Partial Word Order Freezing in Dutch

Gerlof Bouma & Petra Hendriks

G. J. Bouma

Department Linguistik, Universität Potsdam, Potsdam, Germany

tel: +49 331 977 2951

e-mail: gerlof.bouma@uni-potsdam.de

P. Hendriks

Center for Language and Cognition Groningen, University of Groningen, Groningen, The Netherlands

tel: +31 50 363 5863

e-mail: p.hendriks@rug.nl

Abstract   Dutch allows for variation as to whether the first position in the sentence is occupied by the subject or by some other constituent, such as the direct object. In particular situations, however, this commonly observed variation in word order is 'frozen' and only the subject appears in first position. We hypothesize that this partial freezing of word order in Dutch can be explained from the dependence of the speaker's choice of word order on the hearer's interpretation of this word order. A formal model of this interaction between the speaker's perspective and the hearer's perspective is presented in terms of bidirectional Optimality Theory. Empirical predictions of this model regarding the interaction between word order and definiteness are confirmed by a quantitative corpus study.

1 Introduction

Even languages with a relatively fixed word order, such as Dutch, allow for variation in word order. In Dutch, the first position in the sentence is usually occupied by the subject, as is illustrated by (1). However, in particular circumstances another constituent may come first in the sentence. Examples (2) and (3) show that the first position can also be occupied by a direct object or an indirect object, respectively.

(1)   Ik heb jou dat verteld.
        I have you that told
(2)   Dat heb ik jou verteld.
        that have I you told
(3)   Jou heb ik dat verteld.
        you have I that told
        'I have told you that'

Why does such variation occur? And what makes a speaker choose either of these forms? In general, there seem to be few restrictions on the occurrence of subjects in first position. The occurrence of other constituents in first position, in contrast, seems to be strongly restricted by context and other factors. However, very little empirical work has been carried out to investigate the exact nature of these restrictions (Jansen & Wijnands, 2004, is a notable exception). Empirical studies investigating word order variation in the postverbal domain in English, as with heavy NP shift or the dative alternation (e.g., Arnold et al., 2000; Bresnan et al., 2007), have shown that such word order variation is driven by factors like givenness, weight, pronominality, definiteness, and animacy. Furthermore, these studies have emphasized the need to consider actual language use rather than examples contrived by linguists. Only by considering actual language use in context using advanced statistical modeling techniques, these authors argue, is it possible to identify and disentangle the highly correlated and often gradient factors that influence language use. The present study takes a similar approach and investigates which factors influence the variation illustrated in (1)-(3) regarding the preverbal position in Dutch, by examining word order in the spoken Dutch corpus *Corpus Gesproken Nederlands* (CGN).

If we know which factors influence the word order variation illustrated in (1)-(3), however, we do not yet have an answer to the question what makes a speaker choose a particular form. Does a particular choice of form benefit the speaker, the hearer, or both? Arnold et al. (2000), in their corpus study and production experiment on heavy NP shift and the dative alternation in English, argue that the observed word order preferences regarding dative alternation are the result of planning processes benefiting the speaker. When the speakers in their experiment were disfluent during the initial part of the utterance, indicating difficulty in production, they were significantly more likely to produce the more salient and more familiar NP first. However, their study focused on a type of variation in word order that does not result in any differences in truth-conditional meaning. The two dative variants *Give the white rabbit the carrot* and *Give the carrot to the white rabbit* have comparable meanings. As it is not obvious why and how the hearer would benefit from choosing one of these word order variants over the other, the results of Arnold et al.'s study do not bear on the question whether word order variation may also help the hearer.

In Dutch, whether the constituent in first position is interpreted as the subject or as the direct or indirect object is crucial for determining the truth-conditional meaning of the sentence. As a consequence, there is a potential benefit for the hearer in the speaker's selection of one form over the other: The speaker's choice may help the hearer to arrive at the intended meaning. By investigating the variation regarding the first sentence position in Dutch, we may therefore be able to determine whether speakers choose a particular word order variant for their own benefit only or also for the benefit of the hearer. To answer this question, we need to distinguish the intentions and choices of the speaker from the intentions and choices of the hearer. A linguistic framework capable of doing so is bidirectional Optimality Theory (Blutner, 2000; Blutner et al., 2006; Hendriks et al., 2010). Bidirectional Optimality Theory models the interplay between the speaker's choice of form and the hearer's interpretation of this form in terms of optimization over a set of potentially conflicting constraints. To express a particular meaning, speakers choose the form that optimally satisfies the constraints of the grammar. In addition, a bidirectionally optimizing speaker will also determine how a hearer will interpret the chosen form. If the resulting interpretation is different from the meaning that the speaker intended to express, the speaker will avoid this form for the intended meaning and use an alternative form instead. Thus, bidirectional optimization allows us to simultaneously model the effects of interacting factors regarding

linguistic choice and the requirement that the chosen form must be such that the hearer can recover the intended meaning.

This paper is organized as follows. In Section 2, we discuss word order in Dutch in more detail. Here, we argue that the attested variation regarding the first sentence position in Dutch may cause the intended meaning of the sentence to be unrecoverable for a hearer. This problem can be circumvented if the speaker restricts word order variation in those situations in which the hearer has no other clue than word order to arrive at the interpretation of the sentence. In Section 3, it is shown that this restriction on word order variation can be captured in terms of bidirectional Optimality Theory. The proposed model predicts word order freezing when the subject is not higher in definiteness than the object. This prediction is tested in a corpus study, which is the topic of Section 4. Section 5 discusses the results of this corpus study and interprets these results in the light of the bidirectional model.

2 Word Order in Dutch

Word order in Dutch is characterized by the fact that in declarative main clauses the finite verb must occur in second position. Apart from this strong constraint on Dutch word order, Dutch does allow for a moderate amount of word order variation with respect to what can appear in front of the finite verb in second position. The canonical word order is SVO, with the subject occurring in first position in roughly 70% of the sentences in Bouma's (2008) study of the spoken Dutch corpus CGN. However, the first position can also be occupied by direct objects, indirect objects and other constituents, resulting in OVS or XVS word order.

Often, multiple clues are available to help the hearer determine what is the subject and what is the object of a transitive sentence. For example, in the sentences in (1)-(3), the use of the nominative pronominal form *ik* ' I ' unambiguously identifies this form as the subject of the sentence, even when it does not occur in first position. Besides case, other clues may be available from givenness, intonation, event likelihood, selection restrictions of the verb, definiteness, and animacy. For example, the selection restrictions imposed by a verb on its arguments specify what semantic properties the arguments must have. Knowing that the verb *drink* requires its direct object to be drinkable helps in determining what is the direct object. Regarding definiteness, the following table (adapted from Bouma, 2008: 107) shows the

relation between grammatical function and level of definiteness in the spoken Dutch corpus CGN.

	Grammatical Function	
Definiteness Level	Subject	Direct object
Indefinite full NP	-1.4245	1.6718
Definite full NP	-0.1777	0.4881
Pronoun	0.1651	-0.7912

Table 1 Association between grammatical function and definiteness, given as pointwise mutual information

Pointwise mutual information (PMI) provides a measure of the information given by one variable when we know the value of the other. A positive PMI means that a combination of values is favored, a negative PMI means that a combination is disfavored, and a PMI of zero means that the chances of a combination of values are exactly what we would expect if the variables were not associated. As Table 1 shows, the higher the definiteness level of the NP, the more likely it is to be the subject. Conversely, the lower the definiteness level of the NP, the more likely it is to be a direct object. This conforms to the universally observed pattern that subjects tend to be definite, while direct objects tend to be indefinite (Comrie, 1979). Similarly, subjects tend to be animate, while direct objects tend to be inanimate.

Surprisingly, when clues such as those regarding overt case and selection restrictions of the verb are absent, and definiteness and animacy do not distinguish between subject and object, the sentence does not become ambiguous. Consider (4):

(4)   Fitz zag Ella.

       Fitz saw Ella

       Only 'Fitz saw Ella' (SVO), although structurally compatible with 'Ella saw Fitz' (OVS)

A hearer encountering (4) could in principle assign an SVO interpretation or an OVS interpretation to this sentence, as both word orders are possible in Dutch. Under the first interpretation, Fitz is the subject in canonical position. Under the second interpretation, Fitz is the fronted object. Given the word order possibilities in Dutch, hearers should be in doubt as to whether the speaker intended (4) to mean that Fitz saw Ella or that Ella saw Fitz. However, presented out of context and in the absence of any intonational clues, Dutch hearers will interpret sentences with a animate definite subject and a similar object as conveying an SVO interpretation (Cannizzaro, 2010). Their preferred interpretation thus reflects the observation that the first constituent of the sentence most likely is the subject.

So in the absence of other clues, hearers prefer the interpretation that is in accordance with canonical word order. This observation about hearers' preference results in the potential problem that certain meanings may be unrecoverable. Suppose that the speaker intends to convey the meaning that Ella saw Fitz. In this case, (4) would be a poor choice because hearers will interpret Fitz in (4) as the subject. So the meaning that Ella saw Fitz is unrecoverable from the form in (4) if no other clues are present. This suggests that the hearer's interpretation may have implications for the speaker's freedom of word order variation, and that the speaker's choice is limited when word order is the only available clue to arrive at the interpretation of the sentence. In the next section, we will propose an account of word order variation in Dutch that formalizes this intuition. The proposed account is compatible with multi-factorial accounts of word order variation such as advocated by Bresnan (see Bresnan et al., 2007, and subsequent work), but is stronger in the sense that it provides an *explanation* for the speaker's choice of form as partially motivated by the strive for communicative success, rather than merely revealing correlations between the choice for one form over the other and particular properties of these forms or their contexts of use.

3 A Bidirectional OT Account of Word Order Freezing

The phenomenon that in potentially structurally ambiguous sentences only one reading surfaces has been referred to in the literature as *word order freezing*. An early published observation of freezing is found in Jakobson (1936/1971) on Russian. Similar claims can be found for languages as different as Hindi, Korean, German, Bulgarian, Russian and Papuan

languages (Lee, 2001, and references therein), Haida, Swedish (Morimoto, 2000) and Japanese (Tonoike, 1980; Kuno, 1980; Flack, 2007).

Theoreticians applying bidirectional Optimality Theory (BiOT) to syntax have recognized that BiOT allows one to capture freezing in a natural fashion (Kuhn, 2003; Vogel, 2004; Morimoto, 2000; Bouma, to appear). In BiOT, the speaker's and hearer's view of a sentence are explicitly modeled by distinct optimization competitions. Grammaticality is defined in terms of a combination of these two competitions. The reason this allows us to model freezing is that we can condition the grammaticality of a non-canonical word order like OVS on the recoverability of the intended interpretation (i.e., the assignment of grammatical functions to the NPs in the sentence) from the surface string.

Lee (2001) discusses Hindi, a free word order, case marking language, in which any order of S, O and V is permissible under the right information structural circumstances. In particular, consider the SOV and OSV word orders in (5):

(5)   a.   Ilaa-ne yah khat likhaa.
                Ila.ERG this.NOM letter.NOM wrote
        b.   Yah khat Ilaa-ne likhaa.
                this.NOM letter.NOM Ila.ERG wrote
                'Ila wrote this letter'

Lee argues that the two ways of expressing the proposition 'Ila wrote this letter', (5a) and (5b), correspond to different information structural situations: in (5a) the topic is Ila, in (5b) it is the letter. That is, the topic occurs in first position. Note that word order is free to encode this information structural dimension, as the grammatical function assignment is derivable from case marking. When case does not distinguish the grammatical functions, however, word order freezes to SOV, as Lee's example (6) shows:

(6)   Patthar thelaa todegaa.
        stone.NOM cart.NOM break.FUT
        'The stone will break the cart' (SOV)
        Not: 'The cart will break the stone' (OSV)

In this example, the OSV reading is unavailable. Likewise, reordering the arguments will result in the inverse, again SOV, interpretation, in which the cart breaks the stone.

Lee explains the contrast between (5) and (6) with strong bidirectional OT (Blutner, 2000). In strong BiOT, a form-meaning pair <f,m> is grammatical iff it is *production* optimal and *comprehension* optimal. As usual, optimality is defined as being most harmonic amongst a set of candidates taken from the set of all possible form-meaning pairs (*Gen*), where relative harmony $>_{con}$ is determined by a list of ranked violable constraints (*Con*). Crucially, production and comprehension optimality differ in the definition of the set of candidates:

(7)   Production optimal:

<f,m> in *Gen* such that there is no *<f',m>* in *Gen*, *<f',m>* $>_{con}$ *<f,m>*

Comprehension optimal:

<f,m> in *Gen* such that there is no *<f,m'>* in *Gen*, *<f,m'>* $>_{con}$ *<f,m>*

In prose: in a grammatical form-meaning pair *<f,m>*, *f* is the best way to express *m* and *m* is the best way to interpret *f*.

The following two, possibly conflicting, constraints are at work in Hindi word order, ranked as in (9), with Top-Left being stronger than Sub-Left.

(8)   Sub-Left: The subject aligns left in the clause
Top-Left: The topic aligns left in the clause

(9)   Top-Left >> Sub-Left

To see why strong bidirectional OT and the grammar in (8)-(9) explain the contrast between (5) and (6), first consider the optimization resulting in the OSV realization (5b):[1]

---

[1]   Since the precise formal semantics of the sentences is not at issue here, we use a pseudo-semantic representation as the input. The pseudo-operators *T* and ? are used to indicate topics and questioned material, respectively. A *T* in front of an argument indicates that this argument is the topic.

(10a)　Production (Hindi)

write(Ila,*T* this letter)	Top-Left	Sub-Left
Ila.ERG this.NOM letter.NOM wrote	*!	
☞　this.NOM letter.NOM Ila.ERG wrote		*

(10b)　Comprehension (Hindi)

this.NOM letter.NOM Ila.ERG wrote	Top-Left	Sub-Left
☞　write(Ila,*T* this letter)		*
write(*T* Ila,this letter)	*!	*

In OT, optimization is typically represented in a tableau, such as production tableau (10a) or comprehension tableau (10b). Here, the top left-hand cell shows the input to optimization, which is a representation of a meaning in (10a) and a form in (10b). Constraints are listed to the right of this input in order of descending strength. Output candidates are given in the first column below the input. An asterisk in a cell indicates that the candidate in that row violates the constraint in that column and an exclamation mark indicates a fatal violation that renders the candidate suboptimal. The optimal candidate, which satisfies the total set of ranked constraints best and hence is the output for the given input, is indicated by a pointing finger (see Kager, 1999; Prince & Smolensky, 1993/2004, for an introduction into OT).

　　　The fact that Top-Left outranks Sub-Left explains why the OSV order is production optimal (10a). Because a violation of Top-Left is more serious than a violation of Sub-Left due to the relative ranking of the two constraints, the SOV order is less harmonic than the OSV order. Hence, the OSV order is optimal in production. In comprehension (10b), Top-Left makes sure that the correct argument is interpreted as the topic, whereas case marking restricts the candidate set to only include interpretations that already correspond to the intended grammatical function assignment.[2]

---

[2]　Instead of restricting the candidate set like this, one could also implement the influence of case through (a) high ranking constraint(s). This does not change the analysis fundamentally.

Now let us turn to the missing OSV reading for the double nominative in (6). As before, production selects the object-initial candidate when the object is the topic:

(11a)   Production (Hindi)

break(cart,T stone)	Top-Left	Sub-Left
cart.NOM stone.NOM will break	*!	
☞ stone.NOM cart.NOM will break		*

However, due to the lack of a case distinction, we have a less restricted candidate set in comprehension. That is, in addition to the interpretations corresponding to OSV, we have the choice of interpreting the production optimal form as an SOV sentence:

(11b)   Comprehension (Hindi)

stone.NOM cart.NOM will break	Top-Left	Sub-Left
break(cart,T stone)		*!
☞ break(T stone,cart)		
break(T cart,stone)	*!	*
break(stone,T cart)	*!	

We have the option to satisfy both constraints Top-Left and Sub-Left in comprehension. Therefore, comprehension selects the topic-initial candidate which is also a subject-initial candidate. The result is that production and comprehension do not agree on the same form-meaning pair, and < stone.NOM cart.NOM will break, break(cart,T stone) > does not constitute a strong bidirectionally optimal pair. That is, the OSV reading of (6) is not grammatical. In effect, the added freedom of interpretation in comprehension in (11b) is the source of the loss of freedom of word order. Lee's bidirectional OT model neatly captures the intuition that word order freezing is related to the lack of case information about grammatical function assignment.

Lee (2001) observes that word order freezing in this bidirectional perspective is a kind

of Emergence of the Unmarked, which refers to the property of OT grammars that 'even dominated constraints may be visibly active, under appropriate circumstances' (McCarthy & Prince, 1994: 363). The constraint Sub-Left in Hindi is outranked by Top-Left, and in production its effects are not noticeable. Topics are put first whether they are subjects or not. In comprehension, however, it may be the case that Top-Left fails to single out a candidate (as in 11b). In this case, the effects of Sub-Left becomes visible and the unmarked SVO interpretation is the optimal candidate. In Lee's analysis, it is the lack of distinctive case marking that causes the emergence of the unmarked SVO word order. If, however, the information from case *is* enough to identify subject and object, Sub-Left is not visible at all in Hindi, in production nor comprehension.

Bouma (to appear) generalizes this analysis of freezing as emergence of the unmarked word order to a more inclusive notion of what constitutes information about grammatical function. In Bouma's adaption of Lee's bidirectional model for word order in Dutch, not only case supplies information about grammatical function assignment in comprehension, but also default associations between grammatical function on the one hand, and definiteness, animacy and information structure on the other. Word order in Dutch freezes when none of these sources of information prefer a non-canonical order of the arguments. Put differently, when there is no other information as to which of the constituents is the subject, word order becomes the determining source of information.

Recall from Section 2 that Dutch main clause word order is structurally ambiguous between SVO and OVS. In principle the order of subject and object is free. Now consider the following examples:

(12)   Fitz zoekt Ella op.
        Fitz looks Ella up
        'Fitz looks up Ella' (SVO), not (or strongly dispreferred) 'Ella looks up Fitz' (OVS)

(13)   a.   Welk meisje zoekt Frank op?
                which girl looks Frank up
                'Which girl looks up Frank' (SVO) or 'Which girl does Frank look up' (OVS)
        b.   FITZ zoekt Ella op.
                Fitz looks Ella up
                'FITZ looks up Ella' (SVO) or 'Ella looks up FITZ' (OVS)

11

c.   Het nummer zoekt Ella op.

The number looks Ella up

'Ella looks up the number' (OVS) and maybe 'The number looks up Ella' (SVO)

Example (12) is a case of word order freezing analogous to (4) in Section 2 and the Hindi example in (6). The examples in (13) are however not frozen to SVO. Here, we shall first informally discuss why the OVS readings are available to begin with. A formal model of the ambiguity of the sentences in (13) and the lack of ambiguity of (12) follows below.

Example (13a) appears in a discussion of word order freezing in Zeevat (2006). The availability of OVS is a surprise under a bidirectional model of freezing that only considers information expressed through surface form, such as case and agreement, in identifying subject and object. In (13a), neither the NPs nor the verb tells us what constituent is the subject. Why do we not see the emergence of the unmarked SVO reading as in (12)? We follow Kaan (1999; 2001) in assuming that Wh-constituents are indefinite. This means, however, that in terms of definiteness, the OVS reading is a less marked reading. It allows us to have an indefinite object and a definite subject (thus conforming to the pattern shown in Table 1 in Section 2). It seems that this preference stemming from definiteness is strong enough to bring the OVS interpretation to the front.

The explanation of (13b) follows a similar reasoning. In (13b), SVO would involve focusing the subject and backgrounding the object, whereas OVS allows us to focus the object and background the subject. The latter is an unmarked situation in terms of information structure (Zerbian, 2007). Finally, in (13c), OVS is an unmarked reading in terms of animacy as an inanimate subject is a marked situation (Aissen, 2003).

We can summarize the discussion regarding (12)-(13) as follows. The available OVS readings in (13), which are marked in terms of argument order, are unmarked in some other linguistic dimension. It is this conflict in markedness that triggers the ambiguity in the sentences in (13). Word order in (12), on the other hand, is frozen because there is no information whatsoever that might promote the OVS interpretation.

The alternative dimensions of markedness are captured in an OT model by adopting the following constraints, which have been motivated independently in the literature (e.g., Aissen, 2003).

(14)   *Sub/Ind   Subjects should not be indefinite

12

*Sub/Inan	Subjects should not be inanimate
Sub-Background	Subjects are in the information-structural background

To capture the variation attested in (13) - and not just the word order marked OVS readings - Bouma uses Antilla's (1997) conception of language particular OT grammars as *stratified rankings*. The ranking of constraints is only partial: Constraints are assigned to strata which are strictly ranked, but the ordering within a stratum is not specified. When two constraints in one stratum {A,B} conflict, both the candidate preferred by A and the candidate preferred by B will be optimal. That is, a stratified ranking A >> {B,C} >> D can be seen as denoting a set of compatible, fully specified rankings {A>>B>>C>>D, A>>C>>B>>D}, and a form-meaning pair is optimal when it is optimal under one of the fully specified rankings in this set. The resulting bidirectional model is *stratified strong bidirectional OT*. Building on the definitions in (7), grammaticality can be defined as (Bouma, 2008; to appear):

(15) a form meaning pair <f,m> in *Gen* is grammatical in *stratified strong BiOT* iff there is a fully specified constraint ranking *Con* in the stratified grammar *StratCon*, such that

there is no <f',m> in *Gen*, <f',m> $>_{con}$ <f,m> (production optimal)
and
there is no <f,m'> in *Gen*, <f,m'> $>_{con}$ <f,m> (comprehension optimal)

In stratified strong bidirectional OT, the variation in (13) can be modeled by putting the constraints governing word order (e.g., Sub-Left, Top-Left) in the same stratum as the constraints preferring an unmarked association between animacy/definiteness/information structure and grammatical function. When the constraints within this stratum conflict, variation results.

　　Let us illustrate Bouma's proposal by working out the ambiguity of the wh-question in example (13a). For clarity of exposition we restrict ourselves to a subset of the constraints mentioned in (14). We refer the reader to Bouma (2008; to appear) for examples and discussion of a larger grammar containing the other constraints. In addition, we assume the high ranked constraint Wh-Left, that forces wh-constituents to be initial irrespective of their grammatical function.

(16) Wh-Left >> {Sub-Left, *Sub/Ind}

The SVO reading of (13a) corresponds to the form-meaning pair in (17a). The OVS reading corresponds to the one in (17b).

(17)   a.   < which girl looks Frank up, look-up(?girl,frank) >
        b.   < which girl looks Frank up, look-up(frank,?girl) >

The SVO reading, (17a), is bidirectionally optimal according to (15) in the subcase of (16) that Sub-Left >> *Sub/Ind:[3]

(18a)   Production (Dutch)

look-up(?girl,frank)	Wh-Left	Sub-Left	*Sub/Ind
☞ which girl looks Frank up?			*
Frank looks which girl up?	*!	*	*

(18b)   Comprehension (Dutch)

which girl looks Frank up?	Wh-Left	Sub-Left	*Sub/Ind
☞ look-up(?girl,frank)			*
look-up(frank,?girl)		*!	

The OVS reading corresponding to (17b) is an optimal form-meaning pair in the subcase of (16) that *Sub/Ind >> Sub-Left.

---

[3]   As Tableau (18a) shows, the constraint *Sub/Ind is, in production, violated by properties of the input alone. It thus appears to be an unusual type of constraint. After all, the input is given and, therefore, violable constraints that only relate to the input do not influence optimization at all. However, the presence of the constraint in a bidirectional grammar makes sense, as it is a constraint on the output in comprehension. For reasons of parsimony, we assume identical grammars between both optimization directions.

(19a)   Production (Dutch)

look-up(frank,?girl)	Wh-Left	*Sub-Ind	Sub-Left
☞ which girl looks Frank up?			*
Frank looks which girl up?	*!		

(19b)   Comprehension (Dutch)

which girl looks Frank up?	Wh-Left	*Sub/Ind	Sub-Left
look-up(?girl,frank)		*!	
☞ look-up(frank,?girl)			*

The two meaning inputs in (18a) and (19a) both result in the same surface form, because of the high ranked Wh-Left. We thus have neutralization of a meaning difference in production. The conflict between *Sub/Ind and Sub-Left in the same stratum causes variation in interpretation of a form in comprehension. Put together, the result is an ambiguous form in our bidirectional model.

The grammar in (16) also predicts that if we make the second NP in (13a) a less likely subject by turning it into an indefinite NP, only the SVO reading should remain: *Sub/Ind does not prefer the OVS reading anymore and the decision is up to Sub-Left. This prediction is borne out in Dutch.

(20)   Welke jongen belt een meisje?
        which boy calls a girl
        'Which boy calls a girl?' (SVO)
        Not: 'Which boy does a girl call?' (OVS)

As before in (12), we see in (20) that when there is no information that says that the marked OVS reading is available, only the unmarked SVO reading emerges.

Central to our model of word order is the claim that the hearer's perspective on language (comprehension) plays as much of a role in word order variation as the speaker's perspective (production) does. This raises the expectation that we should be able to find

15

evidence of the influence of the hearer's perspective in production data, that is, in a corpus. In the next section, we shall see that we can indeed find such evidence in a corpus of spoken Dutch: Speakers more often use non-canonical word order when the hearer would be able to retrieve the correct grammatical function assignment on the basis of definiteness of the arguments.

4 A Corpus Study of Word Order Freezing

4.1 Empirical predictions of the BiOT model

The bidirectional model of word order laid out in the previous section predicts that word order may vary as long as there is enough information independent from surface order to let the hearer correctly infer the grammatical function assignment. In this section, we present a corpus study of the influence of one of these information sources on word order variation in spoken Dutch. We show that non-canonical word order is positively correlated with the possibility of recovering the intended grammatical function assignment on the basis of the definiteness of subject and object. This provides novel empirical support for the central thesis of the strong bidirectional model, that is, that both the hearer and the speaker perspective need to be taken into account in a model of grammar.

Let us consider how we might investigate the predictions made by the bidirectional model of Bouma (to appear) in a corpus. The model identifies situations in which word order may deviate from the canonical SVO order, and when it may not. Note that the model does not state that, when there are no *hearer* reasons to freeze word order, word order is always non-canonical. Non-canonical word order requires that there are certain reasons (*speaker* reasons) for putting an object in front of the subject. The absence or presence of such reasons results in variation of word order. The model as it stands is a discrete model. It says that when there is no information whatsoever to guide the hearer to the correct interpretation, word order freezes. However, we cannot straightforwardly apply this prediction in a quantitative empirical evaluation of the model by means of a corpus study. First, some of the information sources involve linguistic dimensions that are hard to identify and process automatically on a larger scale. These include animacy and prosody/information structure. Annotating these in a corpus involves considerable manual effort and no existing Dutch corpus has this information

at the time of writing. Definiteness, on the other hand, is a dimension that can be meaningfully approximated on a large scale by looking at the form of the NP. That is, if we aim to do a large scale corpus investigation, we are at this time restricted to concentrating on just one of these linguistic dimensions.

Secondly, there is the danger of immunizing ourselves against empirical falsification by the "no information whatsoever to guide the hearer" part of the prediction: If we fail to find evidence against the bidirectional model - that is, if we find non-canonical word order where we would predict freezing - it would be possible to explain this away by claiming that there must have been other sources of information present to help the hearer. After all, sentences are uttered in rich contexts and there is a wealth of world knowledge available to hearer. These constitute sources of information that in principle would lend themselves for incorporation into the bidirectional model of Bouma (to appear).

We therefore investigate a quantitative variant of the predictions of the bidirectional model. We shall say that we expect that the absence of *one* source of information is related to a decrease of the proportion of non-canonical word order in the corpus. Concretely, the non-canonical word order that we investigate is direct object fronting in Dutch, and the source of information that we consider is the definiteness of the subject and object NPs. Object fronting is the only frequent way of getting an object-before-subject word order in a Dutch main clause (Bouma, 2008), so studying this type of word order variation suffices to study the relative order of subject and object. By using the form of NPs as an approximation for their level of definiteness, we can distinguish three definiteness levels automatically: pronominal NPs, definite full NPs and indefinite full NPs. As mentioned in Section 2, these are associated with grammatical function in the following way (Comrie, 1979; Aissen, 2003; see also Table 1):

(21)   Typical Subject   Typical Object
        pronominal   >   definite full NP   >   indefinite full NP

The abstract hearer who relies on this information to assign the grammatical roles, will thus assume that of two NPs, the one higher on the definiteness scale is the subject. The constraint *Sub/Ind thus captures a part of this scale. Let us refer to the situation where the intended subject is indeed higher on the definiteness scale as *definiteness superiority*. In the case of superiority, definiteness counts as information in the hearer perspective. The reverse situation is *definiteness inferiority*. In the case of inferiority, definiteness is misinformation for the

17

hearer. The remaining case is *definiteness equality*, where the hearer cannot be (mis)guided by definiteness at all. If the freedom of the speaker to choose a non-canonical word order is constrained, in a gradual sense, by the chance that the hearer has enough information to correctly understand an utterance, we may expect the following pattern regarding object fronting in the corpus:

(22)   Relative Definiteness Hypothesis:

Superiority	Equality	Inferiority
object fronting more frequent	(no effect)	object fronting less frequent

To see how this prediction follows from a quantitative re-interpretation of our bidirectional OT model of word order, consider the case in which a speaker has some reason to prefer a non-canonical order of subject and object. If we have definiteness superiority, the speaker is free to use this word order because the chance that the hearer will be confused is low. Definiteness inferiority, however, constrains the speaker because the chance that the hearer will be confused is higher.

4.2 The corpus data

Table 2 shows corpus data on direct object fronting and its relation to definiteness of subject and object. The data comes from 16,146 transitive V2 main clauses extracted from the spoken Dutch corpus CGN, which is a mixed genre corpus with speakers from the Netherlands and Belgium. The argument NPs are assigned to one of three definiteness levels on the basis of surface form characteristics: indefinite full NPs (common nouns with indefinite article or no determiner), definite full NPs (common noun with definite article or universal quantifier, proper names), pronouns (personal or demonstrative). Although some pronouns are marked for case and thus supply unambiguous information about their subject status, we include pronouns in our corpus study because far from all Dutch pronouns show case. Case syncretism is found at least for: *ze* 'they'/'them' (weak), *je* 'you' (sg, weak), *u* 'you' (formal), *jullie* 'you' (pl), *het* 'it' and all demonstrative pronouns. For details and technical background of the extraction method, we refer the reader to Bouma (2008).

Subject Definiteness	Object Definiteness			
	Indefinite full NP	Definite full NP	Pronoun	Total
Indefinite full NP	162	88	114	363
OVS (%)	*2 (1.2)*	*1 (1.1)*	*37 (32.4)*	*40 (11.0)*
Definite full NP	644	477	373	1514
OVS (%)	*13 (2.0)*	*9 (1.9)*	*113 (30.3)*	*135 (8.9)*
Pronoun	5421	2875	5972	14268
OVS (%)	*171 (3.2)*	*300 (10.4)*	*2541 (42.5)*	*3012 (21.1)*
Total	6247	3440	6459	16146
OVS (%)	*186 (3.0)*	*310 (9.0)*	*2691 (41.7)*	*3187 (19.7)*

Table 2 Object fronting by definiteness of subject and direct object in transitive clauses

A few general trends can be seen in Table 2, if we confine ourselves to the row and column totals. First, we may note that subjects tend to be pronominal, but that many objects are pronominal, too. This is a fact about spoken language in general. We may add that many of the pronouns are 1st or 2nd person. However, given the overwhelming amount of pronouns in spoken language, pronominal objects are in fact much rarer than expected by chance. Secondly, we point out that object fronting is relatively frequent in two (not mutually exclusive) circumstances: when the object is pronominal and when the subject is pronominal. The former is caused by the large proportion of demonstrative pronoun objects. Demonstrative pronouns have a strong tendency to appear in initial position. The latter is caused by the large proportion of personal pronouns as subjects: personal pronouns have a tendency to avoid the first position. When the subject is a personal pronoun, the object is thus freer to move into first position (Bouma, 2008).

Now let us turn to the predictions in (22). In Table 2, we have highlighted the cases of definiteness superiority with a darker gray and the cases of definiteness inferiority with a lighter gray. Table 3 summarizes the relation between object fronting and relative definiteness by combining cases with the same relative definiteness.

	Superiority	Equality	Inferiority
All word orders	8940	6611	575
OVS (%)	*484 (5.4)*	*2552 (38.6)*	*151 (26.2)*

Table 3 Counts and proportions of object fronting, per relative definiteness level

A first look at the average fronting percentages suggests that the predictions in (22) are not met at all. At around 5%, the average proportion of object fronting in the superiority data is many times lower than in the equality data (39%) and in the inferiority data (26%). If we look at the raw percentages like this, however, we ignore a crucial assumption in the reasoning that led to these predictions. We considered the influence of relative definiteness given a hypothetical, fixed tendency to front the object. That is, the effect of relative definiteness on object fronting follows *other things being equal*. In the data of Table 2 and 3, however, other things are not equal: we know from existing corpus investigations that constituent fronting in Dutch depends, amongst other things, on the grammatical function, the level of definiteness and the complexity of the constituent (see Bouma, 2008, and references therein). For instance, indefinite full NPs are fronted a lot less frequently than definite ones. The proportions we see in Table 2 are thus not just the result of relative definiteness which guides the hearer in recovering the intended meaning (i.e., the hearer's reasons for fronting), but also of independent effects having to do with the appropriateness of the sentence in the discourse and ease of sentence planning (i.e., the speaker's reasons for fronting). In terms of the bidirectional model, we may say that evaluating the relative definiteness hypothesis directly on the basis of Tables 2 and 3 falls into the trap of ignoring the speaker perspective.

To answer the question about the role of relative definiteness in addition to these independent factors, we fit a logistic regression model that incorporates information about relative definiteness (three levels) as well as other known factors in fronting: complexity of subject and object (as the natural logarithm of the number of words), and NP form of subject and object (six levels each). The result of fitting a logistic regression model to the data summarized in Table 2 is given in Table 4. The model is a good predictor of OVS (c-index

20

0.927) and bootstrap resampling shows no sign of overfitting (Harrell et al., 1999).[4] The parameter estimates of the non-freezing related factors (i.e., complexity of subject and object, and NP form) are in line with earlier work and thus do not indicate any problems with the model: Other things being equal, the more complex an object is, the lower the chance that it will be fronted. Likewise, objects that are higher on the definiteness scale are more likely to be fronted, with the exception of personal pronominal objects, which are fronted even less often than indefinite full NP objects. Interestingly, the model suggests that the absolute level of definiteness of the subject is only relevant when subjects are demonstrative pronouns, in which case the odds of direct object fronting are drastically lowered. We speculate that this is because in these cases the preferred option is to put the subject itself in initial position. The model including the three level factor relative definiteness is a significantly better fit than the same model without this factor (G2 = 11.1, df = 2, p = .004). We conclude that relative definiteness is a factor in predicting direct object fronting.

---

[4]   Model fitting and inspection was done with the "Design" library (http://cran.r-project.org/web/packages/Design/) of the R language for statistical computing (http://www.r-project.org).

Parameter		Estimate	Odds ratio interval (lo-hi)		p
Intercept		-4.729			
Subject Complexity		0.083	0.75	1.57	0.653
Object Complexity		**-0.721**	0.4	0.58	<0.001
Subject Form	Indefinite full NP (base)				
	Bare nominal	-0.722	0.43	1.37	0.171
	Definite full NP	-0.220	0.43	1.49	0.480
	Proper name	-0.450	0.32	1.24	0.184
	Demonstrative pronoun	**-2.498**	0.03	0.23	<0.001
	Personal pronoun	-0.295	0.56	3.19	0.503
Object Form	Indefinite full NP (base)				
	Bare nominal	**1.063**	2.12	3.94	<0.001
	Definite full NP	**1.742**	4.41	7.38	<0.001
	Proper name	**1.949**	4.83	10.19	<0.001
	Demonstrative pronoun	**5.459**	101.96	541.06	<0.001
	Personal pronoun	**-2.228**	0.02	0.44	0.002
Relative Definiteness	Superiority	**1.214**	1.55	7.31	0.002
	Equality (base)				
	Inferiority	-0.437	0.35	1.2	0.163

Table 4 Logistic regression model predicting OVS vs. SVO/XVSO. Parameter estimates in boldface indicate significant contribution of the parameter according to Wald's test; the

associated p-values are reported in the last column. The estimated odds ratio intervals are at the 95% level.

Inspection of the parameter estimates of the logistic regression model now allows us to evaluate the relative definiteness hypothesis in (22). The hypothesis predicts that definiteness superiority should have a positive correlation with object fronting, and definiteness inferiority a negative one. The parameter estimates partially follow this picture. As Table 4 shows, the parameter estimate of superiority is significantly positive. The odds of observing object fronting in the case of superiority are at least 1.55 times higher than in the case of equality. The difference between inferiority and equality fails to reach significance, that is, the model does not suggest that inferiority has an additional negative impact on direct object fronting beyond its being a case of non-superiority.

Because of the definitional dependencies between the factors Subject/Object Form and Relative Definiteness, one can expect strong multicollinearity to be present in the model. Indeed, inspection shows high variance inflation factors for superiority and inferiority (VIFs of 43 and 7, respectively). We do not consider this problematic, however. First and foremost, the significant effect of superiority is found *in spite of* the inflated variance. Secondly, there is little or no sign of instability in the parameter estimates – the parameter estimates are all linguistically plausible and the estimates of the Subject Form and Object Form parameters only change moderately upon inclusion. It may be that the size of our dataset mitigates some of the problems associated with multicollinearity. All in all, we conclude that the high VIFs do not indicate vicious multicollinearity (see O'Brien, 2007, for discussion of interpreting VIFs).

Our analysis of the corpus data shows that speakers use non-canonical word order more freely when definiteness of subject and object provides the hearer with information about the intended grammatical function assignment. The importance of relative definiteness in word order and the precise way in which it should influence word order are derived from the bidirectional model of word order laid out in the previous section. The results of our corpus analysis thus provide empirical support for the bidirectional model of word order developed on the basis of data from language users' intuitions.

5 Discussion

The aim of this study is to investigate which factors influence the observed variation regarding the occupation of the first position in Dutch sentences, and what makes the speaker choose SVO or OVS. Factors known to influence constituent fronting in Dutch are grammatical function, complexity of subject and object, and the form of subject and object. A quantitative investigation of transitive sentences in the spoken Dutch corpus CGN revealed that complex objects and personal pronouns tend not to be fronted, whereas demonstrative subjects as well as demonstrative objects have a strong tendency to occur in first position (Bouma, 2008). These factors relate to aspects of sentence production such as the appropriateness of the sentence in the discourse and ease of sentence planning, and may therefore benefit the speaker without necessarily also benefiting the hearer.

In addition to these speaker-oriented factors in constituent fronting, we hypothesized that there are also hearer-oriented factors that influence the speaker in deciding whether to front the object. The hearer's aim is to arrive at the meaning the speaker intended to convey. To avoid a situation in which the hearer arrives at a different meaning than the intended one, the speaker may choose a particular form over another form. The BiOT model developed in Bouma (2008; to appear) predicts that speakers use canonical SVO order in such cases, and use non-canonical OVS order only when hearers have other ways than word order to determine the grammatical functions of the NPs in the sentence. As subjects tend to be definite and objects tend to be indefinite, one factor that may guide hearers in their interpretation of non-canonical word order is definiteness. Our analysis of transitive sentences in the spoken Dutch corpus CGN confirmed that speakers use non-canonical word order more freely when the subject is higher in definiteness than the object. We thus conclude that Dutch speakers take into account their hearers when choosing a particular word order.

We related the logistic regression model of OVS occurrence presented in Section 4 to a bidirectional OT model. However, the logistic regression model can also be related to a unidirectional OT production model (see Bresnan et al., 2007, for discussion, and Jäger & Rosenbach, 2006, for a constraint-based model closely related to logistic regression). For instance, the model parameters regarding object definiteness can be seen as constraints demanding the fronting of objects of different definiteness levels (e.g., OBJ/FULLDEF-LEFT, OBJ/PROPNAME-LEFT, etc.), the weights of the parameters can be seen as specifying a ranking (e.g., OBJ/PROPNAME-LEFT >> OBJ/DEF-LEFT), and the observation that OVS in general is the

less frequent case corresponds to the SUBJECT-LEFT constraint introduced in Section 3. This informal correspondence to a unidirectional production model prompts the question whether the data really support a bidirectional model. Could we not achieve the effects observed in the corpus with a unidirectional model? We argue that the bidirectional model is to be preferred for two reasons: First, the constraints that one needs to capture freezing effects in a bidirectional model are less complex than in a unidirectional model. Second, the bidirectional model is able to generate further predictions about trends in a corpus of Dutch and preferences during Dutch hearers' online processing of word order.

To start with the first reason, the constraints that one needs to capture freezing effects in a unidirectional model would have to be more complex than in the bidirectional model. The relative definiteness parameters of the logistic regression model would correspond to conjoined OT constraints that punish object fronting to different degrees, depending on the relative definiteness of subject and object. The levels of relative definiteness themselves are also conjunctions of constraints that mention the absolute definiteness levels of subject and object. One ends up with a set of multiply conjoined constraints of the form *Subj/X&*Obj/Y&Subject-Left, which *prohibit* object fronting in particular circumstances and whose position relative to those constraints that *promote* object fronting determines how relative definiteness restricts object fronting. Although technically unproblematic, these constraints are not much more than a listing of possible scenarios and their particular impact on word order. Without further assumptions, these constraints could be ranked such that object fronting in the context of definiteness superiority is punished more severely than in the context of definiteness inferiority (e.g. *Subj/Pron&*Obj/Ind&Subject-Left >> *Subj/Ind&*Obj/Def&Subject-Left). However, we are unaware of the existence of languages that behave like this. In the bidirectional model of word order that we have presented here, on the other hand, this situation cannot arise: The kind of impact that different levels of relative definiteness have on word order is predicted by the bidirectionality of grammar and the harmonic alignment of definiteness and grammatical function. So even though a production model involving the multiply conjoined constraints as outlined above would be able to capture word order freezing, it does not provide us with any insight into why and when word order freezes.

This brings us to a second advantage of the bidirectional model over a unidirectional model: The BiOT model allows us to formulate and test further empirical predictions regarding word order variation in Dutch and its freezing in particular situations. For example,

the BiOT model predicts that animacy and givenness will have similar effects on the speaker's choice of word order as definiteness, as subjects are generally not only higher in definiteness than objects, but also tend to be higher in animacy, more given, and intonationally less prominent than objects (which is reflected by the OT constraints in (14) in Section 3). As a consequence, relative animacy, givenness and prosody may also provide the hearer with cues about the intended word order, and hence allow the speaker to use non-canonical word order. Thus we expect object fronting to be more frequent if (i) the subject is higher in animacy than the object, (ii) the subject is more given than the object, and (iii) the subject is intonationally less prominent than the object. We have not been able to test these predictions yet because of the unavailability of a sufficiently large corpus of Dutch that is annotated for information such as animacy and givenness. A further prediction of the BiOT model is that the relative definiteness effect we observed in the corpus of spoken Dutch, as well as the expected trends of relative animacy and relative givenness, are the indirect result of the hearer's interpretational preferences. That is, Dutch *speakers* select a particular word order because Dutch *hearers* prefer subjects to be definite, animate and given. Indeed, definiteness and animacy have been found to be important sources of information for Dutch hearers in resolving temporary subject-object ambiguities during online sentence comprehension (e.g., Kaan, 1999; Lamers, 2005).

This study related the factors found to influence word order variation in a corpus of spoken Dutch to a formal model of grammar that distinguishes the speaker's perspective from the hearer's perspective. Whereas the first position in Dutch sentences may be occupied by subjects and objects, OVS word order is dispreferred in those situations where hearers have no other sources than word order to determine the grammatical functions of the arguments. This pattern of partial word order freezing provides evidence that the speaker's linguistic choices are at least partially driven by their aim to avoid potential misunderstanding by the hearer.

Acknowledgments

References

Aissen, J. (2003). Differential object marking: iconicity vs economy. *Natural Language and Linguistic Theory*, 21, 435-483.

Arnold, J., Wasow, T., Losongco, A. & Ginstrom, R. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, 76, 28-55.

Blutner, R. (2000). Some aspects of optimality in natural language interpretation. *Journal of Semantics*, 17, 189-216.

Blutner, R. de Hoop, H. & Hendriks, P. (2006). *Optimal Communication*. Stanford, CA: CSLI Publications.

Bouma, G. (2008). *Starting a Sentence in Dutch: A Corpus Study of Subject- and Object-Fronting*. Dissertation, University of Groningen.

Bouma, G. (to appear). Production and comprehension in context: The case of word order freezing. To appear in A. Benz & J. Mattausch (Eds) *Bidirectional Optimality Theory*. John Benjamins.

Bresnan, J, Cueni, A., Nikitina, T. & Baayen, R. H. (2007). Predicting the dative alternation. In G. Bouma, I. Krämer, & J. Zwarts (Eds) *Cognitive Foundations of Interpretation*, (pp. 69-94). Amsterdam: Royal Netherlands Academy of Science.

Cannizzaro, G. (2010). Animacy and early word order. In: J. Costa, A. Castro, M. Lobo & F. Pratas (Eds) *Language Acquisition and Development: Proceedings of GALA 2009*. Newcastle upon Tyne, UK: Cambridge Scholars Publishing.

CGN (2004). Corpus Gesproken Nederlands, v1.0. Electronic Resource. See: http://lands.let.ru.nl/cgn/home.htm.

Comrie, B. (1979). Definite and animate direct objects: A natural class. *Linguistica silesiana*, 3, 13-21.

Flack, K. (2007). Ambiguity avoidance as contrast preservation: Case and word order freezing in Japanese. In L. Bateman, M. O'Keefe, E. Reilly, & A. Werle (Eds) *UMass Occasional Papers in Linguistics 32: Papers in Optimality Theory III* (pp. 57-89). Booksurge Publishing.

Harrell, F. E., Lee, K. L & Mark, D. B. (1996). Tutorial in biostatistics, multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicin*, 15(4), 361-387.

Hendriks, P., de Hoop, H., Krämer, I., de Swart, H., & Zwarts, J. (2010). *Conflicts in Interpretation*. London: Equinox Publishing.

Jakobson, R. (1936). Beitrag zur allgemeinen Kasuslehre. Gesamtbedeutungen der russischen Kasus. In Travaux du Cercle Linguistique de Prague 6 (pp. 240-288). Consulted in Word and Language, volume 2 of Selected Writings, 1971 (pp. 23-72). Den Haag/Paris: Mouton.

Jansen, F. & Wijnands, R. (2004). Doorkruisingen van het links-rechtsprincipe. Neerlandistiek.nl.

Jäger, G. & Rosenbach, A. (2006). The winner takes it all - almost. Cumulativity in grammatical variation. *Linguistics,* 44(5), 937-971.

Kaan, E. (1999). Sensitivity to NP-type: Processing subject-object ambiguities in Dutch. *Journal of Semantics*, 15(4), 335-354.

Kaan, E. (2001). Subject-object order ambiguities and the nature of the second NP. *Journal of Psycholinguistic Research*, 30(5), 527-545.

Kager, R. (1999). *Optimality Theory*. Cambridge: Cambridge University Press.

Kuno, S. (1980). A note on Tonoike's intra-subjectivization hypothesis and A further note on Tonoike's intra-subjectivization hypothesis. In Y. Otsu & A. Farmer (Eds) *Theoretical Issues in Japanese Linguistics* (MWPL 2). MIT Working Papers in Linguistics (pp. 149-157, 171-185).

Lamers, M. (2005). The on-line resolution of subject-object ambiguities with and without case-marking in Dutch: Evidence from event-related brain potentials. In M. Amberber & H. de Hoop (Eds) *Competition and Variation in Natural Languages: The Case for Case* (pp. 251-293). Elsevier.

Lee, H. (2001). Markedness and word order freezing. In P. Sells (Ed.) *Formal and Empirical Issues in Optimality Theoretic Syntax*, Volume 5 of Studies in Constraint-based Lexicalism. Stanford, CA: CSLI Publications.

McCarthy J. J. & Prince, A. (1994). The emergence of the unmarked: Optimality in prosodic morphology. In M. González (Ed.) *Proceedings of the North East Linguistics Society 24* (pp. 333-379). Amherst, MA.

Morimoto, Y. (2000). 'Crash vs yield': On the conflict asymmetry in syntax and phonology. Manuscript Stanford University.

O'Brien, R.M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity* 41(5), 673-690.

Prince, A. & Smolensky, P. (1993/2004). *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden, MA: Blackwell.

Tonoike, S. (1980). Intra-subjectivization; and More on intra-subjectivization. In Y. Otsu & A. Farmer (Eds) *Theoretical Issues in Japanese Linguistics* (MWPL 2). MIT Working Papers in Linguistics (pp. 136-148, 157-171).

Zeevat, H. (2006). Freezing and marking. *Linguistics*, 44(5), 1095-1111.

Zerbian, S. (2007). Subject/object-asymmetry in Nothern Sotho. In K. Schwabe & S. Winkler (Eds) *On Information Structure, Meaning and Form*, Linguistik Aktuell 100 (pp. 323-347). Amsterdam: John Benjamins.