Disambiguation of syntactic functions in Norwegian: modeling variation in word order interpretations conditioned by animacy and definiteness *

Lilja Øvrelid
Göteborg University, Sweden
lilja.ovrelid@svenska.gu.se

Abstract

This paper deals with a functional ambiguity which obtains in simple transitive sentences in Norwegian and examines facets of argument realization apart from stricly structural ones. A corpus study of Norwegian simple transitives investigates the relative animacy and definiteness of the arguments in such constructions and the influence of these properties on word order variation. By emphasizing a notion of typological markedness, the observations from the corpus study may be modeled successfully within the framework of stochastic Optimality Theory, where variation between different interpretations of the ambiguous constructions result from different rankings of a set of universal markedness contraints, whose rankings are derived directly from the relevant data.

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 presented here. 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 SVO word order, whereas in (1b), we see a transitive sentence where the object has topicalized, giving us an OVS word order. The work presented here has originated from a problem of

*This paper is based on the work presented in my MA-thesis (Øvrelid, 2003). I would like to thank my two supervisors, Prof. Janne Bondi Johannessen and Prof. Helge Lødrup, for their insightful guidance during the writing of the thesis, as well as the participants at the 20th Scandinavian Conference of Linguistics in Helsinki, January 2004, and the participants at the 8th Workshop on Optimality-Theoretic Syntax in Berlin, Aug. 2004, for helpful comments.

automatic disambiguation of syntactic functions in Norwegian, and a need to discover facets of word order and argument realization apart from solely structural criteria. 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. However, with no overt case marking, the question of which factors that contribute to the disambiguation of the syntactic functions of subject and object in sentences like the above becomes a pertinent one, and 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.

We will start out by describing a corpus investigation of Norwegian transitives, which examines how two lexical properties internal to the noun phrase, namely animacy and definiteness, contribute towards the resolution of the aforementioned ambiguity, and do so in a way that reflects typological generalizations regarding argument realization. Following a wave of work expressing typological tendencies as a set of universal, ranked constraints, a small set of constraints which models the data at hand will be proposed. In particular, we will attempt to motivate the use of a stochastic grammar rather than a categorical one, in order to model the variation in interpretations of these types of sentences, and seeing that these typological principles obtain as strong statistical tendencies in Norwegian. Also, a gradience in grammaticality seems to be proportional with the degree of markedness associated with a certain construction along the dimensions of animacy and definiteness.

1.1 A note on Norwegian case and word order

There are a few properties of the expression of syntactic functions in Norwegian which contribute to the observed ambiguity. First of all, Norwegian makes very limited use of case. Only personal pronouns are marked for case, but even here syncretism and variation abounds (Faarlund et al., 1997). Also, nominative case is usually preferred when the pronoun is modified, regardless of syntactic function Johannessen (1998)[1]:

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

b. Dette gjelder i tillegg *de* som håndterer ..
this concerns in addition they-NOM who handle ..
'This also concerns those who handle..'

Norwegian is a V2-language, i.e. the finite verb is always the second constituent in declarative main clauses. The canonical word order in Norwegian is SVO, however, pretty much all types of constituents may, in principle, be sentence-initial. Through the process of topicalization constituents other than the subject may come to occupy sentence-initial position, thus enabling

[1]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.

the marked word order OVS. Due to all of these properties of Norwegian, then, the observed functional ambiguity obtains in simple transitive sentences[2].

1.2 Animacy and definiteness?

As mentioned above, the corpus study performed for Norwegian focuses on the influence of the relative animacy and definiteness of the arguments in these ambiguous structures. This might seem an odd choice, as animacy, in particular, has not recieved much mention in earlier linguistic literature when it comes to argument realization in Norwegian. In typological literature, prominence hierarchies expressing properties such as relative animacy or definiteness have been 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. The prominence hierarchies from Aissen (2003) are presented below:

Syntactic function scale: Subject $>$ Object

Animacy scale: Human $>$ Animate $>$ Inanimate

Definiteness scale: Personal Pronoun $>$ Proper Noun $>$ Definite NP $>$ Specific indefinite NP $>$ Non-specific indefinite NP

The prominence hierarchies presented above are often related by participation in different phenomena. 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:

> 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)

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

[2]Note that for transitive sentences containing a complex verb phrase, the structural positioning of the arguments indicate their syntactic function, hence there is no ambiguity.

(1) a. *Jenta* har skrevet *brevet*
girl-DEF has written letter-DEF
'The girl has written the letter'

b. *Brevet* har *jenta* skrevet
letter-DEF has girl-DEF written
'The letter, the girl has written'

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. So-called *freezing effects* on word order have been observed in several languages, mainly in two slightly different, but related circumstances:

1. When properties of the arguments are maximally marked in some sense(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)

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? These are some of the relevant questions as we now turn to look at the data.

2 A corpus study of Norwegian simple transitives

There is no corpus annotated for both animacy and definiteness available for Norwegian[3]. One thousand simple transitive sentences were therefore sampled randomly from the Oslo Corpus, a corpus of Norwegian texts of approximately 18,5 million words.[4] The corpus has been morphosyntactically annotated automatically, and the functionally ambiguous sentences were sampled using the following criteria:

- They were declarative main clauses with a transitive matrix verb
- They contained two arguments - subject and object - of the same verb, i.e. no raised arguments, topicalization from subordinate clause/prepositional modifier etc.
- They did not contain clausal arguments
- They did not contain reflexive objects
- Any pronouns were ambiguous pronouns, i.e. either pronouns that are syncretic across nominative and accusative case, or pronouns that were modified, hence stressed.

Some examples from this material include (3a), a SVO sentence and (3b), an object-initial sentence with a morphologically ambiguous pronominal subject following the finite verb:

(3) a. Knut Wickstrøm solgte huset på Seiersbjerget
Knut Wickstrøm sold house-DEF on Seiersbjerget
'Knut Wickstrøm sold the house on Seiersbjerget'

b. Boten på 2000 kroner betalte han uten å kny
fine-DEF of 2000 crowns he-NOM paid without to cry
'The fine of 2000 crowns, he paid without complaining'

[3]Dahl and Fraurud (1996), however, perform a similar study for Swedish, where they examine the effect of animacy on the syntactic distribution of grammatical relations in Swedish.

[4]The Oslo corpus is available at `http://www.hf.uio.no/tekstlab/`

The obtained data were annotated with the values from the animacy and definiteness scales given above, with the only modification that the specificity distinction was not maintained within the group of indefinite NPs[5].

2.1 Results

2.1.1 Animacy

The results for animacy are shown in table 1 below.[6] Dahl and Fraurud (1996) conclude that

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 1: Distribution of 1000 transitive clauses according to animacy of subject and direct object

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:

- 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%**

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, as in other languages[7].

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, and a relevant question then, pertains to whether these sentences have any common characteristics. One interesting generalization is that the

[5]Due to limits of time, the data sample contained only single sentences without their linguistic context, and the notion of specificity is highly context-dependent.

[6]Due 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.

[7]For instance, in an inverse language, like Navajo, the inverse, more marked form, expressed by the verbal affix *bi* must be employed when the subject is lower in animacy than the object, hence overtly marking the non-canonical argument pattern.

verbs in this small group of sentences are largely psych-verbs[8] and causative verbs, which all have the property of *causation* in common. According to Dowty (1991), causation is a crucial Proto-Agent property, which might clarify the mapping to syntactic functions in these cases, as exemplified by (4a) and (4b) below:

(4) a. Spørsmålet plager Espen
question-DEF bothers Espen
'The question bothers Espen'

b. Den nye ordningen rammer brukerne
the new arrangement-DEF affects users-DEF
'The new arrangement affects the users'

2.1.2 Definiteness

The results for definiteness are less conclusive than those for animacy, and are presented in table 2. Due to a more fine-grained scale we find a more scattered distribution of arguments

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

than for animacy. The picture becomes clearer when looking at the hierarchical relationship between the subject and object with regards to definiteness. There does seem to be a tendency towards avoiding a subject that is less definite than its object:

- Subject higher than or equal to object in definiteness: 823 sentences - 82.3%

[8]These verbs are characterized by taking Theme subjects and Experiencer objects, and commonly do not passivize in Norwegian (Lødrup, 2000).

- 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 topic tends to be given information, thus often definite in form, this tendency is not at all surprising.

Indefinite subjects 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). Our data sample includes 113 sentences with indefinite subjects. Quite a few of these, however, are clearly either specific, as in (5), or generic, as in (6):

(5) En nabo oppdaget innbruddstyven
a neighbour discovered burglar-DEF
'A neighbour discovered the burglar'

(6) 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'

Had we operated with the distinction in Aissen between specific and non-specific indefinites, or better even, weak vs. strong (Mikkelsen, 2002), the number of maximally marked subjects would certainly have been smaller. 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. A possible reason for the difficulty of ascertaining specificity 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. Norwegian also allows for several strategies for avoiding a weak indefinite subject, such as presentational constructions and clefts.

2.1.3 Animacy, definiteness and word order variation

We have seen that the arguments in the structurally ambiguous sentences seem largely to adhere to the markedness relations expressed by the interaction of the prominence scales above. We have also mentioned that when arguments are marked along one of these dimensions, i.e. an inanimate subject and an animate object, this may in many languages be expressed overtly through morphological marking or so-called word order freezing. A pertinent question at this point is therefore whether Norwegian word order variation between SVO and OVS is at all influenced by the relative animacy/definiteness of the arguments and if so, in what way?

In table 3 we see the distribution of the sentences in the sample according to word order pattern: The majority of sentences (approximately 90%) in the sample exhibit the canonical

Word order	N	%
SVO	913	91.3
OVS	97	9.7
Total	1000	100.0

Table 3: Distribution of 1000 transitive clauses according to word order

SVO word order, whereas around 10% front a topicalized object. We may attempt to answer

the question of the influence of argument animacy and definiteness on word order variation by looking at the sentences in the sample that exhibit OVS word order. Interestingly, we find that, among these, there is not a single sentence in the sample, where the object is higher in animacy than the subject. This finding points towards a possible freezing effect on word order which is sensitive to animacy i.e. when the subjects and objects are maximally marked along some dimension we find an avoidance of a marked (topicalized) word order. This is certainly the case for the sentences with inanimate subject and animate objects, where we find that all of these display canonical SVO word order. This is not to say that a topicalized version of these sentences would be strictly ungrammatical. Rather, the boundaries of grammaticality seem gradient with regards to topicalization in these sentences. The example in (7) illustrates this gradience, providing topicalized versions of the sentences in (4), i.e. a topicalized animate object and an inanimate subject[9]:

(7) a. ?? Espen plager spørsmålet
Espen bothers question-DEF
'Espen, the question bothers'

b. ?? Brukerne rammer den nye ordningen
users-DEF affects the new arrangement-DEF
'The users, the new arrangement affects'

These sentences are certainly in need of a much higher degree of contextual licensing, indicating a higher degree of markedness than sentences that are fully acceptable in isolation. Interestingly, a sentence like (7a) becomes much more acceptable when used with an accusative pronoun as topicalized object, thus unambiguously marking its grammatical function morphologically:

(8) Meg plager spørsmålet veldig
me bothers question-DEF very
'Me, the question bothers very much'

The questions of gradedness in grammaticality will not be pursued further in the present context, but will be left for future work. We simply note that there exists a possible relationship between the relative markedness of a construction and its grammaticality (or acceptability), however, in order to further investigate such a claim, methods other than corpus investigations will be a necessity[10].

Moving on to the definiteness of the arguments in the OVS sentences, we find that:

- 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%

The fact that definiteness does not seem to clearly differentiate the arguments in the OVS-sentences is really 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

[9]Note that these are constructed sentences created by varying the word order of the actual sentence from the data sample in (4).

[10]For instance, psycholinguistic experiments using so-called ERP-technology have examined the influence of animacy on processing effort (Weckerly and Kutas, 1999).

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, as in (9) below:

(9) a. *Det* lovet Leif
It promised Leif
'Leif promised it'

b. *Dette* sier en skuffet Miss Norway
this says a disappointed Miss Norway
'This is what a disappointed Miss Norway says'

To sum up, then, we have seen that for the sentences with non-canonical word order, arguments tend to be canonical or unmarked along the dimension of animacy. We have also related this observation to the phenomenon of word order freezing, as found in many languages in exactly this circumstance, expressing an avoidance of the "worst of the worst" (Lee, 2002b), or, a marked argument constellation in combination with a marked word order.

2.1.4 Equally animate arguments

We saw above that animacy, in particular, plays an important part in differentiating between the subject and the object in our structurally ambiguous transitive sentence. However, a good portion (36%) of the sentences in the sample contain arguments that are equally animate:

(10) a. Gro møtte PLO-lederen
Gro met PLO-leader-DEF
'Gro met the leader of PLO'

b. Alle vognene på Marina har slike hjul
all wagons-DEF on Marina have such wheels
'All the wagons on Marina have wheels like that'

There are obvious lexical differences between the verbs in these examples, (10a) contains a reciprocal verb whereas (10b) does not. However, if we focus only on animacy and definiteness, a relevant question might be whether definiteness figures in the disambiguation of these sentences? We find that for these sentences, subjects are higher than their objects in definiteness in 78% of the sentences. This actually constitutes a notable difference from the results for the sample as a whole, where this figure is around 50%.

With regards to word order, the sentences with equally animate arguments seem to resist topicalization at a significantly higher frequency than the sample as a whole (2.35% vs. 13.9% for sentences with arguments of unequal animacy). This is remniscient of the second circumstance in which word order freezing typically takes place, as mentioned above, when arguments are indistinguishable along some important disambiguating dimension. 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, in contrast to the rest of the sample, where the results for definiteness in OVS-sentence were around 50-50. The example sentence in (11) illustrates this; it contains a topicalized, indefinite object and a pronominal subject:

(11) Barnet bærer hun i sin egen kropp
child-DEF carries she-NOM in her own body
'The child, she carries in her own womb'

2.1.5 Result summary

The main findings from the corpus investigation of these sentences may be summarized under the following 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 constellations where the arguments are maximally marked is present in Norwegian as a strong statistical tendency.

Ambiguity and word order Sentences with equally animate arguments resist non-canonical word order. If they do topicalize, however, the arguments are disambiguated by definiteness.

Whereas it certainly is true that most constituents in Norwegian may topicalize, we see that they are not associated with the same probability. Sentences with a maximally marked 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.

3 Proposed Optimality Theory model

In the following section we will look at how the data presented above may be modeled by appealing to a small set of markedness constraints expressing typological generalizations. Optimality Theory (OT)[11] is in many ways an implementation of markedness and involves the competition between universal, violable constraints expressing markedness relations ranked language-specifically with respect to eachother. This means that Universal Grammar is thought to consist of these universal constraints, whereas language-specific variation is a product of varying rankings of these constraints across different languages.

The proposed model involves competition between different (functional) *interpretations* of the same surface string, so, input consists of information directly inferrable from a surface sentence, i.e. lexical properties inherent in the arguments, e.g. animacy and definiteness, as well as the ordering of the arguments within the sentence, schematically represented below:

Input: $[\mathrm{GF}_{1\mathit{Inan/Def}}\ \mathrm{V}\ \mathrm{GF}_{2\mathit{Anim/Indef}}]$

Candidate 1: $[_{\mathit{IP}}\ \mathrm{Subject}_{\mathit{Inan/Def}}\ [_{\mathit{I'}}\ \mathrm{V}\ [_{\mathit{VP}}\ \mathrm{Object}_{\mathit{Anim/Indef}}]]]$

Candidate 2: $[_{\mathit{CP}}\ \mathrm{Object}_{\mathit{Inan/Def}}\ [_{\mathit{C'}}\ \mathrm{V}\ [_{\mathit{IP}}\ \mathrm{Subject}_{\mathit{Anim/Indef}}]]]$

[11] See Kager (1999) for a good intrduction to Optimality Theory.

3.1 Constraints

Earlier 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. This is not an isolated tendency, but one that has categorical consequences in several languages. What we want to express is a constraint on the relative markedness of the arguments, i.e. the animacy of arguments in relation to eachother.

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[12]. 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. The harmonic alignment of the hierarchies from section 1, gives us the following constraint subhierarchies:

(12) a. *SU/INAN $\gg$ *SU/ANIM $\gg$ *SU/HUM

b. *OJ/HUM $\gg$ *OJ/ANIM $\gg$ *OJ/INAN

The constraint subhierarchy in (12a) above reads as follows: "an inanimate subject is more marked than an animate subject, which in turn is more marked than a human subject". In order to express the importance of the markedness relation that holds between the subject and object of a transitive sentence, we make use of the formal technique of *local conjunction* [13] between the harmonically aligned scales for grammatical functions and animacy (Aissen, 2003). The product partial order obtained is generalized in the subhierarchy in (13)(Morimoto, 2000), a ranking which is also supported by our data frequencies[14]:

(13) *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."

We have looked at the interplay between properties of the arguments in a transitive sentence and the preferred word order in the sentence. A markedness constraint like (14) expresses the unmarked, canonical order (Choi, 2001):

(14) CANON$_{GF}$ - Grammatical functions remain in their canonical positions

[12]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 ... > z on its elements. The harmonic alignment of D1 and D2 is the pair of harmony scales:
H_x: X/a > X/b > ... > X/z
H_y: Y/z > ... > Y/b > Y/a
The constraint alignment is the pair of constraint hierarchies:
C_x: *X/z $\gg$ ... $\gg$ *X/b $\gg$ *X/a
C_y: *Y/a $\gg$ *Y/b $\gg$ ... $\gg$ *Y/z
(Prince and Smolensky 1993, as quoted in Aissen (2003, p. 441))

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

[14]As a reminder, the following frequency distribution obtains in our data, S>O(Anim): 61.5%, S=O(Anim): 36.1% and S<O(Anim): 2.4%.

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 (15) below, with a topicalized adverbial, to violate CANON$_{GF}$:

(15) 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.

A hypothesized ranking supported by data frequencies, is obtained by interpolating CANON$_{GF}$ into the subhierarchy from (13):

(16) *S<O(ANIM) $\gg$ CANON$_{GF}$ $\gg$ *S=O(ANIM) $\gg$ *S>O(ANIM)

We may see the interaction between the proposed constraints in the example tableau in (17).

(17)

	[GF$_{1Inan}$ V GF$_{2Anim}$]	*S<O(ANIM)	CANON$_{GF}$	*S=O(ANIM)	*S>O(ANIM)
	a. [$_{IP}$ S$_{Inan}$ [$_{I'}$ V [$_{VP}$ O$_{Anim}$]]]	*!			
☞	b. [$_{CP}$ O$_{Inan}$ [$_{C'}$ V [$_{IP}$ S$_{Anim}$]]]		*		*

For the input of an inanimate argument followed by an animate argument, we see from the above tableau that the optimal candidate is candidate b., i.e. a construction where the inanimate argument is realized as a topicalized object and the animate argument as the subject. The reason for this is the high ranking of the constraint *S<O(ANIM) [15]. However, as we have seen in the data analysis, candidate a., with an inanimate subject and an animate object, should certainly also constitute a possibility, albeit a less frequent one. This is where stochastic OT comes in, as we shall see shortly.

Following from the tentative ranking of the constraints in (16), OVS order will be excluded completely as output for candidates where the arguments are of equal animacy, as we see from the tableau in (18):

[15]Constraints in an OT grammar are ranked in a hierarchy of dominance, related through *strict domination* (Kager, 1999, p. 22), where the seriousness of a violation, i.e. the ranking of the violated constraint(s), is crucial for the outcome of an evaluation.

(18)

	$[GF_{1Anim}$ V $GF_{2Anim}]$	*S<O(ANIM)	CANON$_{GF}$	*S=O(ANIM)	*S>O(ANIM)
☞	a. $[_{IP}$ S_{Anim} $[_{I'}$ V $[_{VP}$ $O_{Anim}]]]$			*	
	b. $[_{CP}$ O_{Anim} $[_{C'}$ V $[_{IP}$ $S_{Anim}]]]$		*!	*	

We saw earlier that the sentences with equally animate arguments often were asymmetric in definiteness in the manner predicted by the harmonic alignment of the scale for syntactic functions with the scale for definiteness employed here, i.e. the subject was more often higher than the object in definiteness in these sentences than for the sample as a whole. Inspired by Bresnan and Nikitina (2003) we may generalize over the picture painted by harmonic alignment as in the constraint in (19):

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

This constraint will, however, have to be ranked quite low, as the generalization does not hold for the data material as a whole. Only about 50% of the sentences in the data sample adhered to this principle. What we wish to express is that in sentences where the arguments are of equal animacy, definiteness plays a disambiguating role. 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):

(20) *S=O(ANIM) & SO-PRIMACY(DEF)

3.2 Stochastic OT

We want to model the observed ambiguity in the data material as a variation between two different interpretations. This variation should be conditioned by linguistic generalizations expressed as constraints and reflect the actual distributions found in the data.

An aspect of traditional OT 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. 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 fact, at

each evaluation the ranking of a constraint ranges over a normally distributed range of values, where the more stable ranking value is its mean. This crucial difference from traditional OT, introduces the stochasticity into the model and allows for the variation in constraint rankings, hence outputs.

In order to obtain the ranking values of our constraints we employed the Praat Software (Boersma, 1999), which is an implementation of the Gradual Learning Algorithm and stochastic Optimality Theory (Boersma and Hayes, 2001). In essence, the Praat software computes the rankings of constraints in a grammar supplied by the user, based on data also supplied by the user. Praat is equipped with a machine learning algoritm, the Gradual Learning Algorithm (Boersma and Hayes, 2001), which "tries to locate an empirically appropriate ranking value for each constraint" (Boersma and Hayes, 2001, p. 51) in an OT grammar. 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. After training on our corpus data, the constraints achieved the rankings depicted in table 4[16]. As we saw from the tableau in (17), for an input of an animate argument followed by

constraint	ranking value
*S<O(ANIM)	109.266
*CANON$_{GF}$	107.104
SO-PRIMACY(DEF) & *S=O(ANIM)	103.452
*S=O(ANIM)	84.340
*S>O(ANIM)	77.023
SO-PRIMACY(DEF)	-83.447

Table 4: Ranking values after training

an inanimate argument, the OVS version always became the optimal candidate 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 - the closer two constraints are ranked, the more chance of them varying. It turns out that the top-ranked of the animacy constraints, *S<O(ANIM), is ranked at approximately 2 units above CANON$_{GF}$, which gives these two constraints 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 subject and an animate object, as in (21). This interpretation would, however, be less probable, mirroring the observed distributions in the data set.

[16]Note that the numbers in themselves are not meaningful, rather it is the distance between them which depicts something about the variation found in the data.

(21)

	$[GF_{1Inan}$ V $GF_{2Anim}]$	CANON$_{GF}$	*S<O(ANIM)	*S=O(ANIM) & SO-PRIMACY(DEF)	*S=O(ANIM)	*S>O(ANIM)
☞	a. $[_{IP}$ S_{Inan} $[_{I'}$ V $[_{VP}$ $O_{Anim}]]]$		*			
	b. $[_{CP}$ O_{Inan} $[_{C'}$ V $[_{IP}$ $S_{Anim}]]]$	*!				*

The stochastic ranking in table 4 also predicts a variation in ranking between *S=O(ANIM) & SO-PRIMACY(DEF) and CANON$_{GF}$. As CANON$_{GF}$ is ranked highest, a majority of interpretations will be SVO, a property which correctly models the data. We saw earlier that sentences with equally animate arguments resisted topicalization at a higher frequency than the rest. However, an OVS interpretation may occur when the rankings of these two are reversed at a single evaluation. This will give us an OVS interpretation in sentences with equally animate arguments only when these are unmarked with regards to the dimension of definiteness, which is a welcome property of our grammar.

An example evaluation, where the input arguments are of equal animacy and the rankings are reversed, will give us the topicalized version as optimal, as we see in the tableau in 22:

(22)

	$[GF_{1Anim/Def}$ V $GF_{2Anim/Pro}]$	*S<O(ANIM)	*S=O(ANIM) & SO-PRIMACY(DEF)	CANON$_{GF}$	*S=O(ANIM)	*S>O(ANIM)
	a. $[_{IP}$ $S_{Anim/Def}$ $[_{I'}$ V $[_{VP}$ $O_{Anim/Pro}]]]$		*!		*	
☞	b. $[_{CP}$ $O_{Anim/Def}$ $[_{C'}$ V $[_{IP}$ $S_{Anim/Pro}]]]$			*	*	

We see, then, that the above stochastic ranking of our previously introduced constraints, models the functional ambiguity observed in the data and its resolution, by appealing to general

markedness constraints on argument realization in terms of animacy and definiteness.

4 Conclusion

Whereas it is clear that animacy and definiteness are not the only conditioning factors on the choice of interpretations in our transitive constructions, pragmatics, lexical semantics etc. clearly also contribute to this process, we have, by focusing on two simple lexical properties of the arguments, uncovered some interesting generalizations regarding Norwegian. Traditional grammars of Norwegian claim that pretty much any constituent may topicalize. Whereas this certainly is true in principle, our corpus study has indicated that topicalization is contingent on the relative markedness of the argument in question. By employing a stochastic grammar and a set of typologically motivated constraints, we have seen that word order variation is closely connected with a notion of typological markedness, which manifests itself as a strong statistical tendency in Norwegian. This observation requires a grammar which may deal with probabilities in an analysis of ambiguity and variation, and a model within the framework of stochastic Optimality Theory has been proposed.

References

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.

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. 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.

Hye-Won Choi. Phrase structure, information structure, and resolution of mismatch. In Peter Sells, editor, *Formal and Empirical Issues in Optimality Theoretic Syntax*, pages 17–62. CSLI Publications, 2001.

Bernard Comrie. *Language Universals and Linguistic Typology*. University of Chicago Press, 1989.

William Croft. *Typology and Universals*. Cambridge University Press, 1990.

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.

David Dowty. Thematic proto-roles and argument selection. *Language*, 67(3):547–619, 1991.

Jan Terje Faarlund, Svein Lie, and Kjell Ivar Vannebo. *Norsk Referansegrammatikk*. Universitetsforlaget, 1997.

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.

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.

Helge Lødrup. Exceptions to the Norwegian passive: Unaccusativity, aspect and thematic roles. *Norsk Lingvistisk Tidsskrift*, 2000.

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.

Lilja Øvrelid. Subject or object? Syntactic disambiguation in Norwegian, stochastic Optimality Theory and application in an automatic system. Master's thesis, University of Oslo, Norway, November 2003.

Jill Weckerly and Marta Kutas. An electrophysiological analysis of animacy effects in the processing of object relative sentences. *Psychophysiology*, 36(5):559–570, 1999.