

Резюмета 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.

Монография

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

Монографията проследява развитието в историята на българския език на клас нееднородни езикови единици, които в научната литература са известни като клитики. Фокусът е върху онези от тях, които претърпяват промени в исторически план. Монографията се състои от предговор и три части. В 1.1. се ограничава обектът на изследване, т.е. групата на клитичните елементи, които ще бъдат разглеждани. В 1.2. се представя накратко езиковият материал, върху който се правят наблюденията и изводите за поведението и особеностите на разглежданите езикови елементи. В 1.3. са уточнени понятията, използвани за описание на поведението на изследваните елементи. В 1.4. се прави обзор на известната ни литература на български език, в която са анализирани някои от разглежданите елементи, за да се аргументира нуждата от по-цялостно изследване на съдбата на клитиките в българския език.

В първата част на изследването „Клитиките: същност и особености“ се разглеждат, в критичен план, особеностите и характера на елементите, които можем да определим като клитични, като акцентът е върху две основни функционални характеристики, които са взаимнообвързани, а именно тяхната дефектност и зависимостта им. Дискусията се разгръща в контекста на два лингвистични закона, които са формулирани още през 19. век – Законът на Вакернагел и Законът на Тоблер-Мусафия – и чието действие може да се проследи и в съвременните езици (включително в българския език). Езиковият материал е извлечен от текстови колекции и корпуси с данни за български. В тази първа част акцентът е най-вече върху някои особености на поведението на разглежданите клитични елементи в съвременния български език, които могат да се проследят в нестандартизирани езикови фрагменти – привлечен е материал от книжовни архаизирани източници в състава на Българския национален корпус, от транскрибираните записи в Корпуса с транскрибирана разговорна реч, от транскрибираните фрагменти на диалектна реч в рамките на интерактивната база от данни *Bulgarian Dialectology as a Living Tradition*. Целта на дискусията в тази първа част е да постави в контекст търсенията, резултатът от които е представен във втората част на настоящото изследване, която разглежда пътя на изследваните клитични елементи в историята на българския език. В първата част се аргументират – от позицията на съвременния носител на езика, който все пак разполага

с предимството на езиковата интуиция – онези особености на разглежданите клитични елементи като проявления, позиции и поведение спрямо околните елементи, които ще бъдат проследени в историческия езиков материал, за да бъде защитена хипотезата за единството на българския език през различните етапи на неговото развитие.

Втората част на монографията „Клитиките в историята на българския език“ разглежда проявленията на разглежданите клитични елементи в историята на българския език. Изводите са резултат от наблюдения върху колекция с примери от текстове, за които се предполага, че отразяват състоянието в три етапа от развоя на българския език – старобългарския (ексцерпирани са текстовете от Мариинското евангелие и Супрасълския сборник, които са част от корпуса TOROT); среднобългарския (ексцерпирани са текстовете на Троянската притча, по Ватиканския препис, и издадените Й. Богдан Влахо-български грамоти, като за целите на сравнението са привлечени и данни от Манасиевата хроника – примерите са ексцерпирани според текстовете от Историческия корпус на българския език); и ранния новобългарски (текстовете в Ловешкия дамаскин и в Троянския дамаскин, които също са част от Историческия корпус на българския език). Анализираните елементи са разделени в три групи, като първо, в 3. глава, са разгледани т.нар. дискурсни клитики же и бо и въпросителната клитика ли. Въпреки че обект на изследването са т.нар. парадигматични клитики, тъй като именно те са се запазили и/или развили до клитични елементи в съвременния български език, дискурсните клитики са включени като най-ранните клитични елементи, които имат своите съответствия в гръцките текстове – приемаме, че именно те указват кои са онези позиции в изречението и/или фразата, които може да приемат клитични елементи. Следващата група са местоименните клитики, като прегледът в 4. глава отново започва от най-ранните клитични елементи – първо- и второличните дателни клитики в единствено число ми и ти, преминава през останалите форми на кратките дателни местоимения, за да стигне до обзора на дателните клитики в среднобългарските и ранните новобългарски текстове. Обзорът на дателните клитики в 4.1. е следван от анализ на позициите и поведението на кратките първолични и второлични прякообектни местоимение в 4.2. и на кратките винителни форми на анафоричното местоимение в 4.3. И двете подгрупи според нас не са пълноценни клитики в ранните текстове, но показват склонност да се появяват в позиции, характерни за клитичните елементи (поради което са известни и като ‘полу(ен)клитики’ в литературата), като се проследява по-нататъшният развой на винителните клитики в примерите от средно- и новобългарските текстове. Възвратните винителни и дателни клитични елементи са разгледани отделно в 4.4., като и тук в ранните текстове те са по-скоро полуклитични, т.е. срещат се в позиции, нехарактерни за същинските клитики. Третата група елементи, които разглеждаме – в 5. глава, са форми на спомагателните глаголи, които показват разнопосочно поведение както в най-ранните текстове, така и по-късно, поради което ги наричаме ‘квазиклитични’. Прегледът на реализациите и позициите на тези елементи започва от примерите от старобългарските текстове, минава през примерите от среднобългарските текстове и завършва с примерите от ранните новобългарски текстове, които в голяма степен отразяват и съвременното състояние, с някои показателни изключения. Разгледани са позициите на клитичните елементи

спрямо обкръжението им и управляващия ги елемент (често това е глаголът, чиито аргументи са), спрямо лявата граница на изречението, в рамките на клитичния комплекс, както и по отношение на конституентите на сложните глаголни форми (където можем да наблюдаваме позициите на спомагателните глаголи и на останалите клитични елементи спрямо тях).

Посоката на развитие, която се очертава от наблюденията върху езиковите данни, е промяната в състава на елементите, т.е. някои елементи като дискурсните клитики отпадат, а се появяват други като някои местоименни и глаголни клитични елементи, при относително запазване на позиционните им възможности в изречението и/или спрямо глагола, който управлява повечето от тях. Наблюдаваните промени са свързани с промени в някои функционални елементи (например отрицанието) и глаголни форми. Същевременно тази стабилност включва и наблюдаваната във всички данни и периоди от развитието на езика вариативност на позициите на клитичните елементи, т.е. освен преобладаващите позиционни модели, в данните се наблюдават редица отклонения от тях.

В третата част на изследването „Анализ“ в 6. глава се прави опит за моделиране именно на вариативността на позициите на клитиките с инструментариума на фазовата теория в рамките на минималисткия синтаксис. Прави се кратък преглед на генеративната синтактична традиция, от която е част използваният модел, представят се две от хипотезите (на Ст. Франкс и на Р. Панчева) за деривацията и позиционирането на клитиките в българския език и в историята на българския език, които използват този инструментариум. Изследването завършва с формулиране на хипотеза за деривацията на клитичните елементи в различните позиции, като се изказва предположение, че клитиките могат да бъдат както зависими от глагола, чиито аргументни изисквания удовлетворяват (такива са например винителните местоименни клитики), така и да се реализират по-скоро на нивото на изречението (каквито в съвременния български език вероятно са дателните ‘етични’ клитики и притежателните клитики, реализирани на равнището на изречението, вместо на фразата).

Монографията завършва със заключение в 7. глава и Приложение с количествени данни от колекцията от извлечени примери, която е свободно достъпна под формата на таблици с данни и описания към данните. Данните могат да се използват за верификация, корекция, както и за целите на други изследвания.

Статии и студии: Езикови ресурси, компютърно моделиране на езика

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>

В статията в 2 се предлага стратегия за идентифициране и аотиране на деривационни отношения в Българския УърдНет, която да представи комплексната морфология на българския език. Методът включва трансфер на (морфо)семантичните релации от Принстънския УърдНет в Българския УърдНет на ниво синонимно множество и на деривационните релации между литералите в българския език. Деривационните релации са аотирани, за да отразят сложността на българската морфология. Въведените релации разширяват възможностите за приложение на УърдНет. Приносът на Цв. Димитрова е в описанието на аотацията на деривационните релации в Българския УърдНет (част 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-](http://dcl.bas.bg/clib/wp-content/uploads/2018/06/CLIB_2018_Proceedings_v1_final.pdf#page=156)

[content/uploads/2018/06/CLIB_2018_Proceedings_v1_final.pdf#page=156](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>

Статиите в 3, 4 и 5 описват организацията на и работата със система за създаване и редакция на УърдНет, създадена от Борислав Ризов. Статиите представят уебинтерфейса за работа с лексикалносемантични бази от данни с релационна структура, подобна на УърдНет, наречен Hydra for Web, който е изграден на основата на системата Hydra – инструмент с отворен код за създаване, редактиране и работа с уърднети. Приложението предлага лесен за използване и същевременно достатъчно мощен и удобен графичен интерфейс, с два режима за визуализация – едноезиков и паралелен – на езиковите съответствия на търсените (и намерените) синсети в уърднет. Интерфейсът е достъпен на адрес: <http://dcl.bas.bg/bulnet/>. Статиите разглеждат потребителския интерфейс и функционалностите му, както и някои проблеми и решения, свързани с адаптирането на ресурсите, които се визуализират (към момента

на публикацията на статиите това са Принстънският уърднет за английски език, Българският уърднет, Румънският уърднет и уърднети със свободен достъп за още 20 езика, които са част от базата от данни). Основният принос на Цв. Димитрова е при описанието на работата със системата.

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-](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=)

[85029188425&origin=resultslist&sort=plf-](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=)

[f&src=s&sid=3d2073faa936b694604aa8486288ec6a&sot=autdocs&sdt=autdocs&sl=18&s=](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=)
[AU-ID%2856288017300%29&relpos=2&citeCnt=1&searchTerm=](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-](https://www.scopus.com/record/display.uri?eid=2-s2.0-85060784706&origin=resultslist&sort=plf-f)

[85060784706&origin=resultslist&sort=plf-f](https://www.scopus.com/record/display.uri?eid=2-s2.0-85060784706&origin=resultslist&sort=plf-f)

В статиите в 6, 7 и 8 е представен подход за разработване на класификация на прилагателните в УърдНет, която се прилага върху Българския УърдНет с цел обогатяването му с допълнителна информация за понятията и връзките между тях.

В статията в 6 се прави критичен преглед на организацията на информацията за прилагателните имена в УърдНет, включително на използваните принципи за включването им в структурата на УърдНет и за свързването им (чрез различни релации) с другите части на речта (съществителни, глаголи и наречия), както и на съществуващите класификации на прилагателните имена в други уърднети и в литературата за български език.

В статията в 7 е представена семантична класификация на прилагателните имена, приложена върху прилагателните в Българския УърдНет, която използва за основа както вече съществуваща класификация на прилагателните в УърдНет за немски език (GermaNet), така и изследвания върху прилагателните имена в българския език. Макар да е разработена за Българския УърдНет, класификацията може да се

прилага и за други лексикалносемантични мрежи, в които информацията е организирана паралелно на или подобно на тази в Принстънския УърдНет. Класификацията разчита на информация, която вече е налична в УърдНет било то в рамките на дефинициите и примерите за употреба, или (най-вече на) релациите към други множества (на съществителни, глаголи и други прилагателни имена), които са свързани със съответното прилагателно чрез лексикалносемантични връзки (например със съществителни и глаголи от кои семантичните класове се свързва съответното прилагателно). Формулирани са 17 семантични класа, които са приписани на 2500 прилагателни имена.

В студията в 8 е представен модел за класификация, при който повече от един семантичен клас може да бъде приписан към едно прилагателно. Работата, представена в студията, надгражда тази, представена в статията в 7, като са класифицирани 3110 прилагателни, към част от които са приписани два или повече семантични класа, от които единият посочва реферанта, за който/когото може да се отнася прилагателното (отнасящо се например до поведение на човек, поведение на животно, тяло на човек / животно / растение и др.).

Работата по класификацията на прилагателните беше част от проекта „Семантична класификация на прилагателните в Българския УърдНет“, по който Цветана Димитрова е била научен ръководител на гл.ас. д-р Валентина Стефанова. Приносите на Цв. Димитрова са свързани с описанието на данните и с обосноваването на класификация на няколко равнища (студията в 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). Wrocław: 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>

Публикациите в 9, 10, 11 и 12 представят работата по формулиране и приписване на допълнителни семантични (и морфосемантични) релации между глаголи и съществителни и между съществителни, както и на допълнителни семантични класове към съществителните в УърдНет. Публикациите представят част от работата по проекта „Семантична мрежа с широк спектър от семантични релации“ (ДН 10-3 / 14.12.2016), финансирана от Фонд „Научни изследвания“ (ФНИ). Приносът на Цв. Димитрова е свързан с подходите за разширяване на релациите, свързани съществителните, както и на семантичните класове на съществителните имена.

В статията в 9 се описва етап от работата по разширяване на набора от морфосемантични релации, които могат да се припишат към двойките синонимни множества глагол – съществително в Българския УърдНет, свързани с релациите Агент (agent) и Търпящ (undergoer). Работата се основава на предположението, че словообразователните процеси при глаголно-именното словообразуване се определят и от тематичната структура на предиката. Ръчно са приписани седем допълнителни релации, формулирани съобразно тематичните роли на аргументите (според някои теории за реализацията на предикатно-аргументната структура на глаголите от определени семантични класове). Най-продуктивните словообразователни наставки при съществителните, свързани с релацията Агент, са подчертано агентивни, докато съществителните с допълнителна роля Експериенцер и Пациенс често са субстантивирани причастия.

Статията в 10 описва стратегия за пренасяне на словообразователните и морфосемантичните релации между съществително и глагол и съществително и прилагателно към (нови) релации между съществителните. Подходът разчита на вече налична информация – както от семантичните класове на свързаните понятия (синонимни множества), така и от съществуващите словообразователни и (морфо)семантични релации между съществително и глагол и съществително и прилагателно. Добавени са морфосемантични релации между съществителните в УърдНет, които са непряко свързани, чрез релации, с глаголи и прилагателни. Наблюденията върху комбинацията между релациите и семантичните класове на съществителните, свързани с тези релации, могат да допринесат за по-пълно описание на семантичните свойства на съществителните, чрез които те могат да участват в различни предикатно-аргументни структури.

В статията в 11 се прави преглед на подходите към включването на причастия в синонимните множества на прилагателни имена в УърдНет (върху материал от Българския УърдНет). Прилагателните синонимни множества, класифицирани като причастия в УърдНет, са много малко на брой, но в синонимните множества има включени литерали, които са всъщност причастия по форма. Затова е разгледан възможен подход за откриване на причастни форми в синонимни множества,

некласифицирани като причастия, като се идентифицират прилагателни синонимни множества, свързани – пряко или непряко (през други прилагателни или съществителни) – с глаголни синонимни множества.

Статията в 12 предлага подход за валидиране на данните, получени в резултат на предишна работа по разширяване на семантичните класове на съществителните в УърдНет чрез съпоставянето им със семантичните типове в рамките на онтологията от Корпусния анализ на моделите (Corpus Pattern Analysis – CPA), използвана от Речника на моделите на английските глаголи (Pattern Dictionary of English Verbs – PDEV). Представени са случаи с примерна верификация в контексти, които трябва да бъдат проверени при валидирането на данните, получени при съпоставянето на двата ресурса.

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.

В студията в 13 е описан Българско-английският корпус със съотнесени (прости) изречения (БулЕнСИ), който е извадка от Българско-английския паралелен корпус в състава на Българския национален корпус (БНК). Изреченията и в българския, и в английския подкорпус са автоматично разделени и съотнесени, като резултатът е проверен от експерти. Границите на изреченията в английския език са определени с помощта на OpenNLP парсер, а анотациите са редактирани допълнително. Сложните изречения са разделени на прости изречения, като е анотиран видът на синтактичната връзка и на позицията на простите изречения спрямо (съответното) главно изречение. Съотнесените изречения, както и съотнесените прости изречения в рамките на съответните двойки сложни изречения са анотирани от експерти. В статията са разгледани конвенциите, следвани при разделянето на изреченията и анотирането им, което е съобразено със синтактичните правила и с граматичната традиция и практиката за анотиране за български и за английски език.

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>

Изследването, представено в студията в 14, предлага единен модел за описание на три основни типа съществителни собствени имена – названия на лица, на географски обекти и на организации. Предлага се структурна и семантична класификация на разглежданите съществителни собствени имена и модел за езиково независимо семантично описание. Представени са основните типове семантични

модели, характерни за названията на лица, географски обекти и организации, и общо описание на синтактичните структури, които им съответстват в българския език.

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-](http://dcl.bas.bg/wp-content/uploads/2022/11/12_DimitrovaStefanovaTodorova_STONTBgRu2022.pdf)

[content/uploads/2022/11/12_DimitrovaStefanovaTodorova_STONTBgRu2022.pdf](http://dcl.bas.bg/wp-content/uploads/2022/11/12_DimitrovaStefanovaTodorova_STONTBgRu2022.pdf)

Студията в 15 разглежда предикативните конструкции за изразяване на състояние в български и в руски език, като са направени наблюдения, анализ и обобщения върху колекция от предикативни конструкции за състояние в български и в руски език, извлечени от паралелните българо-руски и руско-български подкорпуси в състава на Националния корпус на руския език. Приносът на Цв. Димитрова е свързан с ексерпцията на данните от Българо-руския и Руско-българския паралелен корпус (паралелните предикативни конструкции) и описанието им за целите на обобщенията и анализа.

Статии и студии: Исторически корпуси, описание на исторически езикови данни

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

Статията в 16 разглежда въпроса за използването на историческите корпуси за целите на изследването на езикови явления. Предимствата и недостатъците на подхода са илюстрирани чрез целева ексерпция на данни от три публично достъпни корпуса, които съдържат текстови източници, отразяващи състоянието на българския език в старобългарския и среднобългарския период, за да се хвърли светлина върху някои аспекти от развитието на т.нар. дискурсни клитики же и бо. В статията се разглежда поведението им, за да се очертаят условията за промяната, през която преминават в историята на българския език (за да изчезнат напълно – както като самостоятелни думи, така и като афикси). Трите корпуса, които са представени в статията, са свободно достъпни онлайн. Приносът на Цв. Димитрова е в ексерпцията, работата с корпусите и анализа на данните в част 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.

Статията в 17 разглежда въпроса за позицията на местоименните (изреченски) клитики в българския въз основа на данни от два новобългарски дамаскина и конструкции в съвременния български. Наблюденията се ограничават до кратките форми на личните местоимения и позициите им спрямо други (изреченски) конституенти, в опит да се установи дали има основание хипотезата, че в българския език клитиките могат да заемат две позиции – втора или приглаголна, както и допускането, че т.нар. „втора позиция“ е активна и се наследява в историческото развитие на езика.

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

В статията в 18 се прави обзор на конструкциите за изразяване на бъдещо действие, от които се развиват аналитичните форми за бъдеще време в съвременния български книжовен език, с акцент върху среднобългарския и ранния новобългарски период от развитието на езика (по данни от Ватиканския препис на Троянската притча – 14. век, Влахо-българските грамоти – 15. – 16. век; и Ловешкия дамаскин и Троянския дамаскин от 17. век). Най-голямо разнообразие на конструкциите се наблюдава във Влахо-българските грамоти, включително запазване на инфинитива, който се конкурира със съкратения инфинитив. Още в този период се появяват да-съчетанията, макар и с много малко (в сравнение с другите конструкции) срещания, но през 17. век именно те надделяват. Съкращаването на *хощѣ* до *щѣ*, започнало много рано, се отразява в явната конкуренцията между двете в Троянската притча, като във Влахо-българските грамоти вече се наблюдава преимущество на *щѣ*.

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

Студията в 19 проследява формирането на клитичния комплекс (клъстър) в българския език от старобългарския през среднобългарския до ранния новобългарски период. Предполага се, че комплексът има две равнища – основно, което се състои от (местоименно) ядро и (глаголна) периферия, и вторично, в което се намират (псевдоклитични) елементи, които проявяват поведение – и в диахронен, и в синхронен план, което се разминава с това на елементите от основното равнище. Езиковият материал от трите корпуса не показва промяна в позициите на елементите в ядрото; наблюдаваните промени в периферията се дължат главно на промени във формалната реализация на елементите (в резултат на преструктурирането на местоименната система и промените в състава на спомагателните глаголи, както и на загубата на някои ранни клитики като например дискурсните маркери).

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-s2.0-84975677603&origin=resultslist&sort=plf-f>

В студията в 20 са проследени етапите на развитие на т.нар. родово-дателен синкретизъм в старобългарския език – явление, което бележи началото на процеса на разпадане на падежната система в историята на българския език. Изследването се основава върху корпусни данни (включващи текстовете на Мариинското евангелие, Зографското евангелие и Супрасълския сборник). Въз основа на анализ на синтактичните позиции на родителните и дателните местоимения се налага хипотезата, че една от конструкциите, в които може да се проследи промяната при елементите в дателен падеж, в старобългарския език е т.нар. конструкция с външен притежател (*dativus sympatheticus*), която позволява дативът да се интерпретира извън именната фраза, но при задължително рефериране към притежателен елемент, генериран в именната фраза. Представени са етапите от реанализа на функциите, който води до установяване на елементи, вътрешни за именната фраза и рефериращи към притежателя, като акцентът е върху ролята на нововъзникващите клитични местоимения и синтактичното им позициониране във втора позиция в изречението като основен стимул за родително-дателния синкретизъм.

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>

Студията в 21 разглежда семантиката и структурата на изреченията, въведени със съюза *макар*, въз основа на данни от български писмени паметници от периода между 15. и 17. век. В този ранен период *макар* е нов съюз, който по-късно разширява употребата си и в съвременния български език е сред най-често употребяваните съюзи за отстъпка. Изследването се основава на данни от извадка от текстове, които са част от базата от данни на Историческия корпус на българския език. Съюзът *макар* се среща във Влахо-българските грамоти и в два новобългарски дамаскина от 17. век (Троянски и Ловешки дамаскин) както като самостоятелен съюз, така и в състава на сложните подчинителни съюзи *макар (и) да*, *макар че*, *макар (и) ако*, *макар и като*, които въвеждат и предикативни, и непредикативни сегменти.

Най-често срещаният вариант на сложния съюз, образуван с частицата *да*, има същите синтактични и граматични свойства като съюза *макар (и) да* в съвременния български език, особено по отношение на контактната позиция на *да* и глагола. В периода, когато това съчетание се установява, да-конструкцията вече е била със стабилна структура: глаголът е следвал непосредствено след *да* (двата елемента са

били разделяни най-вече от местоименни клитики и отрицателната частица, както е и в днешния български език), а между *макар* (*и*) и *да* са могли да се позиционират други изреченски конституенти.

Отделно се разглежда и семантиката на съюзите за отстъпка с *макар*, като в разглеждания период изразяването на отстъпително значение е било в ранен етап. Подчинените изречения, въведени с *макар*, са могли да изразяват не само прототипното значение на условие за отстъпка, както е в съвременния език, но са имали и други интерпретации в рамките на значенията за отстъпка. Подчинителните изречения с *макар* са имали и допълнителна паратактична семантика.

Приносът на Цв. Димитрова е свързан с ексцерпцията на данни от историческите – среднобългарски и новобългарски – източници и анализа им.

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>

В статията в 22 се предлага минимален набор от критерии за сегментиране на изречения в средновековните текстове (сегментацията на изречения е задължителен етап от обработката и аотирането на корпуси, особено по отношение на синтактичната аотация). В контекста на прегледа на различните дефиниции на изречението (изреченската единица) и подходите за сегментиране на изреченията се разглеждат различни критерии: структурни, тематични, графични, въз основа на примерни изречения с цел определяне на минималните критерии. Различните фактори се илюстрират с примерни изречения от два текста от 14. и от 17. век. Предложените критерии вземат под внимание предимно структурните характеристики на езиковите данни, като оставят настрана текстовата и семантичната интерпретация (доколкото е възможно). Последните обаче също могат да доведат до някои проблеми при формулирането на решенията, тъй като анализът на (синтактичната) структура е неизбежно свързана с интерпретацията на (семантичното) съдържание.

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>

Статията в 23 разглежда възможен подход за комбиниране на електронно описание, електронно издание и аотиран корпус от средновековни славянски текстове. В първата част на статията се акцентира върху практическите въпроси, свързани с аотирането на различни видове метаданни като информация за текста, атрибуция, датиране, светци, изображения. Във втората част, в която е и основният принос на Цв. Димитрова, се разглежда подход за аотиране на анафорично свързани елементи, които са определящи за организиране на структурата на текста и връзките към конкретни елементи, които ще бъдат аотирани на ниво текст. Примерите

включват както фрагменти от пергаментни ръкописи, така и такива от ръкописи от различни периоди на развитието на южнославянската и новобългарската книжнина. Приносът на Цв. Димитрова е в описанието на подхода към лингвистичната анотация (част 3).