

# Topics First! In- and Outside of Bulgarian *Wh*-Interrogatives

Florian Jaeger

Stanford University

Proceedings of the 10th International Conference on  
Head-Driven Phrase Structure Grammar

Michigan State University

Stefan Müller (Editor)

2003

pages 181–202

Jaeger, Florian. 2003. Topics First! In- and Outside of Bulgarian *Wh*-Interrogatives.  
In Müller, Stefan (Ed.), *Proceedings of the 10th International Conference on Head-  
Driven Phrase Structure Grammar, Michigan State University*, 181–202. DOI:  
10.21248/hpsg.2003.14.



## Abstract

In Jaeger (to appear) I have described clitic doubling in Bulgarian *wh*-interrogatives which constitutes a type of Superiority violation that cannot be accounted for by any existing analyses. By showing that clitic doubling of object *wh*-phrases marks topicality, I raised the hypothesis that many (or maybe all) so called Superiority effects in Bulgarian are due to topic-fronting of *wh*-phrases. Here, I provide further support for this hypothesis and show that there is also evidence for topic-fronting of non-object *wh*-phrases. Differences between colloquial and formal Bulgarian are restricted to how topical objects have to be realized at the site of the extraction (i.e. the VP), which also makes the account readily extendable to other multiple fronting languages. The complex ordering constraints on the left periphery are captured in a Linear Syntax approach (similar to but different from Kathol 2000).

## 1 Introduction

Superiority in multiple *wh*-interrogatives has been an ongoing topic in generative grammar for at least thirty years. Within the literature on Slavic syntax, Bulgarian has received special attention with regard to Superiority since the complex constraints that govern the ordering of fronted *wh*-words in Bulgarian multiple *wh*-interrogatives have been taken to be of great theoretical significance within GB/MP research (Bošković 1993; Chomsky 1973; Pesetsky 1987; Richards 1997). Still, there is considerable disagreement over the acceptability of certain examples and overall, over the stability of the Superiority effects, just as much as about the best account for the ordering constraints on Bulgarian *wh*-questions.<sup>1</sup>

In this paper, I present a formal account that differs substantially from the above-mentioned ones, most crucially in that I take so called ‘Superiority effects’ to be – at least in large part – due to topicality. This paper thus aligns with others who have raised doubt about the Superiority as a syntactic axiom (e.g. Ginzburg and Sag 2000:247f. for English; King 1995:56f. for Russian; among many). The current work is then motivated by the question ‘What is Superiority?’. In addressing this question, I use earlier work as a starting point.

The paper is structured as follows. In section 2, I provide the relevant background on topic- and focus-fronting, clitic doubling, multiple *wh*-interrogatives and so called ‘Superiority effects’. Section 2.3 discusses clitic doubling of *wh*-phrases and links it to topic-fronting (cf. Jaeger to appear). In section 3, I develop an analysis for topic-fronting in- and outside of *wh*-interrogatives, including the data introduced in section 2.3. Finally, section 4 contains a summary and conclusions.

---

<sup>0</sup>My heartfelt thanks go to Ivan Sag, Veronica Gerassimova (without them, this paper would not have been possible), Loren Billings, Mila Dimitrova-Vulchanova, Mariana Lambova, Mila Tasseva-Kurktchieva for their critical feedback and valuable discussions. I also would like to thank Lev Blumenfeld, Elisabeth Norcliffe, and the audiences at the FASL-12 and HSPG 2003 conferences. The usual disclaimers apply.

<sup>1</sup>To name a few papers with conflicting claims regarding Bulgarian Superiority data: compare Billings and Rudin (1996, 1998) vs. Bošković (1998b,a) vs. Grewendorf (2001) vs. Pesetsky (1987).

## 2 Background

In this section, I briefly introduce some relevant background. Section 2.1 describes multiple topic- and focus-fronting and its relation to clitic doubling (henceforth CD) in declarative clauses. Section 2.2 summarizes the relevant claims made in the literature about Superiority in multiple *wh*-interrogatives. The reader familiar with the literature on Bulgarian syntax will not miss anything by skipping over these two sections. In section 2.3, I summarize the data from Jaeger (to appear), showing CD in *wh*-interrogatives.

### 2.1 Discourse Function Fronting and Clitic Doubling

In Bulgarian, certain discourse functions (topic and focus) are marked in syntax by means of fronting of the respective constituents I will refer to this process as *discourse function fronting* (DF-fronting). In Bulgarian and other Slavic languages, fronted topics precede fronted foci. Examples for Bulgarian and Russian are given below:

- (1) Decata MAMA šte vodi na cirk. [Bulgarian]  
 children-the<sub>TOP</sub> mom<sub>FOC</sub> will take to circus  
*The kids, MOM will take to the circus.* [Lambova 2003b:1]
- (2) Ja k ANNE prišel. [Russian]  
 I to Anna arrived  
*I visited ANNA.* [King 1995:207]

In colloquial Bulgarian and some other languages (e.g. Albanian and Greek; Kallulli 2001) *topic*-fronted object constituents are CDed, i.e. they are doubled by a clitic somewhere lower in the clause agreeing in person, number, gender and case.<sup>2</sup> CD is well-known from Romance languages (e.g. Rumanian, Italian, French, and Spanish) and the languages of the Balkan Sprachbund (e.g. Albanian, Bulgarian, Greek, Macedonian) among others. Although many different functions have been proposed for Bulgarian CD (for an overview, see Jaeger 2002), the literature clearly converges on the claim that CD marks topicality (e.g. Alexandrova 1997; Dimitrova-Vulchanova and Hellan 1995/1999; Jaeger and Gerassimova 2002; Leafgren 1997; Rudin 1997). Example (3), in which the fronted *Todor* is extracted out of a sentential subject, shows that topic-fronting is a long distance dependency. (4) shows that fronting and doubling of several constituents is possible. DOC stands for the direct object clitic and IOC for the indirect object clitic. Topic-fronted constituents and clitics are underlined.

- (3) Todor e jasno, [če Ivan go e vidjal]  
 Todor<sub>TOP</sub> is clear that Ivan DOC<sub>3.SG.MASC</sub> is seen  
*Todor, it is clear that Ivan has seen him.*

---

<sup>2</sup>I restrict myself to object CD and ignore subject CD which is also possible in several of the above-mentioned languages.

- (4) Na Ivan knigata az mu ja dadox.  
 to Ivan<sub>TOP</sub> book-the<sub>TOP</sub> I IOC<sub>3.SG.MASC</sub> DOC<sub>3.SG.FEM</sub> gave  
*I gave the book to Ivan.* [Dimitrova-Vulchanova and Hellan 1998:xviii]

While CD of topic-fronted constituents is *obligatory* for colloquial Bulgarian (i.e. (4) would not be acceptable without the clitics), more formal registers of Bulgarian do generally avoid CD, as e.g. in (1) above. This variation will fall out of the analysis proposed here (cf. section 3.1).

## 2.2 Multiple *wh*-Interrogatives

Bulgarian *requires* all *wh*-phrases in non-echo questions to be extracted to the left periphery of the clause. In the case of embedded questions, *wh*-phrases can be extracted to the front of the embedding clause or to the front of the embedded clauses. In both cases they follow topics (Dimitrova-Vulchanova and Hellan 1995/1999; Rudin 1985). An example of an embedded question is given below:

- (5) Čudja se kâde kogo da izpratja.  
 wonder<sub>1.SG</sub> REFL where whom to send  
*I wonder whom to send where.* [Pavlov 2000:134]

Multiple *wh*-interrogatives have often been discussed under the keyword Superiority (Chomsky 1973). As in the case of many other languages (e.g. English and Russian), in Bulgarian, too, Superiority has been taken to enforce certain ordering restrictions on fronted *wh*-phrases. However, it is still unclear to which extent Superiority applies to Bulgarian *wh*-interrogatives. Many competing hypotheses have been proposed since Rudin (1985) who was the first to address the topic (for Bulgarian) within a generative framework. Before I proceed, I summarize three influential hypotheses with conflicting predictions (see also Jaeger to appear).

In (6), the subject *wh*-phrase supposedly has to precede the direct and indirect object *wh*-phrases, but the latter two can order freely in the second and third position. This is taken to also hold for sentences without a subject *wh*-phrase.

- (6) a. Koj kogo kak e celunal?  
       who whom how is kissed  
       *Who kissed whom how?*  
       b. Koj kak kogo e celunal?  
       c. \*Kogo koj kak e celunal?  
       d. \*Kak koj kogo e celunal?

### Bošković (1993, 1998b,a) and Lambova (2003b)

- (a) The *first wh*-phrase in *wh*-interrogative is subject to Superiority.  
 (b) In a multiple *wh*-interrogatives, all *wh*-phrases after the first order freely.

However, Grewendorf (2001:97) gives the following example to show that non-subject *wh*-phrases can actually order freely if there is no subject *wh*-phrase:

- (7) a. Kakvo na kogo e dal Ivan?  
           what to whom is given Ivan  
           *What has Ivan given to whom?*
- b. Na kogo kakvo e dal Ivan? [Grewendorf 2001:97]

**Grewendorf (2001:97)**

- (a) *Subject wh*-phrases are subject to Superiority.  
 (b) In a multiple *wh*-interrogatives, all other *wh*-phrases order freely.

This claim is further revised by Billings and Rudin (1998:5-6) who introduce examples of sentences with non-external subjects, such as (8), and examples of psych verbs with obligatory clitic doubling, such as (9), to show that animate object *wh*-phrases can sometimes precede subject *wh*-phrase.

- (8) a. Kakvo kogo e udarilo?  
           what whom is hit  
           *What hit whom?*
- b. Kogo kakvo e udarilo? [Billings and Rudin 1998:5]
- (9) a. Koj na kogo mu xaresva?  
           who to whom IOC pleases  
           *Who likes whom?*
- b. Na kogo koj mu xaresva? [Billings and Rudin 1998:6]

**Billings & Rudin (1996:46,1998)**

- (a-1) *External* [+human] subject *wh*-phrases are subject to Superiority.  
 (a-2) If there is no external subject, [+human] *wh*-phrases precede [-human] *wh*-phrases.  
 (b) All remaining *wh*-phrases order freely.

### 2.3 Clitic Doubling in *wh*-Interrogatives

In this section, I present data that constitute a systematic violation of Bošković's claim and cannot be accounted for by Billings and Rudin's animacy hypothesis either. These data were first introduced and discussed in more detail in Jaeger (to appear). I first summarize the phenomenon and then describe the analysis proposed in Jaeger (to appear).

### 2.3.1 The Phenomenon

As already mentioned, colloquial Bulgarian exhibits CD, which in some cases (e.g. for topic-fronted objects) is obligatory. But aside from the well documented cases of CD in non-interrogatives, *wh*-phrases in interrogatives can be CDed, too. Interestingly, CD in *wh*-interrogatives licenses a clear violation of Superiority, as shown in (10a) and (11a). In both examples, the object *wh*-phrase precedes the subject *wh*-phrase – contrary to what is predicted by any of the analyses discussed in the previous section. Note that the direct object clitic (DOC) is obligatory. In the default order, given in (10b) and (11b), the subject *wh*-phrase precedes the object, and CD is unacceptable or at least not preferred (compared to the (a)-variants).<sup>3</sup>

- (10) a. Kogo koi ženi \*(go) poznaxa?  
 whom which<sub>PL</sub> women–<sub>DEF</sub> DOC<sub>3.SG.MASC</sub> recognized<sub>3.PL</sub>  
*Whom did which women recognize?*

b. Koi ženi kogo (?go) poznaxa?

- (11) a. Kogo kakvo \*(go) ubi?  
 whom what DOC<sub>3.SG.MASC</sub> killed<sub>3.SG</sub>  
*Whom did what kill?*

b. Kakvo kogo (?go) ubi?

The effect of CD is further illustrated by (12) which contains two 3.SG *which*-phrases. With the DOC the first *wh*-phrase, *koj máž*, is interpreted as object. Without the DOC the first *wh*-phrase is interpreted as subject. Note that the argument status of the *which*-phrases in (12) cannot be determined by means of gender or case. The verb form in (12) does not mark gender and *which*-phrases – just like almost all NPs in Bulgarian – do not have overt case marking.

- (12) Koj máž koja žena (go) običa?  
 which man which woman DOC<sub>3.SG.MASC</sub> loves  
 Without DOC: *Which man<sub>SUBJ</sub> loves which woman<sub>OBJ</sub>?*  
 With DOC: *Which woman<sub>SUBJ</sub> loves which man<sub>OBJ</sub>?*

CD of a *wh*-phrase is neither limited to certain kinds of verbs (e.g. there are no Aktionsart restrictions) nor is it dependent on the animacy of the arguments (the latter is illustrated by (10) above).

<sup>3</sup>This generalization seems to be less clear for overtly D-linked *wh*-phrases (so called ‘which’-phrases), which seem to be acceptable with CD even if they are not fronted.

### 2.3.2 The Function: Marking of Topicality

The analysis of the above data put forward in Jaeger (to appear) states, in a nutshell, is that CD of *wh*-phrases, like CD of other types of fronted objects, marks topicality. The topic of a question is what the question *primarily requests information about* (for topics in interrogatives, see also Leafgren 1997:127; Steedman 2000:659). The claim that CD of *wh*-phrases marks topicality is supported by a range of arguments that are discussed in detail in Jaeger (to appear). Although topicality in questions may – on the first sight – appear to be an odd claim, it has nonetheless been argued for under labels like ‘D-linking’ for e.g. Rumanian (Comorovski 1996), Russian (Scott 2003), and German (Grohmann under review).<sup>4</sup>

In other words, I have argued that CDed *wh*-phrases are topical and that CD in *wh*-interrogatives works just like CD outside of *wh*-interrogatives. A possible objection to this claim could be that it has been argued that only *one* *wh*-phrase can be CDed (Dimitrova-Vulchanova and Hellan 1998, 1995/1999), whereas I have shown above that Bulgarian declaratives can have multiple fronted topics and that all fronted objects are CDed in the colloquial register. Dimitrova-Vulchanova and Hellan (1998:xxi) cite (13a) to show that “in constituent questions with many *wh*-items, one, but not more than one, clitic may occur agreeing with the respective *wh*-constituent”. In addition, Dimitrova-Vulchanova and Hellan (1995/1999:37) mention (13b) to illustrate that, more generally, only one constituent (regardless of whether it is an *wh*-phrase or not) in a *wh*-interrogative can be doubled.

- (13) a. Na kogo kakvo mu (\*go) dadoxa?  
to whom what IOC<sub>3.SG</sub> DOC<sub>3.SG.NEUT</sub> gave<sub>3.PL</sub>  
*What did they give to whom?* [Dimitrova-Vulchanova and Hellan 1998:xxii]
- b. Knigata na kogo (\*mu) ja dadoxa?  
books-the<sub>FEM</sub> to whom IOC<sub>3.SG</sub> DOC<sub>3.SG.FEM</sub> gave<sub>3.PL</sub>  
*To whom did they give the books?* [D.V. and H. (1995/1999:37)]

However, it turns out that questions with more than two *wh*-phrases are much more compatible with CD of *two* object *wh*-phrases (Mila Vulchanova, p.c.):

- (14) ?Na kogo kakvo koga mu go dadoxa?  
to whom what when IOC<sub>3.SG</sub> DOC<sub>3.SG.NEUT</sub> gave<sub>3.PL</sub>  
*To whom did they give what when?* [Mila Vulchanova, p.c.]

To sum up, although topic-marking in *wh*-questions is subject to some additional constraints<sup>5</sup>, in principle multiple topic-fronting *is* possible for *wh*-interrogatives.

<sup>4</sup>For a more general discussion of topicality and D-linking in *wh*-interrogatives, see also Kuno and Takami (1993); Grohmann (1998); Boeckx and Grohmann (2003).

<sup>5</sup>Recall that, after all, (13a) is possible with both clitics in declaratives, that is if both *wh*-phrases are substituted by lexical NPs, as in (4).

## 2.4 Summary

In this section, I have provided a brief summary of the overall configuration of the left periphery in the Bulgarian clause. I have paid particular attention to what I take to be topic-fronting of CDed object *wh*-phrases. The type of Superiority violations mentioned in section 2.3 cannot be accounted for even by those analyses that predict *some* violations of strict Superiority (e.g. Billings and Rudin 1996, 1998; Grewendorf 2001; Pesetsky 2000).

The remainder of the paper lays out a formal analysis of the left periphery, especially topic-fronting (within as much as outside of *wh*-interrogatives). I also gather further support for an extension of the above-stated hypothesis to non-object *wh*-phrases. Whereas CD provides a way of identifying topical object phrases, topical non-object phrases do not have a comparable morphological marking in Bulgarian. There is, however, some support for topic-fronting of non-object *wh*-phrases, which I discuss in section 3.2.

## 3 The program

An adequate analysis of the left periphery of the Bulgarian clause (which is the target of such phenomena as DF-fronting and *wh*-fronting) has to account for the following issues: (A) the correct order of fronted constituents, i.e. (A-1) topics precede non-topics; (A-2) topic-fronted non-*wh*-phrases can precede the *wh*-cluster in Bulgarian *wh*-questions (cf. (13b) in section 2.3.2); it also has to account for the facts that, in colloquial Bulgarian, (B-1) topic-fronted objects *must* be CDed and (B-2) focus-fronted objects *cannot* be CDed. For colloquial Bulgarian, this in turn raises the following questions: (C) what information object clitics contain in their lexical entry, and (D) how this information is passed from the clitics to the constituents on the left periphery of the clause. Taken together, questions (C) and (D) address the question of how an analysis can guarantee that object clitics have to agree with the topic-fronted constituent they double (see above) and that the constituent an object clitic agrees with must be topical.

Questions (B-1), (C) and (D) are addressed in section 3.1. The issues raised under (A) turn out to be quite intricate. They are discussed in detail in section 3.2. The remaining point (B-2) is addressed in section 3.3. I provide the formal constraints on the constructions of the left-periphery (e.g. the *wh*- and topic-clause types) and briefly sketch how the different parts of the proposed analysis interact. I will assume familiarity with the framework proposed in Ginzburg and Sag (2000) as well as with the idea of Linear Syntax (Reape 1994; Kathol 1995, 2000).

### 3.1 The Extraction Site: Colloquial $\neq$ Formal Bulgarian

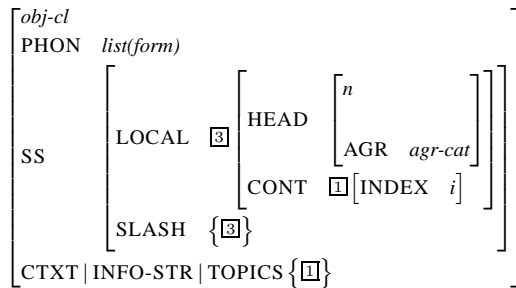
As I have already pointed out above, formal and colloquial Bulgarian seem to behave fairly similar much alike with respect to DF-fronting – except for the fact that colloquial Bulgarian requires CD of topic-fronted object constituents. In other



words, in terms of the relation between the ‘extraction site’ (i.e. the site from which something is extracted) in the clause and the ‘extraction target’ (i.e. the left periphery), object clitics have the same distribution as gaps. In light of this, I suggest the following. First, Bulgarian object clitics (in their function as discussed here) should be treated as phonetically non-empty gaps. Second, the ordering on the fronted field should be defined in terms of topicality rather than with direct reference to CD, which is only *one way to mark topicality* (i.e. for objects in colloquial Bulgarian). The second point will receive more attention in the next section, but I ask the reader to keep it in mind while reading the current section.

What does this mean for clitics? Somehow the lexical entry of a clitic introduces an element into the VP’s SLASH set and state that this element is topical and that it must have the right agreement features (i.e. the agreement features that are expressed in the clitic). Here I do not wish to discuss whether clitics in Bulgarian are adjoined to the verb in morphology or in syntax.<sup>6</sup> For simplicity’s sake, let us assume that clitics are adjoined to the verb in syntax<sup>7</sup>. Furthermore, given that, whenever a topical object is extracted in colloquial Bulgarian it has to be CDed, I postulate that colloquial Bulgarian (unlike more formal registers) has no way of introducing topical object gaps. Note that this is the answer to (B-1) raised at the beginning of section 3, i.e. ‘Why do fronted topics have to be CDed?’ A clitic identifies its own LOCAL value as the only element of its SLASH set and further determines that the CONTENT of this element is a member of the TOPICS set. The template for an object clitic is given in (15).

(15) Schematic template for object clitics



I assume a construction which identifies clitics with items on the verb’s ARG-ST. Thus whichever fronted constituent fills the ‘gap’ introduced by a clitic will be identified as a specific argument of the verb. The SLASH value percolates up to the clausal level due to the non-LOCAL Amalgamation Constraint (Ginzburg and

<sup>6</sup>This still appears to be an unresolved issue in the literature and is not relevant for this paper (see Franks and King 2000 for an overview over mostly syntactic approaches; for a recent morphological approach, see O’Connor 2002; for a similar approach in HPSG, see Miller and Sag 1997).

<sup>7</sup>In Bulgarian, object clitics are part of the so called *predicate clitic cluster* which is always verb adjacent. One could therefore propose a construction that combines the verb with all clitics to form the predicate clitic cluster. The construction identifies clitics with elements of ARG-ST and cancels the corresponding COMPS in the resulting predicate clitic cluster phrase.

Sag 2000:398), which collects all daughters' SLASH sets into the head's SLASH set. Therefore, the CDed phrase has to be topical since the clitic identifies its SLASH element as topical.<sup>8</sup> Any element of TOPICS is passed up to the clause by the *Information Structure Principle* (ISP), which is defined as a constraint on headed phrases (i.e. the type *hd-ph*; cf. Ginzburg and Sag 2000):

(16) Information Structure Principle (ISP)

*For each information structural feature  $F$  (such as e.g. TOPICS or FOCI), the value of  $F$  of a headed phrase's (*hd-ph*) mother is the union of all its daughters'  $F$  values:*

$$hd-ph \Rightarrow \left[ \begin{array}{c} DTRS \left\langle \begin{array}{c} [CTXT | INFO-STR | F \ \boxed{\Sigma_1}], \dots, \\ [CTXT | INFO-STR | F \ \boxed{\Sigma_n}] \end{array} \right\rangle \\ [CTXT | INFO-STR | F \ \boxed{\Sigma_1} \cup \dots \cup \boxed{\Sigma_n}] \end{array} \right]$$

Since clitics identify their CONTENT to be a member of TOPICS, the ISP ensures that this information is passed up to the clausal level. In section 3.3, it will become clearer precisely how this in turn forces the extraction target to be a member of TOPICS. In sum, colloquial Bulgarian has only one way to realize the extraction site of a topic-fronted object, namely via an object clitic. I have sketched the information provided by clitics (agreement, topicality of co-indexed item, and indirect argument identification). Formal Bulgarian, on the other hand, does not have clitics because but allows topical object gaps. In other words, colloquial and formal Bulgarian differ at the extraction *site*. Note that I have refrained from introducing a CLITIC feature (cf. Avgustinova 1997). Instead the absence or presence of CD is represented indirectly. If an object is CDed it is deleted from the COMPS list and required to be topical. This approach is a priori preferable to one that employs a CLITIC feature, and will in addition prove elegant once I provide the analysis for the extraction *target* in section 3.3.

### 3.2 The left periphery of the Bulgarian clause

In section 2.1, I showed that Bulgarian has two types of DF-fronted constituents, namely topics and foci. The former always precede the latter. Similarly, in questions, CDed *wh*-phrases, which have been argued to be topical, have to precede the non-CDed *wh*-phrases. Thus we already know that  $[-wh; +top] \prec [-wh; -top]$  (i.e. non-*wh*-phrase 'topicalization') and  $[+wh; +top] \prec [+wh; -top]$  (i.e. *wh*-phrase 'topicalization').<sup>9</sup> We also know that  $[-wh; +top] \prec [+wh; -top]$  (i.e. non-*wh*-phrase topic-fronting before the *wh*-cluster). Note that we do not know whether  $[-wh; +top] \prec [+wh; +top]$  (i.e. topic-fronted non-*wh*-phrases precede topic-fronted *wh*-phrases) simply because this combination is very difficult or

<sup>8</sup>Note that I treat topics in a slightly different way from that proposed in Engdahl and Vallduví (1996) in that I take topics to be *semantic objects* (i.e. of type *sem-obj*; cf. Ginzburg and Sag 2000:387) rather than signs (see also Jaeger and Oshima 2002).

<sup>9</sup>I use  $[+/- \alpha]$  purely as a convenient notation for the *descriptive generalizations*. The sign  $\prec$  denotes a linear precedence relation (a  $\prec$  b if 'a must precede b').

even impossible to get.<sup>10</sup> The same difficulty holds for the relative order between non-*wh*-foci and *wh*-phrases. Rudin (1985:89) argues that focus-fronting is not possible in *wh*-questions. Pavlov (2000:142) provides (17) to show *wh*-fronting before a focus-fronted phrase (marked by the focus particle *li*) within a yes/no-question:

- (17) Kâde VČERA li bjaxa xuknali v tozi stud?  
 Where yesterday<sub>FOC</sub> FOC were rushed in this cold  
*Where had they rushed YESTERDAY in this freezing weather?*

However, in the default order for (17) the *wh*-phrase *kâde* would follow the focus phrase *včera li* (Veronica Gerassimova, p.c.). Thus [+wh] | [-wh;+foc].<sup>11</sup> Note that, strictly speaking, no focus feature is needed to describe this order constraint. A preliminary version of the left periphery precedence constraints is given in (18).

(18) **Left periphery precedence constraints (preliminary version)**

[-wh;+top] | [+wh;+top] < [+wh;-top] | [-wh;-top]

The precedence relations in (18) constitute the issue raised at the beginning of section 3 under point (A). Next, I will discuss *splitting of the wh-cluster*, a phenomenon, which, I argue, reveals further evidence for the existence of topic-fronted *wh*-phrases.

Consider the following data, in which a phrase splits the *wh*-cluster. Lambova (2003c), building on Rudin (1988), shows that emphatic particles, parentheticals, and adverbs (both sentential and manner adverbs) can occur after the first but not after the second *wh*-phrase. Below I give one of her examples, where *navjarno* ('perhaps') splits the cluster of fronted *wh*-phrases. Lambova (2003a,c) has taken these data as evidence that the first *wh*-phrase (sometimes) does not form a constituent with the remaining *wh*-phrases:

- (19) a. Koj, navjarno, kâde koga šte porâča tortata?  
 who perhaps where when will order cake-the  
*Who will perhaps have the cake made where when?*  
 b. \*Koj kâde, navjarno, koga šte porâča tortata?  
 c. \*Koj koga, navjarno, kâde šte porâča tortata? [Lambova (2003c)]

Lambova (2003c) also gives several examples illustrating that the same phrases that can split a *wh*-cluster after the initial *wh*-phrase *cannot* do that if the question

<sup>10</sup>I leave it open as to whether there may be sentences containing both 'normal' and *wh*-topics, since I do not have enough data to decide this point. For some data that could possibly be taken to support that [+wh;+top] < [-wh;+top] *cannot* hold, see Lambova (2003c) who argues that non-*wh*-topics cannot follow the first *wh*-phrase.

<sup>11</sup>I use 'a | b' to indicate that 'a and b can order freely'.

is preceded by a topic-fronted non-*wh*-phrase. The translations have been slightly changed to match the way other examples in this paper have been translated. Topic-marking is indicated by underlining (not given in the original examples):

- (20) a. Kakvo, kazvaš, koga iska šefât?  
 what<sub>TOP</sub> you-are-saying when wants boss-the  
*What, you're saying, does the boss want when?* [Lambova (2003c)]
- b. \*Šefât, kakvo, kazvaš, koga iska?  
 boss-the<sub>TOP</sub> what<sub>TOP</sub> you-are-saying when wants  
*The boss, what, you're saying, does (he) want when?* [Lambova (2003c)]

Although Lambova does not consider topic-fronting of *wh*-phrases, she provides examples showing that the same types of phrases that can split the *wh*-cluster (henceforth SPP for splitter-phrases) can also appear after fronted non-*wh*-topics (Lambova 2003a). For multiple topic-fronting as well, speakers seem to prefer SPPs between the topics and the *wh*-cluster (rather than after the first topic-fronted constituent):<sup>12</sup>

- (21) Na Maria (?obiknoveno) tortite (obiknoveno) koj i gi dava?  
 to Maria<sub>TOP</sub> usually cakes-the<sub>TOP</sub> usually who IOC DOC gives  
 Roughly: *To Maria the cakes, who (usually) gives (them) (to her?)*

I propose the following analysis. The SPPs in the above examples occur between topic-fronted constituents and non topic-fronted *wh*-phrases. *Wh*-phrases preceding an SPP are topic-fronted.<sup>13</sup> This claim predicts that SPPs should be able to occur after an initial CDed object *wh*-phrase, since they are topical. This is indeed the case:

- (22) a. Kogo, naj-verojatno, koj \*(go) obra?  
 whom<sub>TOP</sub> most-probably who DOC<sub>3.SG.MASC</sub> robbed  
 Intended: *Whom did most probably who rob?*
- b. Koj, naj-verojatno, kogo \*(go) obra  
 who most-probably whom<sub>TOP</sub> DOC<sub>3.SG.MASC</sub> robbed  
 Intended: *Who did most probably rob whom?*

I thus take this to be evidence for the hypothesis stated above that SPPs *can* appear between CDed topical *wh*-phrases and the remainder of the *wh*-cluster.<sup>14</sup>

<sup>12</sup>The data seem to be far more complex since judgements depend on the kind of SPP that is chosen (Veronica Gerassimova, p.c.). Here, it only matters that there seem to be certain SPPs which occur after the topic-cluster. I also do not discuss occurrences of SPPs *following* the *wh*-cluster.

<sup>13</sup>Since SPPs can also adjoin to VPs, one has to be careful, since because of this a single non topic-fronted *wh*-phrase also 'precedes an SPP'. Above, I refer to SPPs that occur directly *before* the *wh*-cluster. For *wh*-interrogatives with two or more *wh*-phrases, this is unambiguously identifiable.

<sup>14</sup>One may ask why it is not possible to have two topical *wh*-phrases or one topical non-*wh*-phrase and a topical *wh*-phrase followed by an SPP. As already discussed above, Dimitrova-Vulchanova

While more data are needed to be certain, it seems plausible that non-object *wh*-phrases, e.g. *koj* in (19a) or *kakvo* in (20a), are topic-fronted, just as CDed object *wh*-phrases are, e.g. *kogo* in (22a). This would simplify the formulation of the left periphery precedence constraints, thereby allowing a uniform analysis for SPP-placement. Furthermore, the proposed analysis of (at least some) initial *wh*-phrases as topics provides an explanation for (at least some) so-called Superiority effects. Rather than restricting *wh*-topic-fronting to CDed object *wh*-phrases, I assume (based on the data presented in this section) that topical *wh*-phrases of any kind precede non-topical *wh*-phrases. This parallels the data known from declaratives where topics precede foci. The revised and simplified version of the left periphery precedence constraints is the following:

(23) **Left periphery precedence constraints (final version)**

[+top] < SPP < [-top]

The next section addresses those parts of (23) that are crucial to multiple *wh*-questions with and without CD.

### 3.3 The Extraction Target: Colloquial = Formal Bulgarian

Below I present an analysis of the linear order constraints on the left periphery described in the previous section. After considering a range of different analyses (some rather hierarchical, some purely linear), I have come to the conclusion that the best analysis makes reference both to linear order constraints and to a hierarchy of phrases on the left periphery. Linear order is needed to provide an elegant description of the phenogrammatical properties of the left periphery, and a hierarchical organization proves necessary in order to capture its tectogrammatical properties.<sup>15</sup> I therefore adopt a version of Linearization-based Syntax (cf. Reape 1994; Kathol 1995, 2000), which makes use of the idea of topological fields. Before I proceed, let me briefly summarize the core of Kathol's proposal and where the approach taken here deviates from his (for further details, see Kathol 2000).

In addition to the standard features, each construction/phrase/word is assumed to contain an ORDER DOMAIN feature (henceforth DOM). I follow Reape (1994) and Donohue and Sag (1999) – and deviate from Kathol (1995:127) and Kathol (2000:99-100) – in that I take the value of DOM to be a list of *signs*. The advantage of this stems from the fact that the information-structural status of DOM elements has to be accessible for ordering constraints (I elaborate on this below). I adopt the idea of topological fields (Kathol 1995, 2000). The fact that a given DOM element has to be realized in a specific topological field is encoded in the type of that element (following Kathol 2000). That is, DOM elements are of type *sign* and of type

---

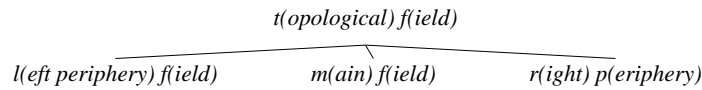
and Hellan (1998, 1995/1999) have shown that it is extremely difficult to get several topics in *wh*-questions. Adding an SPP does not make the sentence less complex – eliciting such question becomes increasingly difficult. Note, however, that the same difficulties hold for two fronted non-*wh*-topics.

<sup>15</sup>For the distinction between 'phenogrammatical' vs. 'tectogrammatical' representations in linguistics, see Dowty (1996).

*topo* and therefore “positionalized signs” (Kathol 2000:77). This also implies that all words must be of a specific *topo* type (i.e. words specify in which topological field they can occur). DOM values are handed up to constructions (although the order within the DOM list may change). Thus the *topo* types of all DOM elements are available at the constructional level and ultimately at the clausal level. Constructions can determine or constrain the *topo* type of any of their daughters. For example, the filler constructions for the left periphery could specify their fillers to be of the left periphery field type. I will come back to this below. At any given point, the actual phonological realization of a construction/phrase corresponds to the order of elements in DOM (Reape 1994:155). The order within DOM is in turn determined by Linear Precedence (henceforth LP) constraints, which are sensitive to topological fields (i.e. the DOM elements’ types). To sum up, Topological LP constraints determine the linear order of phonological elements in a sentence, thereby accounting for phenogrammatical restrictions. At the same time, constructions/phrase types constitute the tectogrammatical structure of a sentence.

The Bulgarian type hierarchy assumed here for topological fields is shown in (24). Although by no means complete, all types *relevant to the current problem* are given. The left periphery contains all elements that are fronted because they bear discourse functions, such as topics and foci (including *wh*-phrases). In other words, a word can only appear in *lf* if it is marked to be part of a topic or focus of a sentence (or some other kind of discourse marking function, as assumed for SPPs). The main field contains everything between the left and the right periphery. The right periphery contains right-dislocated elements such as antitopics (cf. Lambrecht 1994), which I will not discuss here further. Even though Bulgarian, unlike German, lacks a ‘Satzklammer’ (sentence bracket), it shares with German the property that the left and right periphery are the target of (discourse function driven) extractions:

(24) The topological fields of the Bulgarian clause



The ordering constraints observed in the previous section are captured by the LP Constraints in (25).

(25) Topological LP Statements for the Bulgarian clause

LP-1 (Bulgarian TF Constraint):  $[lf] \prec [mf] \prec [rf]$

LP-2 (Topics-First! Constraint):  $\left[ \begin{array}{cc} lf & \\ \text{CONT} & \boxed{\phantom{x}} \\ \text{TOPICS} & \text{set} \uplus \{\boxed{\phantom{x}}\} \end{array} \right] \prec \text{SPP: } [lf] \prec \left[ \begin{array}{cc} lf & \\ \text{CONT} & \boxed{\phantom{x}} \\ \text{TOPICS} & \text{set} - \{\boxed{\phantom{x}}\} \end{array} \right]$

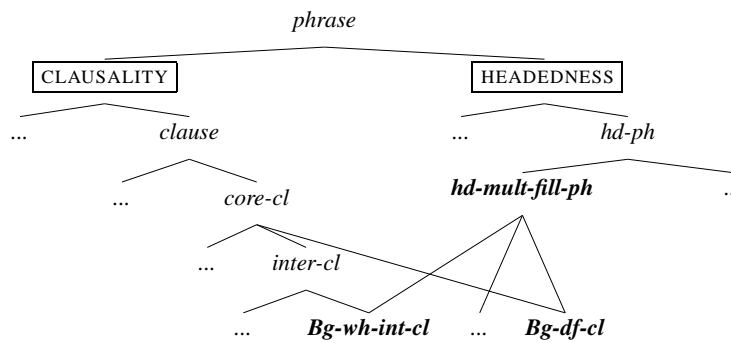
LP-1, the Bulgarian Topological Fields Constraint, states that elements in the left periphery precede elements in the main field, which in turn precede elements in the

right field. LP-2, the Topics-First! Constraint (henceforth TFC), is more complex in that it does not only make reference to topological fields. The TFC is effectively restricted to the left periphery (since it only states precedence constraints on elements of type *lf*). Within the left periphery, the TFC enforces the order that has been described in (23) in the previous section (i.e. topics have to precede non-topics, and SPPs appear after topics but before non-topics). Given the TFC, it also becomes clearer why it is advantageous that DOM elements be of type *sign*. As shown throughout this paper, especially in the previous section, the linear order of elements in the left periphery of the Bulgarian clause is clearly sensitive to information structure, most clearly to topicality. It thus seems as good or better a way to encode this *linear* order constraint directly by means of LPs (such as the TFC) rather than, for example, in the tectogrammatical component of the grammar (i.e. by means of phrase structure in the widest sense). Another way to capture the fact that topics precede foci would be to assume two left periphery fields (cf. Dimitrova-Vulchanova and Hellan 1995/1999) and associate one with topics and the other one with focus. This approach would not be incompatible with the one that I adopt here but I prefer the latter because of its conceptual clarity. Furthermore, as I show further down, the approach taken here reduces the number of constructions that are required in order to describe the left periphery.

Note that the LP rules do not make direct reference to *wh*-phrases. While the correct *ordering* of all fronted phrases is achieved via the TFC, I have yet to provide the tectogrammatical structure that explains how the extracted phrases on the left periphery (e.g. *wh*-phrases) are combined with the remainder of the clause.

In order to do that, I sketch the type hierarchy for the constructions of the left periphery and show how the extracted elements in, for example, a ‘topicalization’ clause or a multiple *wh*-question are combined with the remainder of the clause. For the reader’s orientation, the proposed type hierarchy for the Bulgarian clause is shown in (26). The two types *Bg-df-cl* and *Bg-wh-int-cl* correspond to the constructions for DF-fronting and *wh*-interrogatives, respectively.

(26) Type hierarchy for the left-periphery of the Bulgarian clause (non-leaf nodes)



I begin the discussion of the new types with *hd-mult-fill-ph*, an extension of the En-

glish *hd-fill-ph* that allows *multiple* fillers instead of just one. A similar type will be needed for any kind of multiple fronting language (e.g. Serbo-Croatian, Russian, Romanian). The *hd-mult-fill-ph*, as defined in (27), describes a flat structure with multiple non-head daughters (cf. Ginzburg and Sag 2000:364). Like the English *hd-fill-ph*, the *hd-mult-fill-ph* is a subtype of *hd-ph* (i.e. *hd-mult-fill-ph* is a headed phrase).

(27) Bulgarian *hd-mult-fill-ph*

$$\left[ \begin{array}{l} \text{hd-mult-fill-ph} \\ \text{SS | SLASH} \quad \boxed{\Sigma_1} \\ \text{DTRS} \quad \left\langle [\text{LOC } \boxed{1}], [\text{LOC } \boxed{2}], \dots [\text{LOC } \boxed{n}], \boxed{0} \right\rangle \\ \text{HD-DTR} \quad \boxed{0} \left[ \begin{array}{l} \text{LOC | HEAD} \quad v \\ \text{SLASH} \quad \{\boxed{1}, \boxed{2}, \dots, \boxed{n}\} \uplus \boxed{\Sigma_1} \end{array} \right] \end{array} \right]$$

The constraints of *hd-mult-fill-ph* are inherited by the type for DF-fronting clauses (*Bg-df-cl*) and the type for *wh*-clauses (*Bg-wh-int-cl*). I discuss those two new types in turn.

The *Bg-df-cl* type is very similar to the English *top-cl* suggested in Ginzburg and Sag (2000:379). It is a head-filler construction, and all its fillers' WH values must be empty (this guarantees that *wh*-interrogative phrases cannot be fillers in *Bg-df-cl*). Each filler has to correspond to a SLASH element of the HD-DTR. The mother's SLASH value is the HD-DTR's SLASH value after all the fillers' LOCAL values have been removed from it. Unlike the English *top-cl*, *Bg-df-cl* enables both topic and focus fronting. The Discourse Configurability Constraint (DCC) in (28) states that the CONTENT values of all non-head daughters of the *Bg-df-cl* must be either a member of TOPICS or FOCI:

(28) Discourse Configurability Constraint (DCC) on *Bg-df-cl*

$$Bg-df-cl \Rightarrow \left[ \begin{array}{l} \text{DTRS} \left\langle [\text{LOC | CONT } \boxed{C_1}], \dots, [\text{LOC | CONT } \boxed{C_n}], \boxed{1} \right\rangle \\ \text{HD-DTR } \boxed{1} \\ \text{CTXT | INFO-STRUC} \left[ \begin{array}{l} \text{TOPICS} \quad \{\boxed{C_1}, \dots, \boxed{C_k}\} \cup \text{set} \\ \text{FOCI} \quad \{\boxed{C_{k+1}}, \dots, \boxed{C_n}\} \cup \text{set} \end{array} \right] \end{array} \right]$$

Note that nothing prevents the daughters from being members of both TOPICS and FOCI. This allows for 'newly introduced topics' (e.g. the optional *set* of TOPICS could in principle contain any of the CONTENT values  $C_{k+1} \dots C_n$ ). 'New topics' (here, also [+top;+foc] elements) are indeed possible in Bulgarian (as in English left-dislocations; cf. Keenan-Ochs and Schieffelin 1976). Note further that instances of *Bg-df-cl* are also subject to the ISP because *Bg-df-cl* is a subtype of *hd-ph*. This implies that the TOPICS and FOCI values of *Bg-df-cl* correspond to the union of the TOPICS and FOCI values of its daughters.



Thus, if both topics and foci (and possibly other discourse functions) are fronted by the same construction, how, one may ask, can the correct order of DF-fronted elements be predicted given that the *Bg-df-cl* does not place any direct restrictions on the order of its DOM elements? This brings us back to the TFC, stated in (25) above. Since the linear ordering is done by the TFC, the *Bg-df-cl* type only has to state that the DOM elements corresponding to its fillers must be of type *lf* (i.e. that the fillers must be realized within the left periphery field). This is achieved by the Left Periphery Domain Condition (henceforth LPDC; for the notion of Domain Conditions, cf. Kathol 2000) formalized in (29) below. While the LPDC states that the DOM value corresponding to filler daughters must be of type *lf*, the TFC orders these DOM elements (and thereby determines the phonological realization), so that topics precede non-topics (e.g. ordinary focus-fronted phrases and *wh*-phrases, as long as the latter are not topic-fronted). Because not only the *Bg-df-cl* but (as I will show below) also the *Bg-wh-int-cl* is subject to the LPDC, I state this constraint on their common supertype *hd-mult-fill-ph*.

(29) The Left Periphery Domain Condition (LPDC) on *hd-mult-fill-ph*

$$hd-mult-fill-ph \Rightarrow \left[ \begin{array}{l} \text{DTRS} \langle \boxed{1} \dots, \boxed{n}, \boxed{0} \rangle \\ \text{HD-DTR} \boxed{0} \\ \text{DOM} \langle \boxed{1}[lf], \dots, \boxed{n}[lf], \boxed{0} \rangle \end{array} \right]$$

At this point one may wonder why the order among topics and foci is not directly encoded via the *Bg-df-cl*. Recall, however, that Bulgarian also allows for topic-fronted *wh*-phrases. These phrases cannot be daughters of *Bg-df-cl*. Instead, like other non-topical *wh*-phrases, they are fillers in the Bulgarian *wh*-interrogative construction (*Bg-wh-int-cl*). If the ‘topics must precede foci’ constraint were postulated on *Bg-df-cl* (and maybe even *Bg-wh-int-cl*) it would not be possible to derive the fact that topics precede foci in the *whole left periphery*.<sup>16</sup> On the contrary, for the account proposed here, this is not a problem at all. As a matter of fact, everything that is necessary to predict the correct ordering of fronted phrases has already been given above.

Like the *Bg-df-cl* type, *Bg-wh-int-cl* inherits the LPDC from *hd-mult-fill-ph*. This predicts that *wh*-phrases in *wh*-interrogatives have to appear in the left periphery where they are subject to the same linear order constraint as DF-fronted phrases (i.e. the TFC). Here, I do not discuss the details of the *wh*-interrogative construction but merely summarize the formal details for the interested reader. I follow in essence what has been proposed in Ginzburg and Sag (2000). I adopt the Interrogative Retrieval Constraint (Ginzburg and Sag 2000:365) which ensures that, in a question, at least one element of PARAMS is retrieved from the HD-DTR’s STORE. Next, I update the Filler Inclusion Constraint (FIC; Ginzburg and Sag 2000:228),

<sup>16</sup>Accounts that rely on separate types for topic- and focus-fronting and encode linear order directly via those types (rather than via Topological LPs) run into similar problems since there is no easy way to predict the correct order of application for the two construction types.

which guarantees that the extracted *wh*-phrases contribute their WH values to the PARAMS set of the mother. The new version, the Multiple Filler Inclusion Constraint (MFIC) given in (30), is compatible with the *hd-mult-fill-ph*. It also differs from the FIC in that it does *not* allow optional retrieval of additional *params* (which could only come from in-situ *wh*-phrases). The PARAMS value of Bulgarian *wh*-interrogatives is determined exclusively by the WH values of its filler daughters.

(30) Multiple Filler Inclusion Constraint (MFIC) on *Bg-wh-int-cl*

$$Bg\text{-}wh\text{-}int\text{-}cl \Rightarrow \left[ \begin{array}{l} SS | LOC | CONT \quad \left[ \begin{array}{l} \text{PARAMS} \quad \{\pi_1\} \uplus \dots \uplus \{\pi_n\} \\ \left\langle \left[ \begin{array}{l} \text{WH} \quad \{\pi_1\} \end{array} \right], \dots, \left[ \begin{array}{l} \text{WH} \quad \{\pi_n\} \end{array} \right], [\dots] \end{array} \right\rangle \end{array} \right] \\ \text{DTRS} \end{array} \right]$$

I also assume a couple of constraints defined on *words* to guarantee that (a) only fillers in filler-extraction constructions can have non-empty WH values, and (b) all *wh*-phrases with non-empty WH values have to be fronted (cf. WHSP, WHC; Ginzburg and Sag 2000:189).<sup>17</sup>

To sum up, the tectogrammatical analysis of Bulgarian *wh*-interrogatives closely resembles the analysis for English *wh*-interrogatives proposed in Ginzburg and Sag (2000). The *params* contributed by the *wh*-phrases' WH features (i.e. the semantic content of the *wh*-phrases) are added to the mother's PARAM value. This and the fact that the mother's CONTENT value is defined to be of type *question* (that is an abstraction over its head daughter's CONTENT value, which must be a *proposition*; cf. Ginzburg and Sag 2000) create the necessary question semantics whenever *wh*-phrases are fronted. The two main differences to Ginzburg and Sag's approach to the left periphery are that (a) Bulgarian has a *hd-mult-fill-ph*, i.e. it allows multiple DF- and *wh*-fronting (a language-specific difference), and (b) the ordering of fronted constituents (including the *wh*-cluster) is achieved by Topological LP Constraints (a theoretical choice which I have motivated above).

I have already stated that the daughters of the *Bg-wh-int-cl* and *Bg-df-cl* constructions are subject to the LPDC. Thus all topic-fronted phrases (*wh*-phrases or not) will be ordered before SPPs (which I assume to be introduced by a separate construction I do not discuss here) by the TFC, as stated above in (25). 'Normal' *wh*-phrases (i.e. non-topical ones) are correctly predicted to follow SPPs just as (non-topical) foci are predicted to follow SPPs.

Finally, let me come back to the claim I made at the end of section 3.1, namely that it would be advantageous to avoid a specific CLITIC feature. Instead, I suggested that colloquial Bulgarian realizes topical object extraction with an object clitic at the extraction site, whereas formal Bulgarian allows topical object gaps.

<sup>17</sup>In addition to the changes just mentioned some additional small changes have to be made: (a) the constraint on English subject *wh*-clauses that handles the gap-filling for subject gaps (cf. Ginzburg and Sag 2000:237) has to be updated to be compatible with *hd-mult-fill-ph*, and (b) the Inversion constraint (INVC; cf. Ginzburg and Sag 2000:231) is irrelevant for Bulgarian. Since I am not concerned with infinitival *wh*-questions here, I will not discuss the necessity of the Optional *Pro* Condition (OPC; cf. Ginzburg and Sag 2000:231).

While *Bg-df-cl* allows both topical and non-topical fillers, only topical ones can be CDed.<sup>18</sup> This is sufficient to capture the fact that, in colloquial Bulgarian, topic-fronted phrases will have to be CDed further down in the clause (since this was the only way to introduce a topical object gap). Thus, the current proposal for the left periphery (i.e. the extraction target) holds unchanged for both the formal and the colloquial registers of Bulgarian.

## 4 Conclusion

I have provided a general account of the Bulgarian left periphery, focusing on clitic doubling (CD) in *wh*-questions. While fronting of constituents bearing discourse functions is well-researched for non-interrogatives (see references in section 2.1), the possibility of topical *wh*-phrases has mostly been ignored in the literature on Bulgarian.<sup>19</sup> Similar ideas have, however, occasionally been mentioned – mostly under the related label of D-linking – for other languages (e.g. Comorovski 1996; Grohmann 1998; Pesetsky 1987; Scott 2003).

After providing an argument for the general possibility of topic-fronting of *wh*-phrases, be they CDed or not (cf. section 3.2), I outlined a formal account of the Bulgarian left periphery (both the syntax and at least to some degree the semantics). The account employs topological fields and Linear Precedence Constraints defined on them, thereby distinguishing between pheno- and tectogrammatical properties of the left periphery. The analysis handles topic-fronting in and outside of *wh*-interrogatives as well as simple *wh*-interrogatives (without topic-fronting). As it stands, the overall framework assumed for the phenogrammatical analysis is a hybrid of Kathol (1995, 2000) on the one hand and Donohue and Sag (1999) on the other hand. What I really had in mind while drafting this analysis is, however, a version of construction grammar in which constructions are – among other things – responsible for organizing the information necessary for the linear ordering of their daughters. Although this is in some respect close to what I have proposed here, the current analysis would benefit from being restated (and refined) within a construction grammar framework of that type.

By basing the order of the fronted periphery on grammaticalized sensitivity to a general pragmatic concept (namely *topicality*), rather than on a morpho-syntactic feature of colloquial Bulgarian (i.e. CD), the present account works both for colloquial and formal Bulgarian and can in principle be extended to other languages with similar left periphery ordering (e.g. Russian, which also seems to allow topic-fronting of *wh*-phrases; cf. Scott 2003).

Finally and maybe most importantly, once we accept the hypothesis proposed in section 3.2 that *wh*-phrases followed by e.g. a parenthetical are topical (including subject *wh*-phrases, as in (19) and (22) above), this sheds new light on

<sup>18</sup>To be precise, CD *defines* whichever filler the clitic agrees with as topical.

<sup>19</sup>Dimitrova-Vulchanova and Hellan (1998, 1995/1999) and Jaeger (2002) mention CD in *wh*-interrogatives without directly relating it to topicality.

what has traditionally been called ‘Superiority effects’. Suddenly, the fact that subject *wh*-phrases occur clause-initially in a large majority of Bulgarian clauses ‘suspiciously’ resembles the fact that, cross-linguistically, subjects have been most frequently observed to be topics (cf. Lambrecht 1994:131f.). Thus I take it to be of crucial importance to investigate to which extent ‘Superiority’ (in Bulgarian as much as in other languages) can be accounted for by semantic and/or pragmatic facts.

## References

- Alexandrova, G. 1997. Pronominal clitics as g(eneralized) f(amiliarity)-licensing agro. In *Formal Approaches to Slavic Linguistics: The Cornell Meeting, 1995*, ed. Browne et al., 1–31. Ann Arbor: Michigan Slavic Publications.
- Avgustinova, Tania. 1997. *Word order and clitics in Bulgarian*, volume 5 of *Saarbrücken Dissertations in Computational Linguistics and Language Technology*. Saarbrücken: DFKI.
- Billings, Loren, and Catherine Rudin. 1996. Optimality and superiority: A new approach to overt multiple-*wh* ordering. In *Formal Approaches to Slavic Linguistics, The College Park meeting, 1994*, ed. J. Toman, 35–60. Ann Arbor: Michigan Slavic Publications.
- Billings, Loren, and Catherine Rudin. 1998. Animacy and focus in Bulgarian *wh* questions. In *Proceedings of Conference on Slavic Morphosyntax, Bloomington, 1998*.
- Boeckx, Cedric, and Kleanthes Grohmann. 2003. SubMove: Towards a unified account of scrambling and D-linking. In *Peripheries*, ed. David Adger; Cécile de Cat and George Tsoulas. Dordrecht: Kluwer.
- Bošković, Željko. 1993. On certain violations of the superiority condition, agro, and economy of derivation. *Journal of Linguistics*.
- Bošković, Željko. 1998a. Multiple WH-fronting and economy of derivation. In *The Proceedings of the Sixteenth West Coast Conference on Formal Linguistics held at the University of Washington*. Stanford: CSLI Publications.
- Bošković, Željko. 1998b. On the interpretation of multiple questions. Ms.
- Chomsky, Noam. 1973. Conditions on transformations. In *A Festschrift for Morris Halle*, ed. St. Anderson and P. Kiparsky. New York: Holt, Reinhart and Winston.
- Comorovski, Ileana. 1996. *Interrogative phrases and the Syntax-Semantics Interface*. Dordrecht: Kluwer Academics.
- Dimitrova-Vulchanova, Mila, and Lars Hellan. 1995/1999. Clitics and Bulgarian clause structure. In *Clitics in the languages of Europe*, ed. Henk van Riemsdijk, volume Empirical Approaches to Language Typology of *Language Typology*. Berlin: Mouton de Gruyter.
- Dimitrova-Vulchanova, Mila, and Lars Hellan. 1998. Introduction. In *Topics in south Slavic syntax and semantics*, ed. Mila Dimitrova-Vulchanova and Lars Hellan, ix–xxvii. Amsterdam and Philadelphia: John Benjamins Publishing.
- Donohue, Cathryn, and Ivan A. Sag. 1999. Domains in Warlpiri. In *Sixth International Conference on HPSG–Abstracts. 04–06 August 1999*, 101–106. Edinburgh. <http://www-csli.stanford.edu/~sag/papers/warlpiri.pdf>.

- Dowty, David. 1996. Towards a minimalist theory of syntactic structure. In *Discontinuous constituency*, ed. Harry Bunt and Arthur van Horck, 11–62. Mouton de Gruyter.
- Engdahl, Elisabet, and Enric Vallduví. 1996. Information packaging in hpsg. *Edinburgh Working Papers in Cognitive Science, Studies in HPSG* 12:1–31.
- Franks, Steven, and Tracy Holloway King. 2000. *A handbook of Slavic clitics*. Oxford: Oxford University Press.
- Ginzburg, Jonathan, and Ivan Sag. 2000. *Interrogative investigations. the form, meaning, and the use of English interrogatives*. Stanford: CSLI Publications.
- Grewendorf, Günther. 2001. Multiple wh-fronting. *Linguistic Inquiry* 32:87–122.
- Grohmann, Kleanthes. 1998. Syntactic inquiries into discourse restrictions on multiple interrogatives. *Groninger Arbeiten zur germanistischen Linguistik* 42:1–60.
- Grohmann, Kleanthes. under review. Top issues in questions. In *tba*. MIT Press.
- Jaeger, T. Florian. 2002. On the optional and obligatory realization of the Bulgarian direct object clitic. Ms, Stanford University.
- Jaeger, T. Florian. to appear. Topicality and superiority in Bulgarian wh-questions. In *Proceedings of Formal Approaches to Slavic Linguistics: The Ottawa meeting, 2003*, volume 12. Ann Arbor: Michigan Slavic Publications.
- Jaeger, T. Florian, and Veronica Gerassimova. 2002. Bulgarian word order and the role of the direct object clitic in LFG. In *Proceedings of the LFG02 Conference, Athens, July 3rd - 5th*, ed. M. Butt and T. King, 197–219. Stanford: CSLI Publications.
- Jaeger, T. Florian, and David Oshima. 2002. Towards a dynamic model of topic marking. In *Pre-proceedings of the Information Structure in Context Workshop, Stuttgart University, 15–16 November 2002*, 153–67. Stuttgart.
- Kallulli, Dalina. 2001. Direct object clitic doubling in Albanian and Greek. In *Comparative syntax of Balkan languages*, ed. M.L. Rivero and A. Ralli. Oxford University Press.
- Kathol, Andreas. 1995. Linearization-based German syntax. Doctoral Dissertation, Ohio State University.
- Kathol, Andreas. 2000. *Linear syntax*. Oxford: Oxford University Press.
- Keenan-Ochs, E., and B. Schieffelin. 1976. Foregrounding referents: A reconsideration of left-dislocation in discourse. In *Proceedings of BLS*, volume 2, 240–57. Berkeley.
- King, Tracy H. 1995. *Configuring topic and focus in russian*. in: *Dissertations in linguistics*. Dissertations in Linguistics. Stanford: CSLI Publications.
- Kuno, Susumu, and Ken-Ichi Takami. 1993. *Grammar and discourse principles: Functional syntax and gb theory*. Chicago: Chicago University Press.
- Lambova, Mariana. 2003a. On information structure and clausal architecture: Evidence from Bulgarian. Doctoral Dissertation, University of Connecticut. DRAFT.
- Lambova, Mariana. 2003b. On the interaction of multiple (non) wh-fronting and multiple topicalization in Bulgarian. In *The 26th Penn Linguistics Colloquium*.

- Lambova, Mariana. 2003c. When is splitting the wh-cluster in Bulgarian possible? In *Formal Approaches to Slavic Linguistics: The Amherst Meeting, 2002*, ed. Wayles Browne, Ji-yung Kim, Barbara H. Partee, and Robert Rothstein, volume 11, ? Ann Arbor: Michigan Slavic Publications.
- Lambrecht, Knud. 1994. *Information structure and sentence form. topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.
- Leafgren, John R. 1997. Bulgarian clitic doubling: Overt topicality. *Journal of Slavic Linguistics* 5:117–143.
- Miller, P., and Ivan A. Sag. 1997. French clitic movement without clitics or movement. *NLLT* 15:573–639.
- O'Connor, Rob. 2002. Clitics and phrasal affixation in Constructive Morphology. In *Proceedings of the LFG02 Conference, Athens, July 3rd – 5th*, ed. Butt and King, 315–332. Stanford: CSLI Publications.
- Pavlov, Ivan V. 2000. The non-argument functional field of Bulgarian clause structure. Doctoral Dissertation, NTNU Trondheim.
- Pesetsky, David. 1987. Wh-in-situ: Movement and unselective binding. In *The representation of (in-)definiteness*, ed. Reuland and Meulen, 98–129. Cambridge: MIT Press.
- Pesetsky, David. 2000. *Phrasal movement and its kin*, volume 37 of *Linguistic Inquiry, Monograph*. Cambridge: MIT Press.
- Reape, Mike. 1994. Domain union and word order variation in German. In *German in Head-Driven Phrase Structure Grammar*, ed. John Nerbonne; Klaus Netter and Carl J. Pollard, number 46 in CSLI Lecture Notes, 151–197. Stanford: CSLI Publications.
- Richards, Norwin. 1997. What moves where when in which language. In *MIT Working Papers in Linguistics*. MIT Press.
- Rudin, Catherine. 1985. *Aspects of Bulgarian syntax: Complementizers and wh constructions*. Columbus, Ohio: Slavica Publishers.
- Rudin, Catherine. 1988. On multiple questions and multiple wh fronting. *Natural Language and Linguistic Theory* 6:445–502.
- Rudin, Catherine. 1997. AgrO and Bulgarian pronominal clitics. In *Formal Approaches to Slavic Linguistics: The Indiana meeting, 1996*, 224–52. Ann Arbor: Michigan Slavic Publications.
- Scott, Tatiana. 2003. Cp-topic parameter and Russian WH. Handout for LSA Annual Meeting 2003, Atlanta, GA.
- Steedman, Mark. 2000. Information structure and the Syntax-Phonology Interface. *Linguistics Inquiry* 31:649–89.