

# Discontinuous VP in Bulgarian

Elisaveta Balabanova

University of Library Studies and Information Technologies  
UniBIT

elisavetabal@yahoo.com

## Abstract

This paper presents Bulgarian discontinuous constituents.<sup>1</sup> Bulgarian is claimed to be a language of relatively free word order. As a typical manifestation of free word order discontinuous constituents in Bulgarian have not been studied so far. The paper discusses and analyzes the freedom in Bulgarian word order and points out the way discontinuity has been treated within BulTreeBank. We show the results of our linguistic analysis of discontinuous VPs and summarize the extent of word order freedom and word order constraints within VP.

## 1 Introduction

It is well known that discontinuous constituents are typical manifestation of free word order. It is also claimed that discontinuity is characteristic of languages with rich morphology. Bulgarian shares both of the above features. Scientists, working on word order problems in Bulgarian, debate whether Bulgarian is a configurational or non-configurational language. Most of them share the belief that Bulgarian is a configurational language, but has some non-configurational features (i.e. the free permutation of the elements within VP) (Penchev, 1991).

By exploring the issue of discontinuity in VP we aim to show the extent of word order freedom in Bulgarian and the restrictions on this freedom, coming from semantics.

---

<sup>1</sup> In our definition „constituent“ is the same as „phrase“. For the parts of the constituent we use the term „elements of the constituent“.

The structure of the paper is as follows: In Section 2 discontinuous constituents and the theories of “free“ word order in Bulgarian are discussed; In Section 3 discontinuous constituents are presented within BulTreeBank; Section 4 deals with the types of discontinuity in VP and Section 5 concludes the paper.

## 2 Discontinuous constituents and the theories of “free” word order in Bulgarian

Researchers on Bulgarian word order have so far noticed that there is a greater word order freedom within the verb phrase than within other phrases (Rudin, 1986; Penchev, 1991). Scientists show that the variety of word order models usually is due to the influence of discourse on word ordering. In the tradition of Bulgarian word order investigations, especially in the first half of the 20<sup>th</sup> century, information packaging was taken as one of the most prominent manifestations of discourse. Thus much of the research in the field of word order was devoted to the connection between word order and information packaging.

Sv. Ivanchev (Ivanchev, 1975) is the first scientist who spreaded the ideas of the Prague linguistic school in Bulgaria. The relation between word order and information packaging is investigated by a number of researchers (Georgieva, 1974; Brezinski, 1995; Avgustinova, 1997, Tisheva, 2003, Tisheva and Djonova, 2002; Tisheva and Djonova, 2004a; Tisheva and Djonova, 2004b; Tisheva, 2013).

The interrelation between intonation, word order and information packaging is a topic of research of another Bulgarian linguist – Jordan Penchev (Penchev, 1980). J. Penchev aims to describe the main intonation types of Bulgarian sentences and

for this reason he uses the information from the relationship between semantics and information packaging. There are several researchers who investigate particular word order constructions, but as a whole we can summarize that all the researchers claim that there is a number of factors from structural (syntactic), discourse and prosodic nature, which affect the word order models in Bulgarian. The different combinations of these factors give rise to a larger number of word order combinations within some Bulgarian phrases (the VP especially), than in others (NP, AP), which is the reason for the researchers to discuss whether Bulgarian is a nonconfigurational language. Scientists deny this hypothesis, showing that the word order freedom in some phrases is an isolated phenomenon, which cannot be taken as a sign of nonconfigurationality. They claim that within the structure of Bulgarian there is a combination of configurational and nonconfigurational features.

Based on the above mentioned assumptions, all the researchers are on the shared opinion of that Bulgarian has a rather free or relatively free word order, but noone has pointed out precisely the extent of the word order freedom and the restrictions on word order. Also noone so far has studied discontinuity in Bulgarian, so our survey is the first attempt to analyze one of the most frequent types of discontinuous VPs and the factors, causing discontinuity.

### 3 Representation of discontinuous constituents within BulTreeBank

Investigating discontinuity within a corpus is a good way to investigate the extent of word order freedom and word order constraints. For this sake we use the corpus of syntactic trees in Bulgarian, namely BulTreeBank, which is a corpus of syntactic trees of Bulgarian sentences. Constituency within the treebank is represented via graphs, which are defined on the basis of mother-daughter relation (Simov and Osenova, 2004). Graphs are chosen as close to the context free-tree representation (Simov and Osenova, 2004). In the syntactic trees the original word order is preserved and discontinuous elements are introduced where necessary (Simov and Osenova, 2004). In the examples from the treebank VPC stands for verb-complement phrase, VPS – for a verb-subject

phrase, VPA is a head-adjunct phrase and VPF – a head-filler phrase, which has an extracted element, realized outside the phrase.

There are three types of discontinuous constituents in the treebank.

#### 3.1 Functional element DiscA

This is when a higher dependent is realized between the head and lower dependent/s. For the word order to be preserved, the higher element is marked up with the functional element DiscA (Discontinuous adjunct) and is annotated at a higher place with the functional element *nid* (nonimmediate dominance). Then the element DiscA and *nid* are connected with the same index, seen as a line in the tree below (Simov and Osenova, 2004).

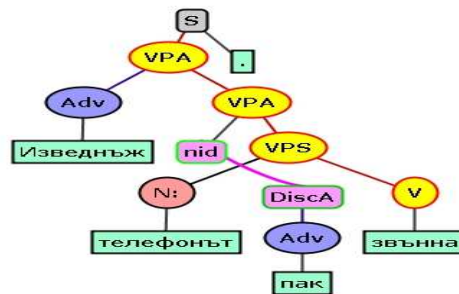
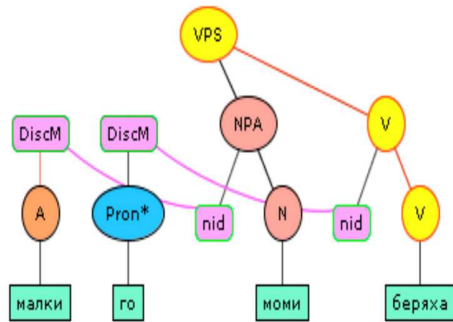


Figure 1: Sentence from the treebank with DiscA element: Izvednazh telefonat pak zvanna. (Suddenly, the telephone again rang

#### 3.2 Functional element DiscM

DiscM stands for discontinuous mixture.

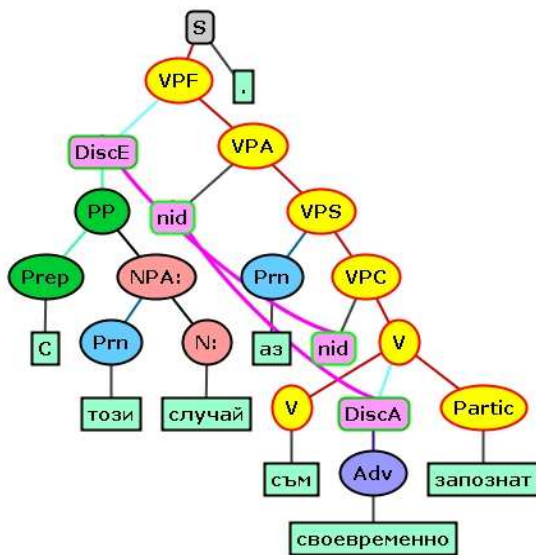
This is a mixture of two constituents. The elements of two constituents are mixed with neither of the two being a governor of the other (Simov and Osenova, 2004). This is a very rare case of discontinuity and has only two or three occurrences in the treebank.



**Figure 2:** Sentence from the treebank with DiscM element: Malki go momi beriaha. (Little it girls picked up; i.e. little girls picked it up)

### 3.3 Functional element DiscE

This is external realization of inner constituent. This is the case of extraction (Simov and Osenova, 2004). Again the element DiscE (Discontinuous extraction) is marked with *nid* with the same index as the phrase where it has been extracted from.



**Figure 3:** Sentence from the treebank with DiscE and DiscA element: S tozi sluchai az sym svoevremenno zapoznat. (With this case I am on time informed.)

## 4 Types of discontinuity, based on the type of the element causing it.

In this part we present our investigation on discontinuity within VP. VP is chosen as the most prominent example of free word order in Bulgarian (see Section 2a). For the completion of the task it was necessary first of all to extract all the sentences, containing discontinuous VPs from the corpus. The

total number of sentences with discontinuity in the corpus is 4160 sentences, which makes about 35% of all the sentences in the treebank. After doing this, we had to select the types of discontinuities within VP. We found out that there are 2 main groups of discontinuity in Bulgarian VP: i) discontinuity, caused by an element which is part of the syntactic structure of the sentence (the element, causing discontinuity in this case is marked up with DiscA in the treebank) and ii) discontinuity, caused by an element, which is not part of the syntactic structure of the sentence (the element, causing discontinuity in this case is marked up with the tag Pragmatic element in the treebank). In this paper we will not deal with discontinuities of the second type. We are focused only on discontinuities, caused by elements, which are part of the syntactic structure of the tree. These elements are: adjuncts; extracted complements of the head verb and the subject. Here we will focus only on discontinuity, caused by adjuncts.

### 4.1 Discontinuity, caused by adjuncts

Within BulTreeBank discontinuities, caused by adjuncts, are the greatest number of discontinuities (67% of all the sentences with discontinuity in VP). In the treebank the sentences are annotated along the lines of HPSG (Pollard and Sag 1994). Thus, according to the theoretical frame we use, adjuncts are attached as sisters of the saturated VP phrase, i.e. when the verb has realized its dependents – complement/s and subject (if there is a subject in the sentence, since Bulgarian is a pro-drop language). Only after the verb has taken its dependents and formed either a VPC (verb-complement phrase) or a VPS (verb-subject phrase), the adjunct is attached to this VPC or VPS phrase, forming a VPA phrase. This is the usual case, when adjuncts are realized linearly without causing discontinuity. In this linear realization of the adjunct the latter modifies semantically the saturated VP. On the contrary, in the cases of discontinuities the adjuncts are realized linearly first and the dependents of the verb (subject and/or complements) – afterwards. In such sentences the projection of the VPA phrase is higher up in the tree and the linear intersection is seen as a line in the graph (see Section 3.1). There are two cases of discontinuity in VP, caused by adjunct: i) The adjunct

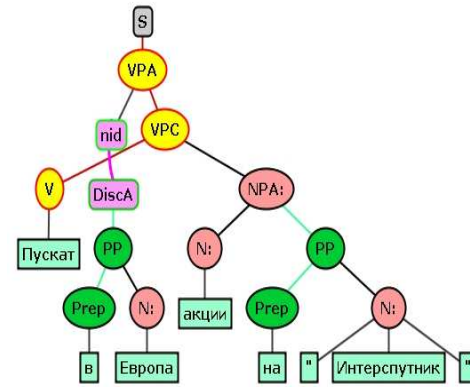
is realized between the subject and the head verb;  
 ii) The adjunct is realized between the head verb and the complement.

Before starting the linguistic analysis, we came across one problem. Namely, the adjuncts were not classified by types in the treebank. Therefore, we needed to have a classification of adjuncts first and then annotate manually all the adjuncts in the sentences of discontinuities, according to this classification. Only afterwards we could extract the sentences with discontinuities by types of the adjuncts. The classification of adjuncts we used is based on GSBKE (GSBKE, 1983) and contains the following types of adjuncts: adjuncts of time, of manner, of quantity and degree, of place, adjuncts of second predication, of condition, of reason, and of aim. In all the sentences with adjunct, causing discontinuity, we annotated the adjuncts manually along this classification. Then it was possible to extract the groups of discontinuities by the type of the adjunct. This allowed us to make conclusions about the reasons, causing discontinuous linear realization of the elements of the VP.

#### 4.1.1 Discontinuity between the subject and the head verb

This is the biggest group of sentences with discontinuities, caused by adjuncts.

Here is the proportional distribution of sentences with adjuncts, causing discontinuity between the subject and the head verb in VP: Adjuncts of time – 45%; Adjuncts of manner – 31%; Adjuncts of quantity and degree – 11%; Adjuncts of place – 5%; Adjuncts of second predication – 3%; Adjuncts of condition – 2%; Adjuncts of reason – 1%; Adjuncts of aim – less than 1%



**Figure 4:** Sentence from the treebank with adjunct between the head verb and the complement: Puskat v Evropa (adjunct of place) akcii na Intersputnik. (They release in Europe shares of Intersputnik).

The information packaging<sup>2</sup> in this type of sentences follows two main patterns:

- 1) The adjunct is part of the Ground.

**Example:** Ground [link[V tzivilizovania sviat] tail [chovek **prez celia si zivot** (adjunct of time)]] Focus [pazi „svetaia svetih“ na svoiata reputacia – svoeto kreditno dosie]. (In the civilized world one, during his whole life, keeps the most precious of his reputation – his credit history.)

In sentences with such discontinuity and communicatively marked word order<sup>3</sup> the adjunct can take the information value of either *tail*, or *link of the tail*. In sentences with communicatively unmarked word order the information value of the adjunct is only a *tail*.

- 2) The adjunct is part of the Focus

<sup>2</sup> For analysis of information packaging we use the methodology of Engdahl and Vallduvi (Engdahl and Vallduvi, 1994; Engdahl and Vallduvi 1996), where **Focus** is the actual information of the sentence, **Ground** is what is presupposed by the information at the output. Sentences have **Ground** only if the context ensures it. The **Link** is the particular place in the sentence for introduction of the new information and the **Tail** points out that there is a need for information update in this part of the discourse.

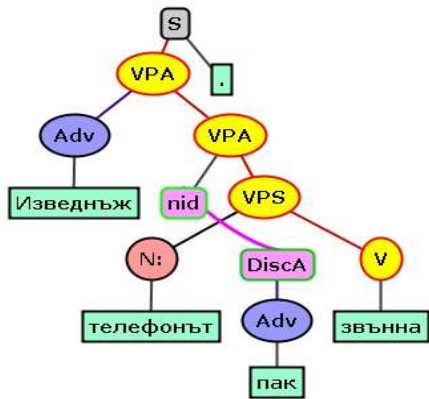
<sup>3</sup> For the relation between word order and information packaging we use the model of T. Avgustinova (Avgustinova, 1997), in which there are 4 types of word order, according to the information packaging: communicatively unmarked, parenthetical, communicatively marked and emphatic.

**Example:** Ground [Vseki opit za ocenka na organiziranata prestapnost v izmereniata na nacionalnata sigurnost] Focus [**zadalzhitelno** (adjunct of manner) predpolaga predvaritelno da se utochniat obhvaty i sadarzhaniето na samoto poniatie. (Any attempt to estimate the organized crime in the context of the national security obligatorily presupposes to define content of the notion itself.)

#### 4.1.2 Discontinuity between the head verb and the complement

The position right after the head verb in VP has been investigated by a number of researchers (Rudin, 1986, Avgustinova, 1997, Penchev in: Boyadzhiev, Kutzarov, Penchev, 1999, Tisheva, 2000, Tisheva, 2013).

Here is the proportional distribution of sentences with adjuncts, causing discontinuity between the subject and the head verb in VP: Adjuncts of manner – 40%; Adjuncts of time – 30%; Adjuncts of quantity and degree – 11%; Adjuncts of place – 9%; Adjuncts of second predication – 4%; Adjuncts of aim – 2%; adjuncts of condition – less than 1%; Adjuncts of reason – less than 1%.



**Figure 5:** Sentence from the treebank with adjunct between the subject and the head verb: Izvendazh telefonat pak (adjunct of manner) zvyinna. (Suddenly the telephone again rang).

In sentences with adjuncts between the head verb and the complement the adjunct becomes part of the focus.

**Example:** Razbira se, choveshko e da se sbyrka, no tuk Focus [znam **mnogo dobre** (adjunct of manner) za kakvo stava дума]. (Of course, it's human to make mistakes, but I know exactly of what we're talking about.)

#### 4.2 Conclusion about the word order models with discontinuity, caused by adjuncts

From our linguistic analysis we can summarize that the factors, which rule the realization of adjuncts within VP are:

1. The information packaging within the sentence, which depends on
2. The semantics of the adjuncts. According to the semantics of the adjunct and according to which part of the sentence the adjunct is syntactically attached, we distinguish<sup>4</sup> two types of adjuncts: i) sentential (they modify semantically the whole sentence) and ii) phrasal (they modify semantically a particular element of the VP).

Most of the adjuncts in Bulgarian modify semantically the whole sentence (these are the adjuncts of time, of place, of condition, of reason and of aim). Syntactically, these adjuncts are realized as sisters of the saturated VP. Thus their linear realization within VP is only a result of the particular information packaging that the speaker chooses to make in his utterance.

The word order realization of the phrasal adjuncts (adjuncts of manner, of quantity and degree and adjuncts of second predication) is restricted by semantic constraints. The semantic scope of these adjuncts – i.e. over a particular element of the VP<sup>5</sup> – demands that they are realized in contact to the element they semantically modify (the contact position can be pre- or postposition).

Since discontinuous constituents are a typical manifestation of free word order, we can summarize that the word order freedom within Bulgarian VP is a result of different information packaging. The constraints on word order, though, come from semantics. This means that whenever adjuncts with narrow semantic scope are realized within VP, their semantics poses restrictions on word order since the adjunct has to be realized in contact to the element of the VP it semantically modifies. The realization of the adjunct in contact to the element it modifies semantically (in pre- or postposition to

<sup>4</sup> This distinction is already known for other languages, but the author is the first one who defines it for Bulgarian.

<sup>5</sup> For this we use the term “narrow semantic scope”.

this element) results in syntactic discontinuity of the phrase.

## 5 Conclusion

In this paper we have reviewed the theories about Bulgarian word order in the limelight of discontinuous constituents. We have shown how discontinuous constituents have been presented within BulTreeBank. We have also pointed out the types of discontinuous constituents and presented our linguistic analysis of the discontinuities, caused by adjuncts. We have described the reasons for linear realization of adjuncts within VP and we have also summarized the factors, which trigger word order freedom and impose word order constraints on the elements of VP, thus pointing out the precise extent of word order freedom in Bulgarian, which had not been studied thoroughly so far.

## References

Tanya Avgustinova. 1997. *Word order and clitics in Bulgarian*. Saarbrücken dissertations in computational linguistics and language technology. Vol. 5. Universität des Saarlandes.

Stefan Brezinski. 1995. *Kratak balgarski sintaksis*. UI „Sv.Kliment Ohridski“. Sofia

BulTreeBank. <http://www.BulTreeBank.org/>

Engdahl and Vallduvi 1994. *Information packaging and grammar architecture: A constraint-based approach*. In E. Engdahl, editor, *Integration Information Structure into Constraint-based and Categorical Approaches*, volume R.1.3.B of DYANA, Edinburgh, 41-79.

Engdahl and Vallduvi 1996. *Information packaging in HPSG*. In Grover and Vallduvi, 1996, 113-128.

Elena Georgieva. 1974. *Slovoed na prostoto izrechenie v balgarskia knizhoven ezik*. Izd.BAN.Sofia

GSBKE, 1983. *Gramatika na savremennia balgarski knizhoven ezik*. 1983 Tom 3. Izd.BAN. Sofia.

Svetomir Ivanchev. 1957. *Nabljudenia varhu upotreбата na chlena*. In: Sv.Ivanchev. *Prinosi v balgarskoto i slavianskoto ezikoznanie*. Sofia. 1978

Jordan Penchev. 1980. *Osnovni intonacionni konturi v balgarskoto izrechenie*. Izd. BAN. Sofia

Jordan Penchev. 1991. *Nekonfiguracionni iavlenia v balgarskia sintaksis*, sp. Balgarski ezik, kn.6, Izd. BAN. Sofia. Bulgaria

Carl Pollard and Ivan Sag. 1994. *Head-Driven Phrase Structure Grammar*. The University of Chicago Press.

Katrin Rudin. 1986. *Aspects of Bulgarian syntax: complementizers and wh-constructions*. Slavica Publishers, Inc.

Kiril Simov and Petya Osenova. 2004. *BTB-TR05: BulTreeBank Stylebook*. <http://www.BulTreeBank.org/TechRep/BTB-TR05.pdf>

Jovka Tisheva. 2003. *Bulgarian yes-no questions with particles nali and nima*. In: *Investigations into Formal Slavic Linguistics. Contributions of the 4<sup>th</sup> European Conference in Formal Description of Slavic Languages*. Peter Kosta, Janna Blaszczyk, Jens Frasek, Ljudmila Geist, Marzena Zygis (eds.). Peter Lang, Europaischer Verlag der Wissenschaften, Frankfurt am Main, 715-729

Jovka Tisheva. 2013. *Pragmaticzni aspekti na ustnata rech*. *Litera et Lingua Series Dissertations*. <http://slav.uni-sofia.bg/naum/liliseries/diss/2013/3>

Jovka Tisheva, Marina Djonova. 2002. *Information structure and clitics in TreeBanks*. In: *Proceedings of the First Workshop on Treebanks and Linguistic Theories (TLT 2002)*. Sozopol. Bulgaria.

Jovka Tisheva, Marina Djonova. 2004a. *Stariat nov topic*. In: *VII Nacionalni slavistichni chetenia*. Sofia.

Jovka Tisheva, Marina Djonova. 2004b. *Za niakoi slovoedni modeli za topicalizacia na razgovornata rech*. In: *Sedma nauchna konferencia po problemite na razgovornata rech*. Veliko Tarnovo.