AN LFG APPROACH TO WORD ORDER FREEZING

Kyle Mahowald
University of Oxford

Proceedings of the LFG11 Conference

Miriam Butt and Tracy Holloway King (Editors)

2011

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

Word order freezing is a linguistic phenomenon by which normally free word order is frozen in the absence of disambiguating case information. It has been said to exist in Russian, Dutch, Korean, and many other languages. Word order freezing has received increasing scholarly attention in recent years, and it is often claimed that the phenomenon should be treated as part of processing or as purely stochastic. Others maintain that it should be treated syntactically, but it has received relatively little attention in LFG. Indeed, word order freezing presents unique challenges within LFG since it resists a purely monotonic structural description. Using a notion of case indeterminacy as in Dalrymple et al. (2009), in this paper I propose a novel analysis of the phenomenon that I believe to be the first full account of word order freezing in a pure LFG framework. I will also compare it to a more intuitive account that I have developed, which uses LFG modified with a single OT constraint.

1 Introduction to word order freezing

The main goal of this paper is to present an LFG account of word order freezing alongside an account that uses LFG in conjunction with Optimality Theory (OT). Before we can do that, however, it is first necessary to give a brief introduction to word order freezing, the phenomenon by which normally free word order is frozen in the absence of disambiguating case information. Word order freezing has received increasing attention cross-linguistically in recent years. Bouma (2008, 2011) notes that in languages like Russian and Dutch, word order can vary freely unless case is ambiguous, at which point word order then steps in to determine how the clause should be interpreted. Tily (2010) discusses the phenomenon at length and points to Lee (2001) for Korean, Potts (2007) for Japanese, and Bouma (2011) for Dutch. I have argued elsewhere (Mahowald, 2011) that it may exist in Old English. That said, word order freezing is a tricky phenomenon to isolate, even in living languages, and it is difficult to conclude with certainty whether it indeed exists in Old English.

Let's take a look at how word order freezing works in Russian. Consider the below examples from Jakobson (1963) and repeated in Bloom (1999). In the examples in (1), both nominals are clearly marked. *Mama* is unambiguously nominative, and *papu* accusative. Thus, it is unproblematic for the object to be topicalized, as in (1b). It remains clear that *mama* is the subject despite the lack of SVO order.

†I gratefully thank my supervisor Mary Dalrymple as well as Louise Mycock, Ash Asudeh, the members of the Oxford Syntax Working Group, and the participants at LFG11 in Hong Kong.

(1) (nominals morphologically distinct)

(a) *Mama* *ljubit* *papu.*
Mother-NOM loves father-ACC
'The mother loves the father.'

(b) *Papu*$_{[TOPIC]}$ *ljubit* *mama.*
Father-ACC loves mother-NOM
'The mother loves the father.'

In (2), however, both nominals are ambiguous between nominative and accusative. As a result, the word order "freezes." The only possible reading is SVO, and (2), with its topicalized object, is said to be impossible.

(2) (all nominals morphologically ambiguous between NOM and ACC)

(a) *Mat'* *ljubit* *doč'.*
Mother loves daughter
'The mother loves the daughter.'

(b) *Doč'* *ljubit* *mat'.*
daughter loves mother
'The daughter loves the mother.'

(c) **Doč'*$_{[TOPIC]}$ *ljubit* *mat'.*
Daughter loves mother.

Although this offers a general overview of how word order freezing works, it may be an oversimplification. King (1995), for instance, notes that precisely the type of supposedly ungrammatical reading shown in (2c) is obtainable given the proper context. Similar context dependence is manifested in Dutch. While Bouma (2011) and others see word order freezing as a clear effect, others have suggested that it does not exist at all or exists only very weakly. Conversation with native Dutch speakers suggests that Bouma's examples referenced in (11) can be easily "unfrozen" through certain intonations much in the same way that a sentence like "Pizza Tom ate" is grammatical in English only in a very specific context with a very specific intonation. For example, imagine that sentence coming as a response to the question "Did Tom eat pizza?" If Tom ate a ton of pizza and the intonation is right, "Pizza Tom ate" would be perfectly grammatical. In a totally unmarked situation, however, it seems quite odd. Similarly, only when Bouma's examples are read with a "hat" intonation pattern is the freezing effect noticeable.

The apparent weakness of the constraint, combined with the fact that word order freezing seems to show up cross-linguistically, has led some to suggest that word order freezing is not a syntactic constraint at all but rather an effect that should be handled stochastically or merely as a processing constraint. Fundamentally, however, it has to do with syntax, and as such Bouma (2011) presents an

interesting problem: “One needs to be able to model word order that is driven by IS [information structure] rather than by grammatical function, but at the same time this freedom has to be taken away when not obviously IS-related syntactic phenomena such as agreement and case are in a certain configuration.” Indeed, while we must recognize that word order freezing is not black and white, it is a phenomenon that a theory of syntax should at least attempt to handle. And that is what I will set out to achieve here in an LFG framework.

One of the things that makes word order freezing difficult to analyze in LFG is the monotonicity requirement, as described in Bresnan (2001) and Dalrymple (2001). The function that maps the c-structure onto the f-structure can only accumulate. Thus, resultant f-structures become more and more specific as we add constraints. Ideally, we would want our rules to say that a word order becomes restricted, perhaps to SVO or SOV, only when both the prospective subject and prospective object have ambiguous case, as in {(↑CASE) = NOM | (↑CASE) = ACC}. LFG and the rules of logic, however, require that only one side of the disjunction can be accepted. At that point, the information that there was ever an ambiguity is lost. As we will see, the notion of case indeterminacy can help us solve the problem.

In proposing an LFG analysis of this complicated phenomenon, I will offer and compare two novel approaches: one within the traditional LFG framework and one that requires a variant of Optimality Theory within LFG in order to explain these phenomena. While the former has a certain appeal in its consistency with the basic tenets of LFG, including monotonicity, I argue that the latter is more elegant.

2 Russian word order freezing in LFG

2.1 Bloom’s account

Before offering my own analysis, it will be useful to understand Bloom (1999) and its implications for Russian—as well as why the analysis is not adequate for other languages. Russian allows quite free word order in transitive sentences since objects can be topicalized and since Russian word order is more related to information-structural categories rather than grammatical function, as in English. Recall the examples from (1) and (2). Bloom assumes that in sentences (1a-b), which have clearly marked case, there is a rule for constructive case attached to each nominal. In Bloom’s analysis for 1a, *Mama* would have (SUBJ ↑) and *papu* would have (OBJ ↑). That is, the former states “I am a subject” and the latter “I am an object.” Moreover, Bloom also attaches a (↑GF) = SUBJ rule to unambiguously nominative nominals and a (↑GF) = OBJ rule to unambiguously accusative ones. He claims that this is necessary to account for agreement, as in Nordlinger (1998).

So Bloom’s lexical entry for an unambiguously accusative noun *papu* ‘father’ is as below:

(3) papu N (↑PRED) = 'FATHER'
(↑GEND) = MASC
(↑NUM) = SG
(↑PERS) = 3
(↑CASE) = ACC
(↑GF) = OBJ
(OBJ ↑)

The c-structure rule for each NP states either that the GF or grammatical function is constructed internally, or it is specified structurally in the c-structure rules. The trees that Bloom gives differ from the one below in that the disjunction is not specified and he leaves out some of the lexical rules attached to each nominal. Having said that, given the information he presents elsewhere, I believe that these trees capture the intuition behind his approach.

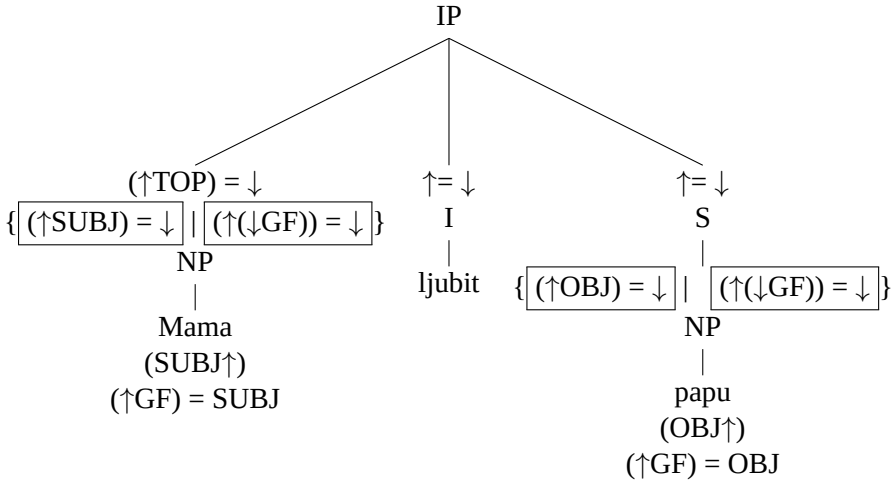
To understand how Bloom represents word order freezing in LFG, it will be useful to pursue the Russian examples along with their corresponding c-structures and f-structure in (4-7). The first two c-structures, (4a) and (5a), show the *differentiated* nominals for nominative and accusative. The rule {(↑SUBJ) = ↓ | (↑(↓GF)) = ↓} on the first nominal in the first two examples states "either I am my mother's subject or "my daughter has transferred a GF upward and I am that GF for my mother." So, in (4a), which is paired with (1a), *Mama* is unambiguously nominative and can thus declare itself a subject. Either side of the disjunction in {(↑SUBJ)= ↑ | (↑(↓GF))= ↓} can be taken with no difference in result. Likewise, *papu* is unambiguously accusative and declares itself an object. Again, either side of the disjunction {(↑OBJ)= ↑ | (↑(↓GF))= ↓} is permissible and a viable option without causing a contradiction.

In (1b), the order of subject and object is reversed. Nonetheless, *mama* still declares itself a subject (SUBJ↑) and *papu* declares itself an object (OBJ ↑). Thus, when faced with the disjunction {(↑SUBJ)= ↓ | (↑(↓GF))= ↓} for the NP headed by *papu*, it cannot be (↑SUBJ)=↓ because that would directly contradict (OBJ ↑). Therefore, the option taken must be (↑(↓GF))= ↓. In this case, (↓GF) = OBJ, so we get (↑OBJ) =↓ for the *papu* NP. A similar logic means that *mama* is still necessarily the subject even though it comes after the verb. All of this can be seen in (5).

Now, let us consider the word freezing instances in (2). *Mat'* and *doč'* are both ambiguous between nominative and accusative. Thus, they cannot project either (↑SUBJ) or (↑OBJ) via constructive case since it could just as easily project one as the other, and that would cause a contradiction. In other words, neither *Mat'* nor *doč* in the current system is able to mark itself as having any particular GF. Because no GF is passed up to the NP, the first respective options of the disjunctions {({(↑SUBJ)=↓ | (↑(↓GF))= ↓} and {(↑OBJ)=↓ | (↑(↓GF))= ↓} *must* be chosen since the second part is no longer an option. As a result, with these ambiguous instances in (2a) and (2b), the order is frozen as SVO. In (6), this works out fine. However,

(4) (Bloom, 1999, 65)

(a)

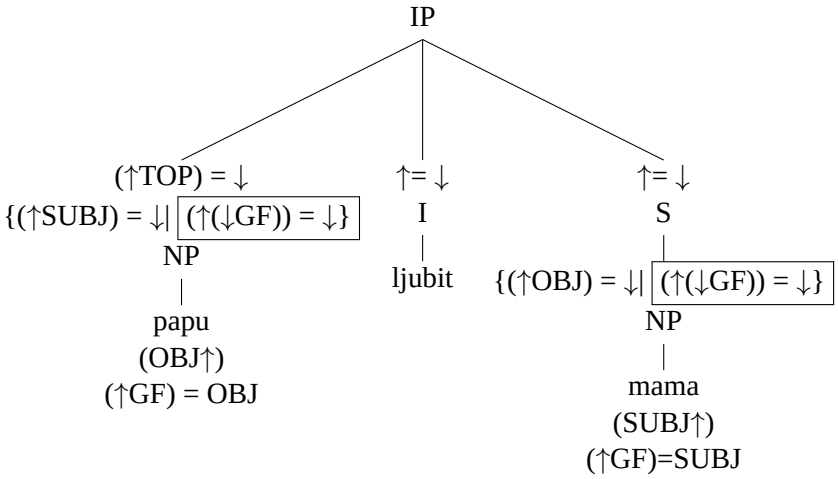


(b)

$$\begin{bmatrix} \text{PRED} & \text{'LOVE}\left\langle \left(\uparrow \text{SUBJ}\right), \left(\uparrow \text{OBJ}\right) \right\rangle \\ \text{TENSE} & \text{present} \\ \text{TOP} & [\ ] \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'mother'} \\ \text{GEND} & \text{fem} \\ \text{PERS} & 3 \\ \text{NUM} & \text{sg} \\ \text{CASE} & \text{nom} \\ \text{GF} & \text{SUBJ} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'father'} \\ \text{GEND} & \text{masc} \\ \text{PERS} & 3 \\ \text{NUM} & \text{sg} \\ \text{CASE} & \text{acc} \\ \text{GF} & \text{OBJ} \end{bmatrix} \end{bmatrix}$$

(5) (Bloom, 1999, 66)

(a)

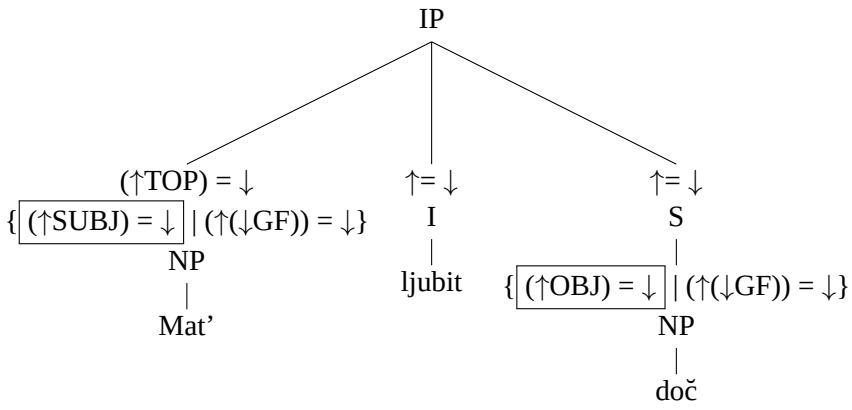


(b)

$$\begin{bmatrix}
\text{PRED} & \text{'LOVE}\left\langle\left(\uparrow\text{SUBJ}\right),\left(\uparrow\text{OBJ}\right)\right\rangle \\
\text{TENSE} & \text{present} \\
\text{TOP} & [\ ] \\
\text{SUBJ} & \begin{bmatrix}
\text{PRED} & \text{'mother'} \\
\text{GEND} & \text{fem} \\
\text{PERS} & 3 \\
\text{NUM} & \text{sg} \\
\text{CASE} & \text{nom} \\
\text{GF} & \text{SUBJ}
\end{bmatrix} \\
\text{OBJ} & \begin{bmatrix}
\text{PRED} & \text{'father'} \\
\text{GEND} & \text{masc} \\
\text{PERS} & 3 \\
\text{NUM} & \text{sg} \\
\text{CASE} & \text{acc} \\
\text{GF} & \text{OBJ}
\end{bmatrix}
\end{bmatrix}$$

(6) (Bloom, 1999, 67)

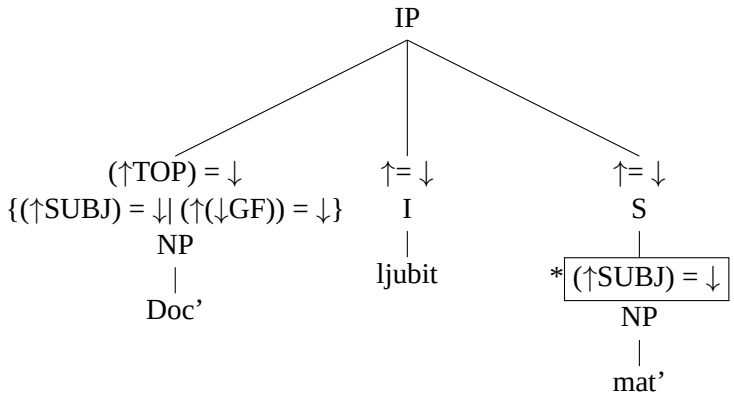
(a)



(b)

$$\begin{bmatrix}
\text{PRED} & \text{'LOVE}\left\langle (\uparrow\text{SUBJ}), (\uparrow\text{OBJ}) \right\rangle\text{'} \\
\text{TENSE} & \text{present} \\
\text{TOP} & [\ ] \\
\text{SUBJ} & \begin{bmatrix}
\text{PRED} & \text{‘mother’} \\
\text{GEND} & \text{fem} \\
\text{PERS} & 3 \\
\text{NUM} & \text{sg} \\
\text{CASE} & \{\text{nom, acc}\}
\end{bmatrix} \\
\text{OBJ} & \begin{bmatrix}
\text{PRED} & \text{‘daughter’} \\
\text{GEND} & \text{fem} \\
\text{PERS} & 3 \\
\text{NUM} & \text{sg} \\
\text{CASE} & \{\text{nom, acc}\}
\end{bmatrix}
\end{bmatrix}$$

(7) (Bloom, 1999, 68)
(a)ungrammatical construction that arises when the second NP in this example is forced to act as the subject



(b)*

$$\begin{bmatrix} \text{PRED} & \text{'LOVE}\left\langle (\uparrow\text{SUBJ}), (\uparrow\text{OBJ}) \right\rangle \\ \text{TENSE} & \text{present} \\ \text{TOP} & [\ ] \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{‘mother’} \\ \text{GEND} & \text{fem} \\ \text{PERS} & 3 \\ \text{NUM} & \text{sg} \\ \text{CASE} & \{\text{nom, acc}\} \end{bmatrix} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{‘daughter’} \\ \text{GEND} & \text{fem} \\ \text{PERS} & 3 \\ \text{NUM} & \text{sg} \\ \text{CASE} & \{\text{nom, acc}\} \end{bmatrix} \end{bmatrix}$$

the c-structure paired with the f-structure in (7), in which the second nominal is the subject, is not possible since that would require taking the first option of the disjunction {(↑SUBJ)=↓ | (↑(↓GF))= ↓}, which we have already said cannot be done. In the c-structures that illustrate this, note that I have boxed the part(s) of the disjunction selected for each construction represented.

Bloom's analysis apparently works satisfactorily for Russian in which an ambiguously marked nom/acc object cannot appear pre-verbally in the topic position even when the other nominal element *is* unambiguously marked.[1] Consider (8a-c), also from Bloom (1999). In these examples, the object is ambiguous in case and, regardless of the subject's case status, it still cannot be topicalized unproblematically. The same cannot be said, however, for other languages. This lacuna will be returned to in Section 3.

(8) (a) *Mal'čik* *videl* *lošad'*
boy-NOM saw horse-nom/acc
'the boy saw the horse'

(b) ?*Lošad'*$_{[TOPIC]}$ *mal'čik* *videl.*
horse-nom/acc boy-NOM saw
? 'The boy saw the horse'

(c) ?*Lošad'*$_{[TOPIC]}$ *videl* *mal'čik.*
horse-nom/acc saw boy-NOM
? 'The boy saw the horse'

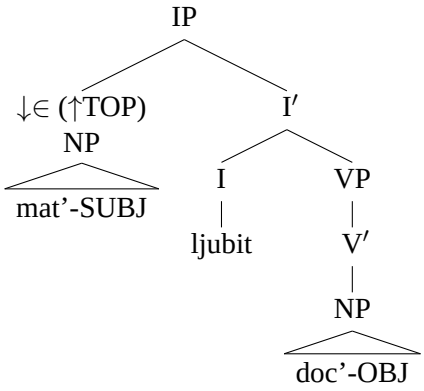
2.2 Information-structural account of Russian freezing

In personal communication, Louise Mycock has suggested that Bloom's analysis fails to give appropriate consideration to information structural factors. That is, she joins King (1995) in rejecting the idea that grammatical functions like SUBJ and OBJ even have structurally assigned positions in Russian c-structure rules. Rather, it is information-structural constraints on discourse functions that determine word order.

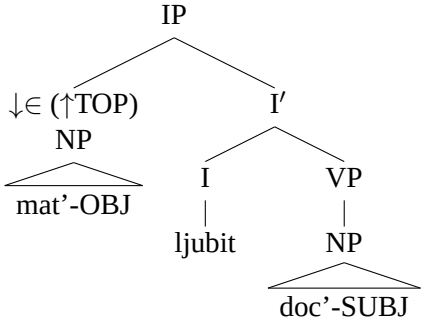
For King (1995), the default order for Russian is actually V-first. The fact that a nominal frequently appears pre-verbally is an artifact of there being a TOPIC position in Spec,IP that is marked ↓ ∈ (↑TOP) and (↑GF) = ↓. These rules suggest that the doubly ambiguous Russian sentences could receive trees as in (9) and (10), which are based on similarly structured trees for different sentences in King (1995, 206, 224).

[1] Again, note the exceptions pointed out in King's work.

(9) SVO



(10) OVS (assumes that the NP1 is marked by context as OBJ)



Contrary to what Bloom claims, the grammatical function in these trees is not determined by c-structure. Rather, particular c-structure positions are associated with particular discourse functions. This still allows a morphologically ambiguous sentence to be disambiguated. In a completely unmarked context in which it is not immediately clear which element is being used as a topic, Lambrecht (1994, 132) provides a way to tell. He posits both that topic followed by comment[2] is a default pragmatic ordering and that subjects are topics in an unmarked context. In Russian, we already know from King (1995) that the topic appears in Spec,IP. We add to that knowledge the fact that, cross-linguistically, topics are usually subjects. We then arrive at a way to disambiguate an ambiguous clause. That is, if it is not clear from context which of two nominals is acting as SUBJ, it can be assumed that it is the one that is also TOPIC. This analysis helps clear up an example like (9): it is natural that the first of the two ambiguous nominals will be taken to be both TOPIC and SUBJ, whereas the second NP will be taken to be the OBJ and part of the comment.

How then can we explain (10)? This is a structure that Bloom's analysis explicitly disallows. But, as King (1995, 2, fn. 2) points out, discourse effects can easily override the freezing effects. If indeed the first nominal is clearly an object from the context of the sentence, then we can conclude that it is both an OBJ *and* a TOPIC since it is in TOPIC position. Even though this is less common and less

[2] *Comment*, in effect, refers to that which is being said about the topic. See Lambrecht (1994).

canonical than a configuration in which the topic is a subject, this construction is perfectly legitimate provided that the context calls for it. There is no need to turn to c-structural GF assignment as in Bloom. But Lambrecht's idea that SUBJ and TOPIC are interwoven in this way makes it easy to see how Bloom comes to associate Spec,IP with (↑GF) = (↑SUBJ).[3]

3 Dutch word order freezing

Unlike with Russian, in Dutch only when *both* nominal arguments of a transitive verb are ambiguous for case *and* there is no other disambiguating information does the word order freeze. Let's examine the following Dutch sentences from Bouma (2011).[4]

(11) (a) *De Rode Duivels verslaan Oranje.*
The Red Devils.PL beat.PL Orange.SG
'The Red Devils beats the Dutch national football team.'

(b) *De Rode Duivels verslaat Oranje.*
The Red Devils.PL beat.SG Orange.SG
'The Dutch national football team beats the Red Devils.'

(c) *België verslaat Oranje.*
Belgium.SG beat.SG Orange.SG
'Belgium beats the Dutch national football team.'[5]
NOT: 'The Dutch national football team beats Belgium.'

Note that, while *De Rode Duivels* and *Oranje* are both themselves ambiguous for case, the distinction between (a) and (b) suggests that the mere presence of disambiguating verb agreement is enough to allow unambiguous meanings. In the case of (c), however, in which the verb is of no help in deciding what is the subject and what is the object, we see word order freezing. The SVO order is frozen.

Bouma (2011) handles this through a rather complicated bidirectional OT analysis. Still, it should be readily apparent that Bloom's LFG analysis cannot be applied to Dutch in the same way that we saw for Russian. Likewise, applying Bloom's c-structure and lexical rules to Old English or Dutch would cause freezing to occur where it need not: for instance, in a sentence with two nom/acc ambiguous nominals but in which verb agreement means that only one can be the subject. Fundamentally, we do not wish to enforce word order freezing only when we have ambiguous case. As I stated previously, this presents a challenge for a monotonic framework like LFG.

[3]See also Alsagoff (1992) for an LFG account of how SUBJ and TOPIC are linked in Malay.

[4]Bouma points out that he is assuming a "hat" (rising on the first nominal and falling on the second) intonation pattern for all three of these sentences. Without that intonation, the freezing effect will not necessarily be achieved.

[5]The Dutch national football team is known as The Orange.

4 Word order freezing in LFG

4.1 Case underspecification

So how can we handle this data in LFG? I will ultimately come down on the side of using a special variant of OT. First, though, I will explore how freezing could be handled in LFG without the use of OT constraints. In doing so, I will draw on Dalrymple et al. (2009), who reject a traditional account of case ambiguity and instead turn to indeterminacy by feature underspecification. That is, instead of representing an ambiguous nominal as {(↑CASE) = NOM | (↑CASE) = ACC}, they represent case by features. The Russian ambiguously nom/acc *mat'* 'mother' would have the features (↑CASE GEN) = − and (↑CASE DAT) = − and so on. The feature values for (↑CASE NOM) and (↑CASE ACC) would remain unspecified if and until other information came along to specify it. For instance, a nominative demonstrative could assign the values(↑CASE NOM) = + and (↑CASE ACC) = −.

If no such disambiguating features are present, then the noun would remain underspecified for case. With this proposal, Dalrymple et al. (2009) solve the so-called transitivity problem exemplified in German in (12). *Hilft* takes a dative object and *Papageien* is ambiguously acc/dat. The noun *Papageien* could meet the verb's dative requirement if its case was (↑CASE) = {ACC, DAT} *and* meet the determiner *die*'s accusative requirement with the same constraint (↑CASE) = {ACC, DAT}. This is problematic.

(12) *Er hilft *die/den Papageien*
he helps *the-ACC/the-DAT parrots-ACC/DAT
'he helps the parrots'

The feature underspecification approach would state initially that *Papageien* is −GEN and −NOM, which would allow it to be either accusative or dative. The presence of a dative verb would then add the feature −ACC. Thus, the determiner in this example could not be *die*-ACC since that would make the nominal impossibly −ACC as well as +ACC. This approach also solves the second-order indeterminacy problem, whose details can be found in the 2009 paper.

4.2 Word order freezing rules

Through feature underspecification, it is possible to construct a series of rules that model word order freezing. For purposes of simplicity, I will address only nom/acc ambiguity in the examples below, but the rules could in theory be extended for any other type of case ambiguity. To summarize, each nominal can be either a) differentiated for case or b) ambiguous between nominative and accusative. We will assume that both nominals are morphologically negative for dative and genitive. Crucially, the freezing effect takes place only if both nominals match the number requirement imposed by the verb—a point we will return to below. A simplified version of the rules attached to a NOM/ACC verb appears in (13). Note that we

could use GF1 and GF2 in these rules in place of SUBJ and OBJ, but it is equivalent and easier to explain the rules when we treat one GF as SUBJ and one as OBJ.

(13) $\left\{ (\uparrow \text{SUBJ CASE acc}) =_c - \;\middle|\; \left\{ (\uparrow \text{OBJ CASE nom}) =_c - \;\middle|\; \begin{array}{c} \neg(\uparrow \text{SUBJ CASE acc}) = - \\ \neg(\uparrow \text{OBJ CASE nom}) = - \\ \text{SUBJ} <_f \text{OBJ} \end{array} \right\} \right\}$

if-then form:
- if $\neg(\uparrow \text{SUBJ CASE acc}) = -$
  - then if $\neg(\uparrow \text{OBJ CASE nom}) = -$
    - then $\text{SUBJ} <_f \text{OBJ}$

A bit of explanation will help clarify these rules. The first part of the first disjunction is $(\uparrow \text{SUBJ CASE acc}) = -$. If this is true (and we are assuming that both nominals are negative for dative and genitive) for the first nominal (GF1), then the only thing that it can rightfully be without contradiction is the subject. This is an instance where there would be no ambiguity. That is, if the nominal represented by GF1 is negative for accusative case, the other disjuncts need not apply since there is no ambiguity possible even if the OBJ is ambiguous for case.

The second half of the main disjunction assumes that the SUBJ is not in fact negative for accusative.[6] That said, as long as $(\uparrow \text{OBJ CASE nom}) = -$ holds, we still escape ambiguity.

However, if it is not the case that $(\uparrow \text{OBJ CASE nom}) = -$, we are forced to choose the right side of the smaller disjunction and the freezing effect occurs. When we have the two-way ambiguous condition, the SUBJ is stipulated through f-precedence as preceding the OBJ.[7]

Note, however, that this still does not solve the problem of how to ensure that the freezing effect obtains only when there is no other agreement information, like number, to disambiguate. By adding extra specifications to the rules, though, we can do just that. An example is shown in (14), where we assume that the rule is attached to a verb that is constrained to have a singular subject:$(\uparrow \text{SUBJ NUM}) =_c$ sg.

(14) $\left\{ (\uparrow \text{SUBJ CASE acc}) =_c \text{-} \;\middle|\; \left\{ (\uparrow \text{OBJ CASE nom}) =_c \text{-} \;\middle|\; \begin{array}{c} \neg(\uparrow \text{SUBJ CASE acc}) = \text{-} \\ \neg(\uparrow \text{OBJ CASE nom}) = \text{-} \\ \left\{ (\uparrow \text{OBJ NUM}) =_c \text{pl} \;\middle|\; (\uparrow \text{OBJ NUM}) =_c \text{sg} \right\} \\ \text{SUBJ} <_f \text{OBJ} \end{array} \right\} \right\}$

[6]The negated construction $\neg(\uparrow \text{SUBJ CASE ACC}) = -$ is used rather than $(\uparrow \text{SUBJ CASE ACC}) = +$ because it is not necessary that a positive value be instantiated in order for ambiguity to occur. It only has to be not instantiated as negative. The only thing that can instantiate ACC = + is the verb, and the presence of other disambiguating factors, like an accusative demonstrative, would serve merely to eliminate all the case options except ACC.

[7]The notation used here indicates f-precedence. F-precedence is defined by Kaplan and Zaenen (2003) as follows: "For any f-structures f and g, f f-precedes g(written $f <_f g$) if and only if all the c-structure nodes that map to f precede all the c-structure nodes that map to g." F-precedence has important theoretical implications for LFG since left-to-right ordering rules typically apply only in the c-structure but not in the f-structure. The use of f-precedence allows ordering constraints to flow between the c-structure and the f-structure. See also Zaenen and Kaplan (1995).

Let's see how this works. For an example in which both nominals are unambiguous, like (1a), the left side of the disjunction applies and there is no freezing effect:

(15) *Mama* *ljubit* *papu*
Mother-NOM loves father-ACC
'The mother loves the father'

$$\left\{ \boxed{(\uparrow \text{SUBJ CASE acc}) =_c \text{-}} \;\middle|\; \left\{ (\uparrow \text{OBJ CASE nom}) =_c \text{-} \;\middle|\; \left\{ \begin{array}{c} \neg(\uparrow \text{SUBJ CASE acc}) = \text{-} \\ \neg(\uparrow \text{OBJ CASE nom}) = \text{-} \\ (\uparrow \text{OBJ NUM}) =_c \text{pl} \;\middle|\; \begin{array}{c}(\uparrow \text{OBJ NUM}) =_c \text{sg} \\ \text{SUBJ} <_f \text{OBJ}\end{array} \end{array} \right\} \right\} \right\}$$

What happens to our Dutch example where number agreement serves to block the freezing effect? Compare (16) and (17). In the former, we are forced to take the "ambiguous" options for SUBJ CASE and OBJ CASE. However, because the object satisfies the constraint (↑OBJ NUM) $=_c$ pl (and we are dealing with a verb that requires a singular subject), the freezing effect does not obtain.

On the other hand, (17) shows a situation in which both nominals are ambiguous but in which (↑OBJ NUM) $=_c$ sg is satisfied. As a result, the f-precedence freezing rule is in effect. In a similar way, any other sort of agreement, like person or animacy, can be taken into account by adding it into the rules.

(16) *De* *Rode* *Duivels* *verslaat* *Oranje.*
The Red Devils.PL beats.SG the Orange.SG
'The Dutch national football team beats the Red Devils.'

$$\left\{ (\uparrow \text{SUBJ CASE acc}) =_c \text{-} \;\middle|\; \left\{ (\uparrow \text{OBJ CASE nom}) =_c \text{-} \;\middle|\; \left\{ \begin{array}{c} \boxed{\neg(\uparrow \text{SUBJ CASE acc}) = \text{-}} \\ \boxed{\neg(\uparrow \text{OBJ CASE nom}) = \text{-}} \\ \boxed{(\uparrow \text{OBJ NUM}) =_c \text{pl}} \;\middle|\; \begin{array}{c}(\uparrow \text{OBJ NUM}) =_c \text{sg} \\ \text{SUBJ} <_f \text{OBJ}\end{array} \end{array} \right\} \right\} \right\}$$

(17) *België* *verslaat* *Oranje.*
Belgium.SG beat.SG Orange.SG
'Belgium beats the Dutch national football team.'

NOT: The Dutch national football team beat Belgium.

$$\left\{ (\uparrow \text{SUBJ CASE acc}) =_c \text{-} \;\middle|\; \left\{ (\uparrow \text{OBJ CASE nom}) =_c \text{-} \;\middle|\; \left\{ \begin{array}{c} \boxed{\neg(\uparrow \text{SUBJ CASE acc}) = \text{-}} \\ \boxed{\neg(\uparrow \text{OBJ CASE nom}) = \text{-}} \\ (\uparrow \text{OBJ NUM}) =_c \text{pl} \;\middle|\; \begin{array}{c}\boxed{(\uparrow \text{OBJ NUM}) =_c \text{sg}} \\ \boxed{\text{SUBJ} <_f \text{OBJ}}\end{array} \end{array} \right\} \right\} \right\}$$

Of course, this could make the rules quite unwieldy. For instance, there can also be accusative/dative ambiguity. Thus, every verb that takes three arguments—a nominative subject, an accusative object, and a dative object—would have an extremely complicated set of rules attached to it. In the next section, I will show how OT can offer a more elegant solution.

It is worth noting here that this analysis is entirely consistent with the information-structural account of word order freezing presented for Russian as an

alternative to Bloom (1999). For Russian, we can easily modify the rules to reflect that Spec,IP (which is typically where the first nominal is found) is not necessarily associated with (↑SUBJ) at all but is simply a TOPIC position. If we then accept the Lambrecht argument that SUBJ and TOPIC are linked by default, the "frozen" SVO order is no longer a mystery but is merely a by-product of the information-structural effect.

5 Word order freezing in LFG with OT

OT points the way to a simpler treatment. The OT that I implement here is not exactly Optimality Theory in the classical sense, as in Prince and Smolensky (2004). Rather, it is based on the XLE implementation of OT constraints. Crouch et al. (2008) describe the OT options in XLE as an instantiation of the "most common mechanism used in Optimality Theory."[8] For word order freezing, I propose that we institute a weak constraint for SUBJ $<_f$ OBJ that could be introduced and applied *only* in instances in which no unambiguous parse is derived for a given sentence. Although this arguably violates LFG's commitment to monotonicity, it is a far simpler way to explain word order freezing.

We can implement this constraint through a simple OT tableau.[9] For each sentence, we can imagine a tableau like the one shown in (18).

(18)

	Consistency	SUBJ $<_f$ OBJ
a		
b		

A and B represent parses of the same sentence. The first column represents the dominant constraint and stands for consistency in the LFG sense. The second column represents the less highly ranked constraint that SUBJ $<_f$ OBJ. In OT terms Consistency »» SUBJ $<_f$ OBJ. Thus, if consistency is violated at all in a given analysis, that analysis is immediately eliminated from contention. If consistency is satisfied for both parses, but one parse has SUBJ $<_f$ OBJ and the other has OBJ $<_f$ SUBJ, the optimal analysis is the one in which SUBJ $<_f$ OBJ.

Let's return to the examples and f-structures from the Dutch examples from above. For a sentence like (19) where number agreement can disambiguate the nominals, we get the two possible f-structures shown in (20). (20a) violates the SUBJ $<_f$ OBJ constraint that states that the subject should precede the object. The f-structure (20b), on the other hand, does indeed have SO order but violates consistency since its verb requries a singular subject but its subject is plural. This, of course, is a crucial violation. Thus, (20a) is accepted and (20b) is dispreferred.

[8] In another departure from traditional OT, XLE allows preference marks alongside dispreference marks.

[9] Besides Prince and Smolensky (2004), also see Kager (1999) for a general introduction and Kuhn (2003) for a specifically syntactic approach to OT.

(19) *De Rode Duivels verslaat Oranje.*
The Red Devils.PL beats.SG the Orange.SG
'The Dutch national football team beats the Red Devils.'

(20) (a)
$$\begin{bmatrix} \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'Orange'} \\ \text{CASE} & \text{nom} \\ \text{PERS} & 3 \\ \text{NUM} & \text{sg} \end{bmatrix} \\ \text{PRED} & \text{'beat}\left\langle(\uparrow\text{SUBJ})(\uparrow\text{OBJ})\right\rangle\text{'} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'Red Devils'} \\ \text{CASE} & \text{acc} \\ \text{PERS} & 3 \\ \text{NUM} & \text{pl} \end{bmatrix} \end{bmatrix}$$
(b) *
$$\begin{bmatrix} \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'Red Devils'} \\ \text{CASE} & \text{nom} \\ \text{PERS} & 3 \\ *\text{NUM} & \text{sg/pl} \end{bmatrix} \\ \text{PRED} & \text{'beat}\left\langle(\uparrow\text{SUBJ})(\uparrow\text{OBJ})\right\rangle\text{'} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'Orange'} \\ \text{CASE} & \text{acc} \\ \text{PERS} & 3 \\ \text{NUM} & \text{pl} \end{bmatrix} \end{bmatrix}$$

(21)

	Consistency	SUBJ $<_f$ OBJ
☞(20a)		*
20b	*!	

(22) shows an altogether different situation. Here, there is no help given by number agreement since the verb and both nominals are singular. Neither f-structure in (23) violates consistency. But (22b) violates the SUBJ $<_f$ OBJ constraint since its object comes before the subject. Because there are no other violations, this becomes a crucial violation and we prefer (22a). This very simple OT analysis achieves what we otherwise need quite complex LFG constraints to handle.

(22) *België verslaat Oranje.*
Belgium.SG beat.SG Orange.SG
'Belgium beats the Dutch national football team.'

(23) (a)
$$\begin{bmatrix} \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'Belgium'} \\ \text{CASE} & \text{nominative} \\ \text{PERS} & 3 \\ \text{NUM} & \text{sg} \end{bmatrix} \\ \text{PRED} & \text{'beat}\left\langle(\uparrow\text{SUBJ})(\uparrow\text{OBJ})\right\rangle\text{'} \\ \text{TENSE} & \text{present} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'Orange'} \\ \text{CASE} & \text{acc} \\ \text{PERS} & 3 \\ \text{NUM} & \text{sg} \end{bmatrix} \end{bmatrix}$$
(b) *
$$\begin{bmatrix} \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'Orange'} \\ \text{CASE} & \text{nominative} \\ \text{PERS} & 3 \\ \text{NUM} & \text{sg} \end{bmatrix} \\ \text{PRED} & \text{'beat}\left\langle(\uparrow\text{SUBJ})(\uparrow\text{OBJ})\right\rangle\text{'} \\ \text{TENSE} & \text{present} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'Belgium'} \\ \text{CASE} & \text{acc} \\ \text{PERS} & 3 \\ \text{NUM} & \text{sg} \end{bmatrix} \end{bmatrix}$$

(24)

	Consistency	SUBJ $<_f$ OBJ
☞(23a)		
(23b)		*!

6 Conclusion

Ultimately, the word order freezing data can be explained effectively through either the OT constraints in Section 5 or through the pure LFG description in Section 2. Nonetheless, questions remain as to whether word order freezing is a syntactic phenomenon at all. Bouma (2011) provides an overview of a debate between those who see it as syntactic and those who see it as a processing constraint. Those that view freezing as a processing constraint point to the fact that it can easily be overcome by prosodic and pragmatic factors. Indeed, it can be a relatively weak constraint. Bouma, however, notes that the effect manifests itself in language as soon as morphological case fails to disambiguate the reading. He proposes a bidirectional OT model of language by which both the speaker and the hearer conspire to avoid meaningful ambiguity. In brief, it more or less states that, given a free word order language, if the hearer does not have morphological or other information to disambiguate the clause, she will assume that the speaker is using word order cues. The speaker will make the same observation, and both parties will arrive at an unambiguous interpretation of the clause.

Tily (2010) provides experimental evidence for this process in Japanese. In a series of timed reading experiments, he finds that word order has little effect on clauses in which case is unambiguously marked. When case markings are removed, however, the reader slows down considerably when faced with a pre-subject object. I do not have space to do justice to the work, but the evidence strongly suggests that comprehenders turn to word order when and only when case marking is ambiguous. The OT approach that I have presented models this process in a way that is quite intuitive. That is, we could say that, only when comprehenders fail to encounter a crucial violation of consistency do they turn to the weaker subject-object constraint. If morphology fails to deliver the necessary information, the speaker and comprehender will turn to word order even during online processing. That said, as Sag and Wasow (2007) and others note, most psycholinguistic evidence favors a model-theoretic approach to language in order to account for the effects seen in online processing—an approach not consistent with the OT approach presented here. In that sense, the "pure" LFG approach is perhaps worth considering, even though it requires a much more complex set of constraints. Further work is needed to assess the two approaches in terms of both psychological reality as well as computational tractability.

Regardless of which treatment emerges as the most satisfactory, I believe that this paper has demonstrated the possibilities available for analyzing word order freezing syntactically in a way that is effective and easily implementable in parsing software like XLE.

References

Alsagoff, Lubna. 1992. *Topic in Malay: The Other Subject*. Ph. D.thesis, Stanford University.

Bloom, Douglas. 1999. *Case Syncretism and Word Order Freezing in the Russian Language*. Masters Thesis, Stanford University.

Bouma, Gerlof. 2008. *Starting a sentence in Dutch: A Corpus Study of Subject- and Object-Fronting*. Ph. D.thesis, University of Groningen.

Bouma, Gerlof. 2011. Production and comprehension in context: The case of word order freezing. In *Bidirectional Optimality Theory*, Amsterdam: John Benjamins.

Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Oxford: Blackwell Publishers.

Crouch, Dick, Dalrymple, Mary, Kaplan, Ron, King, Tracy, Maxwell, John and Newman, Paula. 2008. *XLE User Documentation*. Palo Alto Research Center.

Dalrymple, Mary. 2001. *Lexical Functional Grammar*. San Diego, CA: Academic Press.

Dalrymple, Mary, King, Tracy Holloway and Sadler, Louisa. 2009. Indeterminacy by underspecification. *Journal of Linguistics* 45, 31–68.

Jakobson, Roman. 1963. Implications of language universals for linguistics. *Universals of Language* pages 263–278.

Kager, René. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.

Kaplan, Ronald and Zaenen, Annie. 2003. West-Germanic verb clusters in LFG. In Pieter A. M. Seuren and Gerard Kempen (eds.), *Verb Constructions in German and Dutch*, Amsterdam: John Benjamins B.V.

King, Tracy Holloway. 1995. *Configuring Topic and Focus in Russian*. Stanford: CSLI publications.

Kuhn, Jonas. 2003. *Optimality-Theoretic Syntax - A Declarative Approach*. Stanford: CSLI Publications.

Lambrecht, Knud. 1994. *Cambridge Studies in Linguistics 71: Information Structure and Sentence Form: Topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.

Lee, Hanjung. 2001. Markedness and word order freezing. In Peter Sells (ed.), *Formal and Emperical Issues in Optimality Theoretic Syntax, volume 5 of Studies in Constraint based Lexicalism.*, Stanford: CSLI Publications.

Mahowald, Kyle. 2011. *An LFG Approach to Old English Constituent Order*. Masters Thesis, University of Oxford.

Nordlinger, Rachel. 1998. Complementizing case in Australian languages: a constructive approach. In *LFG98 Conference, University of Queensland, Brisbane*.

Potts, Kathryn Flack. 2007. Ambiguity avoidance as contrast preservation: Case and word order freezing in Japanese. In *UMass Occasional Papers in Linguistics 32: Papers in Optimality Theory III*.

Prince, Alan and Smolensky, Paul. 2004. Optimality Theory: Constraint interaction in generative grammar. In John J. McCarthy (ed.), *Optimality Theory in phonology: A reader*, Oxford: Blackwell.

Sag, Ivan and Wasow, Tom. 2007. Performance-Compatible Competence Grammar. *Non-Transformational Syntax* pages 359–377.

Tily, Harry. 2010. *The Role of Processing Complexity in Word Order Variation and Change*. Ph. D.thesis, Stanford University.

Zaenen, Annie and Kaplan, Ronald. 1995. Formal Devices for Linguistic Generalizations: West Germanic Word Order in LFG. In Mary Dalrymple, John Maxwell, Ronald Kaplan and Annie Zaenen (eds.), *Formal Issues in Lexical Functional Grammar*, pages 215–239, Stanford: CSLI Publications.