ON THE OPTIONAL AND OBLIGATORY REALIZATION OF
THE BULGARIAN DIRECT OBJECT CLITIC

Topic Agreement Marker and Default Pronoun - A Formal Account

T. Florian Jaeger
Linguistics Department, Stanford University

Qualifying Paper, Stanford 12/06/2002
ABSTRACT – This paper provides a Lexical Functional Grammar (LFG) account of the Bulgarian direct object clitic's interaction with information structure (i.e. topic-focus structure) and word order. I show that the direct object clitic has at least two functions (it is both a topical object agreement marker and default pronoun), which are licensed by different environments. The analysis presented accounts for the distribution of the two different functions of the direct object clitic by means of general principles of LFG (i.e. Uniqueness, Completeness, Extended Coherence). Unlike in most – if not all – existing formal accounts (e.g. Rudin 1997, Avgustinova 1997), this analysis predicts when overt realization of the direct object clitic is possible and when it is obligatory.

In order to achieve this, I allow for two different ways to identify a topical object. A topical object can either be realized in grammaticalized TOPIC-position, in which case it the has to be doubled, or remain in its default position and optionally be doubled. Which of the two strategies is chosen depends on the discourse situation. The proposed analysis is based on extensive evidence (an online experiment, as well as the data provided by Leafgren 1997a,b, 1998, and Avgustinova 1997), and incorporates recent findings on the discourse-configurationality of the left periphery in Bulgarian clauses (Rudin 1997, Arnaudova 2001, Dimitrova-Vulchanova & Hellan 1998). While covering a much broader range of data from spoken Bulgarian than previous formal accounts, the analysis makes the right predictions about possible word orders and the optional, or obligatory presence/absence of the direct object clitic. Furthermore, the analysis does not rely on the assumption of configurationality, which has been shown to be problematic for Bulgarian (Gerassimova & Jaeger 2002).

Table of Contents

1 INTRODUCTION ...................................................................................................................................................3
  1.1 Data sources.................................................................................................................................................5
  1.2 Clitic Doubling ............................................................................................................................................6
  1.3 Overview of the paper .................................................................................................................................7
2 AN INTRODUCTION TO BULGARIAN ....................................................................................................................8
  2.1 Dependent Marking.....................................................................................................................................8
  2.2 Word order...................................................................................................................................................9
  2.3 Head Marking and the Predicate Clitic Cluster ..........................................................................................9
  2.4 Two types of Pronouns..............................................................................................................................11
  2.5 Configurationality ......................................................................................................................................12
3 PREVIOUS ANALYSES OF THE DOC...................................................................................................................13
4 DF-CONFIGURATIONALITY AND THE DOC IN CLITIC DOUBLING .................................................................20
  4.1 Phrase structure positions of Discourse Functions ...................................................................................20
  4.2 Obligatory doubling of TOPIC-fronted direct objects................................................................................24
  4.3 Optional vs. obligatory Clitic Doubling....................................................................................................26
  4.4 Functionally determined default word order.............................................................................................32
5 THE DOC AS DEFAULT PRONOUN ..................................................................................................................34
6 ISLAND DATA: WHEN CAN WHICH TYPE OF DOC OCCUR? ...........................................................................36
7 CONCLUSIONS AND OUTLOOK ......................................................................................................................41
  7.1 Summary of the DOC's functions and typological conclusions...............................................................41
  7.2 Theoretical Implications............................................................................................................................43
  7.3 Open question and possible further research ............................................................................................43
8 REFERENCES .....................................................................................................................................................45
9 APPENDIX: GF-(NON-)CONFIGURATIONALITY IN BULGARIAN .................................................................51
1 Introduction

In some contexts, contemporary colloquial Bulgarian allows for overt realization of object clitics. The object clitics can occur as the only realization of the object, as in (1), or double an NP or a long form pronoun, as in (2). The latter construction is commonly called clitic doubling. Bulgarian has both direct and indirect object clitics, which have an almost identical distribution. Here, however, I restrict myself to the investigation of the direct object clitic (henceforth DOC). Throughout the paper, the DOC and the coreferential object (if present) are underlined.

(1) Decata ja običat. The children love her.

(2) Decata ja običat Marija/neja. The children love Maria/her.

1 This QP is for a great part based on two papers I wrote together with Veronica Gerassimova. Section 2, 3, 4, 5, and 6 are loosely based on Jaeger & Gerassimova (2002a). The Appendix (i.e. section 9) contains a summary of the second paper, Gerassimova & Jaeger (2002).

* My special thanks go out to Peter Sells for encouraging me that the research presented here is worthwhile pursuing, thanks to Joan Bresnan, Veronica Gerassimova, Tracy H. King, Chris Manning, Elizabeth Traugott, Arnold Zwicky for their advice and support all throughout the progress of my research. I want to thank Mary Dalrymple, Tracy H. King (again), and Jonas Kuhn for their help with formal aspects of LFG. I also want to very much thank Ruth Kempson for making me aware of several interesting questions and providing good ideas how to approach them, as well as Iskra Iskrova for discussing the relevant data with us. Last but not least, I benefited from the questions and suggestions from the anonymous reviewers of the LFG02 abstracts, as well as Shiao-Wei Tham, Judith Tonhauser, Andrew Koontz-Garboden, and especially Lev Blumenfeld. We are also grateful for the great feedback I got at the LFG02 and at an earlier presentation of my work for the Linguistics Department, Stanford. Thanks to Tracy H. King (yet again) and Lev Blumenfeld for feedback on the final draft of this paper. All remaining mistakes remain mine and must not be reproduced without my permission, :-). I could not have benefited from the LFG02 conference, ESSLLI2002, and the DGfS & LSA summer school 2002 without financial support. Attending those conferences was made possible by grants from the DGfS, the ESSLLI committee, the School of Humanities and the Linguistics Department at Stanford University.

1 I use the term direct object clitic (DOC) to refer to the set of linguistic forms of the direct object clitic, not their meaning. These are the following forms: SG – 1st me, 2nd te, 3rd masc./neut. go, fem. ja; PL – 1st ni, 2nd vi, 3rd gi.

2 The following glosses are used: 1, 2, 3 – first, second, and third person FEM – feminine, MASC – masculine, NEUT – neuter; PL – plural, SG – singular; REFL – reflexive pronoun, IOC – indirect object clitic, DOC – direct object clitic, AUX – auxiliary clitic; SBJ – subjunctive; +/-DEF – presence/absence of definite suffix, -DEF – indefinite specific article; +/-SPEC – specific or non-specific nominal material. SMALL CAPS indicate emphatic accent and # indicates an intonational break.

In addition, the following abbreviations are use: CD – clitic doubling, DOC – direct object clitic, DF – discourse function, GF – grammatical function.
In this paper, I argue that there are at least two different uses of the DOC, which are licensed by different environments. I investigate the DOC’s interaction with syntax (especially word order) and information structure and provide a formal account that makes the correct predictions about the distribution of the overt realizations of the DOC in its two uses. First, in the clitic doubling construction, the DOC is a non-anaphoric direct object topic-agreement marker. Second, the DOC is the default direct object pronoun. Third, the DOC is an intrusive direct object pronoun in extractions.

This last use probably is a variant of the function (just like in English where the default pronouns can be used to indicate extraction "traces" under certain conditions). To the best of my knowledge, this function has not been considered in the literature at all. Although this is not clear on the first view, I argue that the DOC’s use as an intrusive pronoun has to be distinguished from its use in clitic doubling (henceforth CD). While (3) shows the DOC as an intrusive pronoun (Sells 1984), (4) is an example of TOPIC-fronting out of a sentential subject. The topic-agreement marker DOC in (4) is obligatory; the DOC in (3) is an optional marker of direct object extraction.

(3) Radioto, koeto Todor otide na plaž [bez da radio+DEF.3.SG.MASC which Todor went3 on beach without SBJ (go) izključi], e na Elena. DOC3.SG.MASC switch-off is of Elena
The radio which Todor went to the beach without switching off is Elena’s.

(4) Todor e jasno, če Ivan *(go) e vidjal. Todor is clear that Ivan DOC3.SG.MASC is seen
Todor, it is clear that Ivan has seen him.

Here, I only discuss the first two functions of the DOC, i.e. the agreement marker and the default pronoun use, and leave aside the intrusive pronoun use. I argue that some of the confusion about these functions in the literature is due to different notions of topic and propose a way to resolve this issue within LFG.

Furthermore, the literature on Bulgarian has for the largest part treated the DOC as if it had one uniform function. I argue that this misconception is at least one of the reasons why the role of the DOC and its interaction with Bulgarian syntax is still unresolved. I also argue that the analysis presented here accounts for a whole range of data that so far have not been explained, including examples of word order variation and its correlation with information

3 However, the difference between the agreement marker use and the intrusive pronoun use will show up in section 6, when I discuss Bulgarian island data (see especially Table 2, p. 36).
packing (Vallduvi 1992, 1993). Specifically, I show how the range of possible word orders depends on the presence or absence of the DOC, and how this can be accounted for in a formal framework. The last two points are especially important since they have by and large either been ignored in the literature or at least have not been treated in sufficient detail (e.g. Rudin 1990/1991, 1997, Avgustinova 1997). Moreover, I use the case of island violations to illustrate that the analysis makes the correct predictions about the distribution of the DOC as default pronoun or topic marker. That is, I show that the two use of the DOC are actually licensed by different constructional environments.

The research presented here relates to the literature on D[iscourse]- and G[rammatical]-F[unction]-configurationality, an issue which has been identified as the primary, unresolved issue in the literature on (South) Slavic Syntax (cf. Siewierska & Uhlířová 1998:143 in their review of the recent literature on the word order of Slavic languages).

In the following paragraphs, I proceed as follows. First, I describe the data sources which the current analysis is based on. Second, I give a short introduction to the clitic doubling construction, and, finally I provide an overview of the remaining paper.

1.1 Data sources

To guarantee a broad coverage of data, I test and compare the predictions of my analysis with the data provided in Avgustinova (1997; elicited question-answer pairs) and Leafgren (1997a,b, 1998, 2001; corpus studies of written/spoken, informal/formal Bulgarian). Avgustinova (1997), among other things being interested in word order variation in Bulgarian, collects possible answers to question pairs of the type "What about the book? Who read it?". The examples given in her thesis have been recorded from native speakers in a number of controlled experiments. Avgustinova's (1997) transcriptions, therefore, contain reliable information on accent assignment. Thus these data are ideally suited for research on discourse-related phenomena, such as object clitic reduplication (cf. section 4). Leafgren (1997a,b, 1998, 2001, 2002) on the other hand bases his studies in part on self-compiled corpora (two sci-fi novels for informal written Bulgarian, and the prescriptive grammar of Bulgarian for formal written Bulgarian) but also makes use of online corpora. He uses
Aleksova's corpus\(^4\) of conversational, very informal, spoken Bulgarian and a corpus of parliament debates\(^5\) for formal, spoken language (cf. Leafgren 2002:3). In addition to Leafgren's and Avgustinova's data, this paper makes reference to examples from the literature, as well as elicited examples (from 3 native speakers of Bulgarian in face-to-face interviews), and in some cases elicited acceptability judgments (from 7 speakers of Bulgarian by means of an online magnitude estimation\(^6\) experiment). By combining those sources and the research presented here, the analysis in section 4 is grounded in and tested against a rich empirical base.\(^7\)

1.2 Clitic Doubling

Next, I briefly clarify my use of the term 'clitic doubling'. With clitic doubling (CD) I refer to the overt doubling of a constituent, usually an argument (here the direct object), by a phonologically weak, syntactically non-projecting\(^8\) lexical element, i.e. a clitic (here the DOC). CD is a prominent topic in the literature on Slavic and Balkan linguistics (e.g. Franks & King 2000, Rudin 1990/1991, 1996, 1997, Dyer 1992, Guentchéva 1994, a.o.), the typology of pronouns, agreement (e.g. Bresnan & Mchombo 1987), configurationality (e.g. Baker's (1991) *pronominal object hypothesis*), and case assignment (e.g. Rudin 1997). The aspects of CD that are addressed here include the following. First, how is coreference between the clitic and the doubled NP established? In MP/GB this comes down to the question whether, for example, fronted objects are moved or anaphorically bound by the DOC. In LFG terms, this corresponds to the issue of functional control vs. anaphoric binding. Second, does the clitic mark a grammatical function (GF) or a discourse function (DF)\(^9\) or both? Third, is the clitic and/or the lexical object NP the object argument? This is interesting

---

\(^4\) This corpus was collected by Krasimira Aleksova during the late 1980s and early 1990s for her dissertation (Aleksova 1994/2000) and made available for scientific use, for example, at [http://www.hf.uio.no/east/bulg/mat/Aleksova/](http://www.hf.uio.no/east/bulg/mat/Aleksova/).

\(^5\) The corpus of Bulgarian parliament debates contains approximately 20,000 words and was transcribed by Ivanka Mavrodieva during the 1990s. It is publicly available at [http://www.hf.uio.no/east/bulg/mat/Parliament/](http://www.hf.uio.no/east/bulg/mat/Parliament/).

\(^6\) See Bard et al. (1996).

\(^7\) There are two more publicly available sources of Bulgarian language data, which are not considered here. First, Cvetanka Nikolova (1987) collected a 50,000 word corpus of spoken Bulgarian. This corpus can be found at [http://www.hf.uio.no/east/bulg/mat/Nikolova/](http://www.hf.uio.no/east/bulg/mat/Nikolova/). Second, as part of the COP Project 106 MULTTEXT-East a Bulgarian corpus of informal, written language containing approximately 100,000 words was made available at [http://nl.ijs.si/ME/CD/docs/mte-d21f/mte-D21F.html](http://nl.ijs.si/ME/CD/docs/mte-d21f/mte-D21F.html).

\(^8\) See Toivonen (2001:Chapter 3) for a typology of non-projecting words.

\(^9\) Note that, with 'discourse function', we do not refer to discourse function as defined in Schiffrin (1988) or Fraser (1988). We comply with the naming convention of LFG and use the term discourse function (DF) to refer to what more precisely could be called f-structure correlate of an information structural role.
since according to some theories (e.g. GB, MP) only one constituent can be assigned the same theta-role (and via that in most cases also CASE). The LFG framework is less restrictive in this respect. As long as UNIQUENESS\(^\text{10}\) (Bresnan 2001:47) is fulfilled, information belonging to the same GF (i.e. theta-role) can be distributed among several syntactic constituents. Nevertheless, translated into LFG, the question of whether the clitic provides information on OBJ PRED (i.e. the PRED value of the object) still remains.

1.3 Overview of the paper

The remainder of the paper is organized as follows. In section 2, I introduce some basic facts about Bulgarian, and formalize some phrase structure rules, which capture the internal order of the predicate clitic cluster and the fact that Bulgarian is non-configurational. The reader who is familiar with Bulgarian may skim over this section just to be familiar with the above-mentioned phrase structure rules. In section 3, I briefly describe some earlier analyses of the DOC’s function. In section 4, I introduce recent findings on the discourse-configurationality in Bulgarian, incorporate them into my analysis, and formalize the DOC’s properties in CD. In this context, I discuss my proposal in light of the known data and show how the interaction of the proposed lexical entry of the DOC and the proposed phrase structure rules make the right predictions about grammaticality of certain word orders and their information structural correlates. The term information structure will be clarified in section 4.3 where I elaborate on the interaction between an information structure module and other parts of the grammar. For now, it is sufficient to understand information structure as a component of grammar, which takes care of the information packaging (i.e. the packaging of the information expressed by the proposition of a sentence into ”What does the sentence say something about”, ”What is the new information provided by the sentence”; Vallduví 1992, 1993). In section 5, I discuss a second function of the DOC, which has so far been ignored in the literature, namely its use as the default pronoun of Bulgarian. In section 6, I briefly survey data from island violations to show how the account makes the right predictions about the distribution of the different uses of the DOC. Last, I will summarize my conclusions and mention some open issues (section 7).

\(^{10}\) The UNIQUENESS condition stated in Bresnan (2001:47) states that every f-structure attribute has exactly one value.
2 An introduction to Bulgarian

Bulgarian is a South-Slavic language spoken by approximately 9 million speakers world wide. If not mentioned otherwise, I will use the term Bulgarian to refer to contemporary, colloquial, spoken Bulgarian. Bulgaria has a strong prescriptive tradition and the differences between written vs. spoken and formal vs. informal Bulgarian seem to be immense. CD is very rare in formal and written Bulgarian. Leafgren (2001:4) shows that the frequency of CD in formal written texts (0.5% of all object occurrences) contrasts sharply with the 10% frequency of CD in informal oral texts. Furthermore, I restrict myself to those dialects of Bulgarian which make productive use of the object clitics, i.e. mostly the Western dialects (cf. Leafgren 1997a:119).

Since Bulgarian is in many respects the most atypical Slavic language and has some typologically uncommon properties, I will briefly sketch those characteristics of Bulgarian that prove relevant to the analysis in section 4.

2.1 Dependent Marking

Unlike all other Slavic languages except for Macedonian, Bulgarian has lost its case marking system (Siewierska & Uhlířová 1998:105, Franks & King 2000:17). Some scholars have argued that the definiteness suffix (singular: masc. -a, fem.: -ta, neut.: -to; plural: -te/-ta) identifies the subject. This is wrong since the definiteness suffix can also be attached to an object and not all subjects have the definite suffix, as shown in (5). Furthermore, the definite suffix attaches to objects of prepositions, as shown in (6).

(5) Boba go izjade momče. (NE momiče)
beans.DEF DOC3.SG.MASC ate boy.DEF not girl.DEF
A BOY ate the beans (not a GIRL).

(6) ... do revoltucitata ...
... before revolution.DEF

12 During an online experiment that we designed to get native speaker judgments on contemporary, colloquial, spoken Bulgarian we first ran into problems since my informants were so strongly influenced by the idea that they had to judge the prescriptive correctness instead of 'what they actually say'.
The only dependent-marking device in Bulgarian is the preposition *na* which among other things identifies the indirect object. In certain environments even this last bit of dependent-marking can be dropped (cf. Vakareliyska 1994).

### 2.2 Word order

Despite the almost complete lack of dependent-marking, Bulgarian has free word order. With different requirements on the context, the intonation and morpho-syntactic marking, all theoretically possible word orders can actually be observed (Siewierska & Uhliřová 1998:107-10 and implicitly Avgustinova 1997:112). While I provide more details on the effect of the DOC on word order in section 4, it is generally true that some word orders are not possible without the DOC. In other words, the DOC seems to 'license' certain word orders. The role of the DOC as a licenser of word order is illustrated by the contrast in terms of the absence/presence of the DOC between (7) vs. (9) and (8) vs. (10) (based on Avgustinova 1997:112).

(7) *Parite vze Olga.
    money_DEF took3.SG Olga

(8) *vze Olga parite.
    took3.SG Olga money_DEF

Intended: Olga took the money.

(9) Parite gi vze Olga.
    money_DEF DOC3.PL took3.SG Olga

(10) vze gi Olga parite.
    took3.SG DOC3.PL Olga money_DEF

Note, however, that Bulgarian shows a clear preference for a SUBJ-V-DO-IO surface order, a tendency noted by several scholars (cf. Leafgren 2002:1, Dyer 1992:63, Avgustinova 1997:114, a.o.). Leafgren (2002:1) argues that averaged over all registers and genres about 80.5% of all sentences are SVO. Dyer (1992) shows that, like in other Slavic languages, SVO in Bulgarian is not only statistically the most common constituent order but also *stylistically neutral*.

### 2.3 Head Marking and the Predicate Clitic Cluster

The lack of stringent word order and case marking is at first surprising. However, Bulgarian has other means to identify grammatical functions, namely intonation and head-marking. Here, I focus on head-marking, more precisely one kind of head-marking in Bulgarian, clitic doubling by the direct object clitics. Before I turn to the interaction of the direct object clitic and word order, I want to briefly mention some other morphosyntactic facts of Bulgarian. First, the sentence predicate agrees with the subject in person and number,
as in (9) and (10) above, and participles (i.e. subjunctives) agree also in gender. The predicate combines with the clausal clitics into the predicate clitic cluster. Because there is an extensive literature on the internal order of the predicate clitic cluster (e.g. Avgustinova 1997, Siewierska & Uhlířová 1998; see Franks & King 2000:234ff. for a summary), I do not discuss this issue here. To understand the examples given later in this paper, it is sufficient to bear in mind schema (R1) for the internal order of the predicate clitic cluster (Englund 1977:109-19). The auxiliary clitic (AUX) precedes the indirect object clitic (IOC) and the DOC if it not 3.SG.

(R1)  AUX (not 3.SG) > IOC > DOC > AUX (3.SG)

Although the issue of how the internal order of the clitic cluster is achieved is still unresolved (cf. Rudin et al. 1999, Franks & King 2000:chapter 13, O’Connor 2002), here I assume that the final ordering of the clitics in (R1) is not syntactically determined. In this approach, syntax only determines the position of the predicate clitic cluster. There are two phrase structure rules compatible with this claim, which are given in (R2) and (R3). In this paper, I will use the variant in (R3) because it does not stipulate any spurious structure and thus is preferred by ECONOMY OF EXPRESSION. Also, ECONOMY OF EXPRESSION will prevent (R3) from being applied iteratively. Thus all clitics in the predicate clitic cluster and the verb itself are generated in a flat V-structure and ordering is achieved outside of syntax. Note that, employing Toivonen’s (2000:Chapter 3) typology, I assume that all clitics in the predicate clitic cluster are non-projecting.

(R2)  V → (CL)  V
      ↑=↓  ↑=↓

(R3)  V → CL*  V
      ↑=↓  ↑=↓

---

13 This is different from the approach suggested in Jaeger & Gerassimova (2002).
14 Note that there has been a lot of confusion about the principle of the ECONOMY OF EXPRESSION. Here, I do not mean Bresnan’s (2001:91) ECONOMY OF EXPRESSION, which evaluates structures for a already given set of phrase structure rules, rather I compare two candidates for phrase structure rules and choose the one that is more economical, i.e. produces less phrase structure nodes while conveying the same information. A notion of economy like the one employed here is not at all uncommon (Jonas Kuhn, p.c.).
15 In (R2) and (R3), CL is a category label variable which subsumes the verbal (V_{CL}) and nominal (N_{CL}) clitics in the predicate clitic cluster. Note that this is a change to earlier versions of this paper (cf. Jaeger & Gerassimova 2002). The current version is more similar to an older version of the rule, which goes back to a discussion with Peter Sells.
The constituent-structure tree fragment in c-structure 1, corresponding to the predicate clitic cluster in (11), illustrates the effect of (R3).16

(11) ... sa go uvolnili
     AUX3.PL DOC3.SG.MASC fired
     ... we fired him.

The clitic cluster as a whole is preverbal except for the cases where this would cause the clitics to be clause-initial, as in (10) above, where go appears after the verb vze. In those cases, the verb is preposed to the clitic cluster. In other words, the positioning of the Bulgarian clitic cluster is subject to the Tobler-Mussafia effect (cf. Tomić 1997, 1996, Rudin et al. 1998:566; for an OT account to typology of clitic positioning, see Billings 2000) and not to Wackernagel's Law (unlike the clausal clitics in almost all other Slavic languages). The object clitics and therefore the DOC, which is the subject of this paper, are clausal clitics. In the case of clitic doubling, the object clitic(s) agree in person, number and gender (only for 3.SG) with the reduplicated object.

2.4 Two types of Pronouns

Unlike the object clitics, which can only occur in the clitic cluster, the second kind of pronouns in Bulgarian, namely the long form pronouns, have the same syntactic distribution as full lexical NPs. The long form pronouns, when occurring alone, mark contrastive or emphatic focus (Avgustinova 1997:116, Vakareliyska 1994:125; see Leafgren 1997a:118 for a table of all clitic pronouns and long form pronouns), in which case they always receive stress (compare (12) and (13) below).

---

16 The following two categories are subcategories of CL and can appear in the CL phrase structure positions: N_{CL} for the pronominal clausal clitics, such as the DOC go in (11) or indirect object clitics and reflexive clitics, and V_{CL} for the auxiliary clitics, such as sa in (11).
To sum up what has been said so far, Bulgarian is a non-case marking, partially head-marking, free word order language with optional clitic doubling of objects.

### 2.5 Configurationality

Another important aspect of Bulgarian, which has been ignored in the literature so far, is the lack of evidence for G[rammatical] F[unction]-configurationality (Nordlinger 1998:43). Although Rudin (1985:25, footnote 13) already mentions that there seems to be no evidence for it, GF-configurationality is assumed and plays a crucial role in most recent analyses of Bulgarian syntax (including those on CD). There are good reasons to question this assumption. Gerassimova & Jaeger (2002) have shown that it is difficult if not impossible to find evidence for GF-configurationality. More precisely, some tests, such as weak crossover tests, variable binding tests, extraction tests, etc., clearly argue for non-configurationality of Bulgarian.\(^\text{18}\)

Therefore, although it does not affect the analysis, GF-configurationality is not assumed here. The annotated phrase structure rule in (R4) captures the lack of evidence for GF-configurationality and describes a flat VP with unordered constituents (cf. Kiss 1995:11 for Hungarian).\(^\text{19}\)

\[\text{(R4)} \quad \text{VP} \to \text{XP}^* \land \text{V} \land \text{XP}^* \\
\uparrow \text{GF} \downarrow \uparrow \downarrow \uparrow \downarrow \text{GF} \downarrow \downarrow \]

In section 4, I show that the flat VP hypothesis in combination with discourse configurationality (Kiss 1995, 2001) is necessary for or at least highly compatible with the formal account of CD and its interaction with possible word orders presented here. Prior to this, I briefly summarize previous analyses of the DOC.

---

\(^{17}\) In examples throughout the paper, I mark emphatic accent/stress with SMALL CAPS. Although only a part of the word receives emphatic accent I will just mark the whole word as prosodically emphasized.

\(^{18}\) For the interested reader, the appendix in section 9 summarizes the relevant aspects of Gerassimova & Jaeger (2002).

\(^{19}\) We use the XP annotated with \(\uparrow \text{GF} \downarrow \downarrow \) to express that all kinds of core arguments can occur in this position (including e.g. COMPs).
3 Previous analyses of the DOC

All of the accounts discussed here have exclusively dealt with C[litic] D[oubling] (sometimes also referred to as clitic replication in the literature) and ignored other uses/functions of the DOC. To the best of my knowledge, the function of the DOC as default pronoun (section 5) and its interaction with the use of the DOC in the CD construction have not been described by anyone yet. The existing accounts of the DOC can be distinguished by their basic hypothesis. In the following paragraphs, each of the accounts will be discussed in the order they are listed below.

(H1) The object clitics mark non-canonical word orders.
(H2) The object clitics mark case (of the doubled constituent).
(H3) The object clitics mark definite objects.
(H4) The object clitics mark specific objects.
(H5) The object clitics mark topical objects.

Both (H1) and (H2), i.e. the word order marker and the case marker hypotheses, suggested in AG (1983,3:187-188, 282-283), Popov (1963:166, 229-230), Cyxun (1968:110) and Georgieva (1974:75), have in common the claim that CD together with word order serves to disambiguate case roles. Leafgren (1997a:124) concludes that under this view sentences with CD should be structurally unambiguous even if both subject and object have the same gender, number, etc. However, this is not the case. Sentences with CD are compatible with different word orders. For example, as shown below, both VSO, as in (14), and VOS, as in (15) word orders are possible with the same stress assignment as long as the clitic is present (based on Avgustinova 1997:112).  

![Image](image-url)

(14) VZE gi Olga parite. took3.SG DOC3.PL Olga money+DEF Olga took the money.
(15) VZE gi parite Olga. took3.SG DOC3.PL money+DEF Olga
Olga took the money.

Interestingly, both (14) and (15) are focus-tail sentences in Avgustinova's (1997) terminology, i.e. they contain a clause initial focused verb followed by two unfocused constituents. This issue is addressed in more detail in section 4.
Furthermore, both the case marker and the word order marker hypothesis cannot explain why the DOC is optional. The word order marker hypothesis also fails to explain why it can occur in both the unmarked and the marked word order.

The definiteness-marker hypothesis, (H3), as proposed in Cyxun (1962:289-290), Minčeva (1969:3), Ivančev (1957:139), Georgieva (1974:75),21 has been shown to be wrong by Ivančev (1968:164) and Kazazis & Pentheradoukis (1976:399-400), since indefinite specific NPs can be doubled (cf. Leafgren 1997a:122), as shown in (16). Edno is an instance of the Bulgarian indefinite, specific article.22

\[
(16) \text{Edno dete go vidjaj da pluva.}
\]
\[
\text{a-certain child DOC3.SG.NEUT saw1.SG SBJ swim3.SG I saw a (certain) child swimming.}
\]

Also, Vakareliyska (2000:3) points out that just marking definiteness would be a redundant function, since Bulgarian (like other languages of Balkan Sprachbund with CD) already has definiteness marking suffixes to do that.

Avgustinova (1997:92-95) is a recent proponent of the specificity-marker hypothesis, (H4).23 She distinguishes between [+/-limited] nominal material and further divides [+limited] nominal material into [+/-specific] and [-limited] nominal material into [+/-generic]. Referring to Shamray (1989:51), Avgustinova (1997:88) describes the division as follows. The feature value [+limited] indicates that "the identifying, i.e. particularising aspect of the respective noun has been activated", whereas [-limited] "would mean that only the categorizing, i.e. the generalising, aspect of the noun is relevant for the communication" (Avgustinova 1997:88-9; emphasis mine. She uses [lim +/-]). Figure 1 illustrates the approach to a semantic classification of nominal material taken by Avgustinova (1997).

---

21 Also see Popov & Popova (1975:48) and Popov (1973:173), who, probably aiming at specificity, require the doubled NP to be 'articulated' (cf. Leafgren 1997a:121).

22 The specific, indefinite article has the following paradigm: Singular: masc. edin, fem. edna, neut. edno 'a certain, a particular'; Plural: edni 'certain' (cf. Vakareliyska 1994:122). More precisely, this article requires an NP not marked by the definiteness suffix. For a formal description of the semantics of edin, see Izvorski (1994) who, among other things, shows that, in her terminology, edin is not always [+specific].

23 See also Kazazis & Pentheradoukis (1976) and Vakareliyska (1994:122).
To clarify Avgustinova's terminology, I provide some examples below, which are all taken from Avgustinova (1997:92-5). In the English translation of each example, the noun which is under consideration for the given features is marked by [ ].

<table>
<thead>
<tr>
<th>Features</th>
<th>Examples</th>
</tr>
</thead>
</table>
| [+limited]       | **Vidjate li knigite/dvete knigi/njakolkoto knigi.** Did you see [the books]/[the two books]/[(the) several books]?
|                  | **Tezi knigi mi trijabvat.** I need these [books].                                         |
|                  | **Trjabvat mi edni dokumenti.** I need [some (specific) papers].                             |
|                  | **Po livadata tiča bosno edno dete.** [A child] is running barefoot through the meadow.      |
| [-specific]      | **Molja dajte mi pet knigi.** Please give me [five books].                                  |
| [+generic]       | **Xorata sa smâți.** [Humans] are mortal.                                                   |
| [-limited]       | **Toj običa winoto i ženite.** He loves [wine and women].                                   |
|                  | **Edna majka vinagi šte poznane deteto si.** [A mother] would always recognize her child.   |
| [-generic]       | **Tuk kupuvam knigi.** I buy [books] here.                                                  |
|                  | **Vleze grupa studenti.** [A group of students] came in.                                    |

Table 1 – Examples of Avgustinova’s (1997:92-5) classification of nominal material

According to Avgustinova, only [+limited, +specific] objects can be doubled. This claim can be motivated by contrast between (16), here repeated as (17), and (19). In (19), the
fronted [-specific] object cannot be doubled although the corresponding sentence with neutral word order and without CD, i.e. (19), is grammatical.

(17) \textit{Edno dete go vidjaj da pluva.}\hspace{1cm} \textit{I saw a (certain) child swimming.}

(18) \textit{*Njakoja po-nova kola iskam da ja kupja.}\hspace{1cm} \textit{Intended: I want to buy (for myself) some newer car.}

(19) \textit{Iskam da si kupja njakoja po-nova kola.}\hspace{1cm} \textit{I want to buy (for myself) some newer car.}

This point is further supported by the exceptions to the generalization that \textit{edni} is [+specific] (see footnote 22, p. 14). In (20) \textit{edni po-iziskani drexi} is [-specific] (Avgustinova 1997:95) and the fronted object cannot be doubled.

(20) \textit{*Edni po-iziskani drexi gi dadoxa na Ivan.}\hspace{1cm} \textit{Intended: Some stylish clothes, they gave (them) to Ivan.}

However, (H4) has proven to be insufficient since generics can and in some cases even must be doubled, as illustrated in (21). Independently of my observations, Alexandrova (1997) and Guentchéva (1994), too, point out that generics and interrogatives can be doubled (for the doubling of interrogatives, see also Jaeger 2002).

(21) \textit{Slonovete *(gi) obučavat xorata.}\hspace{1cm} \textit{The elephants, (the) people train.}

So far I have shown that [-limited, +generic], e.g. (21), and [+limited, +specific] object NPs, e.g. (16), can be doubled while [+limited, -specific] object NPs cannot be doubled, as shown in (17) and (20). This raises the question if [-limited, -generic] object NPs can also be doubled. As for the examples above, I use the object fronting construction to test this.\(^{25}\) The

\(^{24}\) (21) is grammatical without the DOC if \textit{slonovete} is realized with emphatic stress and thus receives the exclusive focus. This is what we would expect since this is a case of FOCUS-fronting (see section 4). In this paper, we are only interested in non-focus object fronting, i.e. object fronting without emphatic stress on the object. Therefore, whenever we star an example with a fronted object that is not given in small caps, we always mean that this example is ungrammatical for fronted non-focused objects.

\(^{25}\) This will become clearer in section 4. In short, a fronted object without focus intonation must be doubled by the corresponding object clitic if this is possible at all. If doubling is not possible (like for e.g. [+limited, -specific] object NPs) the resulting clause is ungrammatical.
examples (22) and (23) are taken from Avgustinova (1997:92). The corresponding CD examples, (24) and (25), are ungrammatical.

(22) Tuk kupuvam knigi.
here buy1.SG books-DEF
I buy books here.

(23) Târsja prijateli.
look-for1.SG friends-DEF
I am looking for friends.

(24) *Knigi tuk gi kupuvam.
books-DEF here DOC3.PL buy1.SG

(25) *Prijateli gi târsja.
friends-DEF DOC3.PL look-for1.SG
Intended: Friends, I am looking for.

To sum up, I have shown that [-limited, +generic] and [+limited, +specific] objects can be doubled, whereas [-limited, -generic] and [+limited, -specific] objects cannot be doubled. In the following, I adopt a slightly different but (at least) equally common classificatory system where nominal material is [+/-generic], and the [-generic] NPs are further divided into [+/-specific]. This is illustrated in Figure 2.

![Figure 2 - Alternative semantic model of nominal material](image)

In this model, the above-mentioned generalization about Bulgarian CD is captured as follows: [-generic, -specific] NPs cannot be doubled. Note that it is typologically common that [+specific] and [+generic] NPs pattern together (Shiao-Wei Tham, p.c.), as observed in the data above. The observations presented here, like those in Alexandrova (1997) and Guentchéva (1994), argue against Avgustinova’s (1997) claim that only [+limited, +specific] NPs can be doubled. The data also rejects Rudin’s (1997) analysis, incorporated, a.o., by Schick & Beukema (2001), that the DOC only doubles (topical) [+specific] NPs.

The model illustrated by Figure 2 seems to be more suitable for the current purpose since the possibility of CD is not affected by the difference between [+limited, -specific] and
[-limited, -generic]. Furthermore, it is hard to understand which difference Avgustinova (1997) has in mind when she distinguishes between those two feature pairs. In the model illustrated by Figure 2, [+limited, -specific] and [-limited, -generic] are grouped together as [-generic, -specific].

Now consider the topic-marker hypothesis, (H5), as formulated in Leafgren (1997a,b, 1998), Avgustinova & Andreeva (1999), and to some extend Ivančev (1974), Georgieva (1974), Minčeva (1969), Popov (1963:167) and the AG (1983,3:188). According to this hypothesis, the above-mentioned restriction on the doubled object is an indirect effect of the requirement that the doubled object has to be topical. Leafgren (1997a:136ff.) further shows that topicality marking in Bulgarian cannot be reduced to agentivity or subjecthood, two scales that correlate with the scale of topicality in many languages (for a discussion of those hierarchies, cf. Givon 1976). However, Leafgren (1997a,b, 1998) does not show how his proposal (i.e. (H5) as stated above) accounts for the contrast between (16) and (17) or the ungrammaticality of (20), (24), and (25). In fact, (H5) turns out to be to drastic in its formulation. Consider examples (26) and (27). Njakolko is a [-definite; -generic, +specific] quantifier, malko a [-definite; -generic, -specific] quantifier. Njakolko, unlike malko, is compatible with and sometimes even requires CD. However, there is no apparent reason why njakolko spisanija in (26) should be a topic and malko spisanija in (27) not. Thus it seems hard to explain the difference between (26) and (27) by (H5).²⁹

²⁶ Thanks to Shiao-Wei Tham for discussing different classificatory systems for the semantics of nominal material with me. Remaining mistakes are, of course, due to the me.
²⁷ The morphological feature [definiteness] is separated from the semantic features by a ";".
²⁸ I am thankful to Ruth Kempson for pointing me to this data and helping me to gather it. I am also very grateful for the patience of Iskra Iskrova who explained and discussed (26) - (28) (and other material) with me in detail.
²⁹ Interestingly, one of my informants pointed out that for her (27) is only grammatical if either only koito 'which' or only the DOC gi 'them' is realized. This relates to the third use of the DOC as an intrusive pronoun in extractions (cf. Sells 1984), which we cannot discuss here due to lack of space. For V.G. and another informant, (27), as given above, is grammatical.
There are two ways out of this problem. One is to take typological evidence to show that there are universal restrictions on the semantics of topics (e.g. that topics have to be "referring expressions", Lambrecht 1994:155-56).\(^{30}\) This approach will result in a notion of topic that will be qualitatively quite different from that of Leafgren (1997a:127) who defines the topic to be 'what the clause is about'. Second, one could claim that CD in Bulgarian has more than one constraint on the semantics and information structural role of the doubled object, as described by (G1) and (G2).

### Two general constraints on Clitic Doubling in Bulgarian

<table>
<thead>
<tr>
<th>Constraint</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(G1)</td>
<td>Doubled objects have to be topical</td>
</tr>
<tr>
<td>(G2)</td>
<td>Doubled objects cannot be [-generic, -specific]</td>
</tr>
</tbody>
</table>

As the comparison between (26) and (28) shows, this is necessary anyway to explain why CD is obligatory in some cases and optional in others. In other words, (G1) is what is marked by the DOC, and (G2) is a constraints on what can be marked by the DOC.

(28) Njakolko spisanija mnogo xora *(gi) xaresvat.  
a-few SPEC journals DEF lots people DOC PL like PL  
*A few (certain) journals, a lot of people like (them)*

Here I am mainly interested in the differences between cases of obligatory and optional CD and therefore do not need to commit myself to either of the two ways sketched above. The account presented in section 4 is compatible with additional constraints on the semantics (e.g. specificity).

I am aware that the inherently vague and widely varying definition of topic is problematic for (H5). Nevertheless, I take this hypothesis as the starting point for a formalization of the properties of the DOC in the CD construction, which I introduce in the next section. In other words, I adopt an approach similar to that in Lambrecht (1994): topics cannot be [-generic, -specific]. I leave the details open to future research. Finally, note that none of the above-mentioned approaches captures the fact that the DOC can also be the default pronoun. It is exactly the interaction between this use and its use as a topical object agreement marker (in CD) that provides interesting evidence for my analysis. I come back to this issue in section 6.

\(^{30}\) See also Givón (1992:308-309) who claims that contrastive topics can be [+referring, +definite], or [−referring, −definite] but not [+referring, −definite].
4 DF-configurationality and the DOC in clitic doubling

This section presents an analysis of the DOC in the clitic doubling construction (CD). The previous section already mentioned that there are certain constraints on the specificity of the doubled object in CD. The current section focuses on the information structural impact of the DOC in CD.

There is good evidence from the extensive literature on the left periphery of the Bulgarian clause that Bulgarian is D[iscourse] F[unction]-configurational (see Dimitrova-Vulchanova & Hellan 1998, Rudin 1994, 1990/1991, 1985, Arnaudova 2001, Lambova 2002, Dyer 1992, 1993, Leafgren 1997c, a.o). By using the term DF-configurational (Kiss 1995, 2001), I refer to the fact that Bulgarian phrase structure provides certain (grammaticalized) constituent-structure positions, which are strictly associated with specific discourse functions, such as TOPIC or FOCUS. As mention above, this section is only concerned with the DF-configurationality of the left periphery of the Bulgarian clause. In the following, I will describe the evidence for three phrase structural positions in Bulgarian, which mark a constituent to be a DF. I limit myself to the three most frequently described DFs in the literature, namely TOPIC, FOCUS, and EXTERNAL-TOPIC (e.g. King 1995 for Russian, Aissen 1992 for Mayan, Kiss 1994 for Hungarian). I do not discuss their precise function in much detail since this has already done by others (see the references below).

4.1 Phrase structure positions of Discourse Functions

Bulgarian allows EXTERNAL-TOPICS (Aissen 1992, going back to Cinque's 1977 hanging topics; see also Kiss 1994:80 for Hungarian, King 1995 for Russian), which are accounted for by the annotated phrase structure rule in (R5).

(R5) \[ \text{EP} \rightarrow \{(\text{NP}, \text{PP}, \text{AP}, \text{SBJP})\} \quad \text{CP} \quad \text{\textup{(\uparrow E-TOPIC) = \downarrow \quad \uparrow = \downarrow}} \]

31 The term discourse function (DF) is used as defined in Bresnan (2001:98), where DFs are taken to be a subset of the grammatical (non-argument) functions, namely the set of grammatical functions which are "most salient in discourse" (ibid). Since they are required to be grammatical (i.e. grammaticalized), DFs are not identical with but, at best, a subset of the discourse function that Schiffrin (1987) and Fraser (1988) talk about.

32 The abbreviation SBJP is used to refer to a subjunctive phrase.
EXTERNAL TOPICS cannot occur in subordinate clauses, a property that distinguishes them from TOPIC- or FOCUS-fronting. EXTERNAL-TOPICS are left-dislocated and clause external, i.e. they are prosodically independent from the sentence and followed by an intonational break, unlike TOPIC- or FOCUS-fronting discussed further below. This is illustrated by (29) where the DOC cannot be realized after Slavejkov because the DOC is an enclitic proper (cf. section 2.3) which needs a phonological host to its left. The intonational break is not a valid host. (30) shows that there is nothing else that prevents the sentence from being grammatical. As soon as the DOC has a valid host, the sentence is grammatical.33

(29) *(Što se otnasja do) Slavejkov # go četat vsički nego. as-for Slavejkov DOC3.SG.MASC. readPL all him3.SG.MASC.ACC
Intended: (As for) Slavejkov/As far as Slavejkov is concerned, everybody reads him.

(30) (Što se otnasja do) Slavejkov # nego go četat vsički. as-for Slavejkov him3.SG.MASC.ACC DOC3.SG.MASC. readPL all
(As for) Slavejkov/As far as Slavejkov is concerned, everybody reads him.

Also, there is extensive evidence for fronted TOPICS in a position preceding the complementizer (in principal an arbitrary number of TOPICS can be fronted; cf. Rudin 1994, 1990/1991, 1985:24-25). (31) is an example of a fronted TOPIC, which is accounted for by the proposed phrase structure rule (R6). Unlike EXTERNAL TOPICS, TOPICS are not followed by an intonational break. Also note that the TOPIC-fronted object in (31) is doubled by a DOC (ja). I use [...]\_F to mark the constituent(s) in the TOPIC position and [...]\_F for constituent(s) in the FOCUS position.

(31) Toj kaza [Marija]_Tče šte ja vidi. He said Marija that will DOC3.SG.FEM see

(R6) \[
\begin{align*}
\text{CP} & \rightarrow [\text{NP, PP}]^{*} \quad C' \\
& \downarrow \epsilon(\uparrow \text{TOPIC}) \quad \uparrow=\downarrow \\
\end{align*}
\]

Finally, Bulgarian has a FOCUS-position following the TOPIC-position. In subordinate clauses the FOCUS – unlike the fronted TOPIC(s) – follows the complementizer. Consider (32), an example of FOCUS-fronting in an embedded clause. The FOCUS-fronted object Marija follows the complementizer če and receives emphatic accent (indicated by small caps). I

33 The symbol # is used to indicate an intonational break.
propose the two phrase structure rules presented in (R7) and (R8) to capture the phenomenon of FOCUS-fronting in Bulgarian. 34

(32) Toj kaza če [MARIJA], šte vidi.
    He said that Marija will see
    He said that he will meet MARIJA.

(R7)  C' → C    IP
      ↑=↓   ↑=↓

(R8)  IP → (NP, PP, AP, SBJP)* I'
      ↓∈(↑FOCUS)  ↑=↓

The TOPIC and the FOCUS position have different functions in terms of their impact on information packaging (Vallduvi 1992). Briefly, the TOPIC-fronted constituents' referents are "what the sentence is about" whereas the FOCUS-fronted constituents' referents present new information provided by the sentence. More discussion of topicality and focality is provided later in this paper. Application of the phrase structure rules (R6), (R7), and (R8) to the examples in (31) and (32) results in the following partial c-structures.

Note that the phrase structure rules for TOPIC-fronting and FOCUS-fronting in (R6) and (R7) capture the fact that not all phrases can be fronted to both of those positions. Subjunctive phrases, for example, can only be fronted if they get assigned emphatic accent as the contrast between (33) and (34) shows.

(33) [DA BJAGA]F običa.
    TO RUN he likes

34 I shortly delay the discussion of the fact that the TOPIC-fronting rule in (R6) allows only for a subset of the syntactic categories that the FOCUS-fronting allows for. I will come back to this issue below.
Below, the absence of emphatic accent as well as the interpretation of the sentence in terms of information packaging is taken as evidence for TOPIC-fronting. PPs and NPs can be TOPIC- and FOCUS-fronted. Examples for object-NP fronting to the TOPIC or the FOCUS positions were given above. Example for PP-fronting into either the TOPIC- or the FOCUS-position is given below.

(35) [V NEDELJA]F Ivan šte dade knigata na Peter.
in Sunday Ivan will give book +DEF to Peter
On Sunday John will give the book to Peter.
in Sunday Ivan will give book +DEF to Peter
On Sunday John will give the book to Peter.
(37) [V GRADINATA]F izpekoxme prase (NE v parka)
in garden +DEF grilled1.PL pig +DEF not in park +DEF
In the GARDEN we grilled a pig (not in the PARK).
(38) [V gradinata]F [PRASE]F izpekoxme.
in garden +DEF pig +DEF grilled1.PL
In the garden, we grilled a PIG.

Whereas FOCUS-fronting of APs is clearly possible, TOPIC-fronting of APs does not seem to be possible. This fact is what is expected since most adverbs do not make topics (recall that topicality is defined in terms of aboutness here).

(39) [DOBRE]F rokljata e ušita.
well dress +DEF is sewn
The dress is tailored well.
(40) ?? [Dobre]T rokljata e ušita.
well dress +DEF is sewn
Intended: The dress is tailored well.

The rules (R6) and (R8) capture the category restrictions on TOPIC- and FOCUS-fronting and are in line with Rudin (1985:24-25).
4.2 Obligatory doubling of TOPIC-fronted direct objects

The preliminary results of an online experiment\(^{35}\) designed together with Veronica Gerassimova suggest that TOPIC-fronted objects (i.e. not the hanging EXTERNAL TOPICS) are always doubled. Note that this still allows for FOCUS-fronted objects. The results are supported by the observations in Dimitrova-Vulchanova & Hellan (1998:xviii), and implicitly Avgustinova (1997:112). In order to capture this fact and Leafgren's (1997a,b, 1998) claim that CD always marks topicality of the doubled object, I propose that the syntactic topic position is assigned the outside-in functional uncertainty equation in (R9), which updates (R6). The squared parentheses are a convention used to express that the bracketed part of the equation is not defining (Dalrymple, p.c.).\(^{36}\) The data presented here, therefore are interesting because they illustrate the empirical consequences of this representational subtlety.

\[(R9)\quad \text{CP} \rightarrow \begin{cases} (\text{NP, PP})^* & \text{C'} \\ \downarrow \in (\uparrow \text{TOPIC}) & \uparrow = \downarrow \\ (\uparrow \text{XP}^* \text{[GF]}) \downarrow V \downarrow \in (\uparrow \text{XP}^* \text{ADJ}) \end{cases} \]

Note that ADJ is not described as constraining but as defining the adjunct. (R9), therefore, makes the correct predictions since fronted adjuncts do not have to be doubled, see (36) and (38) above. Actually, they cannot be doubled because no lexical item similar to the DOC exists for the doubling of adjuncts. (R9) also predicts that a TOPIC-fronted subject has to be "doubled" by something. This is achieved by assuming that the subject-verb agreement morphology on the verb, in LFG terms, has an optional PRED 'PRO'. This is a strong claim because it predicts that the optional PRED 'PRO' will be realized if nothing else prevents it. In other words, without a TOPIC-fronted subject, the PRED 'PRO' may in principle be realized. There are two cases to consider. First the subject NP (i.e. an element that is both identified as bearing the SUBJ function and has a PRED value) may be somewhere else in the sentence. In this case, UNIQUENESS (as defined in footnote 10, p. 7) prevents the PRED 'PRO' from being realized on the verb because one f-structure can only have one PRED value. Second, there may be no subject NP in the sentence. In this case, nothing prevents the PRED 'PRO' introduced by the verbal agreement morphology to be realized. This predicts subject

---

\(^{35}\) Human Subjects Application #0102-655, approved by the Human Subjects Panel, Stanford. The experiment can be found at [http://symmsys.stanford.edu/experiment/](http://symmsys.stanford.edu/experiment/). In this experiment, subjects were asked to judge Bulgarian sentences after being primed for colloquial spoken language. All judgments were elicited using magnitude estimation, i.e. subjects were asked to assign a gradual value for the "goodness" of each sentence with respect to an always present reference sentence.
drop, or, more precisely, subject pro-drop. Indeed, Bulgarian has subject pro-drop as the following examples taken from Avgustinova (1997:48) show. Thus (R9) makes the correct predictions.

(41) Bxaj _ ja_ vidal.
    AUXIMPERFECT.1.SG DOC3.SG.FEM seenSG.MASC
    I had seen her.

(42) Ne _ šte li_ ni_ kažeš?
    NEG will Q IOC 1.PL tell2.SG
    Won't you tell us?

The DOC is identified as the direct object by its lexical semantics and the phrase structure rule for the predicate clitic cluster (see phrase structure rule (R3), on p. 10). The agreement between the DOC and the doubled object guarantees that no spurious ambiguities are predicted, even in the case of multiple object fronting. Below I give a representative lexical entry for ja, the 3.SG.FEM form of the DOC.

ja: NCL - DOC
   (↑OBJ) = ↓
   ↓PERS = 3
   ↓NUM = SG
   ↓GEN = FEM

**Figure 3 - Simplified lexical entry for the DOC ja (preliminary version)**

The proposal predicts that fronted objects must be doubled since the DOC is the only way to define the object function without violating UNIQUENESS (ignoring UNIQUENESS, one could wrongly generate a second object in the VP to satisfy EXTENDED COHERENCE and COMPLETENESS). As an example, consider the sentence in (43) with subject pro-drop and a fronted topical object. The corresponding c- and f-structure are given below. Due to ECONOMY OF EXPRESSION (Bresnan 2001:91), which is repeated here as (G3), the IP between the CP and the VP is "pruned away".

36 Note that (↑XP* [GF])=↓ = (↑[XP* GF])=↓.
37 EXTENDED COHERENCE requires every syntactic function (e.g. grammatical functions, such as SUBJ, OBJ, etc., and discourse functions, such as TOPIC, FOCUS) to be "integrated appropriately into the f-structure" (Bresnan 2001:63). Argument functions (SUBJ, OBJ, etc.) are integrated if they are licensed by a PRED (usually of a verb), discourse functions (TOPIC, FOCUS, etc.) are integrated if they are functionally or anaphorically bound by an integrated function, and adjuncts are integrated if their immediate f-structure contains a PRED (ibid).
38 COMPLETENESS requires "that every [syntactic] function designated by a PRED be present in the f-structure of that PRED" (Bresnan 2001:63, [ ] added by me).
ECONOMY OF EXPRESSION (Bresnan 2001:91)

(G3) All syntactic phrase structure nodes are optional and are not used unless required by independent principles (Completeness, Extended Coherence, Semantic Expressivity).

I leave it to the readers to convince themselves that the f-structure is the only one predicted.39

(43) Knigata ja e kupil.
bookPL.DEF DOC3.SG.FEM AUX3.SG bought
The books, he has bought.

4.3 Optional vs. obligatory Clitic Doubling

Crucially, my proposal captures the intuition that it is the absence or presence of a fronted topical object that causes obligatory CD. By contrast, Leafgren argues that the following two generalizations hold (the second point is also supported by Vakareliyska 1994:125):

Generalizations about CD in Bulgarian stated by Leafgren (1997a,b)

(G4) All doubled objects are topics.
(G5) Object doubling is always optional.

In other words, CD is just one way of identifying an object as topical.40 Unfortunately, Leafgren (1997a,b,c, 1998, 2001, 2002) does not formalize his working definition of 'topic'

39 Recall that the subject function is defined through the verbal subject agreement morphology (including an optional PRED PRO since subject drop is common in Bulgarian), so that subjects, too, can be in the fronted position.
any more precisely than 'What a clause provides or requests information about' (Leafgren 1997a:127, referring to Sgall 1975:303; see also Sgall 1993). Leafgren gives the following example to illustrate his topic definition. He claims that \textit{tezi ne\v{s}ta} is the topic of the clause.\footnote{Note one important detail in Leafgren's definition. The topic is defined on the level of a clause, not a sentence. This allows for topics in, for example, subordinate clauses. Examples like (31) above clearly show that this is necessary.}

\begin{equation}
\text{(44) } \text{Vanja, ne ja, válnuvat [tezi ne\v{s}ta] ...}
\text{Vanja NEG DOC3.SG.FEM worry3.PL these things}
\end{equation}

These things don't worry Vanja ...

The generalization in (G5) conflicts with Dimitrova-Vulchanova & Hellan's (1998) and my own observations. Leafgren's work is based on a corpus study of more than 7,000 object occurrences in written texts (1997a,b; including \textasciitilde 200 cases of CD), more than 3,000 object occurrences in spoken texts (1998: including \textasciitilde 200 cases of CD), and a comparative study of 1,200 object occurrences each in informal oral, formal oral, and formal written texts (Leafgren 2001). In light of such extensive evidence, it is important to resolve the mismatch between Leafgren's and my observations.

There are two main sources for this mismatch aside from the apparent problem with topic definitions. First, although Leafgren (2001) considers informal oral texts, (G5) is based on Leafgren's (1997a,b) work on a written corpus (consisting of 2 novels and 2 short stories). In contrast, the online experiment done by Jaeger & Gerassimova (2002) aims at judgments about informal contemporary spoken Bulgarian. Secondly, and more importantly, Leafgren does not control for fronted \textit{focused} phrases. In fact, Leafgren (1997a:132) explicitly allows for topics to be "focused" (in his terminology). Here I argue that Bulgarian speakers have two ways involving the DOC to handle a topical object.\footnote{Alternative means of topical object marking depend on the register. In informal, spoken Bulgarian, speakers may also use marked word order (i.e. object-topic fronting) or intonation or just not mark the topicality of the object when the context unambiguously identifies the object to be the topic (cf. Leafgren 1997b:128). In more formal registers, passivization or impersonal reflexive constructions can be used to indicate that the semantic object is topical (cf. Leafgren 2001).} First, there is a grammaticalized \textit{TOPIC} position, which requires \textit{obligatory} CD in the case that the fronted constituent is an object. Second, topical object may \textit{optionally} be doubled even if they are not fronted.

The crucial point, which Leafgren misses, is that Bulgarian has two sentence initial positions, here labeled \textit{TOPIC} and \textit{FOCUS} (see above). These two positions are independently...
motivated because of the different stress contours that go along with them. One of those two positions, namely TOPIC, requires CD if it is filled by an object. The TOPIC position phrase structure rule (R9) captures obligatory CD of TOPIC-fronted objects without any further stipulations. Leafgren does not distinguish between TOPIC- and FOCUS-fronting and, therefore, mistakenly believes that CD of fronted objects is optional because indeed FOCUS-fronting does not require CD.

The following paragraphs outline a model of the interfaces to I[information] S[tructure], which are relevant for the current discussion. More specifically the model proposed below will be used to explain the distribution of optional vs. obligatory realization of the DOC in CD.

Similarly to other authors (Choi 1999:12, King 1997:8), I assume an IS-component which has interfaces not only to f-structure but also to the Prosodic Structure (PS) and the Lexical Structure (LS), see Figure 4. Note that one reason to postulate an IS-component that is not totally determined by syntax/FS is the fact that syntactic constituency does not map one-to-one on informational structural constituency (cf. King 1997, p.c.).

I am not interested in the interface between PS and IS but in the interface between LS and IS. Interfaces that are not directly relevant to IS are not shown in Figure 4 (e.g. the

---

43 More precisely, there is a motivation for a formal distinction, which I choose to capture in terms of c-structure position.
44 King (1997) proposes an IS which interfaces only c-structure and s-structure. However, the problems with the representation of DFs in FS that lead her to this conclusion (ibid: 4-7) can also be handled by distinguishing grammaticalized DFs, which are handled in FS, and non-grammaticalized IS-role assignment, which is handled in IS (see below).
interface between PS and CS). The dotted lines in Figure 4 stand for indirect relations between the different parts of the grammar interfacing with IS. Figure 4 is by no means intended to give a complete picture of either those relations or of the overall interfacing architecture of a grammar. Next, I show that the model presented in Figure 4 provides the ground for an interesting account of the Bulgarian data.

First, I define that TOPIC and FOCUS, which are parts of f-structures, be mapped to at least their corresponding IS-roles, namely topic and focus. Somewhat more formally, this constraint can be captured by (G6) and (G7).

\[ (G6) \quad X \in \text{TOPIC} \Rightarrow x \in \text{topic}, \text{where } X \text{ is the f-structure correspondence of a linguistic form } w, \text{ and } x \text{ is the denotation of } w. \]

\[ (G7) \quad X \in \text{FOCUS} \Rightarrow x \in \text{focus}, \text{where } X \text{ is the f-structure correspondence of a linguistic form } w, \text{ and } x \text{ is the denotation of } w. \]

In case one wants to define other DFs than TOPIC and FOCUS, (G8) describes the same generalization for an arbitrary DF, where, for a given input X, the function ISR(X) maps the f-structure feature X to the IS-role corresponding to the DF, which X is an element of.

\[ (G8) \quad X \in \text{DF} \Rightarrow x \in \text{ISR}(X), \text{where } X \text{ is the f-structure corresponding to a linguistic form } w, \text{ and } x \text{ is the denotation of } w. \]

Similarly to EXTENDED COHERENCE (cf. Bresnan 2000), I formulate a constraint INFORMATION PACKAGING COHERENCE (IPC) that guarantees that the generalization in (G6) holds for all DFs of an f-structure.

\[ \text{INFORMATION PACKAGING COHERENCE (IPC)} \]

\[ (G9) \quad \text{An f-structure } F \text{ fulfills IPC iff every discourse function DF in } F \text{ fulfills (G8).} \]

45 In a model like the one presented here, encoding of IS through CS (and therefore within LFG through F-structure, FS) corresponds to what is commonly called DF-configurationality.

46 Again, here I am not concerned with details of the function of those two IS-roles. Note also that what follows is, in principle, compatible with a theory of IS that claims more IS-roles than just topic and focus (e.g. external topic or contrastive topic).

47 Throughout the paper, SMALL CAPS are used for D[iscourse] F[unction]s, which are part of the f-structure, and lowercase for IS-roles. Most of the time, IS-roles will be indicated by the prefix IS-, such as in IS-topic or IS-focus.
Again, note that (G8) states only a one-way implication. This one-way implication works to the advantage of the account presented here. While every phrase that is a fronted TOPIC must be an IS-topic, non-fronted topics are not required to show up as an f-structure TOPIC. Thus with (G8) in combination with Figure 4, it is possible to distinguish between grammaticalized DF-positions and other non-grammaticalized information structural effects of, for example, prosodic and lexical structure. The one-way implication formally described in (G8) is illustrated in Figure 5, which shows an FS-to-IS mapping of an f-structure containing a TOPIC and a FOCUS element. The IS-component's representation is a set of IS-roles for any given clause. The elements in the IS-role sets are semantic objects of any type.  

Note that the arrows are only pointing from FS to IS, not vice versa.

\[ \text{Figure 5 - FS-to-IS mapping of DFs} \]

Now with the proposed model of IS, LS, and FS interaction in mind, we can restate Leafgren's generalizations in (G4) and (G5) more precisely as (G10) - (G12).

**Final revised generalizations about CD in Bulgarian**

<table>
<thead>
<tr>
<th>Generalization</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(G10)</td>
<td>All doubled objects are IS-topics.</td>
</tr>
<tr>
<td>(G11)</td>
<td>TOPIC objects must be doubled.</td>
</tr>
<tr>
<td>(G12)</td>
<td>IS-topic objects can be doubled.</td>
</tr>
</tbody>
</table>

In order to predict that CD implies that the doubled object is part of the IS-topic, i.e. to guarantee (G10), we have to slightly modify the lexical entry of the DOC(s). Again, the 3.SG.FEM form *ja* is given as a representative example in Figure 6. The upwards-pointing

---

48 For an argument why – at least – IS-topics should be represented that way, see Jaeger & Oshima (2002).
49 Here I do not address the pragmatic factors that determine in which contexts speakers tend to make use of this mechanism (CD to mark IS-topicality of the object). See Givón (1987) for a general discussion of this.
arrow with the subscript 'IS' indicates that the referent identified by the DOC is mapped onto information structure (where it is identified as a part of the IS-topic).\(^{50}\)

\[
\text{ja: } \text{NCL - DOC} \\
(\uparrow \text{OBJ}) = \downarrow \\
(\downarrow \text{PERS}) = 3 \\
(\downarrow \text{NUM}) = \text{SG} \\
(\downarrow \text{GEN}) = \text{FEM} \\
\uparrow \iota \in \text{topic}_{\text{IS}}
\]

\textbf{Figure 6} - Revised lexical entry for the topic-marking DOC \textit{ja}.

The lexical entry in Figure 6 together with the revised annotated phrase structure rule for the \textsc{topic} position in (R9) captures all of the above-mentioned generalizations, (G10) - (G12), and therefore resolves the apparent conflict between Leafgren's work and e.g. Dimitrova-Vulchanova & Hellan's claims. That is, the current account predicts \textit{optional} CD for FOCUS-fronted objects as long as they are part of the IS-topic. If one adopts a two-dimensional IS-component\(^{51}\) (e.g. Choi 1999; Leafgren 1997a,b, Dalrymple 2001) this is not surprising at all. Indeed, reduplication of FOCUS-fronted objects is observed in Bulgarian. First, CD of fronted object \textit{wh}-phrases, which are usually considered to sit in the FOCUS position in Bulgarian (Rudin et al. 1998), is possible (Jaeger 2002; Dimitrova-Vulchanova & Hellan 1998:xxi-xxii).

\begin{equation}
\text{(45)} \quad \begin{array}{ll}
\text{Kogo} & \text{what} \\
\text{kakvo} & \text{doc}_{3,\text{SG,MASC}} \\
\text{go} & \text{suprised}_{3,\text{SG}} \\
\text{iznenada?} & \text{suprised}_{3,\text{SG}} \\
\text{Whom did what surprise?}
\end{array}
\end{equation}

Second, it seems that of all non-\textit{wh}, fronted and emphasized constituents, only contrastive topics/focus can be doubled. Although this issue needs further research, my proposal as it stands predicts this. Avgustinova (1997) cannot straightforwardly account for since she employs the one dimensional IS-component proposed in Vallduví (1992, 1993).

As mentioned in the introduction, one aim of this paper is to provide a formal account for CD and its interaction with word order and IS. The current section has done exactly this. Second, I wanted to resolve the discrepancy between the different empirical approaches to Bulgarian CD and the theoretical literature. In part, this has already been achieved by distinguishing between two independently motivated phrase structural positions and their

\(^{50}\) The notation $\uparrow_i$ for mapping from CS to IS is taken from King (1997).
correspondences in the IS-component. The analysis is able to capture both the generalization from the extensive empirical work and predicts the right restrictions on CD resulting from certain word orders (i.e. obligatory CD of TOPIC objects). Next, I use the second source of data for spoken Bulgarian mentioned above, Avgustinova's (1997) elicited question-answer pairs, to briefly test if the presented proposal makes correct predictions about possible word orders beyond the fronted TOPIC construction.

4.4 Functionally determined default word order

It is beyond the scope of this paper to provide a detailed analysis for all of the patterns (i.e. word order-intonation-information structure mappings) described by Avgustinova (1997:112). Although this issue is open for further research, I suggest that Bulgarian has some kind of 'default ordering' within the flat VP (see (R4) in section 2). Among other features, such as definiteness, person, referentiality, etc., topicality of a phrase seems to be one – maybe the major – determining factor for the constituent order in the VP. Leafgren (1997c:5f) shows that topic-before-comment is the more important ordering mechanism in Bulgarian than subject-before-object or agent-before-patient, both in terms of frequency, and in that all violations of the two other conditions serve to satisfy the topic-before-comment condition or another discourse or information structure constraint (e.g. CD and object fronting). The assumption of a default order that is similar to the one suggested by the Prague school (cf. Functional Sentence Perspective, henceforth FSP; Sgall 1993) but applied to the flat VP instead of the whole clause explains why a certain default constituent order can be observed in Bulgarian while, at the same time, only a few strict rules (like the above-mentioned object TOPIC-fronting) seem to hold. I ask the reader to keep in mind the notion of default ordering as just described during the discussion of Avgustinova's (1997) data.

Apart from direct object fronting, which results in OSV and OVS orders (for the sake of simplicity, I only consider transitive verbs here), there is one other word order that has a strong preference for CD, namely VOS. According to Avgustinova (1997), VOS is possible

51 By two dimensional, I mean that there is not only one dimension along which information structural roles differ, e.g. topic-comment or link-tail-focus (cf. Vallduví 1993, 1992). Instead information structural roles differ along two dimensions, e.g. they can be [+/- prominent] and [+/- given] (cf. Choi 1999).

52 Note that this is not at all uncommon. It has long been known that scrambling in languages like e.g. German or Japanese is sensitive to the above-mentioned categories. Furthermore, especially topicality of phrases has been shown to play a role in determining the word order in several languages (cf. Choi 1999 for German and Korean; Ishihara 2000 for Japanese).
with either $V_{\text{FOCUS}}O_{\text{topic}}S_{\text{topic}}$ or $V_{\text{Otopic}}S_{\text{focus}}$. Here, $V_{\text{Otopic}}S_{\text{focus}}$ is predicted by FSP default word ordering working on a non-configurational VP. The same reasoning applies to $V_{\text{FOCUS}}O_{\text{topic}}S_{\text{topic}}$. Given this, one should expect $V_{\text{FOCUS}}S_{\text{topic}}O_{\text{topic}}$ to be equally acceptable, if the object and the subject are equally topical. Indeed, Avgustinova's data set contains examples of sentences with this word order. For both cases of a focused verb, topical subject and topical object, the clitic marks which of the NPs is the object. Furthermore, for both SOV and VSO the clitic is at least possible (if not preferred) if and only if the object is part of the topic.

Thus, in addition to what I said above, the proposal presented here accounts for the experimentally elicited word orders listed in Avgustinova (1997). Although a detailed syntactic analysis of all possible word orders has to be left to further research, I have sketched an analysis of the DOC in CD and its interaction with word order and information structure. I will refer to this use as '(direct object) topic agreement marker' usage. This label makes reference to Bresnan & Mchombo (1987) who distinguish grammatical and anaphoric agreement markers. I now turn to a second function of the DOC, its use as 'default' pronoun, and then show that my proposal makes the right predictions for the distribution of those two different uses of the DOC.

53 Topic-before-comment ordering holds in 91.0%, subject-before-object in 89.5%, and agent-before-patient in 88.3% of the cases. The correlation between the three scales explains why the numbers are so close.

54 Recall that SMALL CAPS are used for DFs (as part of the f-structure) and lowercase letters for IS-roles. Since Avgustinova (1997) does not make a comparable distinction, the annotation is my translation of her classification.
The DOC as default pronoun

Although this is not a salient topic in the literature on the Bulgarian object clitics (for an exception, see Vakareliyska 1994:125), there is no doubt that the DOC has another use as the default pronoun. That is, the DOC is the pronoun that is used in cases where the speaker does not want to convey any information structural implications. To further clarify what I mean by default and to illustrate the relation between the two types of pronouns, consider the following dialogue, where (47) but not (48) is a possible continuation of (46) if no contrast is intended:

(46) "Karl sreštna onazi tancjorka včera"
   Karl met3.sg that dancer yesterday
   Karl met that dancer yesterday.

(47) "Ivan sâšto ja (*neja/*NEJA) poznava."
   Ivan too DOC3.sg,FEM her/HER knows3.sg
   Ivan knows her, too.

(48) "Ivan sâšto poznava neja/NEJA."
   Ivan too knows3.sg her/HER
   Intended: Ivan knows her too.

First consider the possibility that all cases of the DOC as default pronoun might be due to (topic) object drop. This apparently has not been tested in the literature. Like English, Bulgarian allows specific and unspecific object drop (depending on the verb, cf. Fillmore 1986). However, even after considering object drop, there are still cases where object drop is not possible and the DOC is the only realization of the object in the sentence. Thus there has to be a second use of the DOC as pronoun, i.e. a DOC with a PRED PRO. For a formal LFG analysis, this raises the question whether there are two entries for each DOC or one with an optional PRED PRO. Consider the hypothesis that there is one DOC with an optional PRED PRO. In that case, the default pronoun use of the DOC would always result in the object (i.e. the clitic itself) being marked as topic. It is not clear whether this is desirable, although one could argue that all pronouns have to be topical in some sense anyway, since their referent is 'salient' (cf. Chafe 1976) most of the time (in order to be identifiable). For now, I posit two separate lexical entries for the DOC, one with an optional PRED PRO (the
default pronoun) and one with the topic equation.\footnote{In this point, the current analysis differs from Dalrymple (1993:135) who assumes an optional PRED PRO for the Bulgarian direct object clitic.} Again, this is illustrated for \textit{ja}. The difference between the two lexical entries is highlighted by \textit{italics}.\footnote{Note that \uparrow IS is not part of the lexical entry's c-structure annotation. The up-arrow therefore does not point to the f-structure containing it (i.e. \uparrow OBJ). It only indicates that the referent of the DOC is an element of IS-topic within IS (cf. Figure 4, p.28).}

The existence of two lexical entries poses the question of how the proposal presented here can guarantee the right use of DOC for a given sentence. So far, because of the functional control established by the TOPIC position, the optional PRED PRO use is ruled out by \textit{unicoveness} whenever a fronted (object) constituent sits in a TOPIC position. Whenever the object is realized within the VP, \textit{unicoveness} rules out two PRED values for the object, since the DOC \textit{defines} the object (instead of just constraining it). With no other object constituent being realized, the DOC is interpreted as object (default pronoun). Here, this is guaranteed by \textit{extended coherence}. In the next section, I show that data from topicalization out of islands further support this analysis.

\begin{itemize}
\item \textbf{ja:} $N_{D}$ - DOC
\hspace{1cm} $\uparrow$OBJ) = $\downarrow$
\hspace{1cm} (\downarrow PERS) = 3
\hspace{1cm} (\downarrow NUM) = SG
\hspace{1cm} (\downarrow GEN) = FEM
\hspace{1cm} $\uparrow \in \text{topicIS}$
\item \textbf{ja:} $N_{O}$ - DOC
\hspace{1cm} $\uparrow$OBJ) = $\downarrow$
\hspace{1cm} (\downarrow PERS) = 3
\hspace{1cm} (\downarrow NUM) = SG
\hspace{1cm} (\downarrow GEN) = FEM
\hspace{1cm} (\downarrow PRED) = "PRO"
\end{itemize}
6 Island data: When can which type of DOC occur?

In this section, I show how the current proposal correctly predicts the distribution of the two uses of the DOC (i.e. as default pronoun and as topic agreement marker).

Bulgarian does not obey most of the common island constraints (Rudin 1985). Leaving aside relativization for now, which is somewhat more complex, the Bulgarian island data for two types of object extractions (wh-extraction and "topicalization") can be summarized as shown in Table 2. Where no judgments were available in the literature, the data given in Table 2 come from my informants. In the table, "DOC" indicates that extraction out of the given extraction site is possible if and only if the DOC is realized in the extraction site (e.g. a sentential subject). "(DOC)" indicates that realization of the DOC in the extraction site is optional, and "*" indicates that the extraction is not possible regardless of whether the DOC is realized at the extraction site or not.

<table>
<thead>
<tr>
<th>Island for</th>
<th>wh-extraction (for question)</th>
<th>&quot;Topicalization&quot; (EXTERNAL TOPIC or TOPIC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sent. Subject</td>
<td>(DOC)</td>
<td>DOC</td>
</tr>
<tr>
<td>Sent. Object</td>
<td>(DOC)</td>
<td>DOC</td>
</tr>
<tr>
<td>Adjuncts</td>
<td>(DOC)</td>
<td>DOC</td>
</tr>
<tr>
<td>Relatives</td>
<td>*</td>
<td>Rudin (1985:182-3)</td>
</tr>
<tr>
<td>Questions</td>
<td>*</td>
<td>Rudin (1985:105)</td>
</tr>
<tr>
<td>PP</td>
<td>*</td>
<td>Rudin (1985:143)</td>
</tr>
</tbody>
</table>

Table 2 - Object Long Distance Dependencies and the DOC

Here, I do not discuss why object "topicalization" out of a relative clause is possible although any kind of extraction out of any kind of NP is not possible in Bulgarian (Rudin 1985:143).57

Consider, for example, the following cases of topicalization:

---

57 Maybe the observed cases go back to an EXTERNAL TOPIC instead of TOPIC-fronting. In this case, the explanation would be straightforward. The same explanation that accounts for the grammaticality of example (52) below (p. 39) would account for this kind of object extraction out of relative clauses.
Sentences (49) and (50) show that TOPIC-fronting out of a sentential subject, in (49), and a relative clause, in (50), is possible. Just as in the case of TOPIC-fronting within one clause, the DOC is obligatory. The grammaticality of (49) and (50) in the presence of the DOC is predicted. The fronted object is functionally controlled by the DOC, which has to be realized because it is the only constituent that can define (not just constrain) the OBJ function and agrees with the TOPIC-fronted constituent (i.e. the object) in person, number, and gender. The DOC with a PRED PRO cannot be chosen because this would violate UNIQUENESS.

More generally, TOPIC-fronting within a clause (e.g. (43), p. 26), from an embedded clause into the matrix clause (see above), and within an embedded clause (e.g. (31), p. 21) are all possible and always require CD. All of these cases are correctly predicted by the account presented so far.

Next, consider topicalization out of an island (here, an NP):

Regardless of whether the DOC is realized, sentence (51) is ungrammatical. According to Bresnan & Grimshaw (1978), filler-gap dependencies (i.e. functional control within LFG), but not anaphoric binding, obey island constraints. The DOC does not repair island-violations. Here, the ungrammaticality of (51) is straightforwardly accounted for. The fronted constituent can only satisfy the outside-in functional uncertainty equation (and thereby EXTENDED COHERENCE) if it is functionally controlled by a GF-bearing constituent further down in the f-structure. The fronted object cannot be functionally controlled by a constituent with a PRED value because this would violate UNIQUENESS. The DOC cannot be realized in the embedding (instead of the embedded) sentence to bind the fronted object since the object function of the embedding sentence already has an object with a PRED value. Finally, the DOC with a PRED PRO (i.e. the default pronoun) could be realized in the embedded clause in order to satisfy COMPLETENESS and COHERENCE. However, the outside-in functional
uncertainty equation of the fronted object would still have to be resolved. This is not possible since the embedded GF (i.e. the direct object) is not accessible – it is located in an island. This is illustrated by f-structure 2, where anaphoric binding is indicated by dotted lines, and functional control by solid lines. The doubled crossed line stands for an island violation. F-structure 3 is out for the same reason – because the functional control violates the island condition (Bresnan & Grimshaw 1978). Note that this also confirms that CD should be formalized by means of an functional outside-in uncertainty on the TOPIC position rather than a inside-out functional uncertainty in the lexical entry of the topical agreement marker DOC since otherwise the default pronoun DOC would wrongly be predicted to be possible within the island.

There is one more phenomenon that supports my analysis: EXTERNAL TOPICS. Example (52) – if uttered with a clear pause between the fronted object and the following sentence – is grammatical. This kind of a detached constituent fulfills the criterion of EXTERNAL TOPIC (see above, Aissen 1992, Cinque 1977, King 1995, Kiss 1994).
In the account presented here, the grammaticality of (52) follows from the fact that the phrase structure rule for EXTERNAL TOPICS is not annotated with an outside-in functional uncertainty equation. In GB terms, EXTERNAL TOPICS are base generated to the left of the sentence rather than moved to their position. Therefore, no movement/functional control violates the island constraint and the DOC with a PRED PRO is realized in the embedded clause. To satisfy COMPLETENESS the default pronoun DOC is realized. Assuming that EXTERNAL TOPICS are subject to EXTENDED COHERENCE, the DOC's PRED PRO anaphorically binds the EXTERNAL TOPIC. This is illustrated by the c- and f-structure given below.

![c-structure 5](image-url)
To sum up, the island data presented above is not only compatible with the current analysis but is also predicted by it. In the next and final section, I summarize the account outlined above and list some open questions.
7 Conclusions and Outlook

In this section, I proceed as follows. First, I summarize the present analysis and the conclusion specific for Bulgarian. Second, I point out several more general points of theoretical interest, and, finally, I mention some open questions and issues for further research.

7.1 Summary of the DOC's functions and typological conclusions

I have shown how that the D[irect] O[bject] C[litic] in Bulgarian has at least two different uses. First, the DOC can be used as the default pronoun of Bulgarian. Second, the DOC can be used in C[litic] D[oubling] as an agreement marker for a topical direct object which is not [-generic, -specific]. I have presented an analysis that predicts how these two functions/uses of the DOC interact. The distribution of optional and obligatory realization of the DOC cannot be explained without accounting for the two different uses of the DOC. That is, as the island data presented in section 6 shows, some environments only license one of the two uses. More precisely, the analysis correctly predicts that the DOC cannot be used in islands and that its use in the case of EXTERNAL TOPIC "doubling" it is actually the default pronoun and not the agreement marker DOC, which "doubles" the left-dislocated phrase.

In contrast to the object marker in Chichewa (cf. Bresnan & Mchombo 1987: 745), the Bulgarian DOC does not mark an f-structure TOPIC but an IS-topic. This, in interaction with the constraint in the specificity of the DOC (see above), predicts when CD is possible. The cases of obligatory doubling of TOPIC-fronted objects are predicted by associating an outside-in functional uncertainty equation with the TOPIC position in Bulgarian. More precisely, the DOC's IS-topic marking function, in interaction with the proposed annotated phrase structure rules (especially, the functional control of fronted topic), accounts for both obligatory TOPIC-fronted object doubling and optional doubling of topical objects in general. By this, the present account predicts when overt realization of the DOC is possible and when it is obligatory, a goal which, to the best of my knowledge, is not achieved by any other existing account.

The optionality of CD in many contexts shows that speakers have different options of coding a topical object depending on the register, genre and maybe other factors (see Leafgren 1997a, 2001, 2002 for similar thoughts). Possible generalizations relating the choice
of forms to their functions and other factors such as register merit further investigation. For example, does the absence of intonation in written language enforce the use of alternative linguistic means (such as more strict case marking given otherwise optional case marking, or more strict word order) to identify GFs and DFs?

Although subject to further testing, the presented analysis is supported by a broad empirical basis. In addition to the native speaker intuitions, the analysis accounts for data from Leafgren's (1997a, 1997b) corpus-based studies, Avgustinova's (1997) elicited question-answer data (more than 20 word order-prosody mappings for a transitive verb), and the data of 7 speakers collected in an online experiment. As mentioned in the introduction, all examples were checked by one to three native speakers of Bulgarian.

To the best of my knowledge, unlike all other formal accounts so far (e.g. Rudin 1997, 1996, 1990/1991, 1985, Dyer 1992, Avgustinova 1997, Dimitrova-Vulchanova & Hellan 1998, Franks & King 2000), the account presented above predicts the obligatory CD in the case of fronted objects and provides a possible explanation for the optionality of CD in other cases. For example, Rudin's (1997) MP analysis of the DOC as an AgrO-head cannot predict why the DOC is obligatory in certain cases yet optional in others. Furthermore, I explicitly addressed the relatively free word order of Bulgarian and predicted the resulting word order depending on the IS-roles assigned to the different phrases. Although empirically attested, many of the word orders discussed at the end of section 4 are ignored in most of the theoretical literature on Bulgarian.

I believe that the present account, by predicting when the DOC can/must be overtly realized, draws a picture that is arguably closer to what is going on in a speaker-hearer's mind while formulating/parsing a sentence with CD. The formal analysis models the chain of reasoning of e.g. a hearer who while parsing a sentence makes use of (a) general knowledge of the language (maybe including universal constraints on language such as COMPLETENESS, UNIQUENESS, and EXTENDED COHERENCE), (b) the overt clues given in a sentence (e.g. intonation, absence/presence of the DOC), and (c) context.

While the analysis accounts for all observed word orders (including predictions about prosody via proposed constraints on the IS, e.g. via IPC, cf. (G9) in section 4), it does not predict spurious parses or ambiguities arising from the lexical ambiguity of those two uses. The account presented here could therefore close the gap between the work on DF-configurationality and free word order in Bulgarian. It also is a first step to resolve the
mismatch between the broad-coverage empirical work on Bulgarian and the literature on its formal aspects.

7.2 Theoretical Implications of the conclusion

More generally, the current account stresses that linguistic forms can have several (independent) functions. This is even more evident when we consider that the DOC has a third function as intrusive pronoun\(^{58}\), as mentioned in the introduction. In other words, Bulgarian speaker-hearers do not have any problems with the fact that the DOC has multiple distinct uses since those uses are licensed by different (constructional) environments. It is important to note, that which function a DOC in a given sentence bears, is determined by the interaction of the available lexical entries (of the DOC) and general principles of grammar. In LFG, those principles included COMPLETENESS, EXTENDED COHERENCE and UNIQUENESS. In a more common terminology, it is the theta-role assignment (a.k.a. theta-criterion), which determines which function of the DOC is chosen. By stating that all GFs in a clause have to be licensed by the verb, and that all licensed GFs have to be realized (i.e. have to have a f-structure predicate) exactly once, the theta-role assignment forces the DOC to be interpreted as either a pronoun or a grammatic agreement marker, depending on which other elements are (overtly) realized in the clause.

This paper also provides one further example of a (non-dependent-marking) language that seems to compensate for the lack of GF-configurationality by morpho-syntactic means (head-marking). In this respect, Bulgarian presents an interesting case, because it has neither case-marking nor GF-configurationality. Instead, the left-periphery of the Bulgarian clauses provides grammaticalized positions, which are sensitive to DFs, rather than GFs (cf. King 1995 for Russian).

7.3 Open question and possible further research

Further research is necessary in order to see how the different functions of the DOC relate to each other. I also think that it is worth investigating if there are further restrictions on the optional or obligatory presence of the DOC in certain contexts. The case, where subject and object cannot be disambiguated by means of the DOC (because they have the same person and number, and, in the case of 3.SG., also the same gender) presents an interesting

---

\(^{58}\) This function is probably an extension of the DOC’s use as default pronoun.
case of a conflict between the functional maxim of clarity ("be clear") and the maxim of consistency/generality ("use the same construction in all cases, where it can apply"). In other words, the language either looses a generalization (CD of TOPIC-fronted objects is not possible)59 or allows for ambiguity (CD of TOPIC-fronted objects is possible). An investigation of which contexts prefer which of those two "solutions" would shed light on strategies of speakers to deal with such situations.

Also, there are still possible mismatches in the observations made by Avgustinova (1997) and Leafgren (1997a,b) regarding the question in exactly which contexts the DOC is obligatory. Above all, it is necessary to investigate whether the notion of topicality can be split up into more precise features that can be attested in a corpus, e.g. one should look at the correlation between CD and different degrees of Assumed Familiarity (Prince 1981) or topic-maintenance vs. topic-shift (i.e. is the DOC used to maintain or to introduce a topic). In this context, especially the issue of contrastiveness or emphasis in the TOPIC position needs further investigation.

Another phenomenon that needs further research is the CD of quantified NPs. While I have shown how quantified NPs confirm that [-specifics] cannot be doubled (cf. section 3), the details of CD of quantified NPs are yet to be worked out. It seems that speakers vary a great deal in which quantifiers they perceive to be marking or to be compatible with specificity. This indeterminacy with regards to determiners is a common phenomenon (Arnold Zwicky, p.c.) on which one may receive interesting insights by a more extended study of CD of quantifier phrases in Bulgarian.

Once we have a better picture of all the factors that determine the possible word orders for a given context, a (stochastic) OT account may be able to combine those factors into a formal description of the data. Related to this, it is very interesting that exactly those dimensions which are strict factors of CD in contemporary Bulgarian (i.e. specificity and topicality), show up as statistic preferences in contemporary Macedonian CD (Čašule 1997). Furthermore, Čašule (1997) argues that, in the diachronic development of Macedonian, first topicality and subsequently specificity was the determining factor of CD.

59 Recall that this case would be ambiguous since CD is also allowed for not fronted, topical object.
8 References

AG, ACADEMY GRAMMAR (edited by Popov, Konstantin)
- (1983): Grammatika na sâvremennija bâlgarski knižoven ezik, t. 3. sintaksix; Sofia, Bâlgarska Akademija na Naukite.

AISSEN, JUDITH

ALEKSOVA, KRASIMIRA

ALEXANDROVA, G.

ARNAUDOVA, OLGA

AUSTIN, PETER

AUSTIN, PETER / BRESNAN, JOAN

AVGUSTINOVA, TANIA

AVGUSTINOVA, TANIA & BISTRA ANDREEVA

BAKER, MARK C.

BARD, E., D. ROBERTSON & A. SORACE.

BILLINGS, LOREN A.

BRESNAN, JOAN
- (1982): The Mental presentation of Grammatical Relations; Cambridge, MIT Press.

BRESNAN, JOAN / GRIMSHAW, JANE

BRESNAN, JOAN / MCHOMBO, SAM A.
ČAŠULE, ILIJA

CHAFE, WALLACE, L.

CHOI, HYE-WON

CHOMSKY, NOAM

CINQUE, GUGLIELMO

CYXUN, GENADIJ A.

DALRYMPLE, MARY

DALRYMPLE, MARY, RONALD KAPLAN, AND TRACY KING

DIMITROVA-VULCHANova, MILA / HELLAN, LARs

DYER, DONALD L.
- (1992): Word Order in the simple Bulgarian Sentence: A Study in Grammar, Semantics and Pragmatics; Amsterdam – Atlanta, Rodovi B. V.

EMBICK, DAVID / IZVORSKI, ROUMYANA

FILLMORE, CHARLES J.

FRANKS, STEVEN / KING, TRACY HOLLOWAY

FRASER, BRUCE

GEORGIEVA, ELENA

GERASSIMOVA, VERONICA A. / JAEGER, T. FLORIAN
On the Optional and Obligatory Realization of the Bulgarian Direct Object Clitic

GIVÓN, TALMY

GUENTCHEVA, ZLATKA

HALE, KENNETH

HELMBRECHT, JOHANNES

ISHIHARA, SHINICHIRO

IVANČEV, SVETOMIR
- (1978): Prinosi v bâlgarskoto i slavjanskoto ezikoiznanie; Sofia, Nauka i izkustvo.

IZVORSKI, ROUMYANA

JAÆGER, T. FLORIAN

JAÆGER, T. FLORIAN & VERONICA A. GERASSIMOVA

JAÆGER, T. FLORIAN & DAVID Y. OSHIMA

KAZAZIS, KOSTAS & JOSEPH PENTHERADOUKIS

KING, TRACY H.

KISS, KATALIN E.
LAMBOVA, MARIANA
- (2002): Multiple topicalization wh-fronting in Bulgarian and the fine structure of the left-periphery; in: The 26th Penn Linguistics Colloquium.

LAMBOCHT, KNUD

LEAFGREN, JOHN R.
- (2002): Register, Mode, and Bulgarian Object Placement; presented at the Balkan Conference 2002, source: leafgren@email.arizona.edu.
- (1997c): Topical Objects, Word Order, and Discourse Structure in Bulgarian; presented at the AAASS 1997, source: leafgren@email.arizona.edu.

LEGATE, JULIE A.

MINČEVA, ANGELINA

NIKOLOVA, CVETANKA
- (1987): Chestoten rechnik na balgarskata razgovorna rech (A Frequency Dictionary of Colloquial Bulgarian); Sofia: Nauka i izkustvo

NORDLINGER, RACHEL

O’CONNOR, ROB

PENČEV, JORDAN
- (1984): Stroev na bâlgarskoto izrečenie; Sofia, Nauka i izkustvo.

PETERSON, P.

POPOV, KONSTANTIN P.

POPOV, KONSTANTIN P. & VENČE S. POPOVA
On the Optional and Obligatory Realization of the Bulgarian Direct Object Clitic

PRINCE, ELLEN

RUDIN, CATHERINE
- (1994): On focus position and focus marking in Bulgarian questions; in: Papers from the Fourth annual meeting of the Formal Linguistic Society of Midamerica, April 15-18, 1993, pp. 252-266, Iowa City.

RUDIN, CATHERINE / KRAMER, CHRISTINA / BILLINGS, LOREN / BAERMAN, MATTHEW

SAG, I. / G. GAZDAR, T. WASOW, AND S. WECHSLER

SCHICK, IVANKA P. & FRITS BEUKEMA

SCHIFFRIN, DEBORAH

SELLS, PETER

SHAMRAY, TATYANA

SIEWIERSKA, ANNA / UHLIŘOVÁ, LUDMILA

SIMPSON, JANE HELEN

SGALL, PETR

**SPEAS, MARGARET**

**TOMIĆ, OLGA MIŠESKA**

**VAKARELIYSKA, CYNTHIA**
- (2000): *Talk on Bulgarian Object Doubling*; available from vakarel@oregon.uoregon.edu

**VALLDUVÍ, ENRIC**
9 Appendix: GF-(non-)configurationality in Bulgarian

Traditionally (cf. Penčev 1973), and in more recent GB (cf. Dyer 1992, Rudin 1988, 1989, 1990/1991, Embick & Izvorski 1994) and MP approaches (cf. Rudin 1994, 1997, Embick & Izvorski 1998, Rudin et al. 1998, Rudin et al. 1999, Arnaudova 2001, Bošcovic 1998), GF-configurationality of Bulgarian has never been questioned, or provided evidence for but implicitly assumed. A number of configurationality test have been proposed in the literature (e.g. Speas 1990:137, King 1995:Chapter 3). The problem with some of the tests is their theory-dependency. As shown below, many of the test results have been argued to be explainable by means other than phrase structure configurationality. Here, we proceed as follows. We apply those tests which are applicable to Bulgarian (i.e. extraction, multiple wh-questions, binding, and weak crossover and other constituency tests, namely ellipsis, pronominalization, fronting and coordination) and discuss their outcome. It will turn out that, according to the results of these diagnostics, the GF-configurational status of Bulgarian is by no means as clear as commonly assumed.

One important characteristic of configurational languages is that they exhibit clear subject-object asymmetries such as Superiority effects, difference in possible coreference, and weak crossover, among others (Speas 1990:137). Such phenomena have been attributed to the distinct phrase structure positions of the subject and the object in configurational languages. Note that in the following we will sometimes speak about VP(-tests) when we mean (tests for) a verb-object constituent. We stick to this common terminology, although it is sometimes confusing.

A. In extraction, traces must be properly governed, according to the Empty Category Principle (ECP). The assumption is that objects are directly governed and assigned a theta-role by the verb in the VP. Thus objects extract freely, whereas subjects are governed indirectly, and therefore obey more extraction constraints (King 1995:54). In English, for example, objects but not subjects may extract from embedded clauses with an overt

---

60 As mentioned in the introduction, this section is for the largest part identical with Gerassimova & Jaeger (2002:Section 3). Minor revisions and corrections have been made here and there.

61 For example, Dyer (1992:70) assumes the following c-precedence hierarchy for the basic word order in Bulgarian: SUBJ – V - Odirect* Oindirect. At the same time, he identifies forty-nine possible word orders for sentences with only a ditransitive verb and its complements (ibid:56).
complementizer. In Bulgarian, however, subjects and objects extract with equal ease. Consider the following examples of a noun modified by a relative clause, which involves extraction from an embedded clause with an overt complementizer:

**Subject extraction**

(53) Ženata, kojato Todor kaza, če _ e vidjala Marija, ...
    woman.DEF who T. said.3.SG that _ is seen.3.FEM.SG M.
    The woman, who Todor said that has seen Maria.

**Object extraction**

(54) Ženata, kojato Todor kaza, če Marija e vidjala _, ...
    woman.DEF who T. said.3.SG that M.     is seen.3.FEM.SG _
    The woman, who Todor said that Maria has seen.

Examples (53) and (54) are cases of subject and object relativization, respectively, and both examples are grammatical. The same lack of asymmetry in extraction between subjects and objects is also observed in wh-questions and topicalization. This leads Rudin (1985) to the conclusion that the ECP is irrelevant for Bulgarian. Thus extraction in Bulgarian fails to provide positive evidence for a structural difference between subjects and objects. According to the ECP, the identical behavior of subjects and objects can be attributed to the fact that both are directly governed by the verb, and are therefore inside the VP. In theory neutral terms, these data suggest that the subject and the object in Bulgarian have identical structural positions.

B. In multiple wh-questions in Bulgarian, all wh-words must be fronted, and they have been reported to exhibit strict ordering with subjects appearing first and the rest of the wh-words in specific positions (Rudin 1990/1991:440, 1988, 1989):


(55) Koj kogo vidja?
    who whom saw.3.SG
    Who saw whom?

(56) *Kogo koj vidja?
    whom who saw.3.SG
    Whom did who see?

This subject-object asymmetry, too, has been accounted for in terms of phrase-structure and the Superiority condition (cf. Chomsky 1973). However, in his analysis of multiple wh-questions, Boškovic (1993, 1998) argues that, in Bulgarian, only the first extracted wh-word is subject to the Superiority condition. Since the following wh-words occur in free order with
respect to each other, they are unaffected by Superiority. Consider the following data taken from Boškovic (1993, cf. 1998:60-61):

**Superiority and multiple wh-questions in Boškovic (1998:60-61, 1993)**

(57) Kogo kak e celunal Ivan.
    whom how is kissed Ivan
    *How did Ivan kiss whom?*

(58) *Kak kogo e celunal Ivan*
    How whom is kissed Ivan

(59) Koj kogo kak e tselunal
    Who whom how is kissed
    *Who kissed whom how?*

(60) Koj kak kogo e celunal
    Who how whom is kissed
    *Who kissed whom how?*

Boškovic (1998:62, 1993) concludes that Superiority can only account for the first fronted *wh*-word in multiple *wh*-questions and argues that Superiority can only be understood as a descriptive generalization. An alternative explanation of these ordering effects can be attributed to a typologically motivated GF hierarchy constraint. Restricted to the first *wh*-expression in the string, it predicts that the least oblique constituents will precede all the rest (cf. Billings & Rudin 1994 for an OT account similar in spirit). Although the question requires further investigation, alternative data suggest that animacy is one factor in the ordering of *wh*-words. (61) and (62) show that when the object but not the subject is of type [+animate], both orders are acceptable. If the [+animate] object precedes the [-animate] subject, as in (62), the DOC is obligatory. (63) indicates that a DOC does not repair the ungrammaticality of (56). Therefore the DOC cannot be the direct cause of the constituent reordering in (62) but is better thought of as an indirect consequence of the *wh*-word reordering, which in turn is licensed by a semantic animacy hierarchy.

**Wh-extraction with reversed animacy of object and subject**

(61) Kakvo kogo iznenada?
    what whom surprised.3.SG
    *What surprised whom?*

(62) Kogo kakvo go iznenada?
    whom what 3.MASC.SG surprised.3.SG
    *What surprised whom?*

(63) *Kogo koj go vidja
    whom who him saw.3.SG
    Whom did who see?
Since examples like (62) violate Superiority, we conclude that in Bulgarian the ordering of *wh*-elements in multiple *wh*-questions cannot be explained by Superiority. Thus multiple *wh*-questions do not provide clear evidence about the configurationality of Bulgarian.

C. It has been proposed in the literature that co-reference relations between a lexical NP and a pronoun differ between subjects and objects. This is accounted for by Principle C of the GB Binding Theory (cf. Chomsky 1981), which states that R-expressions (e.g. lexical NPs and *wh*-traces) must be free. That is, they cannot be co-indexed with a c-commanding antecedent. For configurational languages, in which the subject, by definition, c-commands the object, Principle C predicts subject-object asymmetries for binding (e.g. Speas 1990:137, King 1995:50). Examples (64) and (65) illustrate co-reference relations between a pronominal argument and a lexical NP embedded in a relative clause:

**Principle C effects on anaphoric binding**

(64) Kotkata, s kojato Todor, živeeše, go napusna.
    cat.DEF with whom.FEM.SG T. lived.3.SG him left.3.SG
    The cat who Todori lived with left him.

(65) *Toji napusna kotkata, s kojato Todor, živeeše.
    he left.3.SG cat.DEF with whom.FEM.SG T lived.3.SG
    *He left the cat who Todori lived with.

In (64) the pronoun *go* is the object of the main clause, and can be coreferential with Todor. In (65), however, the pronoun is the subject of the main clause, and a co-referent reading is impossible. Following Principle C, the contrast between (64) and (65) can be explained in terms of the different structural positions of *go* and Todor: in (64), the lexical NP and the pronoun do not c-command each other, whereas in (65) the pronoun is the subject, and therefore c-commands Todor. Consider the above sentences with the order of the lexical NP and the pronoun reversed, as in (66) and (67):

**Precedence effects**

(66) *Napusna go, kotkata, s kojato Todor, živeeše.
    left.3.SG him cat.DEF with whom.FEM.SG T. lived.3.SG
    The cat who Todor, lived with left him.

(67) Kotkata, s kojato Todor, živeeše, toj ja napusna.
    cat.DEF with whom.FEM.SG T. lived.3.SG he her left.3.SG
    He, left the cat who Todor, lived with.

In the parallel Russian examples, King (1995:52) assumes that in examples like (66) the subject NP is right-adjoined to VP, and in (67) the object is left-adjoined to IP. As a result, the pronoun and the lexical NP do not c-command each other, and the contrast in
grammaticality is due entirely to their linear ordering. While this analysis is valid for Bulgarian in (67), in (66) the facts are more complex since the object pronoun is realized as a clitic, and not as a free pronoun. We follow Rudin (1997:238) in that go in fact commands Todor, which accounts for the ungrammaticality of (66).

However, these data can be accounted for entirely in terms of linear precedence: in both grammatical examples above go does not precede Todor, while in the ungrammatical ones the pronoun precedes the lexical NP. This follows from the fact that in the typically assumed right-branching structure dominance directly correlates with linear precedence. Thus from these data it is not clear whether c-command is a necessary condition on binding in Bulgarian, or whether the relevant generalizations have to do with linear and syntactic prominence, as proposed by Bresnan (1994). The discussion of weak crossover in the next section provides a more definite answer to this question.

D. Weak crossover is another standard test for configurationality. It is based on the observation that a pronoun cannot appear on the path between a trace and its operator. One explanation for this fact in English is the Leftness Condition, which says that a pronoun cannot be coindexed with a variable to its right (Chomsky 1976). It rules out cases like (68), in which a pronoun is coindexed with a wh-trace to its right, and licenses cases like (69), in which the wh-trace is to the left of the pronoun:

**Weak crossover in English**

(68) *Who does his mother love?  
   who₁ does his₁ mother love t₁

(69) Who loves his mother?  
   who₁ t₁ loves his₁ mother?

In Bulgarian, the weak crossover data are slightly more complex. Like Russian (cf. King 1995:55) and unlike English, Bulgarian has reflexive and non-reflexive possessive pronouns. Reflexive possessives must be bound in their minimal domain. (70) is ungrammatical regardless of whether the reflexive is co-indexed with the object wh-word or not. In contrast, (71) is grammatical if and only if the reflexive is coreferential with the subject wh-word. Such examples show that reflexive possessives do not pattern together with the English possessives, and cannot be used to test crossover.

62 Whereas Rudin (1997) assumes that go is the head of AgrO, Jaeger & Gerassimova (to appear) use an LFG-account where go does not c-command but f-commands (a command notion that does not rely on phrase-
**Reflexive possessives**

(70) *Kogo_1 običa majka si_j?  
Whom love.3.SG mother his-refl  
*Whom, does his, mother love?*

(71) Koj_1 običa majka si_j?  
Who love.3.SG mother his-refl  
Who, loves his, mother?

Unlike reflexive possessives, non-reflexive possessives cannot be bound in their minimal domain. Because of that, example (72), the counterpart of the English (69), is ungrammatical under coreference.

**Non-reflexive possessives in subject wh-questions**

(72) Koji običa majka mu_i/j?  
Who love.3.SG mother his  
*Who, loves his, mother?*

Some evidence relating to weak crossover comes from non-subject wh-questions like (73), corresponding to (71) above. Although the preferred reading for it is non-coreference between kogo and the possessive pronoun, co-indexation is also possible. In both cases the sentence is good, showing lack of crossover effects:

**Non-reflexive possessives in object wh-questions**

(73) Kogo običa majka mu_i/j?  
Whom love.3.SG mother his  
Whom, does his, mother love?

The declarative variants of the object wh-question are shown in (74) and (75), with different word orders. Note that (74) is ungrammatical in the case of coreference, although there are no binding violations. The reason for this is that the pronoun precedes the lexical NP. In contrast, (75), in which the pronoun neither precedes nor c-commands the antecedent, is grammatical:

**Non-reflexive possessives in declarative sentences**

(74) Majka mu_i/j običa Ivan_1.  
mother his love.3.SG I.  
His*i/j mother loves Ivan_i.

(75) Ivan_2 običa majka mu_i/j.  
I. him love.3.SG mother his  
Hisi/j mother loves Ivan_i.

In sum, the grammaticality of (73), which is the counterpart of the ungrammatical English (68), shows that, unlike English, Bulgarian lacks weak crossover. Instead, the data in this section can be explained entirely in terms of the conditions on Binding in Bulgarian: a pronoun may not precede its antecedent. This suggests that weak crossover can be better explained in terms of linear and/or syntactic prominence, as proposed in Bresnan (1994), Dalrymple et al. (2001).

In conclusion, we have established that subject-object asymmetry tests, such as Superiority, binding and weak crossover, seem to suggest that Bulgarian is not GF-configurational. Next, we apply constituency tests such as VP-pronominalization, fronting, and coordination (cf. Speas 1990) to Bulgarian.

E. Bulgarian lacks VP-pronominalization. In other words, it is impossible to use an overt anaphoric element to refer to a preceding verb phrase.64 Concerning VP-fronting, Bulgarian has no construction resembling it, either.65

- **VP-fronting**

  (76) *Ivan obešta, če šte vzeme izpita, 
  I. promised.3.SG that will take.3.SG exam.DEF
  i da go vzeme toj šte.
  and SBJ it take.3.SG he will

  Intended: Ivan promised to pass the exam, and pass it he did.

F. Finally, there is only one test, namely VP-coordination, which can be interpreted as evidence for the existence of a verb-object constituent in Bulgarian. Coordination tests are based on the observation that if two items can be coordinated, they are both constituents (Sag et al. 1985, Peterson 1981). However, since this generalization has been first stated, many exceptions have become known.

---

63 The presence of the direct object clitic in (75) is due to obligatory clitic doubling of topicalized objects (cf. Dimitrova-Vulchanova & Hellan 1998).
64 Note, however, that VP-ellipsis presents an interesting question. In contrast to English, it is impossible to gap a VP to the exclusion of the auxiliary in Bulgarian whereas it is possible to gap VPs which are not governed by an I.
65 Examples like the following show that in some cases it is possible to front a subjunctive verb and its object. However, the fronted constituent is not the matrix VP but an subjunctive argument of trudno with a controlled subject:

  [Da vzeme izpita] šte e trudno za Ivan
  SBJ take.3.SG the-exam will is difficult for I.

  To take the exam will be hard for Ivan.
Nevertheless, the relevant data from Bulgarian is given here. As (77) shows, verb-object phrases can indeed be coordinated in Bulgarian.

**Coordination of verb-object and verb-subject: [ ] indicates the coordination structure**

(77)  
Cjal den štjax [da piša pisma i da četa knigi].
whole day would.1.SG SBJ write.1.SG letters and SBJ read.1.SG books
All day, I would write letters and read books.

(77) could be taken as an indicator that objects and verbs form a constituent together, which the subject is not a part of. Note however that (78) indicates that subject-verb phrases may also be conjoined:

(78)  
[Az običam, a ti mraziš] sladoled.
I love.1.SG and you hate.2.PL ice-cream
*I love, and you have ice-cream.*

While (78) could very well be a case of right node raising, there seems to be only one intonational boundary, namely after *mraziš*, in which case (78) would most likely not be a case of right node raising. However, this remains subject to further testing.

To sum up, Bulgarian does not show conclusive evidence for GF-configurationality (i.e. the majority of the tests for subject-object asymmetries suggest that Bulgarian has a VP that contains subject and object at the same phrase structure level)\(^66\). Theories of grammar which are based on the assumption of a Universal Grammar (UG) specific to natural language (cf. Chomsky 1995, 1985 for the MP and GB respectively) are basically forced (or at least biased) to analyze all languages as configurational. That is, according to those theories, since there are languages which identify GFs by means of phrase structure (e.g. English) and since all languages are alike, all languages must be at least underlingly GF-configurational.\(^67\) Since LFG does not assume a universally shared c-structure, the data discussed in this section does not posit a problem to the account adopted here.

---

\(^66\) It remains open for further research to investigate how the apparent mismatch of the outcome of the different tests can be resolved. Here we cannot provide a full conclusive account of the data.

\(^67\) We are very well aware that relations, such as MOVE and AGREE, between the universal deep-structure and the specific surface realization of it can be seen as metaphors of the theory, so that the absence of configurationality does not disprove a program like the MP. Here we only want to raise the point that for languages like Bulgarian, Jiwarli and others there seems to be no evidences for configurationality if we do not assume a UG.