ПЛОВДИВСКИ УНИВЕРСИТЕТ „ПАИСИЙ ХИЛЕНДАРСКИ“ – БЪЛГАРИЯ
НАУЧНИ ТРУДОВЕ, ТОМ 53, КН. 1, СБ. А, 2015 – ФИЛОЛОГИЯ,
PAISII HILENDARSKI UNIVERSITY OF PLOVDIV – BULGARIA
RESEARCH PAPERS, VOL. 53, BOOK 1, PART A, 2015 – LANGUAGES AND LITERATURE

AUTOMATIC RECOGNITION OF VERB FORMS IN BULGARIAN

Svetlozara Leseva, Ivelina Stoyanova, Svetla Koeva
Institute for Bulgarian Language, Bulgarian Academy of Sciences

The paper presents a pattern-based method for verb form recognition. We discuss the morphological features of Bulgarian verb forms from the point of view of computational linguistics with a focus on analytical forms, their structure, word order and the possibilities for insertion of external elements between the components of the forms. The patterns defining the verb forms use the lemma, the part of speech and grammatical features. Our future work is focused on improving the method, as well as on integrating the verb form recognition in various NLP applications, such as clause splitting, parallel alignment at various linguistic levels, semantic role labelling, etc.

***Key words*:** pattern-based syntactic analysis, verb form recognition

1. Introduction

The analysis of the syntactic structure of a sentence is important for many Natural Language Processing (NLP) tasks, such as information extraction, text to speech systems, word-to-word and phrase-to-phrase alignment for the purposes of machine translation. Practically, it is a prerequisite for the understanding of the meaning of any language unit both by humans and by computers.

Complete syntactic analysis is proved to be more efficient for certain tasks, such as semantic role labelling (Surdeanu, Turmo 2005). The full (and consistent) syntactic parsing of a sentence is a complex task, and to the best of our knowledge, the known parsers do not achieve good precision on an arbitrary text (Oepen et al. 2014). An alternative to full parsing is partial syntactic analysis, also called shallow parsing, or chunking. Various approaches to the task have been developed.

The method presented by Abney (1996) involves finite-state cascades which use a rule-based model for chunking and introduces a limited degree of hierarchy by cascading application. Kermes and Evert (2001) use regular grammars for both partial syntactic analysis and corpora queries which incrementally build flat annotations of syntactic constituents with additional check for agreement and identification of invalid phrases. Grover and Tobin (2006) discuss the development of a rule-based chunker

with a view to sustainability and reusability for new domains and data, as well as the possibility for developing a chunker based on Machine Learning if training data of sufficient quality and quantity are available.

In this paper we describe a pattern matching method for partial syntactic analysis with a focus on verb form recognition. Since the pattern matching operates on lemmas and part-of-speech (POS) tags, a preparatory POS and grammatical annotation is performed. The precise recognition of verb forms can improve the results of various NLP applications, such as clause splitting.

2. Bulgarian Verb Forms

Bulgarian is an analytical language with rich morphology and relatively free word order. These characteristics influence the system of verb forms as well – most of them are analytical and allow word order variations, which poses a challenge for their proper identification and annotation.

2.1. Morphology of Verbs

Table 1 shows the list of categories describing the verb forms and their respective sets of values. Some 3rd person forms differ in structure from the other forms, and/or exhibit different word order variations. These peculiarities are dealt with in the following sections.

Table 1. Grammatical categories of verbs

Category	**Value**	**Example**
Tense (9 members)	Present, Aorist, Imperfect (simple forms)	*чета, четох, четях*
	Perfect	*чел съм*
	Pluperfect	*бях чел*
	Future	*ще чета*
	Future Perfect	*ще съм чел*
	Future in the Past	*щях да чета*
	Future Perfect in the Past	*щях да съм чел*
Polarity (2 members)	**Affirmative**	*чета, щях да чета*
	Negative	*не чета, няма да чета, не ще чета, нямаше да чета*
Voice (2 members)	**Active**	*чета, щях да чета*
	Passive	*четен е*
Mood (3 members)	**Indicative**	*чете, ще чете*
	Conditional	*бих чел*
	Imperative	*чети, да чете, не чети, недей да четеш, недей чете*
Evidentiality (4 members)	**Testimonial**	*чете, ще чете*
	Conclusive	*четял съм, четял е*
	Renarrative	*четял съм, четял*
	Dubitative	*четял съм бил, четял бил*

Several points of clarification are in order with respect to the verb categories that have been subject to different interpretations, particularly Mood and Evidentiality.

Mood denotes the speaker's attitude toward the objective reality of a state of affairs or the possibility for something to happen. The moods, described traditionally for Bulgarian, are the Indicative, the Imperative and the Conditional (Andreychin et al. 1983; Kutsarov 2007; Nitsolova 2008). Other members of the category, such as the Renarrative (Andreychin et al. 1983) and the Conclusive (Kutsarov 2007), as well as more complex mood systems (Nitsolova 2008) have also been proposed.

Evidentiality (*вид на изказването – type of utterance* (Kutsarov 2007), *модус на изказването – mode of utterance* (Gerdzhikov 2003), *хипернаклонение реалис I и реалис II/ – hypermood realis I, II* (Nitsolova 2008) is a category that expresses the attitude of the speaker to the information he/she renders, primarily whether the speaker is the source of information or whether he/she reports someone else's statement (renarrative). It may also deal with whether the statement is based on direct observation or on the basis of the speaker's conclusions (conclusive), and may show the speaker's uncertainty towards the truth of a reported statement (dubitative).

For the purposes of verb form recognition we adopt the widely accepted three-member mood system and the four-member evidentiality system proposed by Gerdzhikov (2003). The latter is based on two distinctive features – renarrativity and subjectivity and includes: the testimonial (unmarked for both), the renarrative (+renarrativity, -subjectivity), the conclusive (-renarrativity, +subjectivity), and the dubitative (+renarrativity, +subjectivity).

With respect to the category of Voice, we adopt the two-member Active – Participial Passive distinction (Kutsarov 2007), excluding *се* passive, as grammatical and lexical reflexivity are hard to distinguish in many cases. The Passive voice forms discussed below follow Nitsolova (2008: 238-239).

2.2. Structure of Bulgarian Verb Forms

In Bulgarian, the simple verb forms are the affirmative indicative active testimonial forms of the Present, the Imperfect and the Aorist, some of the affirmative imperative forms, and some of the affirmative renarrative forms in the 3rd p. sg.

Analytical verb forms make up for the most part of verb forms and include the forms for the rest of the tenses, the Passive voice, the Conditional (the synthetic Conditional is obsolete), the imperative forms with *нека*[1] and *недей*, most of the non-evidential forms, as well as all the nega-

[1] The semantics of the forms with *нека* is not truly imperative.

tive forms[2]. An analytical verb form consists of a head verb, one or more auxiliary verbs and/or particles (*ще, не, да*) and/or the conjunction *да*.

The auxiliaries *съм, бъда*, *ща, бивам*, as well as the head verb agree in number and person with the subject. The aorist and the imperfect active participles and the past passive participle agree with the subject in gender and number. The analytical verb forms headed by a finite verb inherit its person and number. The analytical verb forms headed by a participle inherit the participle's gender and number and the auxiliary verb(s)' person.

A consequence of the relatively free word order in Bulgarian is that many of the analytical verb forms allow regular word-order variations and the possibility for insertion of external elements and whole phrases between their elements. Below we do not aim at providing a full grammatical account but rather at describing the general rules for the word-order and insertion constraints. Exceptions to the proposed rules may be found, especially in older texts and colloquial speech.

2.3. Word Order of Bulgarian Verb Forms

2.3.1. Auxiliary and head verb

When a verb form, built by the auxiliary *съм* and a participle (aorist, imperfect or passive), is at the beginning of the sentence, the auxiliary follows the head verb – e.g. ***ходил съм*** (Perfect), ***четял съм*** (renarrative Present and Imperfect), ***четен е*** (passive Present and Perfect); in all other positions the auxiliary precedes the head verb – e.g. ***съм ходил, съм четял, е четен***.

The Imperfect forms of *съм* (*бях*) may precede the head verb regardless of its position in the sentence or the clause, but inverted forms are also found in poetry, in colloquial speech, etc. (Nitsolova 2008: 292-301) – e.g. ***бях чел, чел бях*** (Pluperfect), ***бе(ше) четен, четен бе(ше)*** (passive Aorist, Imperfect, Pluperfect). The non-evidential forms containing the aorist participle of *съм* (*бил*) – ***бил чел, чел бил*** – exhibit the same behaviour.

The passive forms that comprise the auxiliaries *бъда* or *бивам* (most of them are obsolete) – ***бъда/бивам*** *четен* (Present), *бидох* – ***биде*** *четен* (Aorist), and *бивах* – ***бивах*** *четен* (Imperfect), do not allow inversion.

In the Conditional the auxiliary typically precedes the head verb participle, but the inverted word order – *дошъл бих*, may be found in older texts, in poetry, etc. (Nitsolova 2008: 396).

The auxiliaries usually precede the *да*-complex (the part of the verb form after the conjunction *да* (Avgustinova 1997)), but, although rarely,

[2] The enumerated forms do not represent an exhaustive list. For instance, the synthetic conditional – *ядвам*, marginal or not widely accepted forms have not been included.

may follow it[3]. Such forms are the negative Future and Future Perfect, the Future in the Past and Future Perfect in the Past and the corresponding non-evidential forms, the Imperatives with *нека (да)* and *недей(те) (да)*.

The future and the negative particles *ще, не,* as well as their combination *не ще* precede the rest of the verb form, so that it cannot have an inverted variant – ***ще*** *пиша, *пиша* ***ще***, ***ще*** *съм писал*, **съм писал* ***ще, не ще*** *съм писал, *писал съм* ***не ще***. The restriction also holds when the negative particle appears in the *да*-complex, immediately after the conjunction: *щях да* ***не*** *съм ги писал, *щях да съм ги писал* ***не***.

The particle *да* in the Imperative forms does not allow inversion, too – ***да*** *вървим, *вървим* ***да***.

2.3.2. Combination of auxiliary verbs

The non-evidential forms may have up to three linearly ordered auxiliaries. The forms containing the aorist participle and the Present tense of *съм* may exhibit different word order of the two auxiliaries if they come before the head verb, e.g – ***бил съм/съм бил*** *дошъл*. If the head verb precedes the auxiliaries, their order is to a great extent fixed: *дошъл* ***съм бил,*** *?дошъл* ***бил съм***. The same holds for the respective passive forms: ***бил съм*** *даден,* ***съм бил*** *даден*, *даден* ***съм бил, ?даден бил съм***. The forms marked with ? are strongly marked as colloquial.

The Detached Auxiliary complex (the part of the verb form before the conjunction *да* (Avgustinova 1997)) in the forms containing the aorist participle of *ща – щял,* together with the aorist participle or the Present tense of *съм*, or both, exhibits very free word order with respect to each other – e.g. ***щял съм бил*** *да дойда,* ***бил съм щял*** *да дойда,* ***щял бил съм*** *да дойда,* ***съм бил щял*** *да дойда.* The same holds for the passive.

2.4. Insertion of External Components

At certain positions analytical verb forms allow insertion of external components. There are three types of positions between the components of an analytical form.

The **first type** does not allow any external components. It is the position between the negative particle *не* or the future particle *ще* and the Present tense auxiliary *съм* (except for the 3rd p. sg. form) – ***не съм*** *чел*, ***не съм*** *четял,* ***ще съм*** *чел,* ***ще съм*** *четен*; between *не* and *ще* (rare forms) – ***не ще*** *чете,* ***не ще*** *е чел*; between the aorist active participle of *няма* (*нямало*) and the Present auxiliary *съм* in the negative of some non-evidential forms – ***нямало съм*** *да нося*; between *ще* and *да* in presumptive forms – ***ще да*** *е ходил*.

[3] Inverted forms may be found rarely, in older texts, colloquial speech and in poetry: *да дойде щеше, като се мръкне.*

The **second type of position** is reserved for pronominal clitics (accusative and dative, including the reflexive and the reciprocal pronominal clitics) and/or the question particle *ли*. There are verbs, such as *смея се*, *въобразявам си*, *състезавам се* where the reflexive (reciprocal) particle is part of the lemma, and other verbs, such as *мързи ме*, *хрумне ми*, *гади ми се*, where the pronominal clitic or both the reflexive particle and the pronominal clitic are part of the verb lemma (Koeva 2010). Such verbs may be called grammatical compounds (and viewed as a class of multiword expressions). The automatic grammatical annotation (POS tagging and lemmatisation) provides information for unambiguous grammatical compounds – e.g. *зазорява се*, *домъчнява ми*, *гади ми се*. For the time being, the ambiguous cases remain unresolved – e.g. *целуват се* (reflexive, reciprocal), *момичето* ***се мие*** (reflexive), *боята* ***се мие*** *лесно* (middle), although some combinations of clitics and particles allow disambiguation: *чете ми се* (optative).

A pronominal clitic complex is either a single pronominal clitic or a combination of an accusative and a dative clitic. The Bulgarian pronominal clitics precede the head verb unless that would place them at the beginning of a sentence or a clause. Their order with respect to the verbal clitics (clitic cluster, cf. Avgustinova (1997)) and to one another has been well studied (Hauge 1976, Even 1979, among others). The dative clitic (including the dative reflexive/reciprocal) always precedes the accusative (including the accusative reflexive/reciprocal). This rule is also relevant for the lexicalised pronominal clitics (particles) which are part of the verb's lemma, as well as for the dative ethic (a stylistic usage of the dative or the dative reflexive which may co-occur with the lexicalised and/or the true pronouns).

The particular components between which the position of the pronominal clitics is licensed are discussed below.

1) **the auxiliary *съм* and the head participle.** The position of the pronominal clitics varies with respect to the auxiliary verb – the clitics follow the Present tense *съм* to the exception of the 3rd p. sg. – ***(не) съм*** *му го* ***чел, щях/нямаше да съм*** *му го* ***чел***, ***(не) съм*** *му* ***четен***, and precede the 3rd p. sg. *съм* – ***(не)*** *му го* ***е чел, (не)*** *му* ***е четен***, and all the forms of *бъда* and *бивам*. The Imperfect of the auxiliary *съм* may either precede or come after the clitic complex – ***(не) бе(ше)*** *му го* ***чел, (не)*** *му го* ***бе(ше) чел***.

2) **between the negative particle *не* and the head verb**. These cases encompass the negative forms corresponding to synthetic affirmative forms – e.g. ***не*** *му го* ***чета*** (Present)*,* ***не*** *му го* ***четох*** (Aorist)*,* ***не*** *му го* ***четях*** (Imperfect), ***не*** *му го* ***чети*** (Imperative)*,* ***не*** *му го* ***чел*** (renarrative).

3) **between the particle *не* and the auxiliary verb.** The auxiliary may be the 3rd p. sg. Present tense *съм* – ***не*** *му го* ***е чел, не*** *му* ***е четен***; the Imperfect tense of *съм* – ***не*** *му го* ***бях чел, не*** *му* ***бе(ше) четен***; the auxiliary

бъда – ***не*** *му* ***бъде четен***; the auxiliary *бивам* – ***не*** *му* ***бива четен***; the aorist participle of *съм* – ***не*** *му го* ***бил четял/чел, не*** *му* ***бил четен***.

4) **between the head participle and the auxiliary verb.** The auxiliary verb may be the 3rd p. sg. Present tense *съм* – e.g. ***чел*** *му го* ***е***, ***четен*** *му* ***е***; the Imperfect tense of *съм* – e.g. ***чел*** *му го* ***бях, четен*** *му* ***бе(ше)***, the aorist participle of *съм* – ***чел*** *му го* ***бил***.

5) **between the future particle ще and the head verb** – the Future tense forms – ***ще*** *му го* ***чета***.

6) **between the future particle ще and the auxiliary verb**. The auxiliary may be the 3rd p. sg. Present tense *съм* – e.g. ***ще*** *му го* ***е чел, ще*** *му* ***е четен***; or the auxiliary *бъда* – ***ще*** *му го* ***бъда чел, ще*** *му* ***бъде четен***.

7) **between the conjunction *да* and the head participle**. These forms include the negative forms of the Future active – ***няма да*** *му го* ***чета***, the affirmative and the negative forms of the Future in the Past – ***щях да*** *му го* ***чета, нямаше да*** *му го* ***чета***, the Imperative – ***нека да*** *му го* ***чета, недей да*** *му го* ***четеш***, non-evidential forms – ***щял съм бил да*** *му го* ***чета, нямало съм да*** *му го* ***чета***.

8) **between the conjunction *да* and the auxiliary verb**. The auxiliary may be the 3rd p. sg. Present tense of *съм* and *бъда* (all forms) ***няма да*** *му го* ***е/бъда чел, щеше да*** *му го* ***е/бъде чел, щял/нямало да*** *му го* ***е/бъде чел***, ***няма да*** *му* ***е/бъде четен***.

9) **between two auxiliary forms**. These forms include the non-evidential forms that contain the Present tense and the aorist participle of *съм* – ***(не) съм*** *му го* ***бил чел/четял***.

The question particle *ли* may appear after all components of the verb forms – the head verb, the participle or the auxiliary with the following restrictions (*ли* does not appear after particles – *ще*, *нека*, *не*, *да* and the conjunction *да*):

- *ли* cannot appear between or after the clitic complex immediately followed by a participle – **съм му* ***ли*** *го дал*, **съм му го* ***ли*** *дал*; after the clitic complex immediately followed by the Present tense of *съм* – **не му го* ***ли*** *е дал*; or with the head participle at the first position – **дал съм му го* ***ли***, **дал му го* ***ли*** *е*, **дал съм му* ***ли*** *го*, **дал му го* ***ли*** *е*;
- imperatives do not allow the interrogative particle *ли*;
- in forms built with the conjunction *да*, the interrogative particle *ли* appears either after the first auxiliary or after the whole verb form – *щях* ***ли*** *да му го дам, щях да му го дам* ***ли***;
- after *съм* the interrogative particle *ли* can appear only in negative forms – *не съм* ***ли*** *му го дал*.

The **third type** of position is between the Detached Auxiliary Complex and the *да-complex*. It is relatively flexible and allows insertion of various phrases – e.g. ***няма*** *[AdvP никъде]* ***да ходя***; ***щял бил*** *[PP на сестра си]* ***да купи*** *рокля*; ***нека*** *[NP Иван]* ***да дойде***, including (rarely) entire clauses. Another position may be opened between the Core Clitic Cluster (the Present tense of *съм* and the pronominal clitics (Avgustinova 1997)) and the participle – *съм му го* ***вече*** *дал* (also possible with the Imperfect tense or the aorist participle of *съм – бях/бил му го* ***вече*** *дал*); before the pronominal clitics when the auxiliary is the Imperfect of *съм – бях* ***вече*** *го чел*; between the 3rd p. sg. Present of *съм* or all the forms of *бъда* or *бивам* and the participle – *е/бъде* ***вече*** *чел, бива* ***скоро*** *възнаграден.* The class of components that may fill this position is limited (adverbs).

While the description in the paper offers generalisations about the variants of word order and the possibilities for insertion of external components in the analytical verb form, the patterns suggested do not exhaust explicitly all permissible combinations due to: (a) the possibility to combine word order and insertion, which on the one hand increases the options, and on the other – imposes certain restrictions; (b) the increase of the number of auxiliaries and the addition of *ли* leads to further combinations and restrictions.

Rather than trying to encompass all the possible cases (for many of which examples are not readily available), we adopt a different approach. As for a given position the maximum number of components is limited, as is the set of classes to which they belong, we define the positions at which word order variation and insertions take place, the classes of elements that may be inserted and the maximum number of these elements. This is done for two reasons. First – the number of patterns will be much larger if all combinations are formulated. Second – patterns match grammatically correct forms.

3. Verb Form Recognition

Here we present a pattern-based method for verb form recognition for the purposes of clause splitting and other NLP applications. The patterns rely on lemma and POS to match the components of the verb forms and possible external elements between the components. Verb form recognition is performed after sentence splitting and POS tagging.

Preprocessing is carried out using the Bulgarian Language Processing Chain (Koeva, Genov 2011) which includes the following tools: sentence splitter, tokeniser, POS tagger, and lemmatiser. Some verbal grammatical compounds are also identified by the lemmatiser.

3.1. Formal Description of Verb Forms

We have developed manually a set of patterns that recognise analytical verb forms and tag them with grammatical features such as tense,

mood, voice (as described in section 2.1). The patterns are devised in such a way as to recognise basic word order variants, such as auxiliary – head verb, head verb – auxiliary, negativity marker – auxiliary – head verb, etc., and to identify classes of inserted elements as external to the verb form (2.2.-.2.3.). The word order of the inserted elements is not strictly defined, thus enabling the recognition not only of more frequent, but also of rarer, substandard or old variants. At present, 834 patterns corresponding to 84 different combinations of grammatical features have been defined for the Bulgarian verb forms.

Example 1 presents a sample of patterns with their formal description and usage illustrations. Each component can be defined in one of the following ways: (i) by their lemma (e.g., negation particle *не*); (ii) by their lemma and grammatical constraints for the particular form (e.g., auxiliary *съм:Vr1* in 1st person, present tense); (iii) by their POS with or without additional grammatical constraints (e.g., *Vqo* non-definite form of the past passive participle of a verb); or (iv) by an exclusion category list (e.g., *{-V,-U,-C}* – any POS other than a verb, a punctuation mark or a conjunction.

The possible insertions of external elements in certain positions are in brackets, noting the maximum allowed number (or 'n' for any number) and a list of possible elements defined by a lemma or a POS with or without additional grammatical constraints or by an exclusion category list.

Example 1. Sample of Verb Form Patterns. The components of the verb form are in bold; the external elements intervening the verb form elements are underlined. Default values for each category (bold in Table 1) are not explicitly listed, e.g. Present = Present, Affirmative, Active voice, Indicative mood, Testimonial.

Grammatical Features	**Pattern**	**Example**
Present	Vr	*Мисля да му* ***дам*** *книгата*.
Present+Neg	не (4,{ли,се,си,P}) Vr	***Не*** *му ли я* ***даде****?*
Present+Pass	съм:Vr1 (3,{се,си,P}) Vqo	*Аз* ***съм*** *си му* ***дадена*** *вече за жена.*
Present+Pass	съм:Vr3 Vqo	*Книгата му* ***е дадена***.
Present+Pass+Neg	не съм:Vr1 (3,{се,си,P,ли}) Vqo	***Не съм*** *ли му* ***дадена*** *вече за жена?*
Present+Pass+Neg	не (4,{се,си,P,ли}) съм:Vr3 Vqo	***Не*** *му ли* ***е дадена*** *вече за жена?*
Future_Perf_in_the_Past	ща:Vd (n,{-V,-U,-C}) да съм:Vr (3,{се,си,P}) Vxo	***Щях*** *ли толкова бързо* ***да съм*** *я* ***прочел?***

3.2. Recognition of Verb Forms and Verb Chunks

The method presented here builds on a previous rule-based approach for the recognition of verb forms (Stoyanova, Leseva, Koeva 2013; Lozanova et al. 2013). The verb form recognition takes as input a tagged sentence (POS, lemma, grammatical information) and applies a set of syntactic patterns to identify all verb forms in the sentence. In the pattern recognition, priority is given to the longest match. External elements are clearly identified as such and are not considered as part of the verb form.

Some of the verb forms are ambiguous as they can be matched by more than one pattern, e.g. certain renarrative forms for the Aorist coincide with the non-evidential forms for Present Perfect. As our purpose at present requires only recognition of verb forms, disambiguation is among our future tasks.

4. Conclusions

The proposed pattern-based method for partial syntactic analysis is general and to a large extent applicable for different languages since it relies on a set of patterns for the identification of verb forms. The identification of verbal multiword expressions will additionally help to split them between syntactic chunks or clauses.

Our future work will be focused, on the one hand, on improving the methods for verb form recognition and clause splitting, and on the other – on integrating the methods in advanced NLP applications, such as anaphora resolution, semantic role labelling, parallel text alignment at phrase and word level, and semantic disambiguation.

REFERENCES

Abney 1996: Abney, S. 1996. Partial Parsing via Finite-State Cascades. // *Natural Language Engineering*, 1 (1), 1995, Cambridge University Press.

Andreychin et al. 1983: Andreychin, L, P. Asenova, E. Georgieva, K. Ivanova, R. Nitsolova, P. Pashov, H. Parvev, R. Rusinov, V. Stankov, S. Stoyanov, K. Cholakova. *Gramatika na savremenniya balgarski knizhoven ezik*. Vol. 2. Morphologiya. Sofia: BAS Press, 1983.

Avgustinova 1997: Avgustinova, T. *Word Order and Clitics in Bulgarian.* Saarbrücken Dissertations in Computational Linguistics and Language Technology, Volume 5. Saarbrücken: DFKI, 1997.

Gerdzhikov 2003: Gerdzhikov, Georgi. *Preizkazvaneto na glagolnoto deystvie v balgarskiya ezik* (in Bulgarian), Sofia: St. Kliment Ohridksi, 2003, ISBN 954-07-1834-1.

Grover, Tobin 2006: Claire Grover and Richard Tobin. Rule-Based Chunking and Reusability. // *Proceedings of the Fifth International*

Conference on Language Resources and Evaluation (LREC 2006). Genoa, Italy, 2006.

Even 1979: Even, R. *A Grammar of Bulgarian Clitics.* Doctoral Dissertation, University of Washington, 1979.

Hauge 1976: Hauge, K. R. *The Word Order of Predicate Clitics in Bulgarian.* Meddelelser No. 10. University of Oslo, 1976.

Kermes, Evert 2001: Hannah Kermes and Stefan Evert. Exploiting large corpora: A circular process of partial syntactic analysis, corpus query and extraction of lexicographic information. // *Proceedings of the CL2001 Conference*, Lancaster University, 2001.

Koeva 2010: Koeva, S. *Balgarskiyat FrameNet 2010.* Sofia, 2010.

Koeva, Genov 2011: Koeva, Sv. And A. Genov. Bulgarian Language Processing Chain. // *Proceeding to The Integration of multilingual resources and tools in Web applications Workshop in conjunction with GSCL 2011*, University of Hamburg, 2011.

Kutsarov 2007: Kutsarov, Georgi. *Теоретична граматика на българския език. Морфология* (in Bulgarian). Plovdiv: Paisiy Hilendarski.

Lozanova et al. 2013: Lozanova, Slavina, Ivelina Stoyanova, Svetlozara Leseva, Svetla Koeva, Boian Savtchev. Text modification for Bulgarian Sign Language users. // *Proceedings on Predicting and Improving Text Readability for Target Reader Populations.* Stroudsburg, PA: The Association for Computational Linguistics, 2013, 39–48.

Nitsolova 2008: Nitsolova, Ruselina. *Balgarska gramatika. Morfologiya* (in Bulgarian), Sofia: St. Kliment Ohridksi, 2008.

Oepen et al. 2014: Oepen, S., M. Kuhlmann, Y. Miyao, D. Zeman, D. Flickinger, J. Hajic, A. Ivanova, and Y. Zhang. SemEval 2014 Task 8: Broad-Coverage Semantic Dependency Parsing. In *Proceedings of SemEval*, 2014, 63–72.

Puskasu 2004: Puscasu, G. A Multilingual Method for Clause Splitting. // *Proceedings of the 7th Annual Colloquium for the UK Special Interest Group for Computational Linguistics (CLUK 2004),* Birmingham, United Kingdom, 2004, 199–206.

Stoyanova, Leseva, Koeva 2013: Stoyanova, I., S. Leseva, S. Koeva. Rule-based Clause Splitting for Bulgarian and English. // *Proceedings of the 32nd International Conference on Lexicon Grammar*. Faro: Universidade do Algarve, 2013.

Surdeanu, Turmo 2005: Surdeanu, M., J. Turmo. Semantic role labeling using complete syntactic analysis. // *Proceedings of the Ninth Conference on Computational Natural Language Learning* (CONLL '05). Stroudsburg, PA, USA: Association for Computational Linguistics, 221–224.