

**ISSUES CONCERNING CONSTRAINTS ON DISCONTINUOUS
NPS IN LATIN**

Liselotte Snijders
University of Oxford

Proceedings of the LFG12 Conference

Miriam Butt and Tracy Holloway King (Editors)

2012

CSLI Publications

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

Abstract

In this paper, two main issues concerning discontinuity of Latin noun phrases (NPs) will be discussed from a lexical-functional grammar (LFG) perspective. It is often assumed that Latin word order is free and that therefore discontinuity is not subject to any constraints. However, it appears that the discontinuity of Latin NPs is in fact constrained. Two constraints on this type of discontinuity form problems for LFG as a theory of syntax. For the first constraint, one on discontinuous prepositional phrases (PPs), an exception to the principle of Economy of Expression is proposed. This paper claims that c-structure should be more constrained in LFG. A second constraint on Latin discontinuous adjuncts shows that the way in which LFG treats adjuncts (with a set notation) is problematic for a proper account of discontinuous adjuncts. This paper provides an initial discussion of this issue.

1 Introduction

Latin is well known to be a ‘free word order language’, a language with no fixed word order patterns. This variation in word order is allowed because the specification of grammatical relations in the sentence is not determined by the ordering of constituents, but rather by morphology. This means that Latin also allows a great deal of discontinuity in its phrase structure (Panhuis, 1982; Spevak, 2010). It is often believed that Latin allows discontinuity of phrases in an unlimited way. However, as shown by Bolkestein (2001), discontinuity in Latin is in fact constrained. Bolkestein found three absolute constraints on the discontinuity of NPs in Latin, two of which will be the focus of this paper.¹ Discontinuity (and Latin word order in general) is highly determined by information structure. In this paper, the focus will not be on these information structural tendencies, but only on strict constraints. LFG is generally capable of accounting for discontinuous phrases, as is shown for example in the work by Simpson (1991) on Warlpiri. However, two of the three constraints found by Bolkestein pose problems for LFG. These two constraints will be addressed in this paper. For the first constraint, involving prepositional phrases, this paper suggests an exception to Economy of Expression. It appears that Economy of Expression allows for too much freedom in word order when it comes to discontinuity of NPs, and this freedom needs to be restricted. A second constraint, involving discontinuous adjuncts, brings forth another problem; it shows that the way in which LFG treats adjuncts is not capable of accounting for discontinuous adjuncts at all in any language. This is highly problematic for LFG as a theory of syntax, and it appears that there is no proper solution at the moment. This paper will discuss the two main issues connected to these two constraints; it will illustrate how they are problematic for LFG and how they can potentially be accounted for.

Before turning to the real issues at hand, a note needs to be made of two issues concerning the use of Latin in linguistic research. Latin is a dead language and therefore only limited data is available. Due to the status of the (classical) Latin that is available to us (literary Latin), we can only draw conclusions on this type of Latin, not on Latin as a whole. Spoken Latin might have been somewhat different from the variety of Latin which is available to us and being used for academic research. However, this type of Latin is interesting in its own right. It is merely important to keep in mind that we are dealing with a specific variety of Latin. Secondly, assignment of grammaticality

[†] I would like to thank Mary Dalrymple for her help and her supervision in the writing of my master’s dissertation (Snijders, 2012), which has led to this paper. Also her book (Dalrymple, 2001) has helped a great deal in the formation of my dissertation and this paper.

¹For a full overview of the constraints and their LFG analysis, please see my master’s dissertation (Snijders, 2012). This paper is an outgrowth of this dissertation. One of the constraints is not discussed here because its analysis is relatively straightforward in LFG and does not present any problems to the theory.

in a dead language is different than assignment of grammaticality in languages spoken today. Any grammaticality judgement of a dead language is based on whether a specific sentence is attested in the data or not. Unattested sentences are assumed to be ungrammatical (only if the data set is large enough). Examples in this paper which are marked as ungrammatical are thus assumed to be ungrammatical because they were not found in the data set.²

2 Discontinuity of NPs in Latin

There are two different types of discontinuous NPs in Latin, as defined by Spevak (2010).³ The first kind is an ‘obligatory’ discontinuous NP in which the intervening element occurs in an obligatory position in the sentence. This happens with certain connective particles (e.g. *autem*, meaning ‘indeed’), which always occur in the second position in the sentence. These particles can therefore split up an NP, if this NP occurs at the beginning of the sentence. The phenomenon of clitics appearing obligatorily in second position is very interesting, especially in relation to how it is involved in discontinuity. An LFG account of second position clitics in another dead language, Ṛgvedic Sanskrit, can be found in Lowe (2011). Because the type of discontinuity caused by an intervening second position clitic is obligatory and appears to be a specific rule for Latin, this is not relevant for the discussion at hand. The other type defined by Spevak, and the one relevant for this paper is the type of discontinuous NP with an ‘alien’ element intervening, something which is neither related to the NP nor appearing in its position obligatorily. This type of discontinuity is strongly determined by information structure. This type of discontinuous NP will be the only type considered in this paper. From here on, when talking about ‘discontinuous NP’s, the term will refer to discontinuous NPs with intervening alien elements.

Discontinuous NPs are quite common in Latin; according to Pinkster (2005), 12% of all NPs in Latin are discontinuous. Here a discontinuous NP refers to an (f-structure) NP with two or more parts which are separated in phrase structure (c-structure). As mentioned, the intervening element is not part of this (f-structure) NP; it is an ‘alien’ element (for example a verb). An example of this in Latin is the following:⁴

- (1) ...*a* *qua* *ego* *nullum* *confiteor* *aetatis* *meae*
of+ABL which.ABL I.NOM no.ACC admit.1SG lifetime.GEN my.GEN
tempus *abhorruisse*...
time.ACC deter.INF.PASS
‘... of which I admit that at no point in my lifetime I have been deterred...’
(Cic. *Arch.* 1)

²Grammaticality in a language such as Latin (classical Latin in this case) is also dependent on the specific author, since some authors will allow certain constructions and ordering of constituents, whereas others may not. Therefore it is always important to keep in mind which data set is used for research of the kind in this paper. Most canonical texts of Classical Latin (often used for linguistic research) are available in the Perseus Digital Library, an open-access corpus which at the moment contains 5.5 million words of Classical Latin (<http://www.perseus.tufts.edu/>).

³Spevak (2010) defines three types of discontinuous NPs, but one of them is one in which the intervening element is a modifier of the larger NP. This is not a real type of discontinuity, and was therefore not taken into account in this paper.

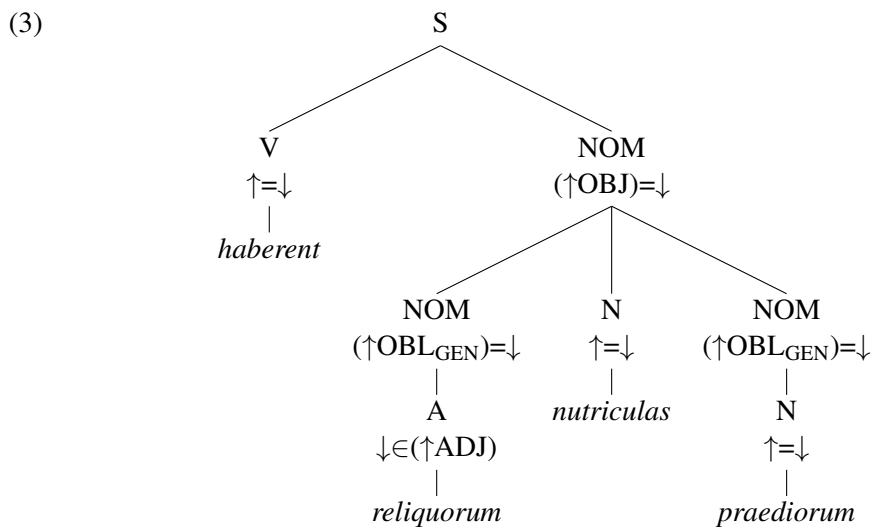
⁴The glossing ‘of+ABL’ means that the preposition *a* has the meaning ‘of’ and takes a complement with ablative case. In Latin prepositions specify the case of their objects.

In this example, the two parts of the NP *nullum tempus* are separated in phrase structure.⁵ The intervening elements are a genitive modifier (which is technically part of the larger NP) and the main verb of the sentence, which is unrelated to the NP. The two parts of the NP can be interpreted to be part of the same (f-structure) NP, shown by the fact that they have the same case marking (accusative in this case). Latin's system of morphology thus allows for discontinuity to occur. In this case the two separated parts of the NP are a determiner and a noun, but they can be two different elements, for example an adjective and a noun.

3 Discontinuity in LFG

In this paper, the approach by Simpson (1991) accounting for discontinuous NPs in Warlpiri is followed, in order to account for discontinuity of phrases in Latin. In her account, Simpson simply assigns all elements of the discontinuous phrase the same category (N' in her case), and annotates these multiple occurrences of this category with the same grammatical function. An example of this for Latin is the tree structure in (3) for the sentence in (2):⁶

- (2) ...*haberent* *reliquorum* *nutriculas* *praediorum*.
 have.3PL.IMPF.CONJ other.GEN.PL foster-mothers.ACC farms.GEN
 '...they might have foster mothers for their other farms.'
 (Cic. *Phil.* 11.12, taken from Bolkestein (2001, p. 253))



As one can see, the discontinuous genitive phrase *reliquorum...praediorum* ('for their other farms') is analyzed as being an argument of the noun *nutriculas* ('foster mothers') because a foster mother always has someone/something which she is taking care of. In the structure in (3), the adjective (A) and the noun (N) each form a separate NOM projection, both of which carry (↑OBL_{GEN})=↓ specification. If a nominal phrase is contiguous, it will form one NOM constituent, which is annotated once for its grammatical function. If it is discontinuous, every separate element

⁵Note that the case of *tempus* is accusative. Most commonly the ending *-us* is used for the nominative alone (as a matter of fact *tempus* is also used for the nominative), but the word *tempus* is of the third declension in Latin, meaning that the nominative and accusative case have the same form, *-us* in this case.

⁶The type of nominal constituent is called 'NOM' here, a term introduced in Snijders (2012) to refer to any type of nominal constituent. This annotation was chosen in order to distinguish it from the classical NP: the structure of Latin nominals appears to be flat, and NPs are generally hierarchical constituents. For the purposes of this paper, the exact annotation is not very important, and 'NOM' can therefore be read as 'NP' or 'nominal constituent'.

or group of elements (if the NOM consists of more than two elements) forms its own NOM. These NOMs, part of the ‘larger’ nominal phrase (one could say), are all annotated with the same grammatical function in c-structure. Here the concept of *unification* is crucial. The annotation of separate elements with the same function ensures that the separate pieces of information of the ‘larger’ NOM appear in the same f-structure matrix: the information from both parts *unifies* in f-structure.

4 Problem 1: Constraint on Prepositional Phrases

4.1 Constraint and Data

One of the three absolute constraints on Latin discontinuous NPs found by Bolkestein (2001) is the following on prepositional phrases:⁷

(4) **Constraint 1 on Latin discontinuous NPs:**

No discontinuity between a P and the NP it governs (yet the NP may be internally discontinuous, meaning that part of the NP may be separated from the P)

This constraint states that there is no discontinuity between a preposition and the noun phrase which it governs. This constraint is similar to saying that preposition stranding is not allowed. Preposition stranding is allowed in some languages, for example in English:⁸

(5) *This is **the man** I told you **about**.*

This type of separation does not occur in Latin. Coming back to the Latin data, an example of a grammatical (attested) sentence is the following:

- (6) *In quo ego accusatore, iudices, primum illud*
 in+ABL which.ABL I.NOM prosecutor.ABL judges.VOC firstly this.ACC
deprecabor, ne quid L. Murenae dignitas illius...
 pray.1SG so-that any.ACC L. Murena.DAT dignity.NOM he.GEN
obsit...
 injure.3SG.CONJ

‘In the case of this prosecutor, gentlemen, I pray this first, that his dignity may not injure L. Murena in any way...’

(Cic. *Mur.* 58, taken from Bolkestein (2001, p. 251), but extended with original text)

In this example, the word *quo* has to appear adjacent to the preposition *in* (alternatively, *accusatore* would have to appear adjacent to *in* if *quo* was not in that position). Cases in which the preposition is completely separate from its NP are not found (in Bolkestein (2001)’s data), and it will therefore be assumed that they are ungrammatical. The NP in this example is internally discontinuous. The order within the NP is irrelevant: either dependent or head (noun) may come first. In (6), the order is P - D - N, and the noun is separated from the preposition. The determiner is also allowed to be separated from the preposition, and it may even occur before the preposition. This is shown in the next example:

⁷Bolkestein (2001) covers data from Cicero, Pliny the Elder and Petronius. In total she used 59 examples of discontinuous NPs from Cicero and 83 examples from the later works of Pliny the Elder and Petronius; 142 examples in total.

⁸Preposition stranding is allowed in English, but the place where the preposition occurs is very restricted.

- (7) *Deinde, si qua ego in re fratri tuo*
 secondly if any.ABL I.NOM in+ABL matter.ABL brother.DAT your.DAT
rei publicae causa restiterim...
 good.GEN public.GEN for-the-sake-of oppose.1SG.PERF.CONJ
 ‘Secondly, if in any matter I have opposed your brother for the sake of the public good...’
 (Cic. *Fam.* 5.2.6, taken from Spevak (2010, p. 25))

Even a three-way split of the NP is allowed, but still part of it occurs adjacent to the preposition:

- (8) *Ac ne in hoc quidem tam molesto tacebant*
 and not in+ABL this.ABL even so unpleasant.ABL be-silent.3PL.IMPF
officio
 duty.ABL
 ‘And they were not even silent by this unpleasant duty.’
 (Petr. *Sat.* 31, taken from Bolkestein (2001, p. 255))

The constraint in (4) implies that examples of the following kind are not grammatical (ungrammatical version of example (6)):

- (9) **In ego quo accusatore, iudices, primum illud*
 in+ABL I.NOM which.ABL prosecutor.ABL judges.VOC firstly this.ACC
deprecabor, ne quid L. Murenae dignitas illius...
 pray.1SG so-that any.ACC L. Murena.DAT dignity.NOM he.GEN
obsit...
 injure.3SG.CONJ

This type of example is indeed not attested in the data. It appears that preposition needs to occur to at least part of its object NP. In the following section, an analysis for this constraint in LFG will be presented.

4.2 Problem for Economy of Expression and Solution

This constraint poses a problem for the LFG notion of optionality of nodes, or in a formalized way, for Economy of Expression. Economy of Expression may be defined in the following way (Bresnan, 2001):

- (10) **Economy of Expression:**
 All syntactic phrase structure nodes are optional and are not used unless required by independent principles (completeness, coherence, semantic expressivity).

This principle makes all nodes, complements and heads, optional, giving the c-structure a great degree of freedom. Economy of Expression works well for cases in which transformational theories have problems, for example when a maximal XP phrase does not dominate a corresponding X head, so-called headless constructions as occur for example in Russian. For examples of this see King (1995).

It appears that Economy of Expression is problematic for the PP-NP adjacency constraint in Latin (described in (4)). Economy of Expression ensures that there are no ‘messy’ empty categories, which is favorable for multiple reasons (Bresnan, 2001). One of these reasons is that empty nodes are problematic for natural language processing, for which syntax should be a model. However, it also means that there will be an overgeneration, or overacceptance of sentences, meaning that certain sentences are incorrectly classified as grammatical.⁹ In other words, Economy of Expression predicts discontinuity in an unconstrained way (not only in Latin), which is contradictory to the data.

In order to overcome this problem, I propose to allow for exceptions to Economy of Expression, making it possible to make certain nodes obligatory. This will be explained in more detail later. Firstly, it must be made clear that the solution to the problem needs to be sought in the formulation of phrase structure rules, since this is a c-structure problem. Therefore I define a set of phrase structure rules for Latin which account for the constraint. These rules are presented below in (11). This set of rules accounts for the constraint on prepositional phrases, but not yet for the constraint on discontinuous adjuncts which will be discussed in the next section. In order to account for this constraint, the rule for S needs to be altered, which will be shown later. For the moment the following set of rules is proposed:¹⁰

(11) Non-final rules:

$$\begin{aligned}
 \text{GF} &\equiv \{ \text{SUBJ} \mid \text{OBJ} \mid \text{OBJ}_\theta \mid \text{OBL}_\theta \mid \text{ADJ} \in \mid \text{OBL}_\theta \text{ OBJ} \mid \text{ADJ} \in \text{OBJ} \} \\
 \text{S} &\rightarrow \{ \text{V} \mid \text{NOM} \mid \text{PP} \}^* \\
 &\quad \uparrow=\downarrow \quad (\uparrow \text{GF}) =\downarrow \quad (\uparrow \{ \text{OBL}_\theta \mid \text{ADJ} \in \}) =\downarrow \\
 \text{NOM} &\rightarrow \{ \text{D} \mid \text{A} \mid \text{N} \mid \text{PP} \}^* \\
 &\quad \uparrow=\downarrow \quad \downarrow \in (\uparrow \text{ADJ}) \quad \uparrow=\downarrow \quad (\uparrow \{ \text{OBL}_\theta \mid \text{ADJ} \in \}) =\downarrow \\
 \text{PP} &\rightarrow \text{P} \ , \ \text{NOM} \\
 &\quad \uparrow=\downarrow \quad (\uparrow \text{OBJ}) =\downarrow
 \end{aligned}$$

Before discussing these rules, it is important to note that the P here may be either a preposition or a postposition, which is shown by the fact that there is a comma between P and NOM. The comma (or shuffle operator, this will be explained later) shows that the order of P and NOM inside the PP does not matter, and therefore that the P can be either a preposition or a postposition. In most cases the P will be a preposition: Latin has predominantly prepositions, and only two postpositions. Unfortunately Bolkestein (2001) does not explicitly mention postpositions; she only mentions prepositions. In order to keep the analysis uniform, it will be assumed that postpositions function in the same way as prepositions, but more research needs to be done on this issue. From this point onward, however, PPs will be referred to as prepositional phrases, but keep in mind that the rule in (11) also takes potential postpositional phrases into account.

As one can see in (11), three phrase structure rules are posited for Latin: one for the sentence S, one for the nominal constituents (NOM or NP) and one for the prepositional phrase PP. For

⁹Within LFG, with phrase structure rules being node admissability conditions, it is better to say overacceptance than overgeneration.

¹⁰GF refers to ‘grammatical function’.

the purpose of this paper, other constituents such as adverbs (or adverbial phrases), auxiliaries, XCOMP and COMP are not directly relevant; therefore they are not included in these rules. The first rule shows that the structure of the sentence S is flat; this is indicated by the curly brackets, which denote disjunction (and not a set). The Kleene star (*) ensures that any number of constituents (as long as they are in the rule) can be daughters of S.¹¹ Verbs, nominal constituents and prepositional phrases (with the function of either oblique argument or adjunct) may occur in any order. As mentioned above, it is important to note that this ordering is largely determined by the sentence's information structure, and that it is not completely random.¹² However, in this paper we are concerned with absolute constraints and not tendencies, and therefore the information structural aspect of these issues will not be elaborated upon. The second rule shows that the structure of the nominal constituent (NOM) is also flat, but it may include hierarchically structured PPs. Determiner and noun may be c-structural co-heads, as signified by the annotation on D and N, $\uparrow=\downarrow$ for both.

The rule relevant for the constraint discussed in this section is the last rule: the one for the PP constituent. Firstly, it needs to be made clear why a PP constituent was posited to start with. Technically it is possible to not posit a PP and say that the S directly dominates prepositions and nominal complements of prepositions. This is possible because nominal complements can be marked as being objects of a preposition, either with $(\uparrow\text{OBL}_\theta \text{ OBJ})=\downarrow$ (for PP arguments) or $(\uparrow\text{ADJ} \in \text{OBJ})=\downarrow$ (for PP adjuncts).¹³ The preposition would simply be marked with its grammatical function. This is possible in Latin because prepositions specify the case marking of their complement. However, it seems a little odd to have a preposition on its own be marked with a specific grammatical function. Also, the fact that discontinuity is relatively unconstrained, but that one of the constraints is that the preposition needs to be adjacent to at least part of its complement, is strong evidence in favor of the existence of the PP category. For this reason, the PP is posited as a constituent. However, if the PP is discontinuous, part of the f-structure prepositional phrase will still need to occur separately as a NOM marked as being the object of the preposition, with the notation $(\uparrow\text{OBL}_\theta \text{ OBJ})=\downarrow$ or $(\uparrow\text{ADJ} \in \text{OBJ})=\downarrow$, as mentioned above. A structural representation of this will be presented in the next section. A similar reasoning as for the PP constituent is used for the existence of the NOM constituent: 12% of all nominal phrases are discontinuous, meaning that contiguity of nominal phrases seems to be the norm. It would be strange to assume that this is a coincidence, and therefore it is assumed that there is a nominal constituent in c-structure (and not just separate adjectives, nouns and determiners annotated to be part of the same f-structure nominal constituent). For a more elaborate explanation of these considerations, see Snijders (2012).

Now let us turn back to the PP rule. This rule, unlike the others, does not involve a notation with Kleene star (and thus optionality) but rather requires that both the P and NOM are present inside the PP. This means that a PP node will always dominate a preposition and a nominal dependent, which is marked as the object of the preposition. By making the PP rule a separate rule, and by saying that in this rule the P and NOM are both obligatory (this nicely contrasts with the Kleene star notation in the other two rules), we make an exception to Economy of Expression. In general Economy of Expression works well (as mentioned, in specific constructions, see King (1995)) and the principle is one of the strong points of a constraint-based theory of syntax. In

¹¹Kleene star traditionally means that zero or more of the constituents are present. The S-rule thus says that an S may contain any number of Vs, NOMs or PPs, in any order. One could also use a Kleene plus (+) here, meaning that at least one of these constituents is present as a daughter of S. However, f-structure constraints should take care of the fact that in a sentence at least one constituent is present: therefore Kleene star suffices.

¹²Whether Latin has an underlying standard word order is debated, see Ledgeway (2011). It is in any case definitely true that Latin allows any ordering of subject, object and verb (Pinkster, 1990).

¹³Here, complement of a preposition and object of a preposition refer to the same thing.

transformational theories, phrase structure is less flexible, meaning that empty categories can be present. Also, making certain categories and levels of structure obligatory means that in the syntax there will be redundant structure. It seems implausible, particularly from the perspective of natural language processing, to posit that empty structure is present. Economy of Expression is a strong principle prohibiting this redundant structure. However, it does cause problems for an analysis of the constraint at hand.

One could alternatively propose to abandon Economy of Expression altogether. It is a crucial concept within LFG, however, and one of the clear points in which LFG differs from transformational accounts, as mentioned. Proposing to abandon Economy of Expression would mean a complete change in theory. Nonetheless, the constraint discussed in this paper has shown that Economy of Expression has weak points, and that we might want to reconsider at least part of it, by positing exceptions when necessary. It appears that c-structure is in some cases not constrained enough and that it will overadmit sentences; it allows for grammatical sentences, but also for certain ungrammatical ones. As explained, f-structure cannot rule these sentences out because the problem is not completeness/coherence of information. In terms of f-structure, all the information is present; it is the ordering that is the problem. Therefore I believe that c-structure should be more strongly constrained than it currently is in LFG.

As a final note, it should be mentioned that one could maintain a principle of economy without positing it explicitly. One simply needs to define the phrase structure rules in such a way that it is clear that in general, all nodes are optional. One would then need to find a proper notation to show that in some cases, nodes are in fact obligatory. As long as the notation is explained properly, this is possible. This is certainly a valid way of going about the issue, but it might become a little messy. Also, the fact that there is a principle such as Economy of Expression makes it explicit that in general, one assumes that there is no redundant structure present in the syntax. Therefore the proposal in this paper is to maintain Economy of Expression, but make exceptions possible.

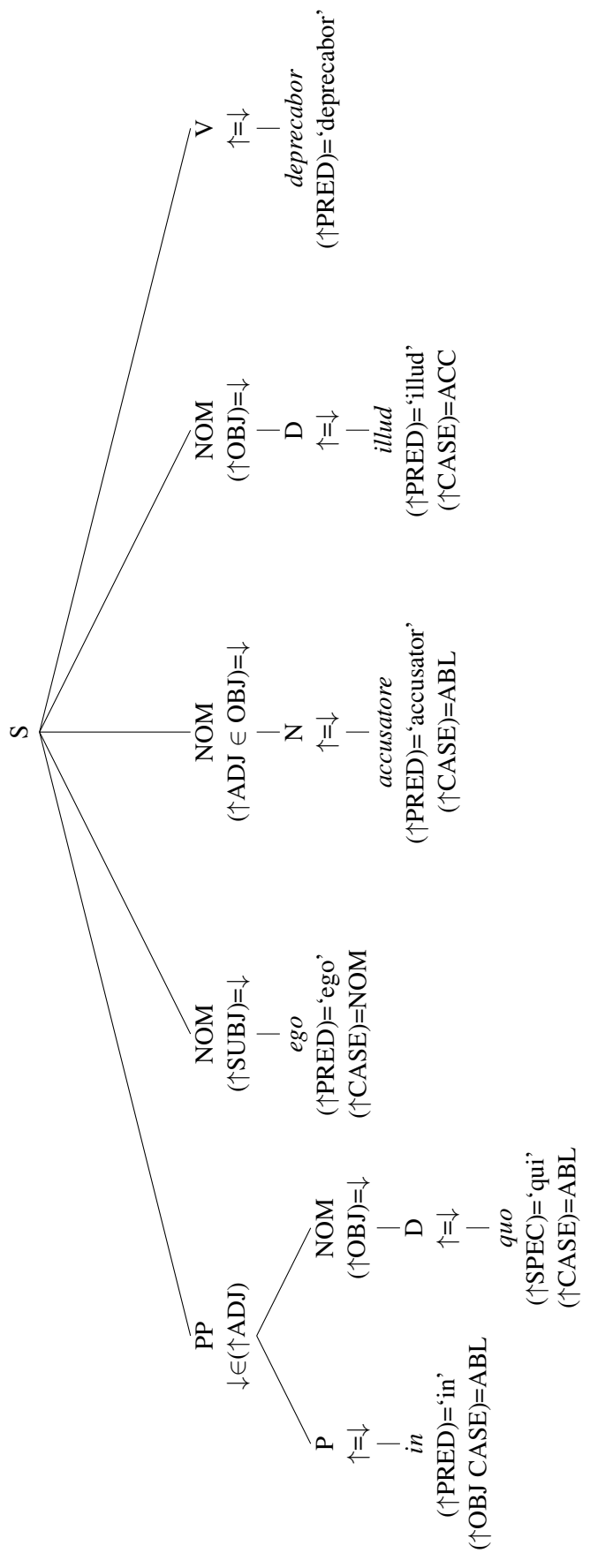
4.3 Analysis of Discontinuous PPs: Structure

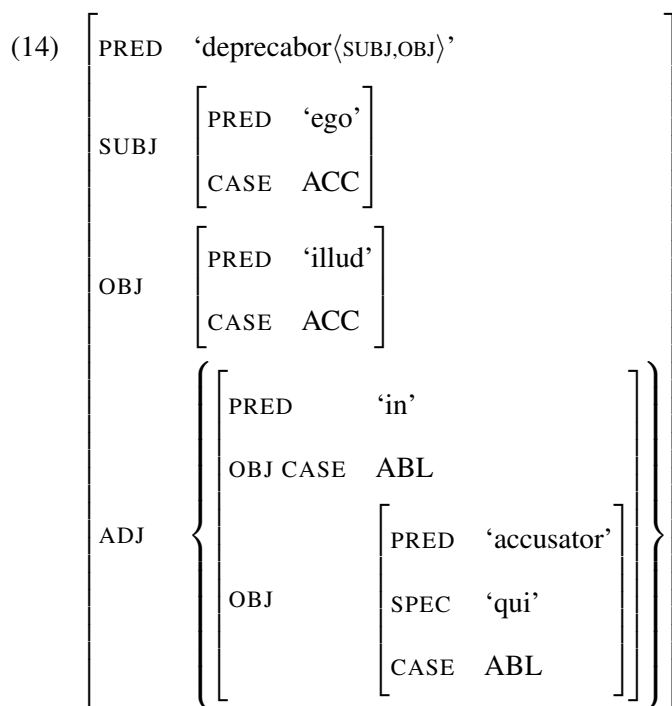
In order to make clear what a Latin prepositional phrase structurally looks like, and to give an illustration of the rules in (11), the structure of a Latin sentence including a PP will be presented here. In this paper it is assumed that an f-structure prepositional phrase which is discontinuous in c-structure consists of two (or more) parts: one PP node with part of the nominal object inside and one (or more) separate nominal object(s). These separate nominal objects are annotated as being the object of the preposition with either $(\uparrow\text{ADJ} \in \text{OBJ})=\downarrow$ or $(\uparrow\text{OBL}_\theta \text{ OBJ})=\downarrow$ (depending on the grammatical function of the PP), with the value θ dependent on the grammatical function of the phrase, often dependent on the specific preposition. Example (6) is repeated below in (12):

- (12) *In quo ego accusatore, iudices, primum illud*
 in+ABL which.ABL I.NOM prosecutor.ABL judges.VOC firstly this.ACC
deprecabor; ne quid L. Murenæ dignitas illius...
 pray.1SG so-that any.ACC L. Murena.DAT dignity.NOM he.GEN
obsit...
 injure.3SG.CONJ
 ‘In the case of this prosecutor, gentlemen, I pray this first, that his dignity may not injure L. Murena in any way...’

The f-structure and c-structure of a simplified version of this sentence are the following (the sentence is simplified to make the analysis presented in this paper clearer):

(13)





In this simplified c-structure and f-structure, we can see that the verb subcategorizes for a subject and an object, and that both are present in c-structure (and f-structure). The prepositional phrase is an adjunct, and therefore ends up in the set of adjuncts. In c-structure, the prepositional phrase consists of one PP node marked as an adjunct, and one NOM node marked as object of an adjunct. The information from both nodes ends up in the same f-structure, namely that of the adjunct in the sentence. In this sentence this notation works fine because there is only one adjunct, but note that problems might arise when the sentence has more than one discontinuous adjunct. This is a serious issue, which will be addressed in the next section. For the purpose of illustration, however, these two structures neatly show how Latin discontinuous PPs can be analyzed.

5 Problem 2: Constraint on Discontinuous Adjuncts

5.1 Constraint and Data

A second constraint on discontinuous nominal phrases which is problematic for LFG is the following, as proposed by Bolkestein (2001):

(15) **Constraint 2 on Latin discontinuous NPs:**

No discontinuous adjuncts when the intervening element is itself an adjunct

This constraint states that in Latin, there are no discontinuous nominal phrases which are adjuncts with the intervening element being an adjunct. This ties in with two findings by Bolkestein (2001) that discontinuous adjuncts are rarer than discontinuous arguments, and that adjuncts as intervening elements are less common than arguments as intervening elements. In general, adjuncts are not involved in discontinuous nominal phrases as commonly as arguments are.

As an illustration of this constraint, see the following two examples:¹⁴

¹⁴The first example contains a so-called *Accusativus cum Infinitivo*-construction (*AcI*) in which a complement clause is formed by a subject in accusative case and verb in infinitive form. As one can see, it is translated with the accusative phrase being the subject and the verb in finite form in English.

- (16) *ternis expeditionem eam mensibus confici*
 three.ABL expedition.ACC this.ACC months.ABL accomplish.INF.PASS
 ‘that this expedition was accomplished in three months’
 (Plin. *Nat.* 7.26, taken from Bolkestein (2001, p. 254))
- (17) **ternis magna celeritate mensibus expeditionem confecit*
 three.ABL great.ABL speed.ABL months.ABL expedition.ACC
 accomplish.3SG.PERF
 ‘?? he finished the expedition very quickly in three months’
 (Taken from Bolkestein (2001, p. 255))

In the first example, the ablative adjunct phrase *ternis...mensibus* (‘in three months’) is discontinuous, and the intervening element is the subject of the complement clause (it has accusative case because of the *AcI*-construction of which it is part, see footnote 14). In the second example, which is of a type not occurring in the data, a discontinuous adjunct is split by another adjunct. This type of split is not attested in the data and we therefore assume that it is not grammatical.

The examples in the literature of adjuncts that are discontinuous only include adjuncts on the clausal level; adjuncts on the nominal level are not mentioned. The constraint found by Bolkestein discussed in this chapter seems to be true only for adjuncts on the clausal level, and because of lack of data we will assume this is the case and provide an analysis only for discontinuous adjuncts on the clausal level.

This constraint is highly problematic for LFG, as will be shown in the next sections. It turns out that not only the constraint itself is problematic, but more importantly it appears that an analysis for any type of discontinuous adjunct (no matter what element is intervening) is problematic. Before turning to the actual discussion of a potential analysis, it will be shown what type of elements may in fact intervene in discontinuous adjuncts, in order to give a complete overview of the data, and therefore to aid in finding a proper solution to this problem.

5.2 Intervening Elements

In finding an analysis for this constraint, an investigation into the type and number of elements splitting up a discontinuous nominal phrase was made. As mentioned earlier, discontinuity of NPs may be caused by obligatory intervening elements, when a specific element needs to have a set position in the sentence (e.g. second position). This is one type of intervening element. More interesting and relevant to the discussion in this paper is the type of discontinuous NP in which the intervening element is ‘alien’ and can be anything which is not directly associated with the discontinuous phrase. In discontinuity caused by alien elements only one element intervenes in the majority of cases. In Bolkestein’s [2001] data from Cicero (containing 59 instances of discontinuous nominal phrases) there are only nine instances (15%) in which there is more than one constituent intervening. In these cases, they are very often only two elements, either the predicate and one of its arguments or two arguments, as in the following example.¹⁵

- (18) *putares... aliquo te cum hoc rei publicae vinculo esse coniunctum*
 think.2SG.CONJ some.ABL you.ACC by this.ABL affair.GEN public.GEN
 responsibility.ABL be.INF bind.PTPC

¹⁵This example is another case of an *AcI*-construction.

‘you would have thought... that you were bound to him by some responsibility for the state’
(Cic. *Mur.* 64, taken from Bolkestein (2001, p. 252))

The intervening elements in this example are the subject of the verb, *te* (‘you’, in accusative case in the *AcI*-construction), and the oblique argument of the verb, *cum hoc* (‘to him’). In Bolkestein’s data set of Cicero there are no occurrences of adjuncts as intervening elements in any of the discontinuous nominal phrases (even when the discontinuous phrase is an argument).

In the works by Pliny the Elder and Petronius the number of intervening elements can be even larger, and adjuncts may intervene:

- (19) *nostram scilicet de more ridebant invidiam*
 our.ACC namely out+ABL habit.ABL laugh.3PL.IMPF envy.ACC
 ‘they namely laughed at our envy out of habit’
 (Petr. *Sat.* 14, taken from Bolkestein (2001, p. 254))

In this example, an adverb, an adjunct PP and the main verb all intervene between the two parts of the object of the sentence. This shows that adjuncts may in fact occur as intervening elements when the discontinuous NP is an argument. Bolkestein (2001) only gives examples of intervening adjuncts which are prepositional phrases, not of, for example, bare ablative phrases (not governed by a preposition).¹⁶ She notes that this needs further investigation. In any case we assume that adjuncts can intervene if the discontinuous nominal phrase is an argument. From the example in (19) it appears that any number of constituents may intervene between the parts of the discontinuous nominal phrase; the two parts of the NOM are located at the edges of the phrase. Since there are examples of this kind, we may conclude that there is no real constraint on the number of intervening elements.

From the work on this issue by Bolkestein (2001) (see also Spevak (2010)) we conclude that in general there may be more than one intervening element, but that in the case of discontinuous adjuncts, rare as they are, only one constituent may intervene, and this is never another adjunct. The intervening element may be either an argument of the verb, or the verb itself. This insight is important for a potential analysis, since one only has to constrain one element (the intervening element). If more intervening elements were allowed, it would be quite difficult to formalize how one of these intervening constituents is not allowed to be an adjunct. Even under the assumption that only one element may intervene, the analysis of this constraint in LFG will not be straightforward. In fact, this is an incredibly difficult, even seemingly impossible, constraint to analyze within LFG, as will be shown in the next section.

5.3 Search for an Analysis of Discontinuous Adjuncts

The constraint on discontinuous adjuncts and the number of intervening elements pose some problems for a potential analysis within LFG. First of all, the S rule which was posited in (11) is problematic. It is repeated below in (20):

- (20) $GF \equiv \{SUBJ \mid OBJ \mid OBJ_{\theta} \mid OBL_{\theta} \mid ADJ \in \mid OBL_{\theta} \mid OBJ \mid ADJ \in \mid OBJ \}$
 $S \rightarrow \{ V \mid NOM \mid PP \}^*$
 $\uparrow = \downarrow \quad (\uparrow GF) = \downarrow \quad (\uparrow \{OBL_{\theta} \mid ADJ \in\}) = \downarrow$

¹⁶Bolkestein does not even mention genitive phrases. Ablative phrases are nearly always adjuncts because of their nature of providing extra information not specified by the predicate, such as place, manner, accompaniment.

As explained in Section 3, this rule allows NOMs to be discontinuous (by positing more than one NOM in c-structure, the information of which unifies in f-structure), but there is no way to constrain what type of elements may intervene between the different parts of the nominal phrase. Anything is allowed to intervene according to the rule in (20). As explained in the previous section, it may be assumed that in most cases this is fine and that in general any number and kind of constituents are allowed to intervene. However, the rule needs to be changed in order to account for the constraint on discontinuous adjuncts.

The closest approximation of a properly working S rule is displayed below in (21). Unfortunately it is not completely capable of accounting for the constraint, or in fact any type of discontinuous adjunct, as will be explained later. However, it is displayed here in order to show what the problem is. The close approximation rule is the following:

(21) Closest approximation but faulty S rule:¹⁷

$$\begin{aligned}
 \text{GF} &\equiv \{\text{ARG-GF} \mid \text{ADJ-GF}\} \\
 \text{ARG-GF} &\equiv \{\text{SUBJ} \mid \text{OBJ} \mid \text{OBJ}_\theta \mid \text{OBL}_\theta \mid \text{OBL}_\theta \text{ OBJ} \mid \text{ADJ} \in \text{OBJ}\} \\
 \text{ADJ-GF} &\equiv \text{ADJ} \in \\
 \text{S} &\rightarrow \left[\left\{ \begin{array}{c} \text{V} \mid \text{NOM} \mid \text{PP} \\ \uparrow=\downarrow \quad (\uparrow \text{ARG-GF}) =\downarrow \quad (\uparrow \{\text{OBL}_\theta \mid \text{ADJ-GF}\}) =\downarrow \end{array} \right\}^* \right], \\
 &\quad [(\text{NOM}) \quad (\{ \text{V} \mid \text{NOM} \}) \quad (\text{NOM})]^* \\
 &\quad (\uparrow \text{ADJ-GF}) =\downarrow \quad \uparrow=\downarrow \quad (\uparrow \text{ARG-GF}) =\downarrow \quad (\uparrow \text{ADJ-GF}) =\downarrow
 \end{aligned}$$

In this rule, the shuffle operator (comma) is used and the constraint on the discontinuous adjunct is expressed in the bracketed group after the operator. The shuffle operator ensures that anything occurring before it in a PS-rule can be freely ordered ('shuffled') in relation to everything occurring after it in a PS-rule. In the rule below, the shuffle operator ensures that anything occurring after it in the rule (for example, an adjunct NOM) can occur anywhere within the first part of the rule before the operator. This means that an adjunct NOM can, for example, occur in between an argument NOM and a verb (these last two are freely ordered as well, shown by the disjunction notation).

The first part of the rule (before the shuffle operator) is very similar to the S rule proposed in (11), except that the NOMs inside this part of the rule are only allowed to be arguments. Adjunct NOMs are represented in the bracketed group after the shuffle operator. This part of the rule ensures that, first of all, adjuncts are optional (denoted by the parentheses, although Economy of Expression technically takes care of this). Secondly, it makes sure that the only type of intervening element in a discontinuous adjunct is either a verb or an argument of the verb.¹⁸ Thirdly, the Kleene star ensures that more than one discontinuous adjunct is allowed per sentence (which should technically be possible).

¹⁷ Note that GF is now divided into argument functions (ARG-GF) and adjunct functions (ADJ-GF); this distinction is used to show that only arguments may intervene in discontinuous adjuncts.

¹⁸ There appear to be no examples in the data of argument PPs occurring as intervening elements of discontinuous adjuncts. If it is the case that argument PPs are allowed to occur in this position, the group $\text{V} \mid \text{NOM}$ can be extended to include PP (with annotation $(\uparrow \text{OBL}_\theta)=\downarrow$).

This rule unfortunately fails to account for the fact that the two adjunct NOMs in the second part of the rule can form a unit in f-structure. With the current annotation, all adjuncts end up in the same set, and there is no proper distinction between possibly different (f-structure) adjuncts. For example, if there are two adjuncts with the same case, number and gender in c-structure, there is no real way to distinguish between the situation in which they form separate f-structure adjuncts (separate elements in the set of adjuncts) or the situation in which they form two different subparts of the same f-structure adjunct (in a discontinuous phrase, as one element in the set of adjuncts). The only thing constraining this is the concept of PRED clash. One adjunct cannot have two PRED values. This is illustrated by the example f-structure below:

$$(22) \left[\begin{array}{l} \text{PRED} \quad \text{'VERB'} \langle \text{SUBJ} \rangle \\ \text{SUBJ} \quad \left[\text{PRED} \quad \text{'SUBJ'} \right] \\ \text{ADJ} \quad \left\{ \left[\text{PRED} \quad \text{'ADJ1'} \right], \left[\text{PRED} \quad \text{'ADJ2'} \right] \right\} \end{array} \right]$$

This example f-structure shows that all adjuncts are in a set, and each adjunct has its own PRED value. If a sentence has two adjuncts and they both provide a PRED value, they will therefore be two separate adjuncts in the set of adjuncts in f-structure. If there is a discontinuous adjunct, consisting of for example a determiner and a noun, the information from both parts of the discontinuous adjunct will unify in f-structure in the same element in the set of adjuncts. The determiner does not provide a PRED value (rather it provides a SPEC value), and it will therefore need to merge with an element that does provide a PRED value (because every separate f-structure adjunct has a PRED value). This unification will only happen if case/number/gender of the determiner and the noun agree. In most cases this will work fine, but a serious issue arises when there is more than one determiner present in the sentence with the same agreement values (for case, number and gender), annotated as being an adjunct. There is in principle no way to link the correct determiners to the correct nouns in this case. In actual language comprehension, this would most likely not be a problem because the reader can infer from the semantics of the words what is going on. However, the core issue remains that in LFG there is no way to mark two parts of a discontinuous adjunct in such a way that it shows that they are part of the same f-structure adjunct. One cannot exclusively link two different c-structure adjunct nodes together in one f-structure element, which does happen with grammatical functions. For example, a sentence exclusively has one subject only, and every node annotated with $(\uparrow \text{SUBJ}) = \downarrow$ will end up in the f-structure of the SUBJ. This is not possible for adjuncts. There is no way to control how different nodes annotated with $\downarrow \in (\uparrow \text{ADJ})$ are assigned a place in the set of adjuncts. This is a major problem in the treatment of (discontinuous) adjuncts. This means not only that the constraint discussed in this section cannot be analyzed properly, but also that in general discontinuous adjuncts are a major problem for LFG.¹⁹

It appears that LFG cannot account for the constraint that a discontinuous adjunct cannot have another adjunct as its intervening element. With the use of the shuffle operator it is possible to constrain what types of intervening elements may occur between the two parts of the discontinuous NOM, but the real problem is the annotation on adjuncts. In LFG currently, adjuncts are marked in

¹⁹As mentioned by Dag Haug at the LFG12 conference, there might be a way to solve this issue in a computational way, by assigning one of the parts of the discontinuous phrase an annotation with functional uncertainty. This is very briefly mentioned in Haug (2011, pg. 8). This can then be implemented as such in XLE, for example. This is not part of my work, but will be left for future research.

such a way that they all end up in the same f-structure set, which is problematic. The problem here is that merging of adjuncts is (relatively) uncontrolled. In languages such as English this is fine because in English one adjunct appears in the c-structure as one node, because they are not allowed to be discontinuous. Each adjunct will become a separate member in the set of adjuncts (with its own PRED value). In the case of discontinuous adjuncts in Latin, however, one cannot ensure that the two different nodes marked as being adjuncts form a unit in f-structure. Merging (unification) of adjuncts is uncontrolled (apart from when there is a PRED clash). This is a serious problem in LFG, for which there is not yet a solution. If there was a way to ensure that the two adjunct NOMs were part of the same adjunct in f-structure, the rule would work fine. Finding a solution might involve marking the two parts of the adjunct in such a way that they ‘belong together’, but an important question here is how this would affect the computational power of the system. This is an important consideration to take into account. Instead of finding a solution for this in LFG, one could also find a different way to account for discontinuity of nominal phrases, other than the one proposed by Simpson (1991). However, it is not clear what an analysis of this would look like. For the moment it may be concluded that LFG’s treatment of adjuncts is problematic and that it cannot account for discontinuous adjuncts in Latin (or in other languages).

6 Conclusion and Future Work

In conclusion, it has been shown that constraints on discontinuity of Latin nominal phrases bring forth two main problems for LFG as a theory of syntax. The first constraint, which states that in discontinuous PPs the preposition needs to occur adjacent to at least part of its nominal object, shows that c-structure in LFG is not sufficiently constrained. The principle of Economy of Expression overgenerates sentences. This paper has shown that c-structure needs to be more strongly constrained in LFG than it is at the moment; much of LFG’s attention has been focused on constraints on f-structure. In this paper, it is proposed to make an exception to Economy of Expression in order to account for the Latin data. The second issue discussed in this paper came to light with a closer investigation into another constraint on discontinuous NPs in Latin, namely the one on discontinuous adjuncts. As shown in this paper, the problem goes beyond the constraint because it turns out that in general, discontinuous adjuncts cannot be accounted for properly within LFG, due to LFG’s set notation for adjuncts. No proper analysis has been found for this problem at the moment. This is an important issue which merits future research.

Future work will involve the issue of adjuncts in LFG. Also, an increased data set would allow for further conclusions to be drawn; for this paper only 142 examples of discontinuous NPs (from three different authors) were used.²⁰ It would be interesting to see how genitive adjuncts come into this, since they were not discussed in the data; adjuncts in Latin are most often bare ablative phrases. In order to acquire a better overview of discontinuity in general, other languages will need to be looked at. Discontinuous nominal phrases are not only found in Latin; they also occur in languages such as ancient Greek, Polish and Russian (Siewierska, 1988). Discontinuity and its limits say something about what is ultimately possible in a language, to the extent that it is still possible for listeners to understand the speaker (or readers understand the writer in the case of classical literary Latin). A cross-linguistic investigation would help give a better insight into the limits of discontinuity, and thereby into the limits of the cognitive capacities of language users in perception and production.

²⁰Considering that according to Pinkster (2005) 12% of all NPs are discontinuous, this is still a reasonable amount of data.

7 References

- Bolkestein, A. Machteld. Random Scrambling? Constraints on Discontinuity in Latin Noun Phrases. In Moussy, C., editor, *De lingua Latina novae quaestiones*, pages 245–258. Peeters, Louvain, 2001.
- Bresnan, Joan. *Lexical-Functional Syntax*. Blackwell Publishers Inc., Malden, MA; Oxford, 2001.
- Dalrymple, Mary. *Lexical-Functional Grammar*, volume 34 of *Syntax and Semantics*. Academic Press, San Diego; London, 2001.
- Haug, Dag. Backward Control in Ancient Greek - Subsumption or Linearization? In Butt, Miriam and King, Tracy Holloway, editors, *Proceedings of the LFG11 Conference*. CSLI Publications, 2011.
- King, Tracy Holloway. *Configuring Topic and Focus in Russian*. Dissertations in Linguistics. CSLI, Stanford, CA, 1995.
- Ledgeway, Adam. Syntactic and morphosyntactic change. In Maiden, M.; Smith, J.C., and Ledgeway, A., editors, *The Cambridge History of the Romance Languages, Volume 1: Structures*, pages 382–471. Cambridge University Press, Cambridge, 2011.
- Lowe, John J. Ṛgvedic clitics and ‘prosodic movement’. In Butt, Miriam and King, Tracy Holloway, editors, *Proceedings of the LFG11 Conference*. CSLI Publications, 2011.
- Panhuis, Dirk. *The Communicative Perspective in the Sentence: A Study of Latin Word Order*. John Benjamins, Amsterdam/Philadelphia, 1982.
- Pinkster, Harm. *Latin Syntax and Semantics*. Routledge, London, 1990.
- Pinkster, Harm. The language of Pliny the Elder. In Reinhardt, T.; Lapidge, M., and Adams, J.N., editors, *The Language of Latin Prose*, pages 239–256. Oxford University Press, Oxford, 2005.
- Siewierska, Anna. *Word Order Rules*. Croom Helm, London, 1988.
- Simpson, Jane. *Warlpiri Morpho-Syntax: a Lexicalist Approach*, volume 23 of *Studies in Natural Language and Linguistic Theory*. Kluwer Academic Publishers, Dordrecht, 1991.
- Snijders, Liselotte. Constraints on Discontinuous Nominal Phrases in Latin from a Lexical-Functional Grammar Perspective. Unpublished Master’s dissertation, University of Oxford, 2012.
- Spevak, Olga. *Constituent order in Classical Latin prose*. Studies in Language Companion Series. John Benjamins Publishing Co., Amsterdam, 2010.