OBJ/DEF ≫ *OBJ/INDEF

Instead of locally conjoining all relevant permutations of the above hierarchies we introduce a single constraint that expresses the relevant generalization. This constraint is modeled around a constraint introduced in Bresnan and Nikitina (2003), DOUBLE-OBJECT PRIMACY, which expresses the fact that in a dative double-object construction “the receiver (strictly) dominates the entity on hierarchies of informational prominence, and the entity (strictly) dominates the receiver/possessor on the reversed hierarchies” (Bresnan and Nikitina, 2003, p. 23):

- (13) S(SUBJECT)O(BJECT)-PRIMACY(DEF) - the subject dominates the object on the hierarchy of definiteness, and the object dominates the subject on the reversed hierarchy.

What this means, then, is that the subject is higher than the object in definiteness, and, symmetrically, that the object is lower than the subject in definiteness.

Definiteness becomes an important input factor for the candidates that are ambiguous with regards to animacy, and will give us a possibility for variation in word order also in these sentences. However, the constraint SO-PRIMACY(DEF) will not suffice when modeling the variation in word order, conditioned by definiteness. As we have seen, the generalization embodied in the constraint SO-PRIMACY(DEF) does not hold for the data material as a whole. Only about 50% of the sentences in the data sample adhered to this principle.¹⁶ What we wish to express is that in sentences where the arguments are of equal animacy, definiteness plays a disambiguating role in cases where the word order is non-canonical. This property of our grammar is obtained by introducing a new constraint through local conjunction of two already introduced constraints, namely *S=O(ANIM) and SO-PRIMACY(DEF):

- (14) *S=O(ANIM) & SO-PRIMACY(DEF) - a sentence where the subject and object are of equal animacy and the subject is not higher than the object in definiteness is more marked than a sentence with only one of these properties.

A variation in ranking between the above constraint and $CANON_{GF}$ will give us a topicalized version only for outputs where SO-PRIMACY(DEF) is not violated. In order to illustrate this we will simulate a possible ranking in the data, where our new constraint, *S=O(ANIM) & SO-PRIMACY(DEF), is ranked above the word order constraint, $CANON_{GF}$.¹⁷

- (15) TOP-L, *S<O(ANIM) & $CANON_{GF}$ \gg *S<O(ANIM) \gg *S=O(ANIM) & SO-PRIMACY(DEF) \gg $CANON_{GF}$ \gg *S=O(ANIM)

An example evaluation, where the input is ambiguous with regards to animacy, but specified additionally for definiteness, will give us the topicalized version as optimal, in contrast to the result in the tableau in (11) above:

¹⁶To be more precise, 529 of the sentences in the data sample, i.e. 52.9%, are sentences where the subject is higher in definiteness than the object.

¹⁷Once again, a low ranking constraint, namely *SO-PRIMACY(DEF) will be excluded from the below treatment.

(16)

V(Anim/Pro, Anim/Def/Top)	TOP-L, *S<O(ANIM) & CANONGF	*S<O(ANIM)	*S=O(ANIM) & SO-PRIMACY(DEF)	CANONGF	*S=O(ANIM)
a. $[IP\ S_{Anim/Pro}\ [I'\ V\ [VP\ O_{Anim/Def/Top}]]]$	*!				*
b. $[IP\ S_{Anim/Def/Top}\ [I'\ V\ [VP\ O_{Anim/Pro}]]]$			*!		*
\Rightarrow c. $[CP\ O_{Anim/Def/Top}\ [C'\ V\ [IP\ S_{Anim/Pro}]]]$				*	*
d. $[CP\ O_{Anim/Pro}\ [C'\ V\ [IP\ S_{Anim/Def/Top}]]]$	*!		*	*	*

What we see from the above tableau in (16), then, is a situation where the topicalized version in c. wins out because the canonical word order in b. violates the higher ranked *S=O(ANIM) & SO-PRIMACY(DEF). Obviously we do not want this to be the output in all cases, as sentences with SVO order that violate this constraint are fairly frequent. What we want is to allow for variation between, in this case, candidates b. and c., at a rate that mirrors the actual frequencies in the data material. The introduction of the above locally conjoined constraint will allow us to do exactly that.

6.3 Stochastic OT analysis

StOT (c.f chapter 3.5) provides us with the opportunity of modeling variation within an Optimality Theoretic grammar. As we have seen, categorical constraints do not suffice in modeling the interplay between variation in word order and properties of the arguments along the scales of animacy and definiteness. We will start out with a short survey of the particular implementation of StOT employed here, followed by a presentation and analysis of the results obtained.

6.3.1 Training towards convergence

Computing the rankings of different constraints is complicated when dealing with variability in the data. A key tenet of StOT is that the ranking distance between the constraints also mirrors the probability of these varying amongst

themselves. In order to obtain rankings that reflect the variation in our data material I have employed the Praat Software (Boersma, 1999),¹⁸ which is an implementation of the Gradual Learning Algorithm (GLA, cf. chapter 3.5) and thus stochastic Optimality Theory. In essence, the Praat program computes the rankings of constraints in a grammar supplied by the user, based on data also supplied by the user. This is where the GLA is put to work, and the algorithm learns the grammar based on successive training on the training data.

We started out by supplying the Praat program with a grammar (cf. Appendix A), i.e. constraints, all ranked at 100 (random number) and tableaux specifying all relevant inputs with all possible candidates and their violations of the specified constraints. We had in total twelve different inputs, six for cases where the arguments differed in animacy and six for ones where they did not. With regards to definiteness the values that we were interested in was whether or not one argument was higher or lower in definiteness than the other argument.¹⁹ This gave us the following inputs:

- Differing in animacy:

- V(Anim/EqualDef/Top, Inan/EqualDef)
- V(Anim/HigherDef/Top, Inan/LowerDef)
- V(Anim/LowerDef/Top, Inan/HigherDef)
- V(Anim/EqualDef, Inan/EqualDef/Top)
- V(Anim/HigherDef, Inan/LowerDef/Top)
- V(Anim/LowerDef, Inan/HigherDef/Top)

- Equal in animacy:

- V(Anim/EqualDef/Top, Anim/EqualDef)
- V(Inan/EqualDef/Top, Inan/EqualDef)
- V(Anim/HigherDef/Top, Anim/LowerDef)
- V(Inan/HigherDef/Top, Inan/LowerDef)

¹⁸The Praat Software may be downloaded from <http://www.praat.org>.

¹⁹An input like V(Anim/HigherDef/Top, Inan/LowerDef), generalize over all of the following instantiations with regards to definiteness:

- V(Anim/Pro/Top, Inan/PN)
- V(Anim/Pro/Top, Inan/Def)
- V(Anim/Pro/Top, Inan/Indef)
- V(Anim/PN/Top, Inan/Def)
- V(Anim/PN/Top, Inan/Indef)
- V(Anim/Def/Top, Inan/Indef)

- V(Anim/LowerDef/Top, Anim/HigherDef)
- V(Inan/LowerDef/Top, Inan/HigherDef)

We also supplied the program with our pair distributions (cf. Appendix A) from our data material. The data file specifies the frequency in the data of the different candidates in the grammar.

The program starts the learning process by sampling from the data file and training the grammar on these sampled input-output pairs.²⁰ This process is then repeated a number of times. In the end, the learning has been successful if the grammar reaches *convergence*, which is a stage where the constraint rankings have reached a certain stability, and the addition of further data does not affect the rankings seriously. There are two important variables in the learning process:

Plasticity - a numerical value representing how greatly each learning datum is to affect the rankings of the different constraints, i.e. by how much constraints should be promoted/demoted during learning.

Evaluation Noise - the numerical value that is added to the ranking value of each constraint at evaluation time. This value represents a standard deviation in a probability distribution and simulates a constraint covering a *range* of values with the ranking value as the mean, i.e. the most probable selection point.

Boersma and Hayes (2001) suggest a few different *training schedules* for learning stochastic grammars. These different schedules may lead to somewhat different results and should be chosen according to the goal of the training and the time available. Regarding plasticity, they state that

A small plasticity value does a better job of matching learning data frequencies in the end, but a large plasticity value nears its goal faster. The virtues of the two approaches can be combined by adopting a learning schedule that decreases the plasticity as learning proceeds. (Boersma and Hayes, 2001, p. 79)

They also conclude that decreasing evaluation noise during training also leads to accurate results. In this case, we adopted a training schedule from Boersma and Hayes (2001, p. 80), and adjusted it to our data and grammar.²¹ The training schedule is provided in table 6.1 below. The algorithm

²⁰The sampling process actually involves the algorithm generating a set number of input-output pairs and evaluating these. The user supplies the program with a number, say 1000, and it samples 1000 inputs from the data distribution file and produces an output for each of these inputs. Evaluation then proceeds by constraint promotion or demotion based on the learning data (cf. chapter 3.5). Given enough data and training, the algorithm should arrive at empirically correct rankings. More on the particular training schedule below.

²¹Boersma and Hayes (2001) suggest that the number of data forms at each stage should usually amount to approximately 1000 times the number of inputs in the grammar. In our case, we had 12 different inputs.

was also set to respect the internal rankings that resulted from harmonic alignment and local conjunction. In the end then, we trained 208,000 input-

Data	Plasticity	Noise
First 12,000	2.0	10.0
Second 12,000	2.0	2.0
Third 12,000	0.2	2.0
Fourth 12,000	0.02	2.0
Last 160,000	0.002	2.0

Table 6.1: Training schedule adapted from Boersma and Hayes (2001)

output pairs on the grammar. This, however, is not to be confused with the frequencies supplied in the data file. As we remember, our data material of transitive constructions only includes 1,000 sentences. The numbers in the training schedules represent the samples taken from this data material by the program, which means that it trains on the same data many times, in order to converge on a grammar which correctly reflects the frequencies in the original data material. This indicates that it is important to provide the algorithm with a data set that is representative and large enough, so that arbitrary or insignificant aspects are not reproduced to any large extent.

6.3.2 Results

After training according to the above training schedule and reaching a state of convergence, the constraints achieved the rankings depicted in table 6.2 below. As we predicted earlier the two top-ranked constraints are TOP-L and

constraint	ranking value
TOP-L	131.776
*S<O(ANIM) & CANON _{GF}	108.465
*S<O(ANIM)	87.048
*CANON _{GF}	84.809
SO-PRIMACY(DEF) & *S=O(ANIM)	81.237
*S=O(ANIM)	61.594
*S>O(ANIM)	57.258
SO-PRIMACY(DEF)	-83.447

Table 6.2: Ranking values after training according to the training schedule in table 6.1

*S<O(ANIM) & CANON_{GF}. Their rankings must be said to be categorical as they are both more than ten units from any of the other constraints and “if two distributions are 5 standard deviations apart, the odds that a “reversed” ranking could emerge are about 1 in 5,000. This frequency would be hard to distinguish empirically, we think, from the background noise of speech

errors.”(Boersma and Hayes, 2001, p. 50) Even more dramatically, if the constraints are 9 standard deviations apart, the chances that the rankings will reverse are at 1 in 10 billion, meaning that such a form would probably not emerge during the course of a whole lifetime. Boersma and Hayes (2001) adopt a standard deviation of 2.0, an arbitrary value. The main point is that this value is identical for all constraints at each evaluation. The chosen standard deviation is the same as the evaluation noise added during training. It is unclear however, how a training schedule like the one employed here affects the overall standard deviation. During training, noise was set at ten units at the first stage and 2 in all the rest. Even so, it is fair to conclude that the constraints TOP-L and *S<O(ANIM) & CANON_{GF} are ranked at a safe, categorical distance from constraints like CANON_{GF} and *S>O(ANIM) with which they could vary with, respectively, and thereby produce variable input.

Another point worth mentioning regarding the two top-ranked constraints, is that they are not ranked at the same numerical value. In fact, they are ranked at more than 20 units apart. What might be the reason for this, then? The Gradual Learning Algorithm is affected by the frequencies in the data it is supplied with, and seeks to mirror these frequencies in the resulting grammar. Obviously, a constraint that is frequently involved in evaluations will influence the grammar more than a constraint that is less frequent. TOP-L is a constraint which is involved in all the evaluations, whereas this is not the case for *S<O(ANIM) & CANON_{GF}. Even so, they are both categorical, as the chance of them varying with any other constraints is vanishingly low.

Otherwise, the ordinal ranking resulting from training is pretty much as predicted. What we are interested in, however, is the numerical ranking of the constraints, as this opens for variation.

As we saw from the tableau in (10) above, the OVS version always became the optimal candidate for an input of a non-topic animate and a topic inanimate argument, due to the fact that *S<O(ANIM) was ranked above CANON_{GF}, as it still is. However, as we are now dealing with stochastic evaluation, the numerical distance between the constraints has important consequences. It turns out that the top-ranked of the animacy constraints, *S<O(ANIM), is ranked at approximately 2.2 units above CANON_{GF}, thus giving them a fair chance of variation. A reversed ranking of the word order constraint and the animacy constraint *S<O(ANIM), then, would give us the opposite result, an inanimate, topicalized subject and an animate object, as in (17) below.²²

²²Once again, the arguments’ value for definiteness is not represented in the tableau in (17), as it is not relevant for the evaluation.

(17)

V(Anim, Inan/Top)		TOP-L, *S<O(ANIM) & CANON _{GF}	CANON _{GF}	*S<O(ANIM)	*S=O(ANIM) & SO-PRIMACY(DEF)	*S=O(ANIM)
	a. $[IP\ S_{Anim}\ [I'\ V\ [VP\ O_{Inan/Top}]]]$	*!				
☞	b. $[IP\ S_{Inan/Top}\ [I'\ V\ [VP\ O_{Anim}]]]$			*		
	c. $[CP\ O_{Inan/Top}\ [C'\ V\ [IP\ S_{Anim}]]]$		*!			
	d. $[CP\ O_{Anim}\ [C'\ V\ [IP\ S_{Inan/Top}]]]$	*!*	*	*		

We also see that the interplay between ambiguity in animacy, definiteness and word order is covered by the ranking resulting from stochastic evaluation. A reversed ranking of *S=O(ANIM) & SO-PRIMACY(DEF) and CANON_{GF}, will give us the sentences with ambiguous arguments with regards to animacy wherein the object has topicalized and the arguments are disambiguated by definiteness. This is the situation depicted in the tableau in (16), in opposition to tableau in (11), where the canonical word order was optimal across the board.

When training a grammar on a set of data or frequency distributions, it is also important that the grammar is able to reproduce the distributions in the data to a reasonable degree. Apart from reaching convergence, this is the most significant indicator that the constraints do in fact model the specific phenomenon correctly. The distributions then, refer to the frequency at which a specific input yields each of its candidates. The Praat program produces outputs for the trained grammar, and a comparison between these and the frequencies supplied by the user should match fairly well, as they did in our case (cf. table 6.3 below).²³ As definiteness is not a decisive factor for the arguments that are asymmetric in animacy, the algorithm is not capable of reproducing differences between these inputs that are reliant only on definiteness. In the following, then, the inputs that are asymmetric in animacy will be collapsed into two categories, thus excluding definiteness. A comparison of the output distributions provided by the program and the

²³Phrase structure is excluded in the candidates in this table due to reasons of space.

actual data distributions is provided in table 6.3 below.²⁴ An ordinal ranking

Input	Candidate	Trained Distr.	Data Distr.
1. V(Anim/Top, Inan)	$[S_{Anim/Top} \vee O_{Inan}]$ $[O_{Anim/Top} \vee S_{Inan}]$	100% 0%	100% 0%
2. V(Anim, Inan/Top)	$[S_{Inan/Top} \vee O_{Anim}]$ $[O_{Inan/Top} \vee S_{Anim}]$	21.4% 78.6%	21.2% 78.8%
3. V(Anim/EqDef/Top, Anim/EqDef)	$[S_{Anim/EqDef/Top} \vee O_{Anim/EqDef}]$ $[O_{Anim/EqDef/Top} \vee S_{Anim/EqDef}]$	100% 0%	100% 0%
4. V(Inan/EqDef/Top, Inan/EqDef)	$[S_{Inan/EqDef/Top} \vee O_{Inan/EqDef}]$ $[O_{Inan/EqDef/Top} \vee S_{Inan/EqDef}]$	100% 0%	100% 0%
5. V(Anim/HDef/Top, Anim/LDef)	$[S_{Anim/HDef/Top} \vee O_{Anim/LDef}]$ $[O_{Anim/HDef/Top} \vee S_{Anim/LDef}]$	100% 0%	100% 0%
6. V(Inan/HDef/Top, Inan/LDef)	$[S_{Inan/HDef/Top} \vee O_{Inan/LDef}]$ $[O_{Inan/HDef/Top} \vee S_{Inan/LDef}]$	100% 0%	100% 0%
7. V(Anim/LDef/Top, Anim/HDef)	$[S_{Anim/LDef/Top} \vee O_{Anim/HDef}]$ $[O_{Anim/LDef/Top} \vee S_{Anim/HDef}]$	89.8% 10.2%	87.5% 12.5%
8. V(Inan/LDef/Top, Inan/HDef)	$[S_{Inan/LDef/Top} \vee O_{Inan/HDef}]$ $[O_{Inan/LDef/Top} \vee S_{Inan/HDef}]$	90.3% 9.7%	90.3% 9.7%

Table 6.3: Output distributions and data distributions compared - stochastic evaluation

of the above constraints would not have given us the variation observed for inputs 2., 7. and 8. Rather, we would instead have arrived at the frequencies provided in table 6.4 below.

Input	Candidate	Trained Distr.	Data Distr.
2. V(Anim, Inan/Top)	$[S_{Inan/Top} \vee O_{Anim}]$ $[O_{Inan/Top} \vee S_{Anim}]$	0% 100%	21.2% 78.8%
7. V(Anim/LDef/Top, Anim/HDef)	$[S_{Anim/LDef/Top} \vee O_{Anim/HDef}]$ $[O_{Anim/LDef/Top} \vee S_{Anim/HDef}]$	100% 0%	87.5% 12.5%
8. V(Inan/LDef/Top, Inan/HDef)	$[S_{Inan/LDef/Top} \vee O_{Inan/HDef}]$ $[O_{Inan/LDef/Top} \vee S_{Inan/HDef}]$	100% 0%	90.3% 9.7%

Table 6.4: Output distributions and data distributions compared - ordinal evaluation

6.4 Topics for further research

In the following we will take a look at some topics that might be interesting to pursue further in the future, in light of the work presented above. First of all, we will take a closer look at some insights provided by recent work on an extension of traditional OT, namely bidirectional OT (cf. chapter 3.8) and review our analysis in the light of this approach to optimization. We will also outline how our present analysis could quite easily be incorporated into the framework of bidirectional (stochastic) OT. Following, we will examine the degree of coverage in our analysis which also provides room for improvement.

²⁴The candidates that do not adhere to TOP-L, and thus may never be optimal candidates under the present grammar, are excluded from the candidate set represented in table 6.3.

6.4.1 An alternative approach through bidirectional OT

Our underlying focus in the above analysis, as well as throughout the thesis, has been that of disambiguation. This thus involves both a disambiguation of syntactic functions, and thereby also their positioning within the sentence. We have therefore been interested in properties that might contribute towards such a disambiguation, and have chosen to focus on animacy and definiteness. As we have seen, these findings also provide us with generalizations regarding the general realization and expression of syntactic functions in Norwegian. In particular, we have seen that these properties condition the placement of syntactic functions within the sentence.

In the above analysis we approximated a notion of topic, in order to motivate word order variation and thus model our data. In actuality, then, we have been looking at a competition between two schematic candidates:

- (18) a. $\alpha_{Subject}$ Verb β_{Object}
b. α_{Object} Verb $\beta_{Subject}$

The above candidates represent the same string, but with different linking of syntactic functions. We wanted to see how far we could get, so to speak, with the dimensions of animacy and definiteness alone, and to examine in isolation the influence these dimensions have on the realization of syntactic functions, as well as their positioning within the sentence. As we have seen above, this perspective has allowed us to model the findings from our data material in a structured and well-founded manner. What we are not claiming however, is that this constitutes the full linguistic picture. Central to an account of syntactic functions and their realization are obviously other properties of the arguments, such as thematic (proto-) roles (cf. chapter 3.6.4). We have focused on modeling the findings of our data material, a set of transitive sentences. We have not assumed a linguistically sufficient representation of the production of these sentences. This is an important point, and one that makes an extension into bidirectionality viable.

Morimoto: ambiguity in animacy

As we have seen, canonical word order is prevalent in transitive sentences where the arguments are of equal animacy. In the above treatment, we have derived canonical word order in these sentences through the word order constraint $CANON_{GF}$. Non-canonical word order has been derived through interplay between this word order constraint and the constraint penalizing a lack of asymmetry along the lines of harmonic alignment with regards to definiteness in sentences with arguments of equal animacy. In the following we will take a look at how bidirectional OT can explain the prevalence of canonical word order in sentences that are ambiguous with regards to animacy.

Bidirectional OT was reviewed in chapter 3.8 above. In essence it involves both the speaker and hearer in optimization, and a structure must be optimal in both directions in order for it to be grammatical. In Morimoto (2000) we see a treatment of topicalization in transitive sentences in Swedish where the arguments are of equal animacy. She does not extend the analysis to other cases, as we have, due to the fact that the point to be explicated is that of ambiguity. As we have noted several places, an asymmetry in animacy is crucial to disambiguation. The analysis in Morimoto (2000) differs in a few points from our treatment, whilst also being quite similar. One key difference resides in the fact that Morimoto (2000) makes use of thematic proto-roles in the input structures for production. This gives Morimoto (2000) a different candidate set than the one in our analysis, and the proto-properties of the arguments will to a large degree determine the surface syntactic functions of the arguments. In her approach then, the competition in production takes place between the following two candidates in (19), in contrast to (18) above:

- (19) a. $\alpha_{Subject}$ Verb β_{Object}
 b. β_{Object} Verb $\alpha_{Subject}$

Morimoto (2000) allows for the topicalized object version to surface as optimal from the production function, due to its difference in input regarding proto-roles. The topicalized object element is marked as topic, as in our system. In the ensuing comprehension process that takes the output, topicalized structure as input, the information on topicality is lost or unrecoverable. Because of this, the output structure from the comprehension based process will be the one that is most harmonic with regards to its respective input, namely the canonical unmarked version.

We will not go through the tableaux presented in Morimoto (2000) in great detail, but rather attempt a schematization of the situation. Readers are referred to Morimoto (2000) for greater detail of the analysis. In essence then, the following process takes place. First, the production function, as in traditional OT, takes the input marked for animacy, thematic role and topicality and returns the optimal structure, which is a sentence where the object has been topicalized:

$$(20) f_{prod}(V(\text{Anim/Proto-Patient/Top}, \text{Anim/Proto-Agent})) \rightarrow [{}_{CP} O_{\text{Anim/P-P/Top}} [{}_{C'} V [{}_{IP} S_{\text{Anim/P-A}}]]]$$

This is where the comprehension function, f_{comp} , takes over. It takes as input the output from f_{prod} and returns the candidate that is most harmonic with regards to the grammar. As information about proto-roles and topicality is now lost, this output structure will be the maximally unmarked, canonical SVO structure:

$$(21) f_{comp}([{}_{CP} O_{\text{Anim}} [{}_{C'} V [{}_{IP} S_{\text{Anim}}]]]) \rightarrow [{}_{IP} S_{\text{Anim}} [{}_{VP} V O_{\text{Anim}}]]$$

An example might help to clarify. Morimoto (2000) operates within the framework of OT-LFG, like Lee (2002b,a) and others. Her input is thus a f-structure:²⁵

$$(22) \left[\begin{array}{l} \text{GF}_1 \\ \text{GF}_2 \\ \text{PRED} \end{array} \left[\begin{array}{l} \text{PRED} \quad \text{'jente' ('girl')} \\ \text{ANIMACY} \quad \text{hum} \\ \text{P-ROLE} \quad \text{P-A} \\ \text{PRED} \quad \text{'gutt' ('boy')} \\ \text{ANIMACY} \quad \text{hum} \\ \text{P-ROLE} \quad \text{P-P} \\ \text{PROM} \quad + \\ \text{NEW} \quad - \\ \text{'like' } \langle - \text{ } - \rangle \end{array} \right] \right]$$

Here we see that the predicate *like* ‘like’ takes the two equally animate arguments, *jente* ‘girl’ and *gutt* ‘boy’. The proto-patient argument *gutt* ‘boy’ is also the topic, i.e. it is prominent and new information. When this feature structure is fed to the production function, we get the following output:

$$(23) \quad \text{a. } [_{CP} O_{Anim/P-P/Top} [C' V [_{IP} S_{Anim/P-A}]]]$$

b. Gutten liker jenta
boy-DEF likes girl-DEF

We see that the production function f_{prod} returns a sentence where the topic, proto-patient argument *guten* ‘boy-DEF’ is realized as a topicalized object. The bidirectional optimization process, however, is not yet finished, and this is exactly the point in order. The comprehension function, f_{comp} receives the overt form in (23b) as input. However, as information represented in the input to f_{prod} on proto-role and topicality is lost here, constraints enforcing canonical word order²⁶ will give us (24) below as output, a canonical SVO structure:

$$(24) \quad [_{IP} \text{Gutten liker } [_{VP} \text{jenta}]] \\ [_{IP} \text{Boy-DEF likes } [_{VP} \text{girl-DEF}]]]$$

As this is not the same structure as the one in (23a) and can thus not have the same underlying form, the input is said to be *ineffable*, i.e. the original input is not recoverable. This is a consequence of the ambiguity in animacy represented by the original input, and “ambiguity, thus, always results in the

²⁵The f-structures in Morimoto (2000) look a little different, but the information displayed is identical. Definiteness is not represented in the original, so we will not add this here, but rather assume that it is present.

²⁶Morimoto (2000) operates with a more advanced set of word order constraints, akin to those presented in Sells (2001), e.g. a constraint penalizing unnecessary C’s, *c. This is not of great importance with regards to the point explained here.

loss of recoverability”(Morimoto, 2000). This provides us with an alternative approach to the observed fact that sentences that are ambiguous with regards to animacy tend to resist non-canonical word order. As we have seen, however, an asymmetry along the dimension of definiteness might help resolve this ambiguity to a certain extent.

An extension of our analysis through bidirectional OT

In our analysis, a crucial component has been that of the stochastic evaluation for modeling variation. As we have seen, an ordinal grammar would not have been able to represent the variation observed in the data material. Recent work on bidirectional optimization (Jäger, 2003), however, has provided an opening for a stochastic component also in bidirectional OT.²⁷

It would be an interesting endeavor for the future to attempt an analysis which took into account additional important linguistic factors, like thematic role properties. A bidirectional approach would be the way to go for such an analysis. Here, the candidates in (18) above would not be competing candidates for the production function, but rather compete in the evaluation executed by the comprehension function. The reason for this, is that the candidates competing in the comprehension-based evaluation are the ones that have the same overt form, but different underlying forms. This would allow for the same constraint set to be employed and the same generalizations to be made as in our above analysis.

Let us examine more closely how such an extension might be carried out. This is meant as a hypothetical and theoretical outline, as limitations of time, as well as the recentness of the stochastic addition to bidirectional OT, has made it impossible to attempt a full implementation of these ideas. For the future, however, an analysis by way of bidirectionality would provide us with an extension of our analysis that seems to be more linguistically plausible. The essential idea, is that the constraint interaction of our above analysis would figure in the comprehension direction, rather than the production direction.

Production As input to the production function we would include information regarding proto-roles, in addition to the features of animacy, definiteness and topicality, as before:²⁸:

(25) V(Proto-Agent/Anim/Pro, Proto-Patient/Inan/Indef)

The strong connections between, on the one hand, subject and proto-agent, and on the other, object and proto-patient are well established (Dowty,

²⁷The Bidirectional Gradual Learning Algorithm, BiGLA, has been implemented by Gerhard Jäger (Jäger, 2003) and may be downloaded from <http://www.ling.uni-potsdam.de/jaeger/evolOT/>.

²⁸On the practical side, proto-roles could be extracted from the data material by approximating the subject to be proto-agent and the object proto-patient.

1991). Constraints expressing these markedness relations are easily derived by harmonic alignment of the hierarchy for syntactic functions and the hierarchy for thematic role (P-A(gent) > P-P(atient))²⁹ One might assume, then, that the constraints pertaining to markedness relations between proto-roles and syntactic function, as well as a constraint like our previous TOP-L would be ranked at a safe categorical distance from the other constraints. A production-oriented evaluation of the above example input in (25), then would in essence only involve a linking of proto-role to syntactic function and a possible topicalization of the object:

(26)

V(P-A/Anim/Pro, P-P/Inan/Indef/Top)	TOP-L	*SUBJ/P-P, *OBJ/P-A	*S<O(ANIM) & CANON _{GF} etc...
a. [S _{P-A/Anim/Pro} V O _{P-P/Inan/Indef/Top}]	*!		
b. [S _{P-P/Inan/Indef/Top} V O _{P-A/Anim/Pro}]		*!	
c. [O _{P-A/Anim/Pro} V S _{P-P/Inan/Indef/Top}]	*!	*	*
☞ d. [O _{P-P/Inan/Indef/Top} V S _{P-A/Anim/Pro}]			

The output from the above evaluation would thus be a topicalized inanimate and indefinite object and an animate pronominal subject. So far, then, we are simply dealing with categorical constraints, and thus an ordinal ranking, i.e. for one input there is only one optimal output.

Comprehension For the comprehension function, however, the information regarding proto-role, syntactic function and (discourse-)topicality is lost, and we are dealing with a surface string where the only available information is the one inherent in the arguments, namely animacy and definiteness. Here we observe the evident similarity to the analysis of the previous sections. In the comprehension-oriented evaluation, then, we arrive at an evaluation much like the above stochastic OT analysis (see for example the tableau

²⁹A harmonic alignment of the scales for syntactic function and thematic role would provide us with the following subhierarchies (Aissen, 1999, 2003):

- (1) a. *SUBJ/P-P >> *SUBJ/P-A
 b. *OBJ/P-A >> *OBJ/P-P

in (10).). The candidates competing here, will be the ones with identical surface form, with regards to the relevant dimensions. As we know, based on the properties of animacy and definiteness alone, there are two possible outputs from the comprehension function, given the output from the production function in the tableau in (26) above, candidate d., as input to comprehension:

- (27) a. [$S_{Inan/Indef}$ V $O_{Anim/Pro}$]
 b. e.g. Epler liker han
 apples like he-NOM
 ‘Apples, he likes’
- a. [$O_{Inan/Indef}$ V $S_{Anim/Pro}$]
 b. e.g. Filmen irriterer han
 movie annoys he-NOM
 ‘The movie annoys him’

As before, the variation between these two would be arrived at through possible overlapping between the numerical rankings of the constraints pertaining to animacy and definiteness, mirroring the actual frequencies observed in the data. These numerical rankings are obtained through repeated stochastic evaluation, as before.

It is unclear, however, how the problem of ineffability, following from ambiguity in animacy, as in Morimoto (2000), has been dealt with in the stochastic implementation of bidirectional OT. However, our findings regarding the central role of definiteness in these cases might improve the situation.

It seems then, that the new approach of stochastic bidirectionality might provide an interesting extension to our analysis of the influence of animacy and definiteness on disambiguation of syntactic functions. In particular, it seems plausible that this evaluation takes place in a comprehension-oriented environment, rather than in production, as one would assume that a speaker has access to information beyond the dimensions of animacy and definiteness. An important point, however, is that our whole analysis of the previous sections stays intact in the envisioned extension.

6.4.2 Coverage

Another problem relates to the level of coverage for our analysis. One point is that our above analysis claims too much. Due to the distributions found in our data material, a topicalized object in a reverse-animacy construction becomes categorically ruled out, i.e. it is ungrammatical. This, however, is not strictly the case. It does seem, however, like sentences of this kind are highly marked and would require heavy pragmatic licensing. Additional constraints of this type might contribute towards mending this shortcoming.

A related problem resides in our solution to the resulting ambiguity when arguments are of equal animacy. Our generalization was that SVO order is predominant, unless the arguments were disambiguated by definiteness. What this means is that a topicalized object would never surface in sentences where the arguments are of equal animacy and they are not disambiguated by definiteness. We did, however, have in our data two examples that contested this claim, and sentences of this type are obviously not ungrammatical. In both of the example sentences, the topicalized object was the inanimate pronoun *det* ‘it’. As we saw in chapter 5, this is an extremely common topicalized object, and functions in a way like a topic marker pointing back to a previous context, thus being a perfect topic. Some kind of exceptional treatment of this lexical item or some features associated by it might be in line.

6.5 Conclusion

Summing up then, we have seen that the variation in word order, conditioned by animacy and, in part, definiteness, is captured to a great extent by the above analysis using stochastic OT. It allows us to model the variation that exists between one input and different realizations of this input, whilst at the same time capturing the generalizations found in the data. The fact that StOT is largely data-driven makes it a useful addition to our data analysis in chapter 5. Being, as it is, also well-founded in the theoretical framework of Optimality Theory, gives the above analysis linguistic and theoretical consequences as well.

Chapter 7

Constraint Grammar application

The starting point for the work of this thesis was the problem of automatic disambiguation of the syntactic functions of subject and object in Norwegian. This has led us to some interesting findings from a linguistic perspective in general. I have performed a data analysis of possibly ambiguous, transitive sentences in Norwegian, and looked in more detail at how the obtained generalizations may be modeled using the particular framework of stochastic Optimality Theory. In the following we will return to the practical aspect of the thesis, attempting to draw some conclusions with regards to the practical consequences of our findings.

7.1 The Oslo-Bergen tagger and the Constraint Grammar (CG) formalism

The Oslo-Bergen tagger is a morphosyntactic tagger, i.e. it provides morphological and syntactic information for running text. It is a purely rule-based tagger¹ and operates with a Constraint Grammar backbone. Constraint Grammar (Karlsson et al., 1995) is characterized by a focus on the relationships between individual words and the placement of these within a string, rather than larger structures, such as phrases. Central to this approach then, are terms like ‘word’ and ‘post- and pre-modifiers’, rather than more theoretical syntactic concepts such as sisterhood and dominance. The CG formalism was developed with automatic analyses in mind, but has much in common with grammar formalisms like Dependency Grammar. CG grammars are different from, for instance unification-based grammars like HPSG, in the way that the rules are very specific in coverage and are not very apt for making generalizations. This obviously leads to quite large sets of rules

¹In contrast to taggers that to a varying extent make use of statistical approaches.

which duplicate a lot of information, but also makes it a lot easier to add new rules without making notable changes to the others.

CG, thus also the tagger, is characterized by an eliminative approach. First, a morphological analysis is performed based on the information provided by the tagger's lexicon, which serves as input for morphological disambiguation rules. In the syntactic analysis that follows the morphological disambiguation, the tagger starts out by administering all possible syntactic functions for each word, a so-called *mapping*. It then proceeds by removing unlikely candidates from each word, following the instructions of the CG rules. As CG operates in a word-to-word, linear fashion, abstract concepts such as phrases are obsolete. This provides us with a somewhat different perspective on syntactic functions. We will look at an example output from the tagger after morphological and syntactic disambiguation has taken place, i.e. after mapping and disambiguation by rules. The sentence we will look at is a sentence with a topicalized object:

- (1) Brevet med det pussige innholdet skrev jenta.
 letter-DEF with the strange content-DEF wrote girl-DEF
 'The letter with the strange content, the girl wrote'

The output from the tagger, given the above sentence, is provided in (2) below. Syntactic tags are differentiated from morphology by the @-symbol.² Notice that the subject and object have not been disambiguated, i.e. both readings are still present in the output:

- (2) "<Brevet>"
 "brev" noun common sing def neuter @obj @subj
 "<med>"
 "med" preposition @adv
 "<det>"
 "det" determiner demonstrative sing neuter @det>
 "<pussige>"
 "pussig" adjective sing def @adj>
 "<innholdet>"
 "innhold" noun common sing def neuter @<p-utfyll
 "<skrev>"
 "skrive" verb past tr1 i1 tr11 pa1 d1 pa5 pa3 @fv
 "<jenta>"
 "jente" noun common sing def fem @obj @subj
 "<.>" PCT
 "\$." clb <<< <punkt>

As we can see, only the noun of a DP receives the syntactic function tags @subj or @obj,³ whereas other modifying elements in the phrase will receive

²Morphological tags are translated to English.

³The tagger does not adhere to the DP-hypothesis, as we have in the above.

modifier-tags, relating them to the noun through arrows ('<' and '>'). This gives us a very extended view of syntactic functions, as we get a larger set of functions. For instance, the complement noun *innholdet* 'the content' in the prepositional phrase *med det pussige innholdet* 'with the strange content' from the above example, is given a unique syntactic tag (@<p-utfyll), whereas the prepositional head, *med* 'with', itself is given the adverbial tag @adv.

With basis in the above example we see that CG clearly operates around a concept of *dependency*, marking the head of a phrase and all its modifiers as relating to this head. In most cases, then, it would not be an impossible task to construct phrase structural representations based on a CG parse.

As far as I am aware little work has been done on the topic of disambiguation of the subject and object from a CG-perspective. There is, however, an article worth mentioning, Anttila (1995), which outlines an approach for determining the subject function in a Constraint Grammar system for English. The basic strategy employed by Anttila might work out quite well for English sentences, and also for a subset of Norwegian ones. It would not, however, handle correctly a sentence with a topicalized object, like the one in (1) above. The reason for this is that the main rule for subjecthood in Anttila's system is one in which a word cannot be given subject status as long as "a transitive or copular verb precedes [it]" (Anttila, 1995, p. 321).

7.1.1 The formalities of CG Rules

The CG rules⁴ perform the disambiguation crucial to the formalism. A CG rule consists of four elements (Karlsson, 1995, p. 57): 1) a domain, 2) an operator, 3) a target, and 4) context condition(s). For syntactic rules there is only one domain available, @w, a variable for any word with the property denoted by the target. The operator defines the type of rule we are dealing with - either a *discard* rule (s0-operator), which discards the reading specified in the target, or a *select* rule (s!-operator), selecting the reading specified by the target as the correct reading and eliminating all the other possible readings. The target, in the case of syntactic rules, is simply one of the syntactic functions, the one that the rule either eliminates or selects according to the operator. The context conditions pose contextual restrictions on the domain word that must be fulfilled in order for the selection or discarding of syntactic functions to take place.

Context conditions can be conceived of as triplets, i.e. ordered lists of three elements: <polarity, position, set> (Karlsson, 1995, p. 58). The polarity property is a binary property, i.e. positive or negative. The position is an indicator of relative position, starting with the domain word in position

⁴The CG rules are often referred to as *constraints*, hence the name Constraint Grammar. Here, however, we will refer to these as *rules* so as to not confuse them with our Optimality Theoretic constraints of the previous chapter.

0. Positions may either be absolute, e.g. -2 (2 words to the left), 1 (the following word) or the position may be unbounded, indicated by an asterisk, e.g. *-2 (somewhere further to the left of the second word to the left), *1 (somewhere following the next word). These different positions all have in common the fact that they revolve around the absolute position of the target (0).⁵ Finally, all properties predicted of the target in a context condition must be declared in a set, where the set name is given in the context condition. Sets may contain any type of property relevant for the target, i.e. base form of a word, morphological feature, part of speech, syntactic function etc.

One limitation associated with the context condition is that it may not include a disjunction, thus separate rules must be written for each alternative in a disjunction.⁶ A final optional argument for context conditions exists, making use of a *linking mechanism* available in the CG formalism:

```
(*1 @obj *R)      ; right of position 1 is an object, LINK
(NOT LR0 animate) ; the element in this position is not animate
```

In the two context conditions above, the *R acts as a link between the two conditions, saying that if somewhere to the right of position 1, an object is encountered, continuing in this position, the element should not be a member of the animate set.

An example of a very simple CG rule is provided below for illustration, a rule which eliminates the subject reading from a complement of a preposition:

```
@w = s0 (@subj) ; discard-rule with target @subj
(-1 prep)      ; context condition, position: -1, set: prep
```

The above rule discards the subject reading from a word if the immediately preceding word is a preposition. The context condition in the last line has positive polarity (null expression), pertains to position -1 and indicates that whatever is situated in this position should have the property ‘prep’, i.e. it should be a preposition.

7.2 Practical implementation explored

The parsed example in (2) above was also chosen in order to illustrate the problem of automatic disambiguation that served as the starting point for this work. We see that the two nouns *brevet* ‘letter-DEF’ and *jenta* ‘girl-

⁵There are some other alternatives, which are not relative to the target position, 0. In addition to the above, we will mention the following for ease of understanding of the presented rules: *R/*L (some position to the right/left), LR0 (the position we are looking at in the structure, not position 0.), R+/L- n, where n ≤ 4 (n positions to the right/left).

⁶The CG formalism described here is that of Karlsson et al. (1995). There is however a new version of the formalism, CG2, which is more apt for generalizations, by e.g. allowing for disjunctive statements etc. However, as the first version of CG is the one that is currently employed in the Oslo-Bergen tagger, this is also what we will limit ourselves to.

DEF' are not fully disambiguated, i.e. they are both abandoned by the parser as subject and object. Needless to say, this is not an ideal situation. A Constraint Grammar system should adhere to the Uniqueness Principle, namely that "a clause may contain at most one instance of each principal function" (Anttila, 1995, p. 341).

An interesting question, then, pertains to how our findings in the previous chapters may contribute towards a solution to this problem of disambiguation. In particular, an introduction of the property of animacy to the system may provide useful, whilst at the same time being a simple lexical property of the arguments.⁷

7.2.1 Practical outline

In essence then, we are looking to come up with an algorithm that will encapsulate our earlier findings, whilst at the same time interacting properly with the system as it exists today. This illustrates one of the advantages of CG; that it is fairly easy to enter the process at a later stage and add rules without risking too much.⁸

A first step towards an implementation of our theory relies heavily on an introduction of the property of animacy to the system. Unfortunately, there is no full-coverage lexicon annotated for animacy available for Norwegian, so we will have to make do with a simulation of such a lexicon. Crucial to such an approach then, are the set declarations central to the CG algorithm. We may simulate a lexicon annotated for animacy by declaring large sets consisting of word forms: words that are animate and words that are inanimate. We wish to simulate an idealized situation where information regarding animacy has full coverage, and we will include in these sets all arguments of transitive constructions in the text to be analyzed.

In order to give our rules a wider coverage, in addition to actually going through the text in search of animate/inanimate elements in the proper contexts, we will also make use of the resources actually available in this respect, the SIMPLE and NorKompleks lexica.

SIMPLE. The SIMPLE (Semantic Information for Multifunctional, Plurilingual Lexica) lexicon⁹ contains an encoding of a rich hierarchy of structured multi-levelled semantic representations for approximately 10,000 words, such

⁷The situation is a bit more complex when dealing with metonymy. However, as we have seen in chapter 4.4, metonymy is in many cases a quite regular phenomenon, and lexica annotated for this type of regular metonymy have been produced.

⁸The reason for this is that the CG-rules are all very specific as to coverage and do not employ mechanisms like, for instance, unification, where changes in one rule might have complex repercussions in other rules.

⁹As adapted from Danish to Norwegian by Seksjon for leksikografi og målføre at the University of Oslo.

as information regarding a verb's varying selectional restrictions on arguments, aspectual information etc. The choice of structure and information for each entry in the lexicon is partly based on Pustejovsky's notion of qualia structure, which specifies different aspects of a word's meaning: its constitutive role, i.e. the way it is made up, its formal role, i.e. how it relates to a larger domain, its telic role, i.e. what it is for, or its function, and its agentive role, i.e. what causes it. Verbs and nominals representing events are also singled out by an individual set of characteristics, event types, which include, among others, 'change', 'state' and 'cause change'. In particular, nouns are specified for animacy¹⁰. The SIMPLE lexicon also singles out causative verbs and these are specified as events of the type 'cause change of state'.

NorKompleks. At the moment, the Oslo-Bergen tagger annotates verbs using codes from the NorKompleks lexicon of verbs.¹¹ It contains information regarding verbal argument structure, theta grid etc. The NorKompleks representation of a prototypical transitive verb, i.e. a verb that takes two NP arguments and distributes the thematic roles of agent and theme to these arguments, is provided below:

(3) `arg_code(trans1,[arg1:su::ag::np,arg2:obj::th::np])`.

In particular, the NorKompleks lexicon will be made use of when discerning some of the reverse-animacy verbs (cf. chapter 3.6)¹² from regular transitive verbs.

When cross-checking the Norwegian equivalents of the psych-verbs or *amuse*-verbs mentioned in Levin (1993) and Grimshaw (1990) wherever possible, I found that these were quite systematically encoded as *trans7* in the NorKompleks codes, taking a Theme subject, and an Experiencer object:

(4) `arg_code(trans7,[arg1:su::th::np,arg2:obj::exp::np])`.

The *concern*-verbs constitute another group of verbs that also takes a Theme subject, but not an Experiencer object. In the NorKompleks codes these are

¹⁰In fact, the annotation of animacy in SIMPLE is a bit more complex than a simple dichotomy between animate and inanimate. Animate elements are specified as human, animal, human group etc. The inanimate elements have even more structured representation, e.g. building, artifact, substance etc. We will not make use of this additional information, but simply operate with our animate/inanimate dichotomy, as before.

¹¹The NorKompleks lexicon of verbs was developed at NTNU in Trondheim.

¹²To refresh our memory, the reverse-animacy verbs are verbs that can take an inanimate subject and an animate object. We differentiated within this group between

- the *amuse*-verbs, a subgroup of the psych-verbs, e.g. *interessere* 'interest'
- the *concern*-verbs, e.g. *gjelde* 'concern'
- the causative verbs, e.g. *drepe* 'kill'

encoded as *trans5* and *trans6*, which seem to represent the same type of argument structure, but differ in their ability to passivize. The verbs coded as *trans5* do passivize well, but the *tr6*-verbs do not:

(5) `argcode(trans5/6,[arg1:su::th::np,arg2:obj::th::np])`.

The NorKompleks lexicon thus provides us with a resource for identifying and listing the reverse-animacy verbs. The set of thematic roles it makes use of (Agent, Recipient, Experiencer, Theme and Location), is different than the roles in for instance Lødrup (2000) (cf. chapter 3.6.1). Recipient seems to bear a much more restricted role here than Benefactive, as employed in Lødrup (2000). It occurs only with ditransitive verbs, and a few transitive verbs like *arve* ‘inherit’ and *fa* ‘receive’, which denote a transaction where a participant is in fact a recipient in the most physical and straight-forward sense. The most important characteristic of these verbs however, seems to be maintained in the NorKompleks codes, namely that the subject is encoded Theme.

It seems then that we should make immediate use of this information when further improving the Oslo-Bergen tagger. Our theory of hierarchy alignment does not account for this group of verbs, but rather the majority of cases. After all a hierarchy does not impose categorical restrictions on output, but rather represents a tendency. Verbs classified as *trans5*, *trans6* or *trans7* should be treated as exceptions to the general rule, due to their specific lexical properties.

As outlined at the end of chapter 5, we envision three possible scenarios with regards to animacy from a linear surface structure like [α Verb β]:

1. Disambiguating asymmetry: $\alpha = \text{hum-anim}$, $\beta = \text{inan}$
2. Topicalization/Deviant asymmetry: $\alpha = \text{inan}$, $\beta = \text{hum-anim}$
3. Equally animate: $\alpha = \text{hum-anim/inan}$, $\beta = \text{hum-anim/inan}$

We may, based on this categorization, outline an approach to automatic, rule-based disambiguation of each of the above types of transitive sentences.

1. Disambiguating asymmetry. These sentences are sentences with an animate first argument and an inanimate second argument:¹³

(6) Dørvakten undersøkte tipset
doorman-DEF investigated tip-DEF
‘The doorman investigated the tip’

¹³All examples in the below treatment are taken from the data material described in chapters 4 and 5, unless otherwise stated.

In cases of disambiguating asymmetry, such as the above, we will conclude that the word order in question is SVO, i.e. that the first argument is the subject and the second argument is the object. The reason for this is that we did not come across any topicalized objects that were animate, when the corresponding subject was inanimate. Even though sentences with a maximally marked animacy assignment such as these are not ungrammatical as such, they are certainly strange, extremely rare and very dependent on contextual factors, such as stress:¹⁴

- (7) ?? Guttene gjelder saken
 boys-DEF concern case-DEF
 ‘The boys, the case concerns’

A simplified version of a rule selecting a subject reading for an animate first argument in cases like these then would be something like the following:¹⁵

(@w = s! (@subj) ; select-rule with target @subj
 (0 animate) ; the word is member of the animate set,
 (*1 @fv *R) ; somewhere to the right of 0 is a finite verb
 (*R @obj *R) ; somewhere to the right of the verb is an object
 (LRO inanimate)) ; the object is member of the inanimate set

Obviously, in a proper functioning rule in the tagger system, further conditions must be added in order to constrain the possible environment for the application of the rule. For instance, only adverbials may occur between the subject and the finite verb. Also, care must be taken to see that the elements in question occur within a proper sentence boundary and that a non-finite verb does not follow the object.¹⁶

¹⁴The example in (7) is a constructed one, as no examples of this type were uncovered in the data analysis in chapter 5.

¹⁵A rule selecting, thus disambiguating, the subject of a transitive sentence assumes an additional rule operating on the basis of the Uniqueness Principle mentioned above, and accordingly disambiguating the object.

¹⁶As we remember, only simple verb phrases are subject to the ambiguity dealt with here, as complex verb phrases with an auxiliary disambiguate their arguments structurally:

- (1) a. *Jenta* har skrevet *brevet*
 girl-DEF has written letter-DEF
 (i) ‘The girl has written the letter’
 (ii) * ‘The girl, the letter has written’
 b. *Brevet* har *jenta* skrevet
 letter-DEF has girl-DEF written
 (i) ‘The letter, the girl has written’
 (ii) * ‘The letter has written the girl’

2. Topicalization/Deviant asymmetry The sentences classified under this category are sentences where the initial argument is inanimate and the following argument is animate. Here we must differentiate between two possibilities - SVO order and OVS order:

- (8) a. Det samme gjelder Black Butler
 the same concerns Black Butler
 ‘The same concerns Black Butler’
- b. Dette sier en skuffet Miss Norway
 this says a disappointed Miss Norway
 ‘This is what a disappointed Miss Norway says’

Our tool for disambiguation here is, first and foremost, the verb. As we have seen in the preceding chapters, the reverse-animacy verbs may take an inanimate subject and an animate object.¹⁷ Some of the reverse-animacy verbs may be extracted from the NorKompleks lexicon as these are uniformly coded there, as we saw above. Causative verbs, however, are not given a unique code in NorKompleks. The SIMPLE lexicon provides us with relevant information on verbs encoded as events of the type ‘Cause Change of State’. This information is however limited, as only around 1500 verbs are represented in total in SIMPLE.

Also, presentational constructions will have to be accounted for, as the tagger does not differentiate between the expletive and pronominal inanimate subject *det* ‘it’. This, however, is not a difficult task, as the verbs which participate in the presentational construction are not tagged exclusively as transitives.¹⁸

An appropriate rule for an SVO reading of a sentence with an inanimate subject and an animate object then, will have to look at the verb and ascertain that this is a reverse-animacy verb. Once again, it will be a select rule

¹⁷Here we abstract away from personifying metaphors (cf. chapter 4.4), as these are not predictable based on the verb alone, if predictable at all.

¹⁸As we remember, the object of the presentational construction is the logical subject of the verb, which is usually intransitive:

- (1) Det kommer en mann i trappa
 it comes a man in stair-DEF
 ‘A man is coming up the stairs’

The verb of a presentational construction may, however, also be transitive, the type of transitive verb which might occur without its object:

- (2) Det synger en mann i trappa
 it sings a man in stair-DEF
 ‘A man is singing in the stairs’

We will not deal with these in the following, only the ones that have uniquely intransitive verbs. The reason for this is that they must be said to be ambiguous, between an expletive and pronominal reading of *det* ‘it’.

targeted at the subject, ensuring that this is inanimate, and that the corresponding object is animate. Also, a rule enforcing the Uniqueness Principle will disambiguate the object, as there can be no more than one subject in a sentence:¹⁹

```
(@w = s! (@subj) ; select-rule with target @subj
  (0 inanimate) ; the word is member of the inanimate set
  (*1 @fv *R) ; to the right of 0 is a finite verb
  (LRO reverse-animacy *R) ; the verb is reverse-animacy
  (*R @obj *R) ; to the right of the verb is an object
  (LRO animate)) ; the object is member of the animate set
```

In the case of the OVS reading we would require that the verb is not a reverse-animacy verb. Here we would target the subject, which in this case is the animate element:

```
(@w = s! (@subj) ; select-rule with target @subj
  (0 animate) ; the word is member of the animate set
  (*-1 @fv *L) ; to the left of 0 is a finite verb
  (NOT LRO reverse-animacy *L) ; the verb is not reverse-animacy
  (*L @obj *L) ; to the left of the verb is an object
  (LRO inanimate)) ; the object is member of the inanimate set
```

Once again, the real rules added to the system would be more complex.

3. Equally animate Sentences with equally animate arguments constitute the third group of possible constellations with regards to animacy:

- (9) a. En nabo oppdaget innbruddstyven
 a neighbour discovered burglar-DEF
 ‘A neighbour discovered the burglar’
 b. Parlamentarismen har sine svakheter
 parliamentarism-DEF has its weaknesses
 ‘Parliamentarism has its weaknesses’

We remember from preceding chapters that this type of sentences less frequently contain a topicalized object. Also, we saw that if the object were to topicalize it very often depended on disambiguation from the relative definiteness of the arguments, i.e. that the subject would be higher ranked than the object on the scale of definiteness. Unfortunately, however, this generalization proves to be difficult to make use of in the present context. In our optimality theoretic modelling in the previous chapter, the stochastic component allowed for variation in output, so that sentences with equally animate arguments and a topicalized object could surface as optimal if they were disambiguated by definiteness. However, many of the SVO sentences with equally placed arguments do not adhere to the definiteness hierarchy,

¹⁹The tagger provides a separate code for indirect objects, @i-obj.

thus making it virtually impossible to differentiate between an SVO sentence of this type and an OVS that disambiguates by definiteness. In example (10) below,²⁰ we see that the only difference between the two sentences is the verb and the prepositional phrase. Where an OVS reading is near obligatory in the sentence in (10a), the second sentence in (10b) is ambiguous between a SVO and an OVS reading. Obviously, there are also lexical factors contributing towards disambiguation here, unfortunately being difficult to encapsulate in an automatic system.

- (10) a. Barnet bærer hun i armene
 child-DEF carries she i arms-DEF
 ‘The child, she carries in her arms’
- b. Barnet liker hun med hatten
 child-DEF likes she with hat-DEF
- (i) ‘The child likes the female with the hat’
 (ii) ‘The child, the female with the hat likes’

Our conclusion then is that we are not able to make use of information regarding definiteness in these cases. We will therefore postulate an SVO word order for all sentences with equally animate arguments. Another possibility in the sentences where the arguments are of equal animacy, is to leave these as ambiguously tagged, reflecting the ambiguity in animacy. This approach, however, will give us a lower *precision*, even though *recall* will remain the same.²¹ As our main aim is to disambiguate the readings as much as possible, we will opt for the former approach, inducing SVO readings in the sentences with arguments of equal animacy. This will give us a margin for error at about 2% of the sentences with equally animate arguments (based on the data analysis of chapter 5.).

Ultimately then, we can make the following generalization: for the purpose of our automatic system, an animate first argument induces SVO word order, independent of the animacy of the second argument. This means that our first group of sentences, under the heading disambiguating asymmetry and sentences with arguments that are both animate may be covered by one and the same rule, as outlined below:

²⁰The second example in (10) is constructed as a minimal pair to the first, which is genuine, i.e. from the data material.

²¹The definitions of precision and recall are given as follows (Karlsson et al., 1995, p. 172):

Recall: The ratio “received appropriate readings / intended appropriate readings”

Precision: The ratio “received appropriate readings / all received readings”

```

(@w = s! (@subj) ; select-rule with target @subj
(0 animate) ; the word is member of the animate set
(*1 @fv *R) ; to the right of 0 is a finite verb
(*R @obj *R) ; to the right of the verb is an object

```

In sentences where both arguments are inanimate, however, we do need to check the animacy of the second argument, before concluding SVO word order. If the second argument is animate, we are dealing with a sentence of type 2 above - Topicalization/deviant asymmetry. Our outline for a rule of this kind then is presented below:

```

(@w = s! (@subj) ; select-rule with target @subj
(0 inanimate) ; the word is member of the inanimate set
(*1 @fv *R) ; to the right of 0 is a finite verb
(*R @obj *R) ; to the right of the verb is an object
(LRO inanimate)) ; the object is member of the inanimate set

```

An example rule from the system

Above we have looked at several outlines for rules embodying the generalizations we wish to express. It might be interesting to see what one of our finished, functioning CG rules looks like. A commented rule is provided below, whereas the rest of the rules can be found in Appendix B:²²

Rule 2: [*Obj_{Inanimate} Verb Subj_{Animate}*]

```

(@w = s! (@subj) ; select the subject
(0 lo#animat) ; word is animate
(NOT 0 lo#mod-@adv-@<p-utfyll) ; word is not adverbial etc.
(*-1 lo#ikke-mod-@<p-utfyll-@adv-@app *L) ; find a non-modifier to the left
(LRO @fv *L) ; word is a finite verb
(NOT LRO lo#reverse-animacy-verb *L) ; word is not reverse-animacy
(NOT LRO lo#kop-pass *L) ; word is not copular or passive
(NOT LRO lo#presentering-verb *L) ; word is not a presentational verb
(*L lo#ikke-mod-@<p-utfyll-@adv-@app *L) ; find a non-modifier to the left
(LRO @obj *L) ; word is an object
(NOT LRO lo#mod-@adv-@<p-utfyll *L) ; word is not an adverbial etc.
(LRO lo#inanimat *L) ; word is inanimate
(*L setn-gr/verb/subj *L) ; find sent.boundary/verb/subj
(LRO lo#top-setn-gr) ; word is valid sentence boundary
(*1 lo#ikke-mod-@<p-utfyll-@adv-@app *R) ; find a non-modifier to the right
(NOT LRO @iv)) ; word is not an infinite verb

```

Comment: An inanimate first argument and an animate second argument induces OVS word order if the verb is not reverse-animacy, causative or a verb that can occur in a presentational construction. The verb phrase is

²²The outlined rules in the above sections make use of simpler set names for the sake of illustration. The genuine rule from the system, however, makes use of set names that are a bit different from these. The set definitions are provided in Appendix B.

simple, i.e. no infinite verb, and there are no verbs or subjects to the left of the subject, within the same sentence.

7.3 Results

In order to test the above practical outline and our theoretical predictions, I wrote a small set of general rules in CG and added to the existing set of rules in the Oslo-Bergen tagger system. A small test corpus²³ was employed for this purpose. In order to simulate an idealized situation where all information regarding animacy is available, all animate and inanimate pronouns, nouns and proper nouns in the test corpus were declared in sets, as described above.

As CG rules are very specialized, it is difficult to generalize with them. Without being able to abstract over linguistic concepts like ‘sentence’ and ‘noun phrase’ and including such abstractions in the rules, writing rules with general coverage is a very time consuming and meticulous affair. For instance, the phenomenon of relative sentences as modifiers of nouns proved to be an extremely difficult task to cover, given the workings of the CG formalism. Without consistent punctuation as delimiters and with an optional relative subjunction *som* ‘who/which’ the patterns for possible relatives multiply quickly.²⁴ However, it seemed like an interesting endeavour to examine how far we could get in terms of disambiguation with only a rather small set of general rules.

Based on the rule patterns presented above, eight rules were written to cope with the basic types of transitive sentences described in this thesis. Two sets of word forms were collected from the corpus and the SIMPLE lexicon in order to simulate annotation for animacy. In addition two sets of verbs were declared. The sets are described further below:

1. Animate word forms

²³This is a corpus of approximately 44300 words, which has been proofread for morphology, syntax and named entity disambiguation.

²⁴In order to further illustrate this point, we can take a look at the patterns for which individual rules coping with modifying relative sentences would have to be written:

- | | | | |
|---|--------------------------------------|---|-------------------------------------|
| - | [@subj som _ @fv, @fv @obj] | - | [@subj som _ @fv @fv @obj] |
| - | [@subj som _ @fv @obj, @fv @obj] | - | [@subj som _ @fv @obj @fv @obj] |
| - | [@subj som _ @fv @iv, @fv @obj] | - | [@subj som _ @fv @iv @fv @obj] |
| - | [@subj som _ @fv @iv @obj, @fv @obj] | - | [@subj som _ @fv @iv @obj @fv @obj] |

Here, the only difference between the two columns is the comma after the relative clause. In addition, there is a possibility for a comma after the head of the clause, i.e. before the modifying relative clause. Needless to say, the above patterns are only the tip of the iceberg. In addition, the extracted item may be object of the relative clause, which gives us optionality with regards to the presence of the relative subjunction marker *som* ‘who/which’. The extracted element may also be the topicalized object of the matrix clause. Also, the relative clause itself may contain all kinds of sentences with an extracted element, such as subject predicating sentences, non-finite complements etc. etc.

- (a) nominative pronouns, determiners e.g. *han* ‘he’, *ingen* ‘nobody’ etc.
- (b) common nouns - both inherently animate and regular metonymies, e.g. *jente* ‘girl’, *departement* ‘ministry’ etc.
- (c) proper nouns - both inherently animate and regular metonymies, e.g. *Peter*, *Norge* etc.

2. Inanimate word forms

- (a) pronouns, determiners like *det* ‘it’, *dette* ‘this’ etc.
- (b) common nouns - e.g. *brev* ‘letter’, *departement* ‘ministry’ etc.
- (c) proper nouns e.g. *EØSavtalen*, *Norge* etc.

3. Reverse-animacy verbs - *amuse*-verbs and *concern*-verbs taken from the NorKompleks lexicon of verbs, as well as causative verbs taken from the SIMPLE lexicon

4. Presentational verbs - verbs that occur in the presentational construction, coded as intransitive verbs. This set contains NorKompleks codes for intransitive verbs.

In the small test corpus we found 285 transitive sentences fulfilling the criteria set up in chapter 4. Above all, these are transitive sentences that may not be disambiguated based on morphological case or word order. When evaluating the performance of our rules on the test corpus then, the following results were obtained:

Comment		No	%	Total %
CORRECT	Both arguments disambiguated	214	75.1%	83.5%
	Only target disambiguated	24	8.4%	
UNCORRECT	Lack of coverage in rules	12	4.2%	16.5%
	Mistakes in input made by tagger	34	11.9%	
	Theoretical mispredictions etc.	1	0.4%	
Total		282	100%	100% %

We see from the above table that our small set of rules already has a pretty high rate of disambiguation. The number of sentences where the target of the rule was correctly disambiguated amount to 83.5% of the sentences. Above we differentiate between sentences that have been fully disambiguated and sentences where only the target of the rule was disambiguated. However, both of these must be considered to be correct, as the rule may only target one argument at a time. Further rules invoking the Uniqueness Principle will contribute towards a fuller disambiguation of the sentences where the rules applied.

The sentences labeled ‘incorrect’ represent a rather diverse group of sentences. First of all, our rules at work here represent only the most common and basic patterns for transitive sentences. Due to the linear fashion in which CG works, simple modifiers such as relative sentences become a large and very time consuming task to represent, as we saw above. Therefore, we cannot assume full coverage for our rules at present.

Secondly, our rules depend on an input that is correct with regards to the morphological component and the workings of the other syntactic rules. However, the input is not always correct, as we see in the example below, where the first word *Baste*, an uncommon proper noun, has received a unique reading as an infinite verb:

- (11) "<Baste>"
 "baste" verb tr1 inf @iv
 "<gjør>"
 "gjøre" verb tr1 pres rl9 pr3 @fv
 "gjø" verb tr1 pres rl4 @fv
 "gjø" verb pres i1 @fv
 "<alt>"
 "all" ent noeyt det kvant @det>
 "<arbeidet>"
 "arbeide" subst appell ent be noeyt @subj
 "arbeid" subst appell ent be noeyt @subj

Obviously, we cannot expect our rules to properly handle cases like these. In fact, if we exclude these sentences from our count, operating under an assumption of idealized input, our rate of correct disambiguation rises to 94.8%.

Finally, we have the cases where our rules applied, but not correctly (0.4%). These were expected, as we did not expect our rules to cover metaphorical use or topicalization occurring in sentences where the arguments were of equal animacy. No metaphoric examples were found, the one example was a sentence with a topicalized object and equally animate arguments:

- (12) Det gjør historien om Ane også
 it does story-DEF about Ane also
 ‘So does the story about Ane’

7.4 Problem areas for further work

Obviously, more work remains before an implementation of our findings in this thesis with full coverage is accomplished. In the following we will take a closer look at some problems that are in need of further work.

7.4.1 Metonymy revisited

The fact that we have been working with an idealized situation in terms of full access to information on animacy limits the immediate application of our rules. In a real world situation, when applying the rules to unknown text, it is not very realistic to assume full access to animacy-based information. In particular, with respect to metonymy²⁵, we might never have an account of all possible metonymic usages. Due to the dual nature of metonymy, these lexical elements are coded as both animate and inanimate, e.g. the proper noun *Norge* ‘Norway’.²⁶

- (13) a. Norge likte kampen
Norway liked match-DEF
‘Norway liked the match’
- b. Kampen likte Norge
match-DEF liked Norway
‘Norway liked the match’
- c. Jenta likte Norge
girl-DEF liked Norway
(i) ‘The girl liked Norway’
(ii) ‘The girl, Norway liked’
- d. Norge likte jenta
Norway liked girl-DEF
(i) ‘Norway liked the girl’
(ii) ‘Norway, the girl liked’

We see that in (13a) and (13b), there is only one possible reading for these sentences, SVO and OVS, respectively. As the other argument in these sentences is clearly inanimate, the animate, metonymic reading of *Norge* surfaces as the only possible one. This is obviously also due to the fact that the verb is not a reverse-animacy verb.

The situation is different, however, in examples (13c) and (13d). Here, the opposing argument is clearly animate, thus causing ambiguity. This is not surprising, as sentences with equally animate arguments are typically ambiguous, as we have seen. In the practical implementation reviewed above, we have chosen to follow the most probable path with regards to these sentences that are ambiguous with regards to animacy, thus inducing canonical word order. How then are we to ensure that we get the readings we want for the above sentences in (13) in the CG output?

²⁵See chapter 4.4 for more on metonymies and their treatment within the boundaries of this thesis.

²⁶The examples in (13a) - (13d) are constructed examples.

One problem is that two of our SVO rules will in reality be applicable in the case of example (13a): the one inferring SVO order in cases of disambiguating asymmetry and the rule inferring SVO in sentences with two inanimate arguments. In both cases however, the resulting disambiguation will be correct. Since our system does not account for topicalized objects in sentences with equally animate arguments, the resulting output for (13c), as well, will be correct irrespective of a metonymic, i.e. animate, or a non-metonymic, i.e. inanimate, reading of *Norge*.

However, a more serious problem emerges in sentences where the duality of these metonymic elements actually have a significance for the output. Depending on the arbitrary ordering that the rules apply in, we may get an uncorrect reading for example (13b). The reason for this then is that *Norge* is coded as both animate and inanimate, thus both the rule for topicalization, interpreting *Norge* as animate subject and *kampen* ‘match-DEF’ as inanimate, topicalized object, and the SVO rule for two inanimate arguments, interpreting both arguments as inanimate, may apply in (13b). A possible solution would be to ensure in the SVO rule for two inanimate arguments, that we are in fact dealing with two clearly inanimate arguments. This could be done by adding an extra context condition to the rule:

(*R @obj *R) ; somewhere to the right of the verb is an object
 (LRO inanimate *R)) ; the object is member of the inanimate set
 (NOT LRO animate) ; the object is not member of the animate set

This would ensure that the topicalized object is not metonymic, i.e. that it is not both in the animate set and in the inanimate set.

A similar problem is found in (13d) above. We see that, depending on the arbitrary ordering of the rules performed by the system, SVO word order or OVS order may be induced in this sentence. The fact that *Norway* has a regular metonymic nature, makes the sentence in (13d) seem ambiguous. In this respect, it is not directly incorrect that such a sentence should in possibility receive two different readings. It is however, a problem that the two readings are arbitrarily assigned and not founded in any linguistic insight. What then, are we to do with sentences of this type? In the example in (13d), the position of the possibly metonymic element in combination with the animacy of the other argument induces uncertainty and ambiguity. Following our earlier assumptions, then, we will enforce a SVO reading on (13d). This can once again be done by ensuring that the rule for OVS does not apply in these sentences, by adding a context condition, like the following:

(*L @obj *L) ; somewhere to the left of the verb is an object
 (LRO inanimate *L)) ; the object is member of the inanimate set
 (NOT LRO animate) ; the object is not member of the animate set

Once again, this would ensure that the topicalized object is not metonymic.

Whether the above solutions are the best is a question that certainly should be addressed in future work. Perhaps it would be better to leave the metonymic element as ambiguous? This is a question that also pertains to our earlier assumption about enforcing SVO order on the sentences that were equally animate. Should we induce SVO order or leave these to be ambiguous? In the above, we have operated with a goal of improving disambiguation as much as possible, and have therefore taken a few risks. The idea has been to see how far one could get with only quite crude and simple measures. A further refinement in the future is certainly a topic that should be addressed for future work on improving the approach outlined in this chapter.

7.4.2 Other areas

Another question worth looking into, is the ordering of the rules. The CG formalism includes a heuristic aspect, which allows for ordering of rules. The rules described in this chapter were all applied at the same heuristic level, namely the first stage of rule application. It makes sense that some rules, for instance, rules inferring SVO order when nothing else applies, should apply at a later stage, letting all the rules pertaining to animacy and special cases, like reverse-animacy verbs apply first. In particular, this becomes important when adding further rules to the system. If we draw a comparison with the stochastic OT analysis performed in the preceding chapter, we saw that the constraint enforcing canonical word order, $CANON_{GF}$, was ranked lower than the constraints dealing with an asymmetry in animacy. This would correspond to a rule applying later on, so that if a candidate has not violated any constraints, or alternatively been taken care of by any rules, this constraint will enforce SVO order. This is a matter that certainly should be looked into.

An obvious problem related to the work described here, resides in obtaining more information on animacy. Even if a lexicon annotated for animacy did exist, there would still be a fair amount of new coinings, compounds etc. that would go unrecognized. The tagger does contain a compound analyzer. Perhaps an extension of this, then, could infer the animacy of a compound expression based on the animacy of its parts, e.g. *biljente* ‘car-girl’ is animate, whereas *jentebil* ‘girl-car’ is inanimate.

In general then, more rules are needed in order to capture the most basic patterns. As we have seen, relative clauses, as modifiers of nouns, constitute one problem. In chapter 5.1 we took a brief look at the two-place unaccusatives, a group of verbs which also should be treated within the system.

7.5 Conclusion

Following from the above results, then, it seems fair to say that the practical experiment was quite successful. We see that our generalizations from the previous chapters with only a few basic rules are readily accommodated in the CG system. Also, as we have seen, they help to disambiguate a large number of the transitive sentences that are ambiguously tagged.

Chapter 8

Conclusion

8.1 Summary

Following from a problem of syntactic disambiguation in transitive sentences, we have arrived at some interesting, novel observations with regards to the realization and expression of the syntactic functions of subject and object in Norwegian.

We started out by establishing a structural ambiguity in simple transitive sentences, as caused by lack of morphological marking and the possibility for the object of such sentences to topicalize. In order to resolve this ambiguity, we saw the need to examine additional properties of subjects and objects, apart from strictly structural ones. A focus on the properties of animacy and definiteness pointed us in the direction of typological linguistics and the universal markedness relations expressed through interaction of prominence hierarchies there. In particular, we saw that animacy and definiteness are central conditioning factors for the expression and realization of subjects and objects in a number of different languages. Our question, then, was whether some of these generalizations might also pertain to Norwegian, manifested as strong usage preferences at a level of performance. Following an analysis of one thousand transitive sentences, we concluded that this was partly the case. In particular, we saw that animacy to a great extent conditions the realization of syntactic functions and the possible word order variation between these in Norwegian. Our main findings may be summarized as follows:

- Subjects are almost exclusively higher than or equally placed in animacy as their corresponding objects.
- Several types of verbs may deviate from this norm and take an inanimate subject and an animate object. These are the reverse-animacy constructions. We have identified three groups of verbs which may partake in this constellation:
 - *amuse*-verbs, e.g. *interessere* ‘interest’

- *concern*-verbs, e.g. *gjelde* ‘concern’
 - causative verbs, e.g. *skade* ‘hurt’
- Even though there are a number of possible reverse-animacy verbs, these constructions are rare in the data, a fact which supports our theoretical predictions of hierarchy alignment.
 - A freezing effect on word order in reverse-animacy constellations is present in Norwegian as a strong usage preference.
 - Sentences with ambiguous arguments with regards to animacy, resist object topicalization.
 - If the object topicalizes, however, sentences with arguments of equal animacy must be disambiguated in some other way, most often by definiteness.

By employing a stochastic OT approach, we were able to express in our grammar the theoretical predictions from typological linguistics through universal markedness constraints. It allowed us to model, in an ordered and well-founded manner, the variation between different structural interpretations of a sentence, as conditioned by animacy and, in part, definiteness.

Finally, our theoretical predictions were tested in practice. We simulated a lexicon annotated for animacy and made use of this additional information in Constraint Grammar disambiguation rules. We then employed the rules to a corpus text, with a high resultant rate of disambiguation.

8.2 Topics for further work

8.2.1 Animacy and word order revisited

Part of our claim in the present thesis has been that there is a connection between the relative animacy of the arguments in a transitive sentence, and the degree of word order variation observed. Due to the fact that the simple transitives, which have been our main focus, do not mark syntactic function structurally, we have concluded that other disambiguating factors, like animacy, must adhere to the markedness relations expressed by harmonic alignment in order for variation to occur. We have seen that this generalization is borne out in the case of reverse-animacy constructions. These constructions do not follow the predicted markedness relations, hence do not topicalize at a level of performance. A possible way of testing these predictions further would be to examine complex VPs, where the structural ambiguity is not present. Do these exhibit the same properties as those of simple transitives or are they more often found with marked arguments? Might it be that an unmarked structural construction is to a larger extent preferred for the

expression of conceptually marked arguments, for instance along the dimension of animacy? Intuitively, a sentence with a reverse-animacy constellation of arguments and a topicalized object in a complex VP sentence, as in (1) below, seems more acceptable than the corresponding sentence with a simple VP:

- (1) Espen har spørsmålet plaget i årevis
Eспен has question-DEF bothered for years
'Eспен, the question has bothered for years'

A related question pertains to sentences with equally animate arguments.

8.2.2 Specificity

Our results for the effect of definiteness on the realization and positioning of the subject and object have been far less conclusive than those pertaining to animacy. In particular, our data set included a relatively high percentage of indefinite subjects (11.3% of all the subjects). We have seen that indefinite subjects are ungrammatical in several languages and subject to strict interpretational restrictions in others (Aissen, 2003). In particular they often incur a specific reading, rendering them as strong indefinites, rather than weak. Aissen (2003) operates with a further differentiation within the category of indefinites, namely that of Specific vs. Non-Specific Indefinite. Might it be that a further distinction within the group of indefinites will provide us with a clearer picture of the effects of definiteness? As mentioned, earlier, it seems difficult to obtain a weak reading of an indefinite subject when in a subject-initial sentence. Part of the reason for this, is obviously that the subject is the default topic, thus to a greater extent assumed to be given in some sense. We have also seen that topicalization rates are lower in sentences containing an indefinite subject. It might be that when the object topicalizes, the weak reading for the indefinite subject is more available due to the loss of topicality to the object. A resistance towards topicalization might therefore be part of a more general strategy for avoiding weak, indefinite subjects, a generalization in line with the markedness relations expressed by harmonic alignment. A notion of specificity, however, is quite context-dependent, and therefore difficult to work with. In our data material, only the context of the immediate sentence was included. It might therefore be necessary to extend the data material in order to gain insight into the above aspects of definiteness as a conditioning factor on the realization and expression of syntactic functions in Norwegian.

8.2.3 Other topics

At the end of chapters 6 and 7, outlines for further work on the OT analysis and the Constraint Grammar application were presented.

At the end of chapter 6, we looked at how our stochastic OT analysis might be incorporated within a bidirectional OT framework. We saw how our analysis might figure in the comprehension-directed optimization process in this approach. Due to the recentness of the stochastic implementation of bidirectional OT, however, this provides an interesting topic for future work.

As outlined at the end of chapter 7, there are several topics in need of further work within the Constraint Grammar application presented there. Firstly, our set of rules is limited to only the most general patterns of transitive sentences. An extension of the rule set should be a topic for further work. Another main challenge with regards to the Constraint Grammar application resides in the limitations of the information regarding the relative animacy of word forms. It might be interesting to look into methods for extending the animacy sets of lexical forms. Finally, the enhanced system should be tested further on larger sets of unknown texts.

Appendix A

OT Grammar and Distributions

The grammar input for the stochastic analysis, as well as the pair distribution file is provided below. For additional information on the workings of the Praat software the reader is referred to the Praat tutorial (Boersma, 1999).

A.1 OT Grammar

The OT grammar supplied to the software is provided below. All 8 constraints are given the same initial value (100) which is an arbitrary value. The algorithm is also set to respect the rankings of the subhierarchies following from harmonic alignment, as well as the rankings resulting from local conjunction. As we remember, a locally conjoined constraint is always ranked above its conjuncts. Finally 12 tableaux are provided, one for each input, specifying the violations incurred by every candidate (0=no violation, 1=violation). The order of the violations are the same as the constraints are given in, so it is important that these match.

```
File type = "ooTextFile"  
Object class = "OTGrammar"
```

```
8 constraints
```

```
constraint [1]: "*S<0(Anim) & Canon" 100 100 ! *S<0(Anim) & Canon  
constraint [2]: "S0-Primacy(Def)" 100 100 ! S0-Primacy(Def)  
constraint [3]: "Canon" 100 100 ! Canon  
constraint [4]: "Top-L" 100 100 ! Top-L  
constraint [5]: "*S<0(Anim)" 100 100 ! *S<0(Anim)  
constraint [6]: "*S=0(Anim)" 100 100 ! *S=0(Anim)  
constraint [7]: "*S>0(Anim)" 100 100 ! *S>0(Anim)  
constraint [8]: "*S=0(Anim) & S0-Primacy(Def)" 100 100 !*S=0(Anim) & S0-Primacy(Def)
```

```
6 fixed rankings
```

```
1 3
```

1 5
8 6
8 2
5 6
6 7

12 tableaux

```
input [1]: "V(Anim/EqDef/Top, Inan/EqDef)" 4
  candidate [1]: "[S_Anim/EqDef/Top V O_Inan/EqDef]" 0 1 0 0 0 0 1 0
  candidate [2]: "[S_Inan/EqDef V O_Anim/EqDef/Top]" 0 1 0 1 1 0 0 0
  candidate [3]: "[O_Inan/EqDef V S_Anim/EqDef/Top]" 0 1 1 1 0 0 1 0
  candidate [4]: "[O_Anim/EqDef/Top V S_Inan/EqDef]" 1 1 1 0 1 0 0 0
input [2]: "V(Anim/HDef/Top, Inan/LDef)" 4
  candidate [1]: "[S_Anim/HDef/Top V O_Inan/LDef]" 0 0 0 0 0 0 1 0
  candidate [2]: "[S_Inan/LDef V O_Anim/HDef/Top]" 0 1 0 1 1 0 0 0
  candidate [3]: "[O_Inan/LDef V S_Anim/HDef/Top]" 0 0 1 1 0 0 1 0
  candidate [4]: "[O_Anim/HDef/Top V S_Inan/LDef]" 1 1 1 0 1 0 0 0
input [3]: "V(Anim/LDef/Top, Inan/HDef)" 4
  candidate [1]: "[S_Anim/LDef/Top V O_Inan/HDef]" 0 1 0 0 0 0 1 0
  candidate [2]: "[S_Inan/HDef V O_Anim/LDef/Top]" 0 0 0 1 1 0 0 0
  candidate [3]: "[O_Inan/HDef V S_Anim/LDef/Top]" 0 1 1 1 0 0 1 0
  candidate [4]: "[O_Anim/LDef/Top V S_Inan/HDef]" 1 0 1 0 1 0 0 0
input [4]: "V(Anim/EqDef, Inan/EqDef/Top)" 4
  candidate [1]: "[S_Anim/EqDef V O_Inan/EqDef/Top]" 0 0 0 1 0 0 1 0
  candidate [2]: "[S_Inan/EqDef/Top V O_Anim/EqDef]" 0 0 0 0 1 0 0 0
  candidate [3]: "[O_Inan/EqDef/Top V S_Anim/EqDef]" 0 0 1 0 0 0 1 0
  candidate [4]: "[O_Anim/EqDef V S_Inan/EqDef/Top]" 1 0 1 1 1 0 0 0
input [5]: "V(Anim/HDef, Inan/LDef/Top)" 4
  candidate [1]: "[S_Anim/HDef V O_Inan/LDef/Top]" 0 0 0 1 0 0 1 0
  candidate [2]: "[S_Inan/LDef/Top V O_Anim/HDef]" 0 1 0 0 1 0 0 0
  candidate [3]: "[O_Inan/LDef/Top V S_Anim/HDef]" 0 0 1 0 0 0 1 0
  candidate [4]: "[O_Anim/HDef V S_Inan/LDef/Top]" 1 1 1 1 1 0 0 0
input [6]: "V(Anim/LDef, Inan/HDef/Top)" 4
  candidate [1]: "[S_Anim/LDef V O_Inan/HDef/Top]" 0 0 0 1 0 0 1 0
  candidate [2]: "[S_Inan/HDef/Top V O_Anim/LDef]" 0 0 0 0 1 0 0 0
  candidate [3]: "[O_Inan/HDef/Top V S_Anim/LDef]" 0 0 1 0 0 0 1 0
  candidate [4]: "[O_Anim/LDef V S_Inan/HDef/Top]" 1 0 1 1 1 0 0 0
input [7]: "V(Anim/EqDef/Top, Anim/EqDef)" 4
  candidate [1]: "[S_Anim/EqDef/Top V O_Anim/EqDef]" 0 1 0 0 0 1 0 1
  candidate [2]: "[S_Anim/EqDef V O_Anim/EqDef/Top]" 0 1 0 1 0 1 0 1
  candidate [3]: "[O_Anim/EqDef V S_Anim/EqDef/Top]" 0 1 1 1 0 1 0 1
  candidate [4]: "[O_Anim/EqDef/Top V S_Anim/EqDef]" 0 1 1 0 0 1 0 1
input [8]: "V(Inan/EqDef/Top, Inan/EqDef)" 4
  candidate [1]: "[S_Inan/EqDef/Top V O_Inan/EqDef]" 0 1 0 0 0 1 0 1
  candidate [2]: "[S_Inan/EqDef V O_Inan/EqDef/Top]" 0 1 0 1 0 1 0 1
  candidate [3]: "[O_Inan/EqDef V S_Inan/EqDef/Top]" 0 1 1 1 0 1 0 1
  candidate [4]: "[O_Inan/EqDef/Top V S_Inan/EqDef]" 0 1 1 0 0 1 0 1
input [9]: "V(Anim/HDef/Top, Anim/LDef)" 4
  candidate [1]: "[S_Anim/HDef/Top V O_Anim/LDef]" 0 0 0 0 0 1 0 0
```



```

candidate [2]: "[S_Anim/LDef V O_Anim/HDef/Top]" 0 1 0 1 0 1 0 1
candidate [3]: "[O_Anim/LDef V S_Anim/HDef/Top]" 0 0 1 1 0 1 0 0
candidate [4]: "[O_Anim/HDef/Top V S_Anim/LDef]" 0 1 1 0 0 1 0 1
input [10]: "V(Inan/HDef/Top, Inan/LDef)" 4
candidate [1]: "[S_Inan/HDef/Top V O_Inan/LDef]" 0 0 0 0 0 1 0 0
candidate [2]: "[S_Inan/LDef V O_Inan/HDef/Top]" 0 1 0 1 0 1 0 1
candidate [3]: "[O_Inan/LDef V S_Inan/HDef/Top]" 0 0 1 1 0 1 0 0
candidate [4]: "[O_Inan/HDef/Top V S_Inan/LDef]" 0 1 1 0 0 1 0 1
input [11]: "V(Anim/HDef, Anim/LDef/Top)" 4
candidate [1]: "[S_Anim/HDef V O_Anim/LDef/Top]" 0 0 0 1 0 1 0 0
candidate [2]: "[S_Anim/LDef/Top V O_Anim/HDef]" 0 1 0 0 0 1 0 1
candidate [3]: "[O_Anim/LDef/Top V S_Anim/HDef]" 0 0 1 0 0 1 0 0
candidate [4]: "[O_Anim/HDef V S_Anim/LDef/Top]" 0 1 1 1 0 1 0 1
input [12]: "V(Inan/HDef, Inan/LDef/Top)" 4
candidate [1]: "[S_Inan/HDef V O_Inan/LDef/Top]" 0 0 0 1 0 1 0 0
candidate [2]: "[S_Inan/LDef/Top V O_Inan/HDef]" 0 1 0 0 0 1 0 1
candidate [3]: "[O_Inan/LDef/Top V S_Inan/HDef]" 0 0 1 0 0 1 0 0
candidate [4]: "[O_Inan/HDef V S_Inan/LDef/Top]" 0 1 1 1 0 1 0 1

```

A.2 Pair distributions

The pair distribution file is provided below. Note that for every input, each of the four candidates are provided with their frequency count from the data material.

```

"ooTextFile"
"PairDistribution"
48 pairs

"V(Anim/EqDef/Top, Inan/EqDef)" "[S_Anim/EqDef/Top V O_Inan/EqDef]" 136
"V(Anim/EqDef/Top, Inan/EqDef)" "[S_Inan/EqDef V O_Anim/EqDef/Top]" 0
"V(Anim/EqDef/Top, Inan/EqDef)" "[O_Inan/EqDef V S_Anim/EqDef/Top]" 0
"V(Anim/EqDef/Top, Inan/EqDef)" "[O_Anim/EqDef/Top V S_Inan/EqDef]" 0

"V(Anim/HDef/Top, Inan/LDef)" "[S_Anim/HDef/Top V O_Inan/LDef]" 315
"V(Anim/HDef/Top, Inan/LDef)" "[S_Inan/LDef V O_Anim/HDef/Top]" 0
"V(Anim/HDef/Top, Inan/LDef)" "[O_Inan/LDef V S_Anim/HDef/Top]" 0
"V(Anim/HDef/Top, Inan/LDef)" "[O_Anim/HDef/Top V S_Inan/LDef]" 0

"V(Anim/LDef/Top, Inan/HDef)" "[S_Anim/LDef/Top V O_Inan/HDef]" 75
"V(Anim/LDef/Top, Inan/HDef)" "[S_Inan/HDef V O_Anim/LDef/Top]" 0
"V(Anim/LDef/Top, Inan/HDef)" "[O_Inan/HDef V S_Anim/LDef/Top]" 0
"V(Anim/LDef/Top, Inan/HDef)" "[O_Anim/LDef/Top V S_Inan/HDef]" 0

"V(Anim/EqDef, Inan/EqDef/Top)" "[S_Anim/EqDef V O_Inan/EqDef/Top]" 0
"V(Anim/EqDef, Inan/EqDef/Top)" "[S_Inan/EqDef/Top V O_Anim/EqDef]" 7
"V(Anim/EqDef, Inan/EqDef/Top)" "[O_Inan/EqDef/Top V S_Anim/EqDef]" 15
"V(Anim/EqDef, Inan/EqDef/Top)" "[O_Anim/EqDef V S_Inan/EqDef/Top]" 0

```

"V(Anim/HDef, Inan/LDef/Top)" "[S_Anim/HDef V O_Inan/LDef/Top]" 0
 "V(Anim/HDef, Inan/LDef/Top)" "[S_Inan/LDef/Top V O_Anim/HDef]" 11
 "V(Anim/HDef, Inan/LDef/Top)" "[O_Inan/LDef/Top V S_Anim/HDef]" 27
 "V(Anim/HDef, Inan/LDef/Top)" "[O_Anim/HDef V S_Inan/LDef/Top]" 0

 "V(Anim/LDef, Inan/HDef/Top)" "[S_Anim/LDef V O_Inan/HDef/Top]" 0
 "V(Anim/LDef, Inan/HDef/Top)" "[S_Inan/HDef/Top V O_Anim/LDef]" 6
 "V(Anim/LDef, Inan/HDef/Top)" "[O_Inan/HDef/Top V S_Anim/LDef]" 47
 "V(Anim/LDef, Inan/HDef/Top)" "[O_Anim/LDef V S_Inan/HDef/Top]" 0

 "V(Anim/EqDef/Top, Anim/EqDef)" "[S_Anim/EqDef/Top V O_Anim/EqDef]" 28
 "V(Anim/EqDef/Top, Anim/EqDef)" "[S_Anim/EqDef V O_Anim/EqDef/Top]" 0
 "V(Anim/EqDef/Top, Anim/EqDef)" "[O_Anim/EqDef V S_Anim/EqDef/Top]" 0
 "V(Anim/EqDef/Top, Anim/EqDef)" "[O_Anim/EqDef/Top V S_Anim/EqDef]" 0

 "V(Inan/EqDef/Top, Inan/EqDef)" "[S_Inan/EqDef/Top V O_Inan/EqDef]" 108
 "V(Inan/EqDef/Top, Inan/EqDef)" "[S_Inan/EqDef V O_Inan/EqDef/Top]" 0
 "V(Inan/EqDef/Top, Inan/EqDef)" "[O_Inan/EqDef V S_Inan/EqDef/Top]" 0
 "V(Inan/EqDef/Top, Inan/EqDef)" "[O_Inan/EqDef/Top V S_Inan/EqDef]" 0

 "V(Anim/HDef/Top, Anim/LDef)" "[S_Anim/HDef/Top V O_Anim/LDef]" 30
 "V(Anim/HDef/Top, Anim/LDef)" "[S_Anim/LDef V O_Anim/HDef/Top]" 0
 "V(Anim/HDef/Top, Anim/LDef)" "[O_Anim/LDef V S_Anim/HDef/Top]" 0
 "V(Anim/HDef/Top, Anim/LDef)" "[O_Anim/HDef/Top V S_Anim/LDef]" 0

 "V(Inan/HDef/Top, Inan/LDef)" "[S_Inan/HDef/Top V O_Inan/LDef]" 145
 "V(Inan/HDef/Top, Inan/LDef)" "[S_Inan/LDef V O_Inan/HDef/Top]" 0
 "V(Inan/HDef/Top, Inan/LDef)" "[O_Inan/LDef V S_Inan/HDef/Top]" 0
 "V(Inan/HDef/Top, Inan/LDef)" "[O_Inan/HDef/Top V S_Inan/LDef]" 0

 "V(Anim/HDef, Anim/LDef/Top)" "[S_Anim/HDef V O_Anim/LDef/Top]" 0
 "V(Anim/HDef, Anim/LDef/Top)" "[S_Anim/LDef/Top V O_Anim/HDef]" 14
 "V(Anim/HDef, Anim/LDef/Top)" "[O_Anim/LDef/Top V S_Anim/HDef]" 2
 "V(Anim/HDef, Anim/LDef/Top)" "[O_Anim/HDef V S_Anim/LDef/Top]" 0

 "V(Inan/HDef, Inan/LDef/Top)" "[S_Inan/HDef V O_Inan/LDef/Top]" 0
 "V(Inan/HDef, Inan/LDef/Top)" "[S_Inan/LDef/Top V O_Inan/HDef]" 28
 "V(Inan/HDef, Inan/LDef/Top)" "[O_Inan/LDef/Top V S_Inan/HDef]" 3
 "V(Inan/HDef, Inan/LDef/Top)" "[O_Inan/HDef V S_Inan/LDef/Top]" 0

Appendix B

Constraint Grammar application

B.1 Constraint Grammar rules

The CG rules from chapter 6 are given below:

Rule 1: [**Subj**_{Animate} **Verb** **Obj**_{Animate/Inanimate}]
(@w = s! (@subj) ; select the subject
(0 lo#animat) ; word is animate
(NOT -1 komma) ; word to left is not a comma
(*1 lo#ikke-mod-@<p-utfyll-@adv-@app *R) ; find a non-modifier to the right
(LRO @fv *R) ; word is a finite verb
(NOT LRO lo#kop-pass *R) ; verb is not copular or passive
(*R lo#ikke-mod-@<p-utfyll-@adv-@app *R) ; find a non-modifier to the right
(LRO @obj *R) ; word is an object
(NOT LRO lo#mod-@adv-@<p-utfyll *R) ; word is not an adverbial etc.
(*R lo#ikke-mod-@<p-utfyll-@adv-@app *R) ; find a non-modifier to the right
(NOT LRO @iv) ; word is not an infinite verb
(*-1 setn-gr/verb/subj *R) ; find sent.boundary/verb/subj
(LRO lo#top-setn-gr)) ; word is a valid sentence boundary

Comment: An animate first argument induces SVO word order regardless of the animacy of the second argument. The verb phrase is simple, i.e. no infinite verb, and there are no verbs or subjects to the left of the subject.

Rule 2: [Obj_{Inanimate} Verb Subj_{Animate}]

```

(@w = s! (@subj) ; select the subject
(0 lo#animat) ; word is animate
(NOT 0 lo#mod-@adv-@<p-utfyll) ; word is not adverbial etc.
(*-1 lo#ikke-mod-@<p-utfyll-@adv-@app *L) ; find a non-modifier to the left
(LRO @fv *L) ; word is a finite verb
(NOT LRO lo#reverse-verb *L) ; word is not a reverse-animacy verb
(NOT LRO lo#kop-pass *L) ; word is not copular or passive
(NOT LRO lo#presentering-verb *L) ; word is not a presentational verb
(*L lo#ikke-mod-@<p-utfyll-@adv-@app *L) ; find a non-modifier to the left
(LRO @obj *L) ; word is an object
(NOT LRO lo#mod-@adv-@<p-utfyll *L) ; word is not an adverbial etc.
(LRO lo#inimat *L) ; word is inanimate
(*L setn-gr/verb/subj *L) ; find sent.boundary/verb/subj
(LRO lo#top-setn-gr) ; word is a valid sentence boundary
(*1 lo#ikke-mod-@<p-utfyll-@adv-@app *R) ; find a non-modifier to the right
(NOT LRO @iv)) ; word is not an infinite verb

```

Comment: An inanimate first argument and an animate second argument induces OVS word order *if* the verb is not a reverse-animacy verb or a verb that can occur in a presentational construction. The verb phrase is simple, i.e. no infinite verb, and there are no verbs or subjects to the left of the subject.

Rule 3: [Subj_{Animate} Verb Obj_{Inanimate}]

```

(@w = s! (@subj) ; select the subject
(0 prop) ; word is a proper noun
(NOT 0 lo#mod-@adv-@<p-utfyll) ; word is not an adverbial
(-1 @subst>) ; word to left is a modifying noun
(-1 lo#animat) ; word to left is animate
(-1 prop) ; word to left is proper noun
(*1 lo#ikke-mod-@<p-utfyll-@adv-@app *R) ; find a non-modifier to the right
(LRO @fv *R) ; word is a finite verb
(NOT LRO lo#kop-pass *R) ; verb is not copular or passive
(*R lo#ikke-mod-@<p-utfyll-@adv-@app *R) ; find a non-modifier to the right
(LRO @obj *R) ; word is an object
(NOT LRO lo#mod-@adv-@<p-utfyll *R) ; word is not an adverbial etc.
(*R lo#ikke-mod-@<p-utfyll-@adv-@app *R) ; find a non-modifier to the right
(NOT LRO @iv) ; word is not an infinite verb
(*-1 setn-gr/verb/subj *R) ; find sent.boundary/verb/subj
(LRO lo#top-setn-gr)) ; word is a valid sentence boundary

```

Comment: Unknown proper names inherit animacy from their modifier. Ex.:
Ole Mahatma@subj likes the car.

The verb phrase is simple, i.e. no infinite verb, and there are no verbs or subjects to the left of the subject.

Rule 4: [*Obj_{Inanimate} Verb Subj_{Animate}*]

(@w = s! (@subj)	; select the subject
(0 prop)	; word is a proper noun
(NOT 0 lo#mod-@adv-@<p-utfyll)	; word is not adverbial etc.
(-1 @subst>)	; word to left is a modifying noun
(-1 animate)	; word to left is animate
(-1 prop)	; word to left is proper noun
(*-2 lo#ikke-mod-@<p-utfyll-@adv-@app *L)	; find a non-modifier to the left
(LRO @fv *L)	; word is a finite verb
(NOT LRO lo#reverse-verb *L)	; word is not a reverse-animacy verb
(NOT LRO lo#kop-pass *L)	; word is not copular or passive
(*L lo#ikke-mod-@<p-utfyll-@adv-@app *L)	; find a non-modifier to the left
(LRO @obj *L)	; word is an object
(NOT LRO lo#mod-@adv-@<p-utfyll *L)	; word is not an adverbial etc.
(LRO lo#inanimat *L)	; word is inanimate
(*L setn-gr/verb/subj *L)	; find sent.boundary/verb/subj
(LRO lo#top-setn-gr)	; word is a valid sentence boundary
(*1 lo#ikke-mod-@<p-utfyll-@adv-@app *R)	; find a non-modifier to the right
(NOT LRO @iv))	; word is not an infinite verb

Comment: Unknown proper names inherit animacy from their modifying noun. OVS-version of rule 3. above. An inanimate first argument and an animate second argument induces OVS word order *if* the verb is not a reverse-animacy verb or a verb that can occur in a presentational construction. The verb phrase is simple, i.e. no infinite verb, and there are no verbs or subjects to the left of the subject.

Rule 5: [*Subj_{Inanimate} Verb_{reverse} Obj_{Animate}*]

(@w = s! (@subj)	; select the subject
(0 lo#inanimat)	; word is inanimate
(NOT 0 lo#mod-@adv-@<p-utfyll)	; word is not an adverbial
(NOT -1 komma)	; word to left is not a comma
(*1 lo#ikke-mod-@<p-utfyll-@adv-@app *R)	; find a non-modifier to the right
(LRO @fv *R)	; word is a finite verb
(LRO lo#reverse-verb *R)	; verb is a reverse-animacy verb
(*R lo#ikke-mod-@<p-utfyll-@adv-@app *R)	; find a non-modifier to the right
(LRO @obj *R)	; word is an object
(NOT LRO lo#mod-@adv-@<p-utfyll *R)	; word is not an adverbial etc.
(*R lo#ikke-mod-@<p-utfyll-@adv-@app *R)	; find a non-modifier to the right
(NOT LRO @iv)	; word is not an infinite verb
(*-1 setn-gr/verb/subj *R)	; find sent.boundary/verb/subj
(LRO lo#top-setn-gr))	; word is a valid sentence boundary

Comment: An inanimate first argument and an animate second argument induces SVO word order *if* the verb is a reverse-animacy verb. The verb phrase is simple, i.e. no infinite verb, and there are no verbs or subjects to the left of the subject.

Rule 6: [**Subj**_{Inanimate} **Verb**_{reverse} **Obj**_{Animate}]

```

(@w = s! (@subj) ; select the subject
(0 lo#inanimat) ; word is inanimate
(NOT 0 lo#mod-@adv-@<p-utfyll) ; word is not an adverbial
(NOT -1 komma) ; word to left is not a comma
(*1 lo#ikke-mod-@<p-utfyll-@adv-@app *R) ; find a non-modifier to the right
(LRO @fv *R) ; word is a finite verb
(LRO lo#reverse-verb *R) ; verb is a reverse-animacy verb
(*R lo#ikke-mod-@<p-utfyll-@adv-@app *R) ; find a non-modifier to the right
(LRO @obj *R) ; word is an object
(NOT LRO lo#mod-@adv-@<p-utfyll *R) ; word is not an adverbial etc.
(LRO prop L-1) ; word is a proper noun - 1
(L-1 @subst> *L) ; word is a modifying noun
(LRO lo#animate *R) ; word is animate
(LRO prop *R) ; word is proper noun
(*R lo#ikke-mod-@<p-utfyll-@adv-@app *R) ; find a non-modifier to the right
(*R lo#ikke-mod-@<p-utfyll-@adv-@app *R) ; find a non-modifier to the right
(NOT LRO @iv) ; word is not an infinite verb
(*-1 setn-gr/verb/subj *R) ; find sent.boundary/verb/subj
(LRO lo#top-setn-gr)) ; word is a valid sentence boundary

```

Comment: Unknown proper names inherit animacy from their modifying nouns. An inanimate first argument and an animate second argument induces SVO word order *if* the verb is a reverse-animacy verb. The verb phrase is simple, i.e. no infinite verb, and there are no verbs or subjects to the left of the subject.

Rule 7: [**Subj**_{Inanimate} **Verb**_{presentation} **Obj**_{Animate}]

```

(@w = s! (@subj) ; select the subject
(0 %det%-pron) ; word is det 'it'
(NOT 0 lo#mod-@adv-@<p-utfyll) ; word is not an adverbial
(*1 lo#ikke-mod-@<p-utfyll-@adv-@app *R) ; find a non-modifier to the right
(LRO @fv *R) ; word is a finite verb
(LRO lo#presentering-verb *R) ; verb is a presentational verb
(*R lo#ikke-mod-@<p-utfyll-@adv-@app *R) ; find a non-modifier to the right
(LRO @obj *R) ; word is an object
(LRO lo#animat *R) ; word is animate
(LRO ub *R) ; word is indefinite
(NOT LRO lo#mod-@adv-@<p-utfyll *R) ; word is not an adverbial etc.
(*R lo#ikke-mod-@<p-utfyll-@adv-@app *R) ; find a non-modifier to the right
(NOT LRO @iv) ; word is not an infinite verb
(*-1 setn-gr/verb/subj *R) ; find sent.boundary/verb/subj
(LRO lo#top-setn-gr)) ; word is a valid sentence boundary

```

Comment: An inanimate first argument and an animate second argument induces SVO word order *if* the verb is a the kind of verb that may occur in the presentational construction. The verb phrase is simple, i.e. no infinite verb, and there are no verbs or subjects to the left of the subject.

Rule 8: [*Subj*_{Inanimate} *Verb* *Obj*_{Inanimate}]

```

(@w = s! (@subj) ; select the subject
  (0 lo#inanimat) ; word is inanimate
  (NOT 0 lo#mod-@adv-@<p-utfyll) ; word to left is not an adverbial etc.
  (NOT -1 lo#animat) ; word to left is not animate
  (*1 lo#ikke-mod-@<p-utfyll-@adv-@app *R) ; find a non-modifier to the right
  (LRO @fv *R) ; word is a finite verb
  (NOT LRO lo#kop-pass *R) ; verb is not copular or passive
  (*R lo#ikke-mod-@<p-utfyll-@adv-@app *R) ; find a non-modifier to the right
  (LRO @obj *R) ; word is an object
  (NOT LRO lo#mod-@adv-@<p-utfyll *R) ; word is not an adverbial etc.
  (*R lo#ikke-mod-@<p-utfyll-@adv-@app *R) ; find a non-modifier to the right
  (NOT LRO @iv) ; word is not an infinite verb
  (*-1 setn-gr/verb/subj *R) ; find sent.boundary/verb/subj
  (LRO lo#top-setn-gr)) ; word is a valid sentence boundary

```

Comment: Two inanimate arguments induce SVO word order. The verb phrase is simple, i.e. no infinite verb, and there are no verbs or subjects to the left of the subject.

B.2 Constraint Grammar sets

The following sets were employed in the CG rules:

%det%-pron (“det” pron))

@fv (@fv)

@iv (@iv)

@obj (@obj)

@subj (@subj)

@subst> (@subst>)

komma (“\$,”)

lo#animat ("øyenvitne" "øyensten" "øyenlege" "østtysker" "østerriker" "ørn" "økonom" "ærfugl" "æresmedlem" "æresborger" "århusianer" "åndemaner" "zulu" "zolog" "yppersteprest" "væpner" "voldsmann" "voldsforbryter" "vitne" "viten-skapsmann" "visesanger" "visepresident" "viseborgermester" "viseadmiral" "vipe" "vinner" "vinhandler" "ving" "vinduspusser" "vinbonde" "villdyr" "villand" "viking" "vikar" "vietnameser" "vi" "vever" "veterinær" "vest-tysker" "vestlending" "vestjyde" "vertshuseier" "vertinne" "vert" "verksmester" "verge" "verdensmester" "venninne" "venn" "vendelbo" "velgjører" "velger" "vekker" "veimuseum" "veileder" "vegetarianer" "vedlikeholdssjef" "vasall" "vaktmester" "vaktmann" "vakt" "vagabond" "vadefugl" "uvenn" "utøver" "utvandrer" "utvalgsformann" "utskudd" "utnytter" "utleier" "utkaster" "ut-giver" "utfordrer" "utenriksminister" "utbytter" "utbyder" "urostifter" "ur-maker" "universitetslærer" "ungkar" "unge" "ungdom" "ungarer" "undulat")

"undervisningsminister" "undertrykker" "undertegnede" "undersått" "underdirektør" "ukrainer" "ugle" "tømmermester" "tømmermann" "tyv" "tysker" "tyrker" "tyrann" "typograf" "tvillingsøster" "tvillingbror" "tvilling" "tviler" "tv-titter" "turistsjef" "turist" "turgåer" "tuareg" "tsjekker" "tsar" "trøster" "tryllekunstner" "tronfølger" "trompetist" "trommeslager" "trollmann" "tri-umfator" "trilling" "trener" "tremening" "trellkvinne" "trell" "trekkfugl" "trekkdyr" "transvestitt" "trane" "trafikant" "toppsjef" "toppscorer" "toppolitiker" "toppleder" "toller" "tolk" "togfører" "tobakkshandler" "tjenestepike" "tjenestemann" "tjenerskap" "tjener" "tilskuer" "tillitsvalgt" "tillitsmann" "tilhører" "tilflytter" "tilbeder" "tigger" "thailender" "terrorist" "terne" "terapeut" "teoretiker" "teolog" "tenorsaksofonist" "tenor" "tennisspiller" "tenker" "telegrafist" "tekstforfatter" "tekniker" "tegner" "teatergjenger" "taper" "tan-tebarn" "tante" "tanntekniker" "tannpleier" "tannlege" "tamil" "taler" "talepedagog" "talentspeider" "taktiker" "søvnjenger" "søster" "søskenbarn" "søsken" "sørjyde" "søramerikaner" "sørafrikaner" "søppelmann" "sønnesønn" "sønn" "sølvsmed" "særling" "sæddonor" "syrer" "synder" "syndebukk" "sympa-tisør" "syklist" "sykkelrytter" "sykkelreparatør" "sykkelhandler" "sykepleierske" "sykepleier" "syerske" "svømmer" "svoger" "svindler" "svigersøster" "sviger-sønn" "svigermor" "svigerinne" "svigerfar" "svigerdatter" "svigerbror" "svenske" "svenn" "sveitser" "sveiser" "svane" "svale" "suppleant" "sultan" "støtte-pedagog" "stær" "styrmann" "styremedlem" "stuepike" "studine" "student" "stråmann" "stripper" "streikevakt" "streber" "strateg" "straffange" "stormester" "stork" "storfavoritt" "storesøster" "storebror" "stoffmisbruker" "stifter" "stesønn" "stemor" "stemor" "steinhugger" "stefar" "stedfortreder" "stedatter" "ste-barn" "statssjef" "statsrevisor" "statsoverhode" "statsminister" "statsborger" "statsadvokat" "statistiker" "statist" "stat" "stasoverhode" "stasjonsmester" "stamkunde" "stamgjest" "stamfar" "stabssjef" "spøkefugl" "spåmann" "spåkone" "spurvefugl" "spurv" "sprøytenarkoman" "språkforsker" "språkbruker" "sprinter" "sponsor" "spion" "spiller" "spillemann" "spesialist" "spekulant" "speider" "sparringspartner" "sparer" "spanjol" "sovjetborger" "sosiolog" "sosialrådgiver" "sosialpedagog" "sosialminister" "sosialarbeider" "sopran" "sommerfugl" "somalier" "solodanser" "soldat" "sogneprest" "snekermester" "snekker" "snegl" "småborger" "smugler" "smeller" "smed" "slåskjempe" "slektning" "slavinne" "slave" "slaktermester" "slakter" "skøyteforbund" "skytter" "skyld-ner" "skulptør" "skuespillerinne" "skuespiller" "skrytepave" "skribent" "skred-der" "skotte" "skorsteinsfeier" "skomaker" "skolevaktmester" "skolepsykolog" "skolelærer" "skolelege" "skoleinspektør" "skoleelev" "skoledirektør" "skogsar-beider" "skogdue" "skjørtejeger" "skjøge" "skjære" "skipsreder" "skipsfører" "skipsbygger" "skipper" "skiløper" "skilsmissebarn" "skilpadde" "skeptiker" "skatteyder" "skandinav" "skald" "sjørøver" "sjømann" "sjøhelt" "sjåfør" "sjefsredaktør" "sjef" "sikkerhetsfolk" "sidemann" "sidekamerat" "sheriff" "sheik" "sexolog" "servitise" "serber" "sensor" "senator" "selvmorder" "sel-ger" "sekundant" "sekretær" "sekretær" "sekretariatssjef" "seiler" "seier-herre" "seer" "scenograf" "sangfugl" "sangerinne" "sanger" "samler" "sam-ferdselsminister" "saksøker" "saksofonist" "saksfører" "sadist" "røyker" "røykdykker" "røver" "rømling" "rådgiver" "råd" "rytter" "russer" "rumener" "rovfugl" "romer" "roer" "rockesanger" "rival" "riksråd" "riksadvokat" "revisor" "restau-ratør" "resipient" "reservelege" "resepsjonist" "representant" "reporter" "rengjøringsas-sistent" "rektor" "rekrutt" "reisende" "reiseleder" "reingjeter" "reineier"

"regjering" "regissør" "regionsykehus" "regent" "redningsmann" "reder" "redaktør" "redaksjonssekretær" "ravn" "pårørende" "påfugl" "pyskoanalytiker" "pyroman" "pusher" "publikummer" "psykoterapeut" "psykiater" "prøysser" "prost" "prosessfullmektig" "promotor" "proletar" "programmerer" "profet" "professor" "produsent" "privatperson" "privatdetektiv" "prinsesse" "prins" "prest" "president" "predikant" "pottemaker" "postmester" "portør" "portugiser" "portier" "politimester" "politimann" "politiker" "politikammer" "politibetjent" "politiassistent" "politi" "polfarer" "polakk" "poet" "pleier" "plattenslager" "plageånd" "pilot" "pilkaster" "pianist" "personalsjef" "perser" "pendler" "pelikan" "pedagog" "pave" "patriark" "pastor" "passasjer" "pasje" "pasient" "partner" "partifelle" "part" "pariser" "pappa" "papegøye" "panser" "panser" "pamp" "palestiner" "pakistaner" "oversykepleierske" "oversetter" "overlærer" "overlege" "overhode" "outsider" "orrfugl" "ornitolog" "orlogskaptein" "organist" "organisator" "organisasjon" "ordstyrer" "orakel" "optiker" "oppvasker" "oppleser" "oppkjøper" "oppfinner" "oppdretter" "operatør" "operasanger" "onkel" "ombudsmann" "oljeselskap" "oldemor" "oldefar" "oldebarn" "offiser" "observatør" "oberstløytnant" "oberst" "nyhetsoppleser" "nybegynner" "normanner" "nordmann" "nordamerikaner" "noen" "niese" "nevø" "nederlender" "navigatør" "naturlege" "naturforsker" "nattevakt" "nattergal" "narr" "narkoman" "nabo" "møller" "møll" "møbelarkitekt" "måke" "myndighet" "mygg" "musvåk" "muslim" "musikkpedagog" "musiker" "musikant" "museumsinspektør" "museumsbetjent" "murersvenn" "murermester" "murer" "munk" "motstander" "moteskaper" "moster" "mormor" "morfar" "morder" "mor" "montør" "mongol" "modell" "misjonær" "minoritet" "ministerpresident" "minister" "militærneker" "mexicaner" "meteorolog" "metallarbeider" "mester" "menneske" "mellomste" "melkemann" "mekaniker" "meis" "meieri" "megler" "medisinmann" "medisiner" "medhjelper" "medarbeider" "matematiker" "matador" "massør" "maskinsjef" "maskinmester" "maskiningeniør" "maskinarbeider" "marokkaner" "marketingsjef" "marinesoldat" "marineoffiser" "mannekeng" "mann" "mange" "man" "mamma" "malersvenn" "malermester" "maler" "makedoner" "major" "magister" "lærling" "lærerinne" "lærer" "låtskriver" "lyriker" "lobbyist" "livvakt" "livredder" "litterat" "litauer" "linjemann" "lillesøster" "lillebror" "likemann" "libaneser" "lerke" "lektor" "leietager" "leiesoldat" "legionær" "legevakt" "legesekretær" "lege" "leder" "lastebilsjåfør" "langturssjåfør" "landslagsspiller" "landsdommer" "landbruker" "landarbeider" "laborant" "københavnner" "kvinne" "kuwaiter" "kusk" "kusine" "kursdeltaker" "kurer" "kurder" "kurator" "kunstner" "kunstmaler" "kunsthåndverker" "kunsthistoriker" "kunsthandler" "kunde" "kulturminister" "ku" "krypskytter" "krykkje" "kronikør" "kroat" "kritiker" "kriminalassistent" "kriger" "kremmer" "kranfører" "kosmonaut" "korsanger" "korespondent" "korporal" "koreograf" "koreaner" "koordinator" "kontrollør" "kontraktspiller" "kontorsjef" "kontorist" "kontorassistent" "kontaktperson" "kontaktmann" "konsulent" "konsul" "konstruktør" "konservator" "konsertmester" "konkurrent" "konge" "konfirmant" "konferansier" "kone" "konduktør" "konditor" "komtesse" "komponist" "kommunestyremedlem" "kommune" "kommunaldirektør" "kommentator" "kommandør" "komiker" "kokkepik" "kokk" "knivstikker" "klubb" "klokker" "klinikkassistent" "klient" "klatrer" "klasselærer" "klager" "kjøttmeis" "kjørelærer" "kjøpmann" "kjøper" "kjøkkensjef" "kjerring" "kjemiker" "kjemiiingeniør" "kirurg" "kiropraktor" "kirkem-

inister "kinogjenger" "kineser" "kidnapper" "keramiker" "kelner" "keiserinne" "keiser" "keeper" "kavaler" "kar" "kaptein" "kantor" "kamerat" "journalist" "jordmor" "jente" "jeg" "jarl" "italiener" "israeler" "islending" "isbjørn" "institusjon" "insekt" "indremisjonsselskap" "idrettsmann" "høne" "hvem" "hustru" "husokkupant" "hun" "hane" "han" "hallodame" "hakkespett" "gårdeier" "gutt" "grønnsakshandler" "gråspurv" "greve" "greker" "gjenbo" "genetiker" "generalsekretær" "generalmajor" "generalløytnant" "generaldirektør" "general" "gendarm" "gast" "gartner" "gardist" "førstekandidat" "fører" "fødselshjelper" "færøying" "fysiker" "fyrvokter" "fyrdirektør" "fyllik" "fyllekjører" "fylkesskolesjef" "funksjonær" "fullmektig" "fugl" "froskemann" "fritidspedagog" "frisør" "franskmann" "fotomodell" "fotohandler" "fotograf" "fotogalleri" "fotgjenger" "fotballspiller" "fostersønn" "fostermor" "fosterfar" "fosterdatter" "forvaltningssjef" "forvalter" "forsvarsminister" "forsvarer" "forstanderinne" "forstander" "forsker" "forsanger" "forretningsmann" "forretningsfører" "formann" "forliksmann" "forlegger" "forfatterinne" "forfatter" "forelder" "forbundskansler" "forbundsformann" "forbrukerombudsmann" "folketingsmedlem" "folketingsmann" "folketingsmann" "folkeskolelærer" "folk" "flyttemann" "flyktning" "flyger" "flue" "flamingo" "fjortis" "fiskeørn" "fiskeskipper" "fiskeriminister" "fisker" "fiskehandler" "fiolinist" "finne" "fink" "finansminister" "filolog" "filmskuespiller" "filmprodusent" "filminstruktør" "fiende" "fetter" "fergemann" "fengselsbetjent" "feltherre" "faster" "fasan" "farmor" "farmer" "farmasøyt" "fargehandler" "farfar" "farbror" "far" "fantast" "fangevokter" "fan" "familiemedlem" "familie" "falk" "fagmann" "fagforeningsleder" "fagfolk" "fagfelle" "fadder" "fabrikkarbeider" "eventyrer" "europæer" "etterlikner" "etnograf" "eskimo" "erobrere" "erkebisp" "erkebiskop" "ergoterapeut" "entusiast" "entreprenør" "entertainer" "enkemann" "enkefrue" "enkedronning" "enke" "englender" "enehersker" "eneboligeier" "emigrant" "embetsmann" "elskerinne" "elsker" "elev" "elektriker" "ektepar" "ektemann" "ektefelle" "ekstremist" "ekspert" "ekspedisjonssjef" "egyptere" "dørvakt" "døgnflue" "dyr" "du" "drømmer" "drosjesjåfør" "drunker" "dramatiker" "donor" "domprost" "dommer" "doktor" "dikter" "diktator" "designer" "designer" "departementssjef" "departement" "demonstrant" "deltager" "dekksgutt" "dekan" "degn" "debutant" "datterdatter" "datter" "danske" "damefrisør" "dame" "dagsommerfugl" "dagmamma" "dagleier" "cubaner" "cowboy" "cellist" "canadier" "bøddel" "bærer" "båtsmann" "bysbarn" "byråd" "byrettsdommer" "byplanlegger" "byggmester" "byggherre" "bydel" "byboer" "bruden" "bror" "brite" "brevvenn" "brevskriver" "brevdue" "brems" "brasilianer" "brannmann" "branninspektør" "botaniker" "borger" "bordkavaler" "borddame" "bondemann" "bonde" "bokbinder" "bodybuilder" "bløtdyr" "blåmeis" "blotter" "blomsterhandler" "bisp" "biskop" "biperson" "biolog" "biokjemiker" "billedkunstner" "billedhugger" "bille" "bilist" "bileier" "bie" "bidragsyter" "bibliotekar" "beundrer" "betjent" "bestyrer" "bestemor" "bestefar" "beslutningstaker" "beskytter" "belgier" "bekjent" "beiler" "begge" "beboer" "bass" "baryton" "bartender" "baronesse" "baron" "barnelege" "barnehagepedagog" "barnebokforfatter" "barnebarn" "barn" "baker" "avtaker" "avsender" "avdelingssykepleierske" "avdelingssykepleier" "atlet" "astronom" "astronaut" "astrolog" "assistent" "arkitekt" "argentinere" "arbeider" "annen" "anmelder" "anleggsgartner" "angriper" "anfører" "andunge" "analysator" "amme" "amerikaner" "alpinist" "alkoholiker" "agent"

"afrikaner" "afghaner" "advokat" "administrator" "adjunkt" "adelsmann"
 "abonnet" "Wangerud" "Vinmonopolkøere" "Vesten" "USA" "Toten" "Torkill"
 "Smirnov" "Skane" "Schipper" "Regjeringen" "Readyjente" "Readyjente"
 "Readygutt" "Politiforbund" "Plesner" "Pilkaster" "Palacio" "Oslo" "Nato"
 "NATO" "Matthew" "Matthaus" "Mannesmann" "Lykambes" "Likudreg-
 jeringen" "Lawrence" "Kors" "Killengreen" "Justisdepartementet" "Jarvoll"
 "Indremisjonsselskapet" "Hemingkjørerne" "Heie" "Gudmunsson" "Gudbrands-
 dalen" "Group" "Gregard" "Gomanbakeri" "Giga" "Folkpartiet" "Faneb-
 ust" "Elsero" "Elgin" "Djurhuus" "Clinton" "Carney" "Byutviklingskomite-
 teen" "Byutviklingskomiteen" "Breiby" "Blumenthal" "Beth" "Bergelevne"
 "Baste" "Alpinanlegg" "Akershus" "AS" "AS" "46åringen" "10åringen" <per-
 son> <org> ("dere" nom) ("ingen" mask) ("ingen" fem) ("de" pron pers)
 ("all" fl) ("en" pron) ("undertegne" fl adjj)

lo#ikke-mod-@<p-utfyll-@adv-@app (@iv @fv @subj @obj @i-obj @s-pred
 @o-pred @loes-np @<sbu @<sbu-rel @s-gr clb inf-merke sbu)

lo#inanimat ("øyenvipp" "øyenskygge" "øyenhule" "øyenbryn" "øyelokk" "øyekrok"
 "øyeple" "øye" "øy" "ørken" "øring" "øregang" "øreflipp" "øre" "øl" "øk-
 ning" "ødemark" "ødeleggelse" "æresbevisning" "årgangsvin" "ære" "årbok"
 "år" "åpningstale" "åpning" "åker" "åk" "Østen" "Økourt" "zoomobjektiv"
 "zloty" "yttervegg" "ytterdør" "yoghurt" "yen" "wobler" "wire" "wiener-
 brød" "whisky" "wc" "walkie-talkie" "væte" "væske" "værmelding" "værhane"
 "våtområde" "vulkan" "vott" "vom" "vollgrav" "vokal" "vogn" "vogge" "vodka"
 "vitaminpille" "vitamin" "visker" "viskelær" "visittkort" "visir" "viser" "vise"
 "virus" "virksomhet" "virkning" "vippe" "vinterhage" "vinstokk" "vinmark"
 "vinkart" "vinge" "vinflaske" "vinedikk" "vindusvisker" "vinduskarm" "vindu"
 "vindrue" "vindjakke" "vindeltrapp" "vinde" "vin" "villmark" "villa" "vik-
 tighet" "vikarbyrå" "vik" "vignett" "vifte" "vielsesattest" "videospiller" "video"
 "vest" "veske" "vertsland" "vers" "vern" "vermut" "verk" "verdipapir" "ver-
 denshjørne" "verdenschav" "verdensdel" "verdenscupseier" "verden" "verb"
 "veranda" "ventilator" "ventilasjonsanlegg" "ventil" "venteliste" "ventehall"
 "venstrehånd" "vennskapsby" "vene" "vending" "velling" "vell" "velferds-
 stat" "velferdsprogram" "vektskål" "vekst" "vekkerur" "veiledning" "veike"
 "veikart" "vei" "veggmaleri" "vegg" "vedtekt" "vedtak" "vedovn" "vederlag"
 "vatt" "vaskeri" "vaskepulver" "vaskemaskin" "vaselin" "vase" "varmeut-
 styr" "varmestue" "varmeapparat" "varetektsfengsling" "varedeklarasjon"
 "vannspeil" "vannseng" "vannrør" "vannpytt" "vannmelon" "vannløp" "vannlilje"
 "vannledning" "vannkran" "vannkant" "vannfarge" "vanndråpe" "vanndamp"
 "vannbakkels" "vann" "vandreklikt" "vandelsattest" "valutareserve" "val-
 uta" "vals" "valnøtt" "valmue" "valgresultat" "vaktpost" "vaksine" "vagina"
 "vadested" "uttalelse" "utstyr" "utstillingsvindu" "utstilling" "utspill" "utsmykn-
 ing" "utskrift" "utsikt" "utrustning" "utropstegn" "utreisetilalelse" "utløp"
 "utlån" "utlegg" "utkikkspost" "utkast" "utgivelse" "utgave" "utflod" "utet-
 thet" "utenriksdepartement" "utbetaling" "urverk" "urtete" "urtehage" "urt"
 "urskog" "urne" "urinrør" "urin" "unntak" "unnskyldning" "universitet"
 "univers" "uniform" "ungdomsskole" "ungdomsklubb" "ungdomsherberge"
 "ungdomsbolig" "undervisningsdepartement" "undertrøye" "undertittel" "un-
 dertekst" "underskriftsinnsamling" "underskrift" "underskjørt" "underret-
 ning" "undermunn" "underliv" "underleppe" "underkropp" "underkjeve" "un-
 derholdsbidrag" "underetasje" "underarm" "ullgarn" "uland" "ukvemsord")

"ukelønn" "ukeblad" "ugress" "tøyle" "tøy" "tørkle" "tørketrommel" "tørkesnor"
 "tønne" "tøffel" "tåteflaske" "tåspiss" "tåre" "tåkelur" "tåke" "tå" "tyveri-
 alarm" "tyvegods" "tyttbær" "tynntarm" "tykktarm" "tvil" "tverrligger"
 "tvangstrøye" "tut" "tusing" "tusenårsrike" "turistbyrå" "turbin" "turban"
 "tunnel" "tungmetall" "tunge" "tuner" "tulipan" "tue" "tråd" "tryne" "trykksverte"
 "trykksak" "trykknapp" "trykkluft" "trykkeri" "trygghetsfølelse" "trutmunn"
 "trussel" "truse" "trosse" "tronstol" "tromsøpalme" "trommel" "tromme-
 hinne" "tro" "trivallitteratur" "triumfbue" "trisse" "treetopp" "trestubbe"
 "trestamme" "tresnitt" "tresko" "treningsdrakt" "trekull" "trefot" "tredje-
 plass" "tredjeland" "tredemølle" "trebein" "tre" "trau" "trapp" "transport-
 bånd" "transformator" "tranebær" "tram" "traktement" "traktat" "trakt"
 "trafikkgrunnlag" "torv" "torso" "torpedo" "torn" "topplue" "topp" "tomt"
 "tommeltott" "tommelfinger" "tommel" "tomatpuré" "tomatpure" "tomat"
 "tolvfingertarm" "tobakk" "toast" "toalettbord" "toalett" "tjære" "tjukkmjølk"
 "tjor" "tjenestebolig" "tittel" "tissemann" "tippekupong" "tinnsoldat" "tin-
 ning" "ting" "tind" "timiankvist" "timeplan" "time" "tiltak" "tilståelse"
 "tilsetningsstoff" "tilpassing" "tillitt" "tillitserklæring" "tilleggsbevilling" "tilla-
 telse" "tilholdssted" "tilgodehavende" "tilfluktssted" "tilbaketrekning" "tidskrift"
 "tidsplan" "tid" "thriller" "testikkel" "testamente" "terte" "terrorbombe"
 "territorium" "terrin" "terreng" "terning" "termostat" "termos" "termorute"
 "terminal" "term" "teppe" "tennplugg" "tennisball" "tenning" "tempera-
 ment" "telstang" "telt" "telegrambyrå" "telegram" "telefonsvarer" "tele-
 fonnummer" "telefonkatalog" "telefon" "tekstil" "tekst" "tekopp" "teknikk"
 "teine" "tein" "tegning" "tegnestift" "tegniserie" "tegnefilm" "tegnebrett"
 "tegnebord" "tegneblokk" "tegn" "tegltak" "teglstein" "teater" "te" "tavle"
 "taustige" "tau" "tatovering" "tast" "tarmkanal" "tarm" "tapet" "tannstikker"
 "tannsett" "tannregulering" "tannpasta" "tannhjul" "tann" "tank" "tan-
 gent" "tampong" "tallerken" "tale" "takvindu" "takstein" "takrenne" "tabu-
 rett" "tablett" "søyle" "søppelpose" "søppelkasse" "søm" "sølvmedalje" "sølv-
 fat" "sædcelle" "sæd" "sår" "såpestykke" "såpeboble" "såpe" "såle" "sytråd"
 "system" "syrin" "syre" "syngespill" "synagoge" "symfoni" "symaskin" "syl-
 tetøy" "sylte" "sylinderlås" "sykehus" "sydpol" "sydhavsøy" "syden" "swing"
 "sweatshirt" "svømmebasseng" "svovlestikk" "sviske" "svingdør" "sverte"
 "svennebrev" "svelg" "svamp" "svalgang" "surmelksprodukt" "surdeig" "suppe"
 "supermarked" "sund" "sump" "sum" "sukkerrør" "sukkerroe" "sukkerlake"
 "sukkerert" "sugekopp" "substans" "støvsuger" "støvel" "støtte" "støttann"
 "støtfanger" "stålrør" "stålampe" "stål" "styrthjelm" "stylte" "stykke" "stut-
 teri" "stump" "stumfilm" "stue" "studenterlue" "stubbe" "strømpebånd"
 "strømpe" "stråhatt" "strå" "strupe" "struktur" "strekkode" "strek" "strede"
 "strandkant" "strand" "straffelov" "straff" "stortå" "stortingsdebatt" "stort-
 ingsbehandling" "storseil" "storby" "stoppeklokke" "stopp" "stolpe" "stol"
 "stokk" "stoff" "stivelse" "stilk" "stiletthæl" "stikling" "stikkpille" "stikkon-
 takt" "stikkelsbær" "stigtrinn" "stige" "stift" "stereoanlegg" "steppe" "stem-
 mebånd" "stellebord" "stekepanne" "stekegryte" "steinsetting" "steinfrukt"
 "stein" "steik" "stedsnavn" "sted" "stav" "statuett" "statue" "statsobli-
 gasjon" "stat" "stasjonsby" "stasjon" "start" "stang" "stall" "stakitt" "stake"
 "stag" "staffeli" "stadion" "stad" "stabel" "spørsmålstegn" "spørreskjema"
 "spøkelseshistorie" "spyttklyse" "sprøytemiddel" "springform" "springbrett"
 "sprengstoff" "sprengladning" "sprellemann" "sprekk" "sprayboks" "sporstoff"

"spore" "sponplate" "spon" "spole" "spjeld" "spisestue" "spisesal" "spiserør" "spiselig" "spisekart" "spisebord" "spinner" "spinn" "spindelveg" "spillolje" "spillemål" "spillekort" "spillebord" "spilleautomat" "spill" "spiker" "spenne" "spekesild" "spekepølse" "speilegg" "speil" "spaserdrakt" "spann" "spagnum" "sovesofa" "sovesal" "sovepose" "sovepille" "sovemiddel" "sosiallov" "sosialkontor" "sosialdepartement" "sopp" "sonate" "solkrem" "solfanger" "solcelle" "solbær" "sokk" "sogn" "software" "sofabord" "sofa" "snøscooter" "snømann" "snøklokke" "snøklokke" "snøfonn" "snøfugg" "snøball" "snute" "snurrebass" "snor" "snittsår" "snitt" "sneip" "snegl" "smørbrød" "smør" "småkake" "smykke" "smutthull" "smokk" "smoking" "sminke" "smilehull" "smie" "smekke" "smaragd" "smaksløk" "slør" "slåbrok" "slyngplante" "slynge" "slutt" "sluse" "slumkvarter" "slukt" "slott" "slips" "slimhinne" "slim" "slikkepinne" "sletteland" "slette" "slep" "slalåm" "slakteri" "slakterforretning" "slaktehus" "slagplan" "slagord" "slagmark" "slag" "skøytesport" "skøyteforbund" "skøyte" "skyttergrav" "skyteskive" "skyskraper" "skylapp" "skur" "skulptur" "skulderveske" "skulderblad" "skulder" "skuespill" "skråning" "skrå" "skrue" "skrivepult" "skrivemaskin" "skrivebord" "skriv" "skrin" "skrifttegn" "skrift" "skrent" "skranke" "skramme" "skolebenk" "skole" "skogbunn" "skogbryn" "skog" "skoese" "sko" "skjøte" "skjøt" "skjørt" "skjønnlitteratur" "skjønnhetsflekk" "skjærgård" "skjærebrett" "skjær" "skjorteerme" "skjorte" "skjold" "skjermterminal" "skjerm" "skjerf" "skjenk" "skjema" "skjelett" "skjeggstubb" "skjegg" "skjede" "skinnklær" "skinnben" "skinn" "skinke" "skilt" "skille" "skildring" "ski" "skattekort" "skattekart" "skatoll" "skap" "skamleppe" "skalle" "skall" "skafott" "sjøkart" "sjøbredd" "sjokoladebolle" "sjokolade" "sjampinjong" "sjakkspill" "sjakkbrett" "situasjon" "sitron" "sitat" "sirene" "silo" "silketrykk" "silkepapir" "silke" "silikon" "sikte" "sikkerhetssele" "sikkerhetsnål" "sikkerhetslenke" "signal" "sigarkasse" "side" "show" "shampoo" "sett" "setning" "sete" "sesongbudsjett" "servitutt" "serviott" "service" "servant" "sennep" "senking" "sengeteppe" "seng" "sene" "sendrektighet" "sending" "senat" "seminar" "sement" "selvportrett" "selvbiografi" "selvangivelse" "selskapslokale" "seletøy" "sele" "seksjon" "sekretariat" "sekret" "sekk" "seil" "seddel" "scooter" "savanne" "saus" "sauna" "satsing" "satelitt" "sarkofag" "sangbok" "sang" "samme" "samling" "samfunn" "samarbeid" "salt" "sal" "sak" "sagn" "saga" "safir" "røyksky" "røyk" "rørsopp" "rørledning" "rør" "røntgenstråle" "røntgenbilde" "rødkål" "rødgran" "rødbete" "rævhull" "ræv" "råstoff" "råmateriale" "råk" "råd" "rynke" "ryggsøyle" "ryggstø" "ryggrad" "ryggene" "rygg" "rydning" "rutsjebane" "rutine" "ruteplan" "rute" "rustning" "rundstykke" "rundstokk" "rundskriv" "rundskriv" " rumpe" "rulling" "rulleskøyte" "rullepølse" "rullekake" "rullegardin" "rull" "rulett" "rugbrød" "rubin" "rot" "rosmarin" "rosenkål" "rose" "roman" "rom" "rolle" "robot" "robe" "rist" "ris" "ripsgele" "ringperm" "ring" "rille" "riksdag" "rike" "rigg" "rift" "rideskole" "riddersal" "ribbevegg" "ribbein" "returbillett" "rettsstat" "rettssal" "rettslokale" "rett" "restaurant" "respirator" "resept" "resepsjon" "republikk" "reol" "renseri" "renne" "renn" "remulade" "remedium" "relé" "reling" "rekvisitt" "reklamebyrå" "reklame" "rekkverk" "reisebyrå" "reisebeskrivelse" "reip" "reim" "regulerings sak" "reguleringsforslag" "regnskog" "regnfrakk" "regionsykehus" "region" "redningsvest" "redegjørelse" "reddik" "redaksjon" "realskole" "reaktor" "restaurant" "rapport" "ransel" "ramme" "raffinaderi" "radio" "radikal"

"radiator" "radar" "rabarbra" "pølse" "pæretre" "pære" "pære" "påvirkning" "påskrift" "påskelilje" "påskeegg" "påmelding" "pålegg" "påle" "påkledningsdukke" "påkledning" "påbygg" "pytt" "pyramide" "pynt" "putevar" "pute" "pustehull" "puss" "puslespill" "purre" "pupill" "pumpe" "pulver" "pult" "pulsåre" "pudder" "puck" "publikasjon" "pub" "prøveprosjekt" "prærie" "prydplante" "provinsby" "proton" "protokoll" "protokoll" "protese" "protein" "prosti" "prosjektil" "prosjekt" "prosessor" "prosent" "propp" "propell" "pronomener" "program" "prognose" "profil" "produkt" "produksjon" "problem" "privatskole" "prisme" "prisliste" "pris" "printer" "primærkommune" "presenning" "preparat" "premie" "prekestol" "preg" "praksis" "potteplante" "pote" "postkort" "postkontor" "postkasse" "posthus" "poster" "postei" "post" "posisjon" "pose" "portiere" "port" "porsjon" "porselen" "pornofilm" "pore" "polypp" "polstring" "politistasjon" "politirapport" "politikk" "politikammer" "politietterretning" "pokal" "poeng" "plombe" "plogfure" "plenklipper" "plen" "pleiehjem" "pledd" "platå" "platte" "platespiller" "plate" "plaster" "plassering" "plassbillett" "plass" "plasma" "plantevernmidde" "planteskole" "plante" "plantasje" "plansje" "plankegerde" "planke" "planetarium" "planet" "plan" "plakat" "pizza" "pitabrød" "pistasjenøtt" "pipe" "pinnebrød" "pinne" "pille" "piletre" "pilekvist" "pil" "pigment" "piggtråd" "piggsveis" "pigg" "pidestal" "pessar" "persienne" "permisjon" "perle" "pergola" "pergamentpapir" "pergament" "pepperrot" "pepperkake" "pensjonat" "penis" "pengestøtte" "penger" "penge" "pendel" "pekefinger" "peisestue" "peis" "pedal" "paviljong" "patte" "patron" "pastinakk" "pasta" "pass" "partitur" "parti" "parsell" "parlament" "parkeringsbot" "parfumeri" "parfyme" "parasoll" "parasitt" "paragraf" "paradis" "parabolantenne" "papir" "pantebrev" "panser" "pannekake" "panne" "panel" "pamflett" "palmeblad" "palme" "palett" "pai" "pagode" "paddehatt" "pacemaker" "p-pille" "ovnsform" "ovn" "oversikt" "oversettelse" "overmunn" "overligger" "overleppe" "overkropp" "overflate" "overenskomst" "overarm" "overall" "osteløype" "ost" "orkide" "organisasjon" "organ" "ordrebok" "ordkrig" "ordbok" "ord" "opus" "opptenningsved" "oppslagsverk" "oppslagstavle" "oppslag" "oppskrift" "oppskrift" "oppholdssted" "oppheng" "oppgradering" "oppgave" "oppgang" "oppfatning" "oppakning" "operasjonsstue" "operasjonsbord" "omstilling" "område" "omelett" "omegn" "oljemaling" "oljefyr" "olje" "oliven" "oliven" "oldsak" "oker" "odde" "odd" "observatorium" "obligasjon" "objektiv" "oase" "nøytron" "nøtteskall" "nøttekjerne" "nøtt" "nøkkerose" "nøkkelhull" "nødutgang" "nærlys" "næringsmelding" "næring" "nåløye" "nytteplante" "nytte" "nysgjerrighet" "nyre" "nummerskilt" "nummer" "novellesamling" "novelle" "nougat" "note" "notatblokk" "notat" "nord" "noe" "nobelpris" "nitrat" "nite" "nisje" "neve" "netthinne" "nesebor" "nese" "nervesystem" "nervepille" "nervemedisin" "nerve" "neon" "nellik" "nekrolog" "nek" "negl" "nedtegnelse" "nedløpsrør" "nedgang" "navneskilt" "navn" "navlestreng" "navle" "naturmedisin" "naturgass" "natur" "nattklubb" "nasjonalscene" "nasjonalsang" "nasjon" "narresmokk" "nakkestøtte" "nakke" "nagle" "nabokommune" "møydom" "møtested" "mønster" "møne" "møllpose" "mølle" "møblement" "møbel" "månedssblad" "måne" "myr" "mynte" "mynt" "mye" "mutter" "muslingskall" "muskulatur" "muskatt" "muskatnøtt" "muskatnøtt" "musikkstykke" "musikkskole" "musikkonservatorium" "murstein" "mur" "munnvik" "munning" "munnhule" "munn" "mumie" "mule" "mousse" "motor" "motiv" "motgift"

"motell" "mosebok" "morkake" "morene" "monument" "montasje" "monopol"
 "monogram" "monografi" "monitor" "molekyl" "modifikasjon" "missil" "mis-
 dannelse" "minnesmerke" "minkfarm" "mineral" "milt" "millionbeløp" "milepæl"
 "mile" "mikstur" "mikroprosessor" "mikroorganisme" "mikrobølgeovn" "midt-
 gang" "metalltråd" "metall" "merkevare" "merke" "merke" "meny" "mening"
 "mengde" "melon" "melon" "mellomrom" "mellomnavn" "melding" "meieri"
 "medlemskort" "medium" "medisterpølse" "medisin" "medikament" "med-
 delelse" "medaljong" "medalje" "matvare" "matte" "materie" "materiale"
 "mat" "masse" "massasjeinstitutt" "maskinpistol" "maskingevær" "mask-
 ineri" "maskin" "maskering" "mascara" "marsipanbrød" "marmelade" "markise"
 "mark" "marionett" "marinade" "margeritt" "margarin" "manual" "mani-
 fest" "mango" "mandel" "mandarin" "malm" "maling" "maleri" "makron"
 "make-up" "mahogni" "magnet" "magesekk" "mage" "løvtre" "løvskog" "løve-
 tann" "løpstikke" "løpeseddel" "løkke" "løk" "løgnhistorie" "løe" "lærreim"
 "læreplan" "lærebok" "lårbein" "lår" "lysstoffrør" "lysning" "lyske" "lyskaster"
 "lysingstid" "lysekrone" "lysbilde" "lys" "lyktestolpe" "lykt" "lykkehjul"
 "lydbånd" "lyd" "lunge" "lund" "luftvei" "luftrør" "luft" "lue" "lov" "lotto"
 "lotteri" "losji" "losje" "lommebok" "lokk" "lokale" "lokalavis" "loggbok"
 "loft" "lodd" "livmor" "liv" "litteratur" "liste" "lisse" "linoleum" "linde-
 tre" "lind" "liming" "lillefinger" "lilje" "lian" "levested" "leverpostei" "lever"
 "leseundersøkelse" "leserbrev" "lerret" "lerk" "leppetift" "leppe" "leopard-
 skinn" "lenke" "lenestol" "lend" "lem" "lekt" "leirbål" "leilighet" "leibolig"
 "legg" "legering" "legemsdel" "legemiddel" "legeattest" "ledningsnett" "led-
 ning" "leddbånd" "lavland" "lava" "lav" "laurbærblad" "laurbær" "land-
 sted" "landsdel" "land" "lampe" "lagerrom" "lagerplass" "lager" "labora-
 torium" "køkkultur" "kz-leir" "kvarter" "kvalme" "kurs" "kupong" "kun-
 stverk" "kunstmuseum" "kunst" "krypt" "krostue" "kropp" "krone" "kro"
 "kritikk" "krig" "krematorium" "krav" "kraft" "kostnad" "kort" "korridor"
 "kor" "kopp" "konversasjonsleksikon" "kontrakt" "kontor" "kontakt" "kon-
 sulat" "konstruksjon" "konsertsal" "konsentrasjonsleir" "konjakk" "konfer-
 anse" "komposisjon" "komponent" "kommunekontor" "kommune" "komman-
 dobro" "kombinert" "kok" "kode" "knekt" "knehasene" "kne" "kløver" "klippe"
 "klinikk" "kleshenger" "kjøttkake" "kjøkkenvifte" "kjøkkenhage" "kjole" "kjen-
 nelse" "kjeft" "kiwi" "kirsebær" "kirke" "kiosk" "kinn" "kilometer" "kasse"
 "kart" "kantine" "kantine" "kanal" "kalkulator" "kake" "kakao" "kafete-
 ria" "jur" "jorde" "jordbær" "jobb" "is" "inventar" "interesse" "institusjon"
 "installasjon" "inspirasjon" "inntrykk" "innslag" "innsjø" "initiativ" "illus-
 trasjon" "idrettshall" "høyhastighetsbane" "hår" "håp" "håndbok" "hånd-
 bak" "hånd" "hytte" "hybelkanin" "hvitløk" "hvite" "hva" "hus" "hurtigløpss-
 port" "hull" "hule" "hud" "hovedfagsoppgave" "hotell" "hospital" "hopp"
 "hop" "hofte" "hode" "hjul" "hjerte" "hjerne" "hjemstavn" "hjemland" "hjemby"
 "hjem" "hjelp" "historie" "hefte" "hede" "hav" "haug" "hatt" "hanske"
 "hank" "hane" "handlingsprogram" "halssmykke" "hals" "hall" "hake" "hag-
 tornhekk" "hagtorn" "grønnsak" "gryte" "gruppering" "gruppe" "grunnstoff"
 "grunn" "grotte" "griljermel" "gressløk" "grense" "grein" "grantre" "gran"
 "godteri" "glassplint" "glass" "gjenstand" "gjenlyd" "geysir" "genteknologi"
 "genser" "gemakk" "garderobe" "gang" "galleri" "første" "føflekk" "føde"
 "fyrstikkeske" "fyrstikk" "fyrrom" "fyr" "furutre" "frukt" "fritidsklubb" "fritid-
 shus" "fritidshjem" "fotsåle" "fotnote" "fot" "fossil" "foss" "forventning"

"forutsetning" "fortelling" "fortegnelse" "fortann" "forsvar" "forslag" "forskrift"
 "forsikringspolise" "forråd" "forretning" "forord" "formålsparagraf" "formu-
 lar" "formue" "forkle" "forkjølelsessår" "forhøyning" "forhåndsomtale" "forhold"
 "forfatning" "fordøyelseskanal" "fordypning" "forbundsstat" "forbundsrepub-
 likk" "forbud" "forbrenning" "forbein" "forandring" "folketing" "folkesagn"
 "folder" "foaje" "fnugg" "flodleie" "flodbredd" "flod" "flis" "flette" "flenge"
 "flekk" "flate" "flass" "flaske" "flanke" "fjør" "fjære" "fjærdrakt" "fjord"
 "fjellkjede" "fjell" "fiskesuppe" "fiskerett" "fiskekort" "fiskekake" "fiskefilet"
 "firkløver" "fippskjegg" "fiol" "finne" "fingerspiss" "finger" "filofax" "film"
 "fille" "fikentre" "fiber" "fettstoff" "fetaost" "feta" "festskrift" "festmåltid"
 "ferielandsby" "fennikel" "femøring" "femtedel" "femogtjueøring" "femmer"
 "feltseng" "feil" "feber" "fauna" "fatle" "fat" "fastland" "fasit" "fasade"
 "fargestoff" "fargefilm" "fang" "familieliv" "fallskjerm" "fakkel" "fajanse"
 "faglitteratur" "fagbok" "fagblad" "fabrikkhall" "fabrikk" "evighet" "even-
 tyr" "evangelium" "etikett" "eter" "etasje" "estragon" "esse" "espalier" "eske"
 "ert" "ernæring" "erme" "erklæring" "eple" "enzym" "entredør" "ensym"
 "enkeltbillett" "enhet" "engangsbeløp" "eng" "endringsforslag" "endring"
 "endetarm" "ende" "emblem" "emalje" "elv" "element" "elektron" "ekvator"
 "ekteseng" "ekstrakt" "einer" "einebær" "eiketre" "eik" "eiendom" "eien-
 del" "egn" "eggstokk" "eggleder" "eggeplomme" "eggevitestoff" "eggehvite"
 "eggcelle" "egg" "egenskap" "effekt" "edelmetall" "eddik" "edb-anlegg" "dørk-
 lokke" "dør" "dødsattest" "dødsannonse" "dåpsattest" "dypfryser" "dyng"
 "dynetrek" "dusør" "duell" "drønn" "drue" "drivhus" "dritt" "dressing"
 "dreieskive" "dreiebok" "drakt" "dragkamp" "drag" "dongeribukse" "domkirke"
 "domisil" "dollar" "dokument" "doktoravhandling" "dokke" "dobbeltseng"
 "dobbeltseier" "distrikt" "diskotek" "diskett" "direktørstol" "direktiv" "diplom"
 "dinar" "dimensjon" "dill" "diktsamling" "dikt" "digitalur" "dieselolje" "diesel-
 motor" "diamant" "dessert" "departement" "deodorant" "del" "deksel" "deklarasjon"
 "dataskjerm" "datamaskin" "dagpenger" "dampmaskin" "dam" "dalføre"
 "daler" "dagsavis" "daginstitusjon" "daghem" "dagbok" "container" "com-
 puter" "collage" "cocktail" "chutney" "chrysanthemum" "chemise" "cham-
 pagne" "celle" "carport" "cardigan" "bøtte" "bøketre" "bøkeskog" "bøk"
 "bærepose" "båtbrygge" "båre" "båndopptaker" "bål" "bygning" "bygging"
 "byggestein" "byggemateriale" "byggekloss" "byggeareal" "by" "butterdeig"
 "butik" "busk" "burgunder" "bunke" "bunke" "buljongtarning" "bukser"
 "brønn" "brødske" "brød" "bryst" "bruksrett" "brøse" "bronkie" "brokkoli"
 "brokke" "brokk" "broderi" "brisk" "bringebær" "bringe" "brilleglass" "brille"
 "brie" "brevsprekk" "brevkort" "brev" "brennetid" "brennesle" "bremse"
 "bregne" "brannså" "brannalarm" "bowler" "bot" "borg" "bordbein" "bord"
 "bomullsgarn" "bombe" "bolle" "bolig" "bokstav" "boks" "bok" "boikott"
 "bog" "bløtkake" "blåbær" "bluse" "blues" "blonde" "blomsterpotte" "blomkål"
 "blodåre" "blodkar" "blindtarm" "blikkboks" "blekk" "bleie" "blazer" "blan-
 kett" "blad" "bjørk" "bjerk" "bjelle" "bjelke" "bitter" "bit" "bistro" "biset-
 ning" "biografi" "binge" "bind" "bilradio" "billykt" "billett" "billedrør" "billed-
 bok" "bildekk" "bilde" "bilag" "bikini" "biff" "bibliotek" "bibliografi" "be-
 tydning" "betong" "betaling" "besvarelse" "bestselger" "bestemmelse" "be-
 standdel" "beskyldning" "beskjed" "beskadigelse" "bergflette" "bergart" "beretning"
 "bensinstasjon" "ben" "beløp" "beliggenhet" "bekledning" "bekkenbunn"
 "bekk" "bein" "beholder" "behandlingstilbud" "befalskole" "bedring" "bartre"

"bart" "barskog" "barnetilskudd" "barnehavetilbud" "barnehageplass" "barnebok"
 "barnebidrag" "barm" "bank" "banestrekning" "banan" "balle" "ball" "bag"
 "avvisning" "avtrekker" "avsnitt" "avkastning" "avis" "avhandling" "avfall-
 spose" "avdrag" "avdeling" "automat" "attest" "atomkjerne" "atom" "as-
 parges" "ask" "artikkel" "armbånd" "arm" "ark" "arbeidsplass" "arbeid"
 "appelsin" "anvisning" "antall" "annonse" "anleggsbrakke" "anlegg" "anemone"
 "andedam" "ammoniakk" "alter" "alt" "alpinrenn" "almetre" "alkohol" "al-
 falfa" "aldershjem" "albue" "akvakultur" "akuttavdeling" "aktstykke" "ak-
 tivum" "akademi" "afghaner" "adresse" "administrasjon" "adjektiv" "ace-
 ton" "Zimbabwe" "Zambia" "Zaire" "Yemen" "Yding" "Wien" "Værløse"
 "Vridsløselille" "Vridsløse" "Vordingborg" "Vojens" "Vogesene" "Vietnam"
 "Viborg" "Vestindia" "Vesten" "Vesten" "Vest-Tyskland" "Vest-Samoa" "Venezuela"
 "Venezia" "Vendsyssel" "Velferdsprogram" "Vejle" "Vejen" "Vatikanstaten"
 "Vatikanet" "Varde" "Vanuatu" "Vanløse" "Vallensbæk" "Valby" "Uruguay"
 "Uranus" "Ungarn" "Ukraina" "Uganda" "USA" "Tønder" "Tyskland" "Ty-
 rol" "Tyrkia" "Tuvalu" "Tunø" "Tunisia" "Tsjekkoslovakia" "Trinidad" "Toten"
 "Torshavn" "Tonga" "Togo" "Tibet" "Tiberen" "Thyborøn" "Thy" "Thurø"
 "Thule" "Thisted" "Themsen" "Thailand" "Tchad" "Tanzania" "Taiwan"
 "Sør-Yemen" "Sør-Jylland" "Sør-Afrika" "Sønderborg" "Sæby" "Syria" "Syd-
 polen" "Syd-Schleswig" "Syd-Korea" "Syd-Amerika" "Swaziland" "Sverige"
 "Svendborg" "Sveits" "Svartehavet" "Svaneke" "Sundeved" "Sudan" "Stubbekøbing"
 "Struer" "Storebelt" "Storbritannia" "Stor-København" "Stockholm" "Stille-
 havet" "Stevns" "Stege" "Sprogø" "Spania" "Sovjetunionen" "Sovjet" "Sorø"
 "Somalia" "Slovenia" "Slovakia" "Slagelse" "Skælskør" "Skåne" "Skottland"
 "Skjern" "Skive" "Skandinavia" "Skanderborg" "Skagerrak" "Skagen" "Sjæl-
 land" "Sjustjernen" "Singapore" "Sinai" "Silkeborg" "Sild" "Sicilia" "Sibir"
 "Shetlandsøyene" "Seychellene" "Serbia" "Senegal" "Seinen" "Schleswig"
 "Saturn" "Sardinia" "Sandvig" "Samsø" "Samoa" "Salomonøyene" "Salling"
 "Sakskøbing" "Rønne" "Rønne" "Rømø" "Rødehavet" "Rødby" "Rwanda"
 "Russland" "Rudkøbing" "Roskilde" "Rommeriket" "Romania" "Roma" "Ring-
 sted" "Ringkøbing" "Ribe" "Rhinen" "Regjeringen" "Randers" "Qatar" "Påskeøya"
 "Pyreneene" "Præstø" "Provence" "Preussen" "Praha" "Portugal" "Pom-
 mern" "Polynesia" "Politiforbund" "Polen" "Polarstjernen" "Pluto" "Peru"
 "Persia" "Peloponnes" "Pedersker" "Paris" "Paraguay" "Panama" "Palestina"
 "Pakistan" "Ostindia" "Oslo" "Oslo" "Orkenøyene" "Ordrup" "Oman" "Ol-
 jedirektiv" "Odense" "Odder" "Oceania" "OLbilletter" "OL-billetter" "Nør-
 resundby" "Næstved" "Nysted" "Nykøbing" "Nyborg" "Nubia" "Norske-
 havet" "Norge" "Nordsjøen" "Nordkapp" "Nordkalotten" "Nord-Yemen" "Nord-
 Schleswig" "Nord-Korea" "Nord-Jylland" "Nord-Amerika" "Nigeria" "Niger"
 "Nicaragua" "Nibe" "Neptun" "Nepal" "Neksø" "Nederland" "Nato" "Napoli"
 "Namibia" "Nakskov" "NATO" "Møn" "Moskva" "Mosambik" "Mors" "Mon-
 golia" "Monaco" "Milano" "Midt-Østen" "Middelhavet" "Mexico" "Merkur"
 "Mellom-Europa" "Mellom-Amerika" "Melkeveien" "Melanesia" "Maureta-
 nia" "Marstal" "Marokko" "Marina" "Maribo" "Mariager" "Mandø" "Mand-
 sjuria" "Malta" "Mallorca" "Mali" "Malaysia" "Malaya" "Malawi" "Make-
 donia" "Løgstør" "Limfjorden" "Lillebelt" "Likudregjeringen" "Kypros" "Kors"
 "Karibien" "Kamerun" "Justisdepartementet" "Indremisjonsselskapet" "Hed-
 mark" "Gudbrandsdalen" "Gudbrandsdalen" "Group" "Gomanbakeri" "Giga"
 "Fossaken" "Folkpartiet" "Flensburg" "Flandern" "Finland" "Filippinene"

"Fiji" "Femø" "Femern" "Farum" "Fanø" "Falster" "Fakse" "EØSavtalen"
 "Europa" "Eurasia" "Etruria" "Etiopia" "Estland" "Esrum" "Esbjerg" "Eng-
 land" "Elsero" "Elsass" "Elfenbenskysten" "Elben" "Drejø" "Dragør" "Dolomit-
 tene" "Ditmarsken" "Damaskus" "Cuba" "Colombia" "Canada" "Californi-
 a" "Byutviklingskomiteen" "Bretagne" "Brasil" "Bolivia" "Bogense" "Bloks-
 berg" "Berlin" "Bergen" "Belgia" "Atlanterhavet" "Atlanterhavet" "Athen"
 "Argentina" "Antillene" "Angola" "Andalusia" "Amerika" "Alpinanlegg"
 "Akershus" "Afrika" "Afghanistan" "Adriaterhavet" "Abessinia" "AS" <org>
 ("denne" pron pers) ("disse" pron pers) ("all" noeyt ent) ("det" pron pers)
 ("god" adj sup be) ("hel" adj pos be ent) ("dette" pron pers) ("den" pron
 pers))

lo#kop-pass ("være" "virke" "hete" "bli" pass)

lo#mod-@adv-@<p-utfyll ("seg" "at" @det> @<det @adj> @adv> @<adv
 @subst> @<subst @kon @<p-utfyll @adv @s-gr clb)

lo#presentering-verb (i2 <s-verb>)

lo#reverse-verb ("ødelegge" "vettskremme" "vedrøre" "vedkomme" "vedgå" "varme"
 "uroe" "tynge" "tilstøte" "tilkomme" "tilhøre" "tilfredsstille" "tilfalle" "tilin-
 tetgjøre" "terge" "særprege" "særmerke" "såre" "symbolisere" "støte" "støkke"
 "stigmatisere" "speile" "smitte" "smigre" "smerte" "smadre" "skyldes" "skuffe"
 "skremme" "skake" "skade" "sjokkere" "sjenere" "sjarmere" "seigpine" "sat-
 isfisere" "ryste" "romme" "reflektere" "påhvile" "pryde" "prege" "plage"
 "plage" "pirre" "pine" "overvelde" "overskygge" "overrisle" "oppta" "op-
 prøre" "oppirre" "oppfisse" "omslynge" "omslutte" "omhandle" "omgi" "om-
 fatte" "nage" "more" "mishage" "mette" "mette" "lindre" "lamslå" "knuse"
 "knekke" "kle" "kjede" "kile" "irritere" "involvere" "interessere" "innramme"
 "inneholde" "innebære" "innbefatte" "incitere" "henrykke" "henrive" "hen-
 føre" "helbrede" "harme" "grip" "glede" "gjennomsyre" "gjennomstrømme"
 "gjelde" "gagne" "frustrere" "forvirre" "forurolige" "forulempe" "foruleilige"
 "fortørne" "forskrekke" "fornøye" "fornærme" "forfjamsse" "forferde" "for-
 bløffe" "forbause" "forarge" "fjetre" "favne" "fascinere" "erstatte" "ergre"
 "enervere" "ekvivalere" "egge" "dupere" "drepe" "destruere" "desillusjonere"
 "deprimere" "blidgjøre" "bevege" "beta" "beskadige" "berøre" "beruse" "berolige"
 "berettigede" "bekymre" "bekomme" "behage" "begeistre" "bedøve" "bedåre"
 "bedrøve" "avkjøle" "avspeile" "atskille" "angå" "affisere")

lo#top-setn-gr (clb <<<<>>> ("at" sbu) <strek>)

prop (prop)

setn-gr/verb/subj ("– – ””-" @subj verb ("at" sbu) @sbu >> << (": "clb)(";"
 clb) (clb konj) (clb <komma>))

Bibliography

- Judith Aissen. Markedness and Subject choice in Optimality Theory. *Natural Language and Linguistic Theory*, 17:673–711, 1999.
- Judith Aissen. Differential Object Marking: Iconicity vs. economy. *Natural Language and Linguistic Theory*, 21:435–483, 2003.
- Arto Anttila. How to recognize subjects in English. In Fred Karlsson, Atro Voutilainen, Juha Heikkilä, and Atro Anttila, editors, *Constraint Grammar: A language-independent system for parsing unrestricted text*, pages 315–358. Mouton de Gruyter, 1995.
- Jennifer E. Arnold, Ryan Ginstrom, Anthony Lasongo, and Thomas Wasow. Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, 76(1):28–55, 2000.
- Douglas B. Bloom. Case syncretism and word order freezing in the Russian language. Master’s thesis, Stanford University, 1999.
- Paul Boersma. Optimality-theoretic learning in the Praat program. *IFA Proceedings*, 23:17–35, 1999.
- Paul Boersma and Bruce Hayes. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry*, 32(1):45–86, 2001.
- Joan Bresnan. *Lexical-Functional Syntax*. Blackwell Textbooks in Linguistics, 2001.
- Joan Bresnan. Lecture notes on stochastic generalization. From course web page: <http://www.phil-fak.uni-duesseldorf.de/summerschool2002/CDV/CDAissen.htm>, 2002.
- Joan Bresnan and Tatiana Nikitina. On the gradience of the dative alternation. ms. - downloaded from <http://www-lfg.stanford.edu/bresnan/download.html>, May 2003.
- Andrew Carnie. *An introductory textbook for syntax*. Blackwell Publishers, 2001.

- Hye-Won Choi. *Optimizing structure in context: Scrambling and information structure*. PhD thesis, Stanford, 1999.
- Bernard Comrie. *Language Universals and Linguistic Typology*. University of Chicago Press, 1989.
- William Croft. *Typology and Universals*. Cambridge University Press, 1990.
- William Croft. *Typology and Universals*. Cambridge University Press, 2nd edition, 2003.
- Östen Dahl. Egocentricity in discourse and syntax. ms. - downloaded from <http://www.ling.su.se/staff/oesten/papers/dlpapers.htm>, 1997.
- Östen Dahl and Kari Fraurud. Animacy in grammar and discourse. In Thorstein Fretheim and Jeanette K. Gundel, editors, *Reference and referent accessibility*, pages 47–65. John Benjamins Publishing Company, 1996.
- Shipra Dingare. The effect of feature hierarchies on frequencies of passivization in English. Master’s thesis, Stanford University, August 2001.
- Cathryn Donohue. Optimizing Fore case and word order. ms. - downloaded from <http://www.stanford.edu/donohue>, 1999.
- David Dowty. Thematic proto-roles and argument selection. *Language*, 67 (3):547–619, 1991.
- Elisabet Engdahl. Relative clause extractions in context. *Working Papers in Scandinavian Syntax*, 60:51–79, 1997.
- John Enrico. Word order, focus and topic in Haida. *International Journal of American Linguistics*, 52(2):91–123, 1986.
- Jan Terje Faarlund, Svein Lie, and Kjell Ivar Vannebo. *Norsk Referansegrammatikk*. Universitetsforlaget, 1997.
- Dan Fass. Metonymy and metaphor: What’s the difference? In *Proceedings of the 12th International Conference on Computational Linguistics*, pages 177–181, 1988.
- Talmy Givón. Topic, pronoun and grammatical agreement. In Charles N. Li, editor, *Subject and Topic*, pages 149–188. Academic Press, 1976.
- Jane Grimshaw. *Argument Structure*. The MIT Press, 1990.
- Gerhard Jäger. Learning constraint sub-hierarchies: The bidirectional Gradual Learning Algorithm. In R. Blutner and H. Zeevat (to appear), editors, *Pragmatics in OT*. Palgrave MacMillan, 2003.

- Janne Bondi Johannessen. Tagging and the case of pronouns. *Computers and the Humanities*, 32:1–38, 1998.
- René Kager. *Optimality Theory*, chapter 1. Cambridge University Press, 1999.
- Fred Karlsson. The formalism and environment of Constraint Grammar parsing. In Fred Karlsson, Atro Voutilainen, Juha Heikkilä, and Atro Anttila, editors, *Constraint Grammar: A language-independent system for parsing unrestricted text*, pages 315–358. Mouton de Gruyter, 1995.
- Fred Karlsson, Atro Voutilainen, Juha Heikkilä, and Atro Anttila, editors. *Constraint Grammar: A language-independent system for parsing unrestricted text*. Mouton de Gruyter, 1995.
- Edward L. Keenan. Towards a universal definition of “subject”. In Charles N. Li, editor, *Subject and Topic*, pages 303–333. Academic Press, 1976.
- George Lakoff and Mark Johnson. *Metaphors we live by*. The University of Chicago Press, 1980.
- Hanjung Lee. Crosslinguistic variation in argument expression and intralinguistic freezing effects. In Tania Ionin, Heejeong Ko, and Andrew Nevins, editors, *MIT Working Papers in Linguistics*, volume 43, pages 103–122. MIT Linguistics dept., 2002a.
- Hanjung Lee. Prominence mismatch and markedness reduction in word order. *Natural Language and Linguistic Theory*, 2002b.
- Hanjung Lee. Referential accessibility and stylistic variation. In Mary Anagnostopoulou et al., editor, *Proceedings of CLS 38*. Chicago Linguistic Society, 2002c.
- Beth Levin. *English Verb Classes and Alternations*. The University of Chicago Press, 1993.
- Beth Levin and Malka Rappaport Hovav. *Unaccusativity: At the syntactic-lexical semantics interface*. MIT Press, 1995.
- Helge Lødrup. The realization of benefactives in Norwegian. In *Papers from the 31. regional meeting of the CLS*, pages 317–328. Chicago Linguistic Society, 1995.
- Helge Lødrup. Exceptions to the Norwegian passive: Unaccusativity, aspect and thematic roles. *Norsk Lingvistisk Tidsskrift*, 2000.
- Cristopher Lyons. *Definiteness*. Cambridge Textbooks in Linguistics. Cambridge University Press, 1999.

- Christopher D. Manning. Probabilistic syntax. In Rens Bod, Jennifer Hay, and Stefanie Jannedy, editors, *Probabilistic Linguistics*, pages 289–341. MIT Press, 2003.
- Katja Markert and Malvina Nissim. Towards a corpus annotated for metonymies: The case of location names. In *Proceedings of the 3rd international Conference on Language Resources and Evaluation (LREC 2002)*, 2002.
- Line Hove Mikkelsen. Reanalyzing the definiteness effect: Evidence from Danish. *Working Papers in Scandinavian Syntax*, 69:1–75, 2002.
- Yukiko Morimoto. "Crash vs. Yield: On the conflict asymmetry in syntax and phonology". ms.- downloaded from <http://www.yukiko-morimoto.de>, 2000.
- Torbjørn Nordgård and Tor A. Åfarli. *Generativ syntaks: ei innføring via norsk*. Novus, 1990.
- John I. Saeed. *Semantics*. Blackwell Publishers, 1997.
- Peter Sells. *Structure, Alignment and Optimality in Swedish*. Stanford Monographs in Linguistics. CSLI Publications, 2001.
- Anna Siewierska. *Word Order Rules*. Croom Helm Linguistics Series. Croom Helm, 1988.
- David Stallard. Two kinds of metonymy. In *31st Annual Meeting of the Association for Computational Linguistics*, pages 87–94, 1993.
- Andreas Sveen. *Norwegian impersonal constructions and the unaccusative hypothesis*. PhD thesis, University of Oslo, 1996.
- Maria Wolters. Working with corpora. Lecture notes for ESSLLI 2002, Trento, 2002.