Резюмета на английски език на публикациите на гл.ас. д-р Цветана Иванова Димитрова, представени за участие в конкурс за заемане на академичната длъжност „доцент“ по професионално направление 2.1. Филология, научна специалност „Общо и сравнително езикознание“, за нуждите на Секцията по компютърна лингвистика, по обява в ДВ, брой 102 от 23.12.2022 г.

Book

1. Димитрова, Цв. Българските клитики: история и настояще. София: Институт за български език „Проф. Любомир Андрейчин“ при БАН, 2023. 256 с. ISBN 978-619-908-845-6

The monograph which consists of a preface and six chapters (organised in three parts), traces the development of the so-called (paradigmatic) clitics and their change in the history of the Bulgarian language. The discussion in 1.1. limits the object of the study as outlined in 1.2. on the basis of language data excerpted for observations on the behaviour and placement of the Bulgarian clitics. In 1.3., the terms used were refined, while section 1.4. gives an overview of the literature on clitics in Bulgarian to back for the need for a more comprehensive study.

The first part of the monograph "The features of clitics", critically examines the features and character of the clitics with a focus on two main functional characteristics of clitics that are interrelated, namely their defectiveness and their dependence. The discussion unfolds in the context of two linguistic 'laws' that were formulated in the late 19th century - Wackernagel's Law and Tobler-Mussafia's Law - whose effect can be traced in modern languages (including Bulgarian). The language data is extracted from text collections and corpora with data on the Bulgarian language. In the first part, the focus is mainly on some features and behaviour of the clitics in the modern Bulgarian language at focus that can be traced in non-standardised Bulgarian language fragments - literary archaic sources that are part of the Bulgarian National Corpus, transcribed records in the Corpus with transcribed colloquial speech, transcribed fragments of dialectal speech that are found in the interactive database *Bulgarian Dialectology as a Living Tradition*. The discussion aims at putting in context the analysis in the second part of the study, which examines the evolution of the clitics in the history of the Bulgarian language. The first part gives the context of the discussion on the features of the clitics (their position and behaviour in relation to other adjacent and dependent elements) to be traced in the historical language data in order to argue for the unity of the Bulgarian language in the different stages of its life.

The second part of the monograph "Clitics in the History of the Bulgarian Language" describes the path of the clitic elements throughout the history of the Bulgarian language. A sample collection of language data was extracted from historical sources, which, supposedly, reflect the developments in the three stages of the history of the Bulgarian language - the Old Bulgarian (reflected in the data excerpted from the Codex Marianus (analysed in comparison with the examples from the Codex Zographensis, Codex Assemanius and Codex Sabbae) and Codex Suprasliensis, which are part of the TOROT corpus); the Middle Bulgarian (data excerpted from the Legend of Troy, and texts of Vlach-Bulgarian charters from the edition of

Ioan Bogdan; for the purpose of comparison, data from Manassius Chronicle have been excerpted - all these texts are part of the Diachronic Corpus of the Bulgarian language); and Early Modern Bulgarian (the texts of two damaskins - Trojanski and Loveški, which can be found in the Diachronic Corpus of the Bulgarian language, as well). The elements under discussion are grouped into three classes. Chapter 3 reviews the data on the so-called discourse clitics же and бо and the interrogative clitic ли. Although the focus of the study is on the so-called paradigmatic clitics which have survived (and/or have evolved into clitic elements) in modern Bulgarian, the discourse clitics were the earliest clitic elements that have correspondences in Greek texts - their placement might have indicated the placement of the (other) clitic elements in the sentence and/or the phrase. The next class are the pronominal clitics: the review in chapter 4 starts from the earliest pronominal clitics - the first- and second-person singular dative clitics ми and ти, and moves through the remaining forms of the short dative pronouns to end with an overview of the dative clitics in the Middle Bulgarian and Early Modern Bulgarian. The survey of dative clitics in 4.1 is followed by an analysis of the placement and behaviour of the short first-person and second-person direct object pronouns in 4.2 and the short clitic forms of the anaphoric pronoun in 4.3. Elements of both subclasses did not function as true clitics in earlier texts, but still showed a tendency to appear in positions expected from the clitic elements (hence, they are also known as 'semi(en)clitics' in the literature), and we trace the further development of accusative clitics in examples from the Middle Bulgarian and Early Modern Bulgarian texts. The reflexive accusative and dative clitics are discussed in 4.4.: in the earliest texts they also tended to be semiclitics, i.e. to occur in positions uncharacteristic of true clitics. Chapter 5 reviews the developments in the forms of auxiliary verbs that showed ambiguous behaviour both in the earliest texts and in later ones - hence, they are named 'quasiclitics'. The review of the behaviour and placement of all these elements begins with the examples from the Old Bulgarian texts, proceeds through the examples from Middle Bulgarian, and ends with the examples from the Early Modern Bulgarian texts, which to a large extent also reflect the contemporary situation, with some notable exceptions. The positions of the clitic elements are examined in relation to their surroundings and their governing element (often the verb whose arguments they are) and to the left clausal boundary, within the clitic complex, and in relation to the constituents of complex verbal forms (hence, the positions of the auxiliary verbs in relation to other clitics).

The most notable development based on the observations on the data is the change in the inventory of the elements, i.e. some elements such as discourse clitics completely disappeared (both as clitics and as affixes in present-day Bulgarian), others such as some pronominal and verbal clitics emerge; however, the positions of the clitics in the sentence and/or relative to the verb that governs most of them, have remained relatively stable. The changes observed are related to changes in some functional elements (e.g. negation) and verb forms. This stability of the syntactic structure, however, also includes the variability of the positions of the clitics as observed in all data and periods of the language development, i.e., in addition to the prevailing placement patterns, there are also a number of deviations from them as observed throughout the data.

The third part of the study "Analysis" in chapter 6, attempts to model the variability of clitic positions within the phase theory in the framework of minimalist syntax. A brief overview of

the generative grammar tradition is given, with a review of two hypotheses (by Steven Franks and by Roumyana Pancheva) about the derivation of the positions of clitics in present-day Bulgarian and in the history of the Bulgarian language. The study concludes with a hypothesis about the derivation of clitics in different positions, suggesting that clitics may be dependent on the verb whose argument requirements they may satisfy (such as accusative pronominal clitics), but also may be derived in the sentence (such dative ethical and possessive clitics in the present-day Bulgarian).

The monograph ends with a conclusion in chapter 7 and an Appendix with quantitative data based on the collection of extracted examples, which are freely available online. The data can be used for verification, correction, and other research purposes.

Language resources, formal description of language for natural language processing

2. Dimitrova, T., E. Tarpomanova, B. Rizov. Coping with Derivation in the Bulgarian Wordnet. – Proceedings of the Seventh Global WordNet Conference. Tartu: University of Tartu Press, 2014, pp. 109–117.
https://aclanthology.org/W14-0115.pdf
https://www.scopus.com/record/display.uri?eid=2-s2.0-84904648388&origin=resultslist&sort=plf-f

The article in 2 presents an overview of a strategy for identification and annotation of derivational relations in the Bulgarian wordnet that aims at coping with the complex morphology of the language in an elegant way. Our method involves transfer of the Princeton WordNet (morpho)semantic relations into the Bulgarian wordnet, at the level of the synset, and further detection of derivational relations between literals in Bulgarian. Derivational relations have been annotated to reflect the complexity of Bulgarian morphology. Introduced literal relations improve the consistency and employability of the wordnet. Tsvetana Dimitrova's main contribution is in section 6. Conventions for Annotation of Derivational Relations in BulNet.

3. Rizov, B., T. Dimitrova. Hydra for Web: A Browser for Easy Access to Wordnets. – In: Proceedings of the Eighth Global Wordnet Conference, Research Institute for Artificial Intelligence – Romanian Academy, 2016, pp. 339–343.
https://aclanthology.org/2016.gwc-1.48.pdf
https://www.scopus.com/record/display.uri?eid=2-s2.0-84962811912&origin=resultslist&sort=plf-f

4. Rizov, B., T. Dimitrova, V. Barbu Mititelu. Hydra for Web: A Multilingual Wordnet Viewer. – In: Proceedings of the 11th International Conference Linguistics Resources and Tools for Processing the Romanian Language (ConsILR-2015). Faculty for Computer Science, Alexandru Ioan Cuza University, Iași, Romania, 2015, pp. 19–30.
http://consilr.info.uaic.ro/2015/Consilr_2015.pdf#page=27

https://www.webofscience.com/wos/woscc/full-record/WOS:000461127000003

5. Rizov, B., T. Dimitrova. Online Editor for WordNets. – In: Proceedings of the Third International Conference Computational Linguistics in Bulgaria (CLIB 2018). Sofia: The Institute for Bulgarian Language, 2018, 146–152.
http://dcl.bas.bg/clib/wp-content/uploads/2018/06/CLIB_2018_Proceedings_v1_final.pdf#page=156
https://www.webofscience.com/wos/woscc/full-record/WOS:000455885400018

Articles in 3, 4 and 5 presents Hydra for Web - a web interface for working with wordnets (and lexical-semantic databases with similar relational structure) which is built on top of Hydra – an open source tool for wordnet development. The tools were built and developed by Borislav Rizov. It is a Single Page Application with a simple but powerful and convenient GUI. It has two modes for visualisation of the language correspondences of searched (and found) wordnet synsets – single and parallel modes. It is used for work on databases with relational structure (including query, creation, change, and linking of relational objects), simultaneous access of multiple user profiles, parallel data visualization and editing of the data on top of single- and parallel mode visualization of the language data. Hydra for web is available at: http://dcl.bas.bg/bulnet/. The main focus of the papers is on the user interface, its functionalities and some issues regarding the ongoing process of adapting the resources that are currently visualised (Princeton WordNet for English, Bulgarian WordNet, Romanian WordNet and more than 20 wordnets of other languages). The candidate's main contribution is in the description of the work with the system.

6. Dimitrova, T., V. Stefanova. Adjectives in WordNet: Semantic Issues. – In: Proceedings of the 12th International Conference Linguistics Resources and Tools for Processing the Romanian Language (ConsILR-2016). Faculty for Computer Science, Alexandru Ioan Cuza University, Iași, 2016, pp. 131–141.
https://www.researchgate.net/publication/310613847_Adjectives_in_Wordnet_Semantic_Issues
https://www.webofscience.com/wos/woscc/full-record/WOS:000461127100014

7. Stefanova, V., T. Dimitrova. Classification of Adjectives in BulNet: Notes on an Effort. – In: Proceedings of the LDK 2017 Workshops: 1st Workshop on the OntoLex Model (OntoLex-2017), Shared Task on Translation Inference Across Dictionaries & Challenges for Wordnets co-located with 1st Conference on Language, Data and Knowledge (LDK 2017), 1899, CEUR-WS, 2017, pp. 188–196.
https://ceur-ws.org/Vol-1899/CfWNs_2017_proc7-paper_2.pdf
https://www.scopus.com/record/display.uri?eid=2-s2.0-85029188425&origin=resultslist&sort=plf-f&src=s&sid=3d2073faa936b694604aa8486288ec6a&sot=autdocs&sdt=autdocs&sl=18&s=AU-ID%2856288017300%29&relpos=2&citeCnt=1&searchTerm=

8. Dimitrova, T., V. Stefanova. Semantic Classification of Adjectives in the Bulgarian Wordnet: Toward a Multiclass Approach. – Etudes Cognitives 18 (2018): 1-17.
https://ispan.waw.pl/journals/index.php/cs-ec/article/view/cs.1709/4456
https://www.scopus.com/record/display.uri?eid=2-s2.0-85060784706&origin=resultslist&sort=plf-f

The articles in 6, 7 and 8 present the work on classification of adjectives in WordNet, which was applied to the Bulgarian WordNet in order to enrich it with information about the concepts (reflected in the synsets) and the relations between them.

The paper in 6 critically reviews the organization of the information on adjectives in WordNet, including the principles used to incorporate them into the structure of WordNet and to link them (through various relations) to concepts that are classified in other parts-of-speech (nouns, verbs and adverbs). Classifications of adjectives in other wordnets and in the scholarly literature on Bulgarian language are also taken into account.

The paper in 7 presents a semantic classification of adjectives applied to the Bulgarian WordNet, which is based both on an already existing classification of adjectives in the German WordNet (GermaNet) and research on adjectives in the Bulgarian language. Although developed for the Bulgarian WordNet, the classification can be applied to other lexical-semantic networks in which the information is organized parallelly to or similarly to the Princeton WordNet. The classification relies on information that is already available in WordNet, either within the definitions and usage examples, or (especially) the relations to other synsets (of nouns, verbs and other adjectives) that are related to the adjective by lexico-semantic relations (e.g. the semantic classes of nouns and verbs the adjective is related to). 17 semantic classes have been formulated and further assigned to 2500 adjectives.

The study in 8 presents a classification model in which more than one semantic class can be assigned to a single adjective. The work builds on the one presented in the paper in 7 (semantic classes were assigned to 3110 adjectives, with some adjectives categorised in two or more semantic classes, one of which specifies the potential referent of the adjective (human behavior, animal behavior, human/animal/plant body, etc.).

The work on the classification of adjectives was part of the project "Semantic Classification of Adjectives in the Bulgarian WardNet", where Tsvetana Dimitrova was supervisor of the principal investigator Assist. Prof. Valentina Stefanova. The candidate's contributions are within the description of the data and the justification of the multiclass classification (paper in 8).

9. Димитрова, Ц. Морфосемантични релации и агентивни съществителни в Българския уърднет. – Български език, 65 (2018), 2: 41–58.
https://www.balgarskiezik.eu/2-2018/3_Tsvetana%20Dimitrova_str.%2041-58_bg.pdf

10. Dimitrova, T., V. Stefanova. On Hidden Semantic Relations between Nouns in WordNet. – In: Proceedings of the Tenth Global WordNet Conference (July 23–27, 2019, Wrocław (Poland). Wroclaw: Oficyna Wydawnicza Politechniki Wrocławskiej, 2019, pp. 54–63
https://aclanthology.org/2019.gwc-1.pdf#page=62

https://www.scopus.com/record/display.uri?eid=2-s2.0-85082459370&origin=resultslist&sort=plf-f

11. Стефанова, В., Цв. Димитрова. За причастията в Българския уърднет. – В: Доклади от Международната годишна конференция на Института за български език „Проф. Любомир Андрейчин" (София, 2020). Ваня Мичева, Диана Благоева, Марияна Витанова, Марияна Цибранска, Сия Колковска, Татяна Александрова (отг. ред.), 2, София: Издателство на БАН „Проф. „Марин Дринов", 2020, с. 224–232.
http://ibl.bas.bg/wp-content/uploads/2020/07/PROCEEDINGS_IBLCONF2020_vol2.pdf
https://ispan.waw.pl/ireteslaw/bitstream/handle/20.500.12528/1540/PROCEEDINGS_IBLCONF2020_vol2.pdf?sequence=1&isAllowed=y

12. Dimitrova, T. On WordNet Semantic Classes: Is the Sum Always Bigger? – In: Proceedings of the Fourth International Conference "Computational Linguistics in Bulgaria" (CLIB 2020). Sofia: Institute for Bulgarian Language – Bulgarian Academy of Sciences, 2020, pp. 177–185
https://aclanthology.org/2020.clib-1.19.pdf
https://www.scopus.com/record/display.uri?eid=2-s2.0-85113143384&origin=resultslist&sort=plf-f

Papers in 9, 10, 11 and 12 present work on formulating and assigning additional semantic (and morpho-semantic) relations between verbs and nouns and between nouns, as well as additional semantic classes to nouns in WordNet. The work is funded under the project "Towards a Semantic Network Enriched with a Variety of Relations" (DN 10-3 / 14.12.2016), financed by the Bulgarian National Science Fund (BNSF). The contribution of Tsvetana Dimitrova is within the methods to extending noun-noun relations and semantic classes of nouns.

The article in 9 describes the effort on expanding the set of morpho-semantic relations that can be assigned to noun-verb pairs linked through Agent and Undergoer relations in the Bulgarian wordnet. We take into account the following information that has been already part of the WordNet: verb and noun semantic classes, sentence frames that outline the predicate-argument structure of the simple sentences that verbs can form (by encoding information about the number of the arguments, their referent - person or animal/object, prepositional phrase, preposition, etc.). The work is based on the assumption that derivation of deverbal nouns may be determined by the thematic structure of the predicate. Seven additional relations formulated according to the thematic roles of the arguments (according to some theories of the predicate-argument structure of verbs of certain semantic classes) were manually assigned. The most productive derivative suffixes in nouns associated with the Agent relation are markedly agentive, while nouns with the additional roles Experiencer and Patience are often substantivized participles.

The article in 10 presents a strategy on transferability of noun – verb and noun – adjective derivative and semantic relations to noun – noun relations. The approach relies on information from semantic classes and existing inter-POS derivative and (morpho)semantic relations between noun and verb, and noun and adjective synsets. Morpho-semantic relations

were added between nouns in WordNet that are indirectly linked via verbs and adjectives. Observations on the combination between the relations and semantic classes of nouns they link, may facilitate further efforts in assigning semantic properties to nouns pointing to their abilities to participate in predicate-argument structures.

The article in 11 reviews the ways in which the participles are included in the adjective synonym sets in Wordnet (drawing upon material from the Bulgarian Wordnet). Synonym sets classified as participles are very few in WordNet but other adjective synonym sets contain many literals in the form of participles. Therefore, we propose a way to identify participles in the synonym sets that are not classified as participles by identifying adjective synonym sets that are related – directly or indirectly (through other adjectives and nouns) – with verb synonym sets.

The article in 12 offers an approach to the validation of the data resulted from a previous effort on expansion of WordNet noun semantic classes by mapping them with the semantic types within the Corpus Pattern Analysis (CPA) ontology employed by the framework of the Pattern Dictionary of English Verbs (PDEV). A case study is presented along with a set of conditions to be checked when validating the combined data.

13. Търпоманова, Ек., Цв. Димитрова. Българско-английският паралелен корпус със съотнесени (прости) изречения (BulEnAC). – В: Езикови ресурси и технологии за български език. София: Академично издателство „Проф. Марин Дринов", 2014, 21, с. 105–126.

The paper presents the Bulgarian-English Sentence- and Clause-Aligned Corpus (BulEnAC) that is an excerpt from the Bulgarian-English Parallel Corpus – a part of the Bulgarian National Corpus (BulNC). Both the Bulgarian and the English parts of the corpus were automatically sentence-split and sentence-aligned, with splitting and alignment manually verified by experts. The clauses in English were splitted using an OpenNLP parser, and the annotation was manually post-edited. The Bulgarian sentences were split into clauses manually. The clause-introducing conjunctions were identified, with the type of relation, the type of the clauses involved in the relation, as well as the direction of the relation. Finally, the parallel clauses within corresponding sentence pairs have been manually aligned. The paper discusses the conventions for sentence and clause splitting and annotation. The annotation is in compliance with the specific syntactic rules and the grammar tradition and annotation practices for the respective languages.

14. Коева, Св., Цв. Димитрова. Особености на семантичната и синтактичната структура на собствените имена за лица, географски обекти и организации в български език. – Известия на Института за български език „Проф. Любомир Андрейчин". София: Издателство на БАН „Проф. Марин Дринов", 2017, с. 304–344.
http://ibl.bas.bg/izvestiya/wp-content/uploads/2018/08/IZVESTIA-IBE-XXX-304-344-Koeva-Dimitrova.pdf

The study in 14 offers a model for description of three classes of named entities in Bulgarian – persons, locations, and organizations – which can be expressed in single words, multiword

expressions, and noun phrases. We offer a structural and semantic classification of the named entities and a language-independent model for description of the semantic and syntactic structure of proper names according to their class. Further, we present the major types of semantic structures for persons, locations, and organizations and a general description of the corresponding syntactic structures in Bulgarian.

15. Димитрова, Ц., В. Стефанова, М. Тодорова. Предикативни конструкции за състояние: корпусно изследване върху български и руски. – В: Онтология на ситуациите за състояние – лингвистично моделиране. Съпоставително изследване за български и руски. София: Академично издателство „Проф. Марин Дринов“, 2022, с. 358–384
http://dcl.bas.bg/wp-content/uploads/2022/11/12_DimitrovaStefanovaTodorova_STONTBgRu2022.pdf

The study in 15 discusses predicative constructions expressing state semantics in Bulgarian and Russian. The observations, analyses and generalizations are made on the basis of a collection of predicative constructions describing states in Bulgarian and Russian excerpted from the parallel Bulgarian-Russian and Russian-Bulgarian subcorpora of the Russian National Corpus. The contribution of Tsvetana Dimitrova is with the excerption of the data from the parallel corpora (parallel predicative constructions) and their descriptions for the analysis and conclusions.

Historical corpora, description of historical language data

16. Dimitrova, T., A. Bojadziev. Historical Corpora of Bulgarian Language and Second Position Markers. – In: Proceedings of the First International Conference Computational Linguistics in Bulgaria. Sofia: Institute for Bulgarian Language, 2014, pp. 55-63
http://dcl.bas.bg/clib/wp-content/uploads/2014/07/CLIB2014_Proceedings.pdf

This paper demonstrates how historical corpora can be used in researching language phenomena. It exemplifies advantages and disadvantages through exploring three of the available corpora that contain textual sources of Old and Middle Bulgarian language to shed light on some aspects of the development of two words of ambiguous class. We discuss their behaviour to outline certain conditions for diachronic change that the two words have undergone. The three corpora presented in the paper are freely accessible online (and offline – for downloading search results, xml files, etc.). Tsvetana Dimitrova's contribution is in the excerption, description of the work with corpora and analysis of the data in section 4.

17. Димитрова, Цв. За втората позиция на (местоименните) клитики в български. – В: Българистични езиковедски четения. Материали от международната конференция по случай 130-годишнината на Софийския университет „Св. Климент Охридски“, 19–20 ноември 2018 г. София: Университетско издателство „Св. Климент Охридски“, 2019, с. с. 310–315.

The article in 17 discusses the position of pronominal (sentential) clitics in Bulgarian based on data from two Early Modern Bulgarian sources (damaskins) and a set of constructions excerpted from Modern Bulgarian corpora. The observations are limited to the short forms of personal pronouns and their positions in relation to other (sentential) constituents in an attempt to argue for a hypothesis that in Bulgarian clitics can occupy two positions - second or preverbal, and the assumption that the so-called "second position" is active while being inherited historically.

18. Димитрова, Цв. Отново за историческия развой на конструкциите за бъдеще време – по корпусни данни. – Съпоставително езикознание 2 (2020): 61–78.

The article in 18 reviews the constructions for expressing future action which developed into analytic future tense forms in the modern Bulgarian literary language, with an emphasis on the Middle Bulgarian and Early Modern Bulgarian periods (based on data from the Vatican copy of the Legend of Troy, ca. 14th century, texts of Vlach-Bulgarian charters, ca. 15th-16th century; and Trojanski and Loveški damaskins from the 17th century). The greatest variety of constructions is observed in the Vlach-Bulgarian charters (including the infinitive in competition with shortened infinitive). In this period, da-constructions emerged, although with very few (compared to other constructions) occurrences, but in the 17th century they prevailed. The shortening of хощѫ to щѫ, which began very early, is reflected in the apparent competition between the two in the Legend of Troya, with the Vlacho-Bulgarian charters showing a preference to щѫ.

19. Dimitrova, T. On the diachrony of the clitic cluster in Bulgarian. – Balcania et Slavia 1 (2021), 1: 67–96
https://edizionicafoscari.unive.it/media/pdf/article/balcania-et-slavia/2021/1/art-10.30687-BES-0-2021-01-003_X8Y30SK.pdf

The article in 19 traces back the formation of the clitic cluster in Bulgarian starting from the Old Church Slavonic through Middle Bulgarian up to the Early Modern Bulgarian and beyond. It offers a hypothetical two-layer structure of the cluster – with the main layer consisting of a (pronominal) core and a (verbal) periphery, and a secondary layer hosting ('quasi-clitical') elements that exhibit, both diachronically and synchronically, a behaviour that is not strictly consistent with that of the clitical elements. The language material from three corpora shows that there was no change in the positions of the elements in the core, while the observed changes in the periphery are mainly due to changes in the set of the elements (as a result of the restructuring of the pronoun system and changes in the auxiliary system, as well as the loss of some early clitics, such as the discourse markers).

20. Krapova, I., T. Dimitrova. The Genitive-Dative syncretism in the history of Bulgarian. Towards an analysis. – Studi Slavistici XII (2015): 181–208.
https://oaj.fupress.net/index.php/ss/article/view/2353/2353
https://www.scopus.com/record/display.uri?eid=2-s2.0-

In this article, we trace the diachronic phases of so-called genitive-dative syncretism in Old Bulgarian, a phenomenon which marks the beginning of the process of disintegration of the Case system in the history of Bulgarian. We base our research on a corpus study (comprising the texts of Codex Marianus, Codex Zographensis and Codex Suprasliensis) and we show, after a careful examination of the available syntactic positions for the Genitive and the Dative, that the main reason for the Dative shift in Old Bulgarian was the so-called External Possessive construction (also known as dativus sympatheticus) which allowed the dative to be interpreted outside the noun phrase but with reference to an inalienable possessive element inside the noun phrase. We discuss each stage of the functional reanalysis leading to the establishment of noun phrase internal dative possessives, and in particular the role of the emerging clitic pronouns and their second position syntactic behaviour as a main trigger of genitive-dative syncretism.

21. Иванова, Е., Цв. Димитрова. Союз макар в болгарских памятниках XV-XVII вв. – Studi Slavistici XVII (2020), 1: 15–33.
https://oaj.fupress.net/index.php/ss/article/download/7496/8028/
https://www.scopus.com/record/display.uri?eid=2-s2.0-85095784693&origin=resultslist&sort=plf-f

The article discusses the semantics and structure of the sentences introduced by the conjunction макар (although) drawing upon data excerpted from Bulgarian monuments dated between the 15th and 17th centuries. In this early period, макар was a newly introduced conjunction which later extended its use to become a widely used concessive conjunction in present-day Bulgarian. The study is based on data from a sample of texts that are part of the database of the Historical Corpus of the Bulgarian Language which can be consulted online (<http://histdict.uni-sofia.bg/textcorpus/list>). In the database, the conjunction макар is found in the Vlach-Bulgarian Charters and in two Early Modern Bulgarian damaskins (Trojanski and Loveški) as an independent conjunction and in the compound subordinating conjunctions макар (и) да, макар че, макар (и) ако, макар и като introducing predicative and non-predicative segments.

The most frequent variant of the compound conjunction formed with the particle да had the same syntactic and grammatical properties as the conjunction макар (и) да in present-day Bulgarian, especially with regards to the contact position of да and the verb. By the time this conjunction was established, the formation of да-constructions was already clear: the verb immediately followed да (the two were separated mostly by pronominal clitics and negation as in present-day Bulgarian), while other clausal constituents were placed between макар (и) and да (in earlier texts, да and the verb could be separated by other clausal constituents as well).

The study also deals with the semantics of the concessive макар-conjunctions. The data shows that in the period considered, the expression of concession was still at its beginning. Therefore, clauses introduced by макар could express not only the prototypical 'causal concession' as in the present-day language, but could also have other interpretations within

the semantics of concession. We also observe that макар clauses had additional paratactic semantics within the hypotaxis.

The contribution of Tsvetana Dimitrova is within the data excerpted from the historical corpus and its analysis.

22. Dimitrova, T. On Sentence Segmentation in Diachronic Texts. – Scripta & e-Scripta 21 (2021): 57–69.
http://e-scripta.ilit.bas.bg/archives/year-2021/issue-21/sentence-segmentation-diachronic-texts

The article discusses a proposal of a minimal set of criteria for sentence segmentation (an obligatory stage in the corpus processing and annotation, especially with respect to the syntactic annotation) of medieval texts. In the context of a review of different definitions of a sentence (unit) and approaches to sentence segmentation, various criteria are discussed: structural, thematic, graphic, on the basis of sample sentences in order to define the minimal criteria. The discussion of the different factors is illustrated by sample sentences from two texts from 14th and 17th c. The proposed criteria aim at considering mainly structural characteristics while trying to avoid textual and semantic interpretation though these can also present challenges because the interpretation of the (syntactic) structure is inevitably related to the interpretation of the (semantic) content.

23. Dimitrova, T., A. Bojadžiev. Electronic edition and linguistic annotation of Slavic fragments. – Scripta & e-Scripta 18 (2018): 35-49.
https://obshtezhitie.net/mks/Dimitrova.pdf
https://www.ceeol.com/search/article-detail?id=690140

The paper presents some issues on a possible approach to combining electronic description, electronic edition and annotated corpus of medieval Slavonic texts. In the first section of the paper, the emphasis is on practical issues regarding the annotation of different types of information such as information about the text, attribution, dates, saints, images. The second part discusses an approach to annotation of anaphorically related elements that are decisive for organising the structure of the text and connections to specific elements that will be annotated at the level of the text (the linguistic annotation follows an already implemented approach). The examples given cover not only parchment fragments but also fragments of manuscripts from different periods of the South Slavonic Medieval and Early Modern Bulgarian tradition. Tsvetana Dimitrova's contribution is in the description of the approach to linguistic annotation (section 3).