БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ

ИНСТИТУТ ЗА БЪЛГАРСКИ ЕЗИК
„ПРОФ. ЛЮБОМИР АНДРЕЙЧИН”

СЪВРЕМЕННИ МЕТОДИ И ПОДХОДИ В ЛЕКСИКОГРАФСКАТА ПРАКТИКА

СБОРНИК СТУДИИ И СТАТИИ

Авангард Прима
СОФИЯ • 2011

Изданието е осъществено с финансовата подкрепа на Фонд „Научни изследвания” при Министерството на образованието, младежта и науката на Република България (договор № ДТК 02-53/2009 г.).

Редактори и съставители на сборника:
доц. д-р Диана Благоева
доц. д-р Сия Колковска

Рецензент:
проф. дфн Лилия Крумова-Цветкова

За текста на публикациите отговорност носят съответните автори.

© Диана Благоева, Сия Колковска, Индра Светла, Албена Рангелова, Надежда Костова, Йорданка Трифонова, Божана Нишева, автори, 2011

ISBN 978-954-323-923-8

Авангард Прима, София, 2011

СЪДЪРЖАНИЕ

ПРЕДГОВОР 5

ДИАНА БЛАГОЕВА, СИЯ КОЛКОВСКА (СОФИЯ) –
КОРПУСНИЯТ ПОДХОД В БЪЛГАРСКАТА ЛЕКСИКОГРАФИЯ –
ПРАКТИКА И ПЕРСПЕКТИВИ 7

ИНДРА СВЕТЛА (ПРАГА) – ИЗГРАЖДАНЕ НА СЛОВНИКА
НА ЛЕКСИКАЛНАТА БАЗА ДАННИ *ПРАЛЕКС* 46

ИНДРА СВЕТЛА, АЛБЕНА РАНГЕЛОВА (ПРАГА) –
ЛЕКСИКАЛНАТА БАЗА ДАННИ *ПРАЛЕКС* –
СТРУКТУРА И ОБРАБОТКА НА ЗАГЛАВНИТЕ ЕДИНИЦИ 61

НАДЕЖДА КОСТОВА (СОФИЯ) – ПРОБЛЕМИ НА
ЛЕКСИКОГРАФСКОТО ПРЕДСТАВЯНЕ НА НЯКОИ ГЛАГОЛИ
ЗА РЕЧЕВА ДЕЙНОСТ В БЪЛГАРСКИЯ ЕЗИК.....77

ЙОРДАНКА ТРИФОНОВА (СОФИЯ) –
СЛАВЯНСКО-БЪЛГАРСКАТА ДВУЕЗИЧНА ЛЕКСИКОГРАФИЯ
С ОГЛЕД НА ИЗПОЛЗВАНЕТО Ѝ В Е-РЕЧНИЦИТЕ 116

БОЖАНА НИШЕВА (СОФИЯ) – НЕОГРАФСКОТО ПРЕДСТАВЯНЕ
НА ПОСТПОЗИТИВНИТЕ СЛОВООБРАЗУВАТЕЛНИ КОМПОНЕНТИ
В БЪЛГАРСКИЯ И ЧЕШКИЯ ЕЗИК (С ОГЛЕД НА СЪВРЕМЕННАТА
ЛЕКСИКАЛНА ДИНАМИКА).....142

ПРЕДГОВОР

Сборникът „Съвременни методи и подходи в лексикографската практика” съдържа студии и статии, посветени на актуални проблеми на съвременната лексикография. Той е резултат от дългогодишното сътрудничество между академичните лексикографски звена в България и Чехия: Секцията за българска лексикология и лексикография към Института за български език „Проф. Любомир Андрейчин” при Българската академия на науките и Секцията за лексикография към Института за чешки език при Чешката академия на науките. Автори в сборника са български и чешки езиковеди от двата института.

Основните проблеми, разгледани в отделните разработки в сборника, са свързани с прилагането на модерни подходи и технологии в лексикографската дейност.

В съвместната студия на Диана Благоева и Сия Колковска се разглеждат практиките и перспективите в приложението на корпуснобазирания подход при работата върху различни жанрове речници на българския език – тълковни, неологични, двуезични и др.

Статията на Индра Светла представя принципите и методите за изграждане на лексикалната база данни *Пралекс*, върху която се работи в Института за чешки език.

В съвместната си разработка Индра Светла и Албена Рангелова се спират върху структурата на лексикалната база данни *Пралекс* и подходите за представяне на заглавните единици в нея.

В работата си Надежда Костова провежда корпусно изследване на съчетаемостта на основните глаголи за речева дейност в българския език с оглед на усъвършенстване на лексикографското представяне на тези лексикални единици в тълковен речник.

Йорданка Трифонова представя анализ на развитието на славянско-българската двуезична лексикография през последния половин век и се спира на възможностите, които предоставят постиженията в тази област за създаване на електронни двуезични речници.

В студията на Божана Нишева се разглеждат типовете постпозитивни словообразователни компоненти според тяхната лексикографска интерпретация в най-представителните неологични речници на българския и чешкия език с оглед на съвременната лексикална динамика.

Сборникът ще бъде полезен за лексиколози и лексикографи, за специалисти по компютърна лексикография, както и за изследователи в областта на славянските езици.

От съставителите

Диана Благоева, Сия Колковска
(София)

КОРПУСНИЯТ ПОДХОД В БЪЛГАРСКАТА ЛЕКСИКОГРАФИЯ – ПРАКТИКА И ПЕРСПЕКТИВИ*

The paper is focused on the various ways of application of corpus-based approach in the contemporary Bulgarian lexicography. The special attention is paid to the results of corpus-based lexicographical techniques in their application in the creation of various types of Bulgarian dictionaries (explanatory, neological, bilingual etc.). The future perspectives for introducing more widely corpus-based method in Bulgarian lexicography are outlined.

Ключови думи: корпусна лексикография, електронни корпуси, речници

Key words: corpus lexicography, electronic corpora, dictionaries

1. Въведение

Корпусната лингвистика възниква през 60-те години на миналия век и отбелязва стремително развитие най-вече през последните три десетилетия. Методите и подходите на изследване, предлагани от това направление, водят до съществени изменения в изследователските практики в редица области на лингвистичната наука (Синклер 1991, Филмор 1992, Байбър и др. 1998, Мейер 2004), като в центъра на вниманието се поставя реалната употреба на езика, т.е. „езикът в действие”. Особено значително е влиянието, което корпусният подход оказва върху лексикографската практика (Филмор и др. 1994, Уи 1998, Мейер 2004, Аткинс, Рындъл 2008, Рындъл, Килгариф 2011). Както справедливо отбелязва Чарлз Мейер: „Компютърните корпуси и свързаният с тях софтуер напълно революционизираха създа-

* Изследването е проведено в рамките на научноизследователския проект „Информационна среда за комплексно лексикографско описание на словното богатство на българския език (от началото на XIX в. до наши дни)”, финансиран от Фонд „Научни изследвания”, договор ДТК 02/53 от 2009 г.

ването на речници” (Мейер 2004: 16).

2. Корпусният подход в лексикографията

Корпусният подход намира приложение в лексикографията (първоначално предимно по отношение на английския език) вече повече от три десетилетия, като първите стъпки в тази посока са свързани с използването на електронния корпус Bank of English, създаден в началото на 80-те години на XX в. в рамките на проекта COBUILD (Collins Birmingham University International Language Database) под ръководството на Дж. Синклер (Синклер 1987, Синклер 1991: 13-26). Въз основа на този корпус в издателство Collins (сега HarperCollins) са изработени и публикувани серия речници и граматики на английския език (Карпова 2004, Кришнамурти 2008). С течение на времето използването на корпуси при изработването на речници се утвърждава като стандарт в лексикографската практика, при това вече далеч не само по отношение на английския език (Мейер 2004: 11-16, Кришнамурти 2008), но и за редица други езици (Мейер 2004: 11-16, Кришнамурти 2008, Рындъл, Килгариф 2011).

В научната литература е възприето въведеното от един от основателите на корпусната лексикография – Дж. Синклер – разграничение между корпуснобазиран (corpus-based) и корпусноизведен (corpus-driven) подход и във връзка с това – между корпуснобазирани и корпусноизведени речници (Синклер 1991, Уи 1998, Тонини-Бонели 2001, Кришнамурти 2008 и др.). Характерно е, че корпусните данни играят различна роля при вземането на лексикографски решения, когато се прилага съответно корпуснобазиран или корпусноизведен подход. В първия случай корпусът се използва като основа за потвърждаване, доразвиване или опровергаване на съществуващи теории и хипотези (Тонини-Бонели 2001). Корпусноизведеният подход пък предполага формулиране на теоретични постановки и хипотези изцяло въз основа на наблюдения върху употребата на езиковите единици в корпуса и на доказателства, извлечени от корпус (corpus evidence) (Тонини-Бонели 2001). Както отбелязва В. Уи, в първия случай подходът е „отгоре

надолу” (т.е. от теорията към конкретните корпусни данни), а във втория – „отдолу нагоре” (от конкретните корпусни данни към теорията), макар че нерядко в практиката двата подхода се прилагат в комбинация (Уи 1998: 51-52).

По отношение на лексикографската практика корпусноизведеният подход намира израз например в това, че извличането на семантични и прагматични данни и дефинирането на отделните значения на лексикалните единици се извършва единствено въз основа на наблюдения върху лексикалния и синтактичния контекст на съответните единици в корпуса[1], като редът на значенията в речниковата статия се определя от честотността на техните срещания в корпуса (Синклер 1991).

Класически пример за корпусноизведени лексикографски справочници са речниците на COBUILD за английския език (например CCELD), като по думите на Т. Хербст това са едни от малкото речници, в които корпусноизведеният подход е реализиран системно и в своя пълен вид (Хербст 1996: 322, цит. по Косем, Кришнамурти 2007). Сходна е предложената от Фр. Чермак методика за изграждане на речниковата структура, при която думите в речника се представят чрез своя лексикален профил, съставен главно на базата на извлечена от корпуси информация (Чермак 2010).

По-широко разпространени обаче си остават корпуснобазираните речници, в които традиционните лексикографски практики, свързани например с използването на лексикални картотеки, други речници и пр. като източник на информация, се комбинират с повече или по-малко системно използване на данни от корпуси (Карлсон 2000).

С оглед на разграничението между корпуснобазирани и корпусноизведени речници трябва да се направи уточнението, че обект на внимание в настоящото изследване са преди всичко възможностите, които дава корпуснобазираният подход в лексикографията, като за краткост по-нататък вместо термина *корпуснобазиран подход* ще бъде употребяван терминът *корпусен подход*.

Използването на електронни корпуси като източник на лексикографска информация в значителна степен допринася за

преодоляване на някои недостатъци и ограничения на традиционните лексикографски подходи. От една страна, корпусите са богати и леснодостъпни лексикални ресурси и в това отношение далеч превъзхождат традиционно използваните при речниковата работа лексикални картотеки с фишове (Филмор 1992, Уи 1998: 47, Аткинс, Рындъл 2008: 53). Боравейки с данни от лексикална картотека, лексикографът има възможност да подложи на анализ само ограничен брой неголеми фрагменти от ексцерпираните текстове, в които присъства съответна лексикална единица, докато корпусът предоставя достъп до всички срещания на съответната единица във включените в корпуса материали с възможност за проучване на контекстуалното обкръжение на единицата в широки граници (включително в рамките на целия текст), а не само в границите на отделно изречение (или няколко изречения). Освен това при ексцерпцията на материал за лексикалните картотеки не са изключени прояви на несистемност и субективност особено в случаите, когато не се прилага т.нар. тотална ексцерпция.

От друга страна, традиционно използваните в лексикографията интроспективни методи, които, както е известно, се основават на знания, недостъпни за пряко наблюдение и проверка, може да бъдат верифицирани чрез корпусния подход (Лийч 1992, Уи 1998, Верлинд, Селва 2001). Използването на корпусни данни позволява на лексикографа да се опре на обективна информация за употребата на лексикалните единици в голямо множество реални текстове на съответния език и така да преодолее до голяма степен субективността и недостатъчната надеждност на метода на интроспекцията.

Разбира се, въпреки че корпуснобазираните речници вече са станали норма в лексикографията (Килгариф 2000, Рындъл, Килгариф 2011), едва ли може да се твърди, че традиционните лексикографски методи и подходи са загубили своето значение. Като най-ползотворно на сегашния етап се очертава съчетаването на традиционни и съвременни (в частност – корпусни) подходи.

Приложението на корпусния подход в лексикографията води до редица промени и нововъведения в процеса на изработване на речници, като засяга елементи както на

макроструктурата, така и на микроструктурата на речника. Възможностите за усъвършенстване на лексикографската работа чрез използването на корпуси може да се търсят в две основни посоки:

2.1. По отношение на качеството на създавания лексикографски продукт. Корпусният подход допринася за подобряване на такива важни характеристики на речника като точност, системност и пълнота на лексикографската интерпретация на лексикалните единици (Аткинс 2002). Речникът като лексикографска картина на определен отрязък от лексикалната система и семантичното пространство на езика има отражателен характер. Използването на емпирични данни, основани на корпусния подход, дава възможност за постигане на по-висока степен на обективност и достоверност на лексикографското отражение, а също и за вземане на по-коректни лексикографски решения. Достъпът до обективна информация от корпусите за реалната употреба на езиковите единици, по думите на Б. Аткинс и М. Рундъл, е фундаментална предпоставка за създаване на надеждни речници (Аткинс, Рундъл 2008: 54–55).

2.2. По отношение на труда на лексикографите. Корпусният подход рефлектира съществено върху изискващата значителен разход на време и усилия лексикографска дейност, като дава възможност за нейното оптимизиране включително чрез автоматизиране на някои трудоемки рутинни операции (в частност сортиране и подбор на лексикален материал). Става възможно да се ускори в известна степен процесът на изработване на лексикографските продукти, като се намали продължителността на отделни етапи от лексикографската дейност (Рундъл, Килгариф 2011).

3. Изисквания към корпусите, използвани за лексикографски цели

В зависимост от жанра и предназначението на съставяните речници в лексикографската дейност намират приложение не само големи общи корпуси, но и специализирани корпуси[2] (които може да бъдат подкорпуси в състава на общи корпуси):

паралелни (за целите на двуезичната лексикография), терминологични (при изработване на терминологични речници), диахронни (за целите на историческата лексикография), диалектни (при съставяне на диалектни речници), корпуси, включващи текстове само на отделен автор (при изработване на речници на език на автор) и пр. Планирането и изграждането на всеки един тип специализирани корпуси се ръководи както от общи, така и от специфични изисквания по отношение на обема на корпуса, типа на включваните текстове, равнищата на лингвистична обработка, необходимостта от специални софтуерни средства и пр.

Тук ще се спрем само върху някои по-общи принципни изисквания, на които трябва да отговарят корпусите, използвани за лексикографски цели. Тези изисквания се отнасят най-вече до параметрите на корпуса (например обем, измерван с брой включени графични думи) и до основните характеристики на корпуса (представителност, балансираност, динамичност / статичност).

Споделяме мнението, изказвано в редица изследвания в областта на корпусната лексикография, че за лексикографски цели е необходимо да се използват (много) големи по обем корпуси[3] (Байбър и др. 1998: 30, Мейер 2004: 14, Аткинс, Рындъл 2008: 61, Чермак 2010 и др.). Големите корпуси предоставят по-достоверна информация за фреквентността на изследваните лексикални единици, за тяхното контекстуално поведение и съчетаемост. На базата на значителен масив от текстове лексикографът може да получи адекватна информация (подкрепена със статистически данни) за употребата не само за лексикалните единици с висока и средна честотност (т.е. за единиците от ядрената част на речниковия състав), но и за по-рядко срещаните думи[4] (например единици от периферията на речниковия състав като излизащи от употреба думи, неологизми и пр.) и за по-редките, нетипични значения и употреби, каквито също следва да бъдат отразявани в общите тълковни речници, в някои диференциални речници (като речници на новите думи, на редките и остарелите думи и пр.).

Важно е използваният за лексикографски цели корпус да бъде отворен, динамичен, а не затворен, статичен. Това озна-

чава той да подлежи на непрекъснато обогатяване, осъвременяване и разширяване с нови текстови материали, което е предпоставка за своевременно откриване и лексикографско отразяване на промените и иновациите в лексиката в конкретни периоди от езиковото развитие.

Като основно изискване към корпусите се формулира принципът за представителност (репрезентативност) по отношение на съответния език или подезик (функционална разновидност на езика) (Байбър 1993, Синклер 1995, Уи 1998, Шимкова 2005 и др.). Представителността на корпуса предполага, че подбраната съвкупност от текстове представлява „умален модел" на съответния език или подезик, чието състояние се отразява, т.е. корпусът, който има краен размер, би трябвало да бъде адекватен представител на потенциално безкрайно множество от текстове от някакъв определен тип в даден хронологичен срез (Захаров 2005, Шаров 2003). Постигането на представителност е доста сложна задача поради изключителното разнообразие от устни и писмени текстове, за които е трудно да се формулира достатъчно прецизна и изчерпателна типология и да се установи точно пропорционално съотношение[5] (Павловски 2003). Решение на този проблем се търси, като принципът за представителност се свързва с принципа за балансираност, като се предполага, че адекватно представяне на реалното многообразие на езика в неговите функционални разновидности може да бъде постигнато чрез определяне на правилно (балансирано) съотношение между включените в корпуса текстове от различните стилове и жанрове, групирани в отделни подкорпуси, всеки един от които е представителен за съответната функционална разновидност на езика (Кенеди 1998: 62).

Разбира се, както вече беше казано, посочените изисквания към корпусите, използвани за лексикографски цели, са принципни, като в зависимост от спецификите на създавания речник (тип, жанр, задачи, насоченост към конкретна аудитория ползватели и пр.) може да бъдат формулирани и по-частни изисквания. Така например особеностите на многотомния академичен *Речник на българския език* (*РБЕ*), който има за цел да отрази лексикалния състав на българския език в неговото

функционално многообразие (книжовна лексика, частично – жаргонна лексика, както и общодиалектна лексика, отразена във фолклорни източници, и пр.) за последните близо два века (Генадиева и др. 1966), обуславят и спецификите на корпуса, който следва да се използва при неговото съставяне. Тези специфики се отнасят до състава на корпуса (обхващане на писмени текстове с широк стилов и жанров диапазон, включително и такива, каквито по правило не влизат в общите корпуси с многопосочно приложение, например фолклорни произведения), необходимостта от покриване на по-голям хронологичен отрязък (обхващане на текстове, публикувани в периода от края на XIX в. насам) и др. (вж. по-подробно Благоева, Колковска 2006, Благоева 2008а).

4. Корпусният подход в българската лексикография

Корпусната лексикография в България е все още в началния етап от своето развитие. До сегашния момент са осъществени (или са в процес на осъществяване) само няколко корпусно-базирани лексикографски проекта. Става дума за речници със следните жанрови характеристики:

Общ тълковен речник – *Речник на българския език* (РБЕ). Речникът е многогодишен проект на Секцията за българска лексикология и лексикография (СБЛЛ) в Института за български език „Проф. Любомир Андрейчин" при БАН. В периода от 1977 до 2008 г. са издадени 13 поредни тома от РБЕ (от буква А до буква П), подготвени са за печат още 2 поредни тома и се работи върху следващите няколко. Същевременно в началото на XXI в. започна и преработка и осъвременяване на първите 8 издадени тома, като вече са отпечатани новите издания на т. 1–3 (2001–2006 г.), подготвен за печат е преработеният т. 4. Променените през последните години условия, свързани преди всичко с развитието на съвременните компютърни технологии, наложиха необходимостта да се пристъпи към осъвременяване и дигитализиране на емпиричната база за изработване на *РБЕ* (Благоева и др. 2009). Във връзка с това в периода 2006–2008 г. в СБЛЛ беше създаден специален Лексикографски електронен корпус, предназначен преди всичко за нуждите на *РБЕ*, но с

широки възможности за приложение и за изработване на други типове и жанрове речници (вж. по-подробно Благоева, Колковска 2007, Благоева 2008а). От началото на 2009 г. корпусът е интегрална част от създадения в Института за български език „Проф. Л. Андрейчин" при БАН Български национален корпус[6] (http://www.ibl.bas.bg/BGNC_bg.htm).

Неологичен речник – *Речник на новите думи в българския език (от края на XX и първото десетилетие на XXI в.)* (РНДБЕ). Това е първият речник, отразяващ новата лексика в българския език от последните около 2 десетилетия, който е изработен въз основа на електронен корпус – Българския национален корпус (Благоева 2008б).

Правописен речник – *Правописен речник на българския език* (ПРБЕ), който се посочва от авторите си като първия български лексикографски продукт, базиран върху езиков материал от голям електронен корпус – корпуса BulTreeBank (http://www.bultreebank.org/) (Симов и др. 2008).

Ще разгледаме някои по-важни аспекти от прилагането на корпусния подход в българската лексикография, свързани с отделните етапи от лексикографската работа и конкретизирани в зависимост от спецификите на различните жанрове речници (тълковни, неологични, синонимни, терминологични, двуезични и пр.). Някои от тези аспекти вече са реалност за българската лексикография, а въвеждането на останалите в родната лексикографска практика е въпрос на недалечно бъдеще.

4.1. Усъвършенстване на методите за изготвяне на речников словник

Използването на корпусен метод за определяне на речниковия словник има място при различни жанрове речници – общи тълковни, неологични, синонимни, антонимни, правописни, а също и при двуезичните речници.

По-особен е случаят с терминологичните речници, при които словникът се определя чрез установяване на основните термини в дадена предметна област, т.е. чрез анализ на съответната понятийна система. Тук корпуснобазираните подходи нямат специално приложение за разлика от други подходи като например

когнитивния. Както отбелязва М. Попова, при следването на този подход (в частност – на прототипната теория) словникът се установява, на първо място, чрез определяне на ядрото (прототипния център) на съответната специална област, което включва термините, създадени в нея и обозначаващи специфичните за нея понятия, а на второ място – чрез определяне на термините, които функционират в дадената област, но принадлежат също така и на други области (Попова 2011: 167).

Софтуерните средства за работа с корпуси дават възможност за автоматично генериране на списъци с регистрираните в корпуса леми (или словоформи), а също и за тяхната подредба по (намаляваща или нарастваща) честота (Коева 2000). Тези списъци отразяват обективни данни за разпространението и фреквентността на лексикалните единици в голямо множество реални текстове и може да бъдат използвани (самостоятелно или като допълнение към създаваните по традиционни начини лексемни списъци) като изходна база за определяне на словника за различни жанрове речници.

Независимо дали се използват самостоятелно, или в съчетание с други източници, корпусните данни позволяват поадекватно да се отрази съвременното състояние на лексиката, оперативно да се проследяват промените, настъпили в нея, да се откриват и извличат лексикални единици, нерегистрирани в съществуващите речници на българския език и в лексикалните картотеки. От друга страна, използването на данни за честотността на думите в корпуса дава възможност на лексикографа да се опре на обективни критерии при подбора на думите за словника. Чрез корпусни данни (включително с прилагане на статистически методи), може да се определи подостоверно, отколкото по интуитивен път, степента на узуалност на лексикалните единици (Савицки, Хлавачова 2003). Пренебрегването на корпусите като източник на данни за реалната употреба на думите в съвременния език може да доведе до включването в словника на неактуални заглавни единици или пропускането на някои обичайно употребявани думи[7].

Комбинация от корпусен и традиционен метод за създаване на словник се следва при работата върху последните ня-

колко тома на многотомния тълковен *РБЕ* – най-значителния и мащабен проект в българската лексикография, в който, както беше посочено по-горе, от няколко години успешно се използва корпусният подход. На сегашния етап словникът на *РБЕ* се определя на базата на списъци от думи, които се изготвят въз основа на лексикални картотеки и съществуващите речници на българския език и се допълват с данни от автоматично генерираните списъци на словоформите в Българския национален корпус (БНК). Опирайки се на информацията от корпуса, която включва и обективни сведения за фреквентността на словоформите и броя на текстовете, в които има техни срещания, лексикографът може да вземе мотивирано решение относно това, кои от присъстващите в изходните списъци думи следва да бъдат включени като заглавни единици в *РБЕ*. Използването на корпусни данни при определяне на словника носи и други предимства с оглед на спецификата на *РБЕ*, който е предназначен да отразява възможно най-пълно лексикалното богатство на българския език. Значителният обем на Българския национален корпус, който в сегашния момент надхвърля 400 милиона графични думи, дава възможност за по-пълно лексикографско обхващане на българската лексика чрез регистриране на лексикални и фразеологични единици, които по една или друга причина отсъстват в съществуващите речници на българския език и не са попаднали в лексикалните картотеки на *РБЕ*.

Значителни са предимствата на корпусния подход при изготвяне на словник на неологичен речник. Ползотворността от този подход бе проверена на практика при създаване на *РНДБЕ.* За целите на този речник бяха използвани двата субкорпуса в състава на Българския национален корпус, съдържащи няколко хиляди текстове, публикувани съответно от 1990 г. насам и в периода от 1945 до 1989 г. Чрез специални процедури и софтуерни средства (вж. по-подробно Благоева 2008, Благоева 2009) беше сравнен лексикалният състав на двата подкорпуса, в резултат на което беше извлечен автоматично списък от словоформи, регистрирани само в масива с по-нови текстове. Тези словоформи, разглеждани като „кандидат-неологизми”, бяха подложени на ръчна обработка за отстраняване на нере-

левантните случаи: собствени имена, погрешно изписани думи, неразпознати от програмата форми на познати леми, думи с чужда графика и пр. Полученият в резултат на това разширен списък с нови единици (между които неологизми, оказионализми, авторски новообразувания и пр.) послужи като основа за подбор на единици за съставяне на словника на *РНДБЕ*. Ограниченията на описания полуавтоматичен метод за извличане на неологизми (в частност – неприложимостта му за откриване на нововъзникнали значения на думите, на нови устойчиви съчетания и фразеологизми) бяха компенсирани с традиционни методи (ръчна ексцерпция на неологизми).

При избора на единици за включване в словника на неологичния речник бяха взети предвид и честотни данни от корпуса с цел да бъдат елиминирани оказионалните образувания и употреби. При това беше отчитана не само абсолютната честота на употреба (общият брой срещания на единицата в корпуса), но и броят на текстовете, в които са регистрирани срещания. Така се избягва опасността да се припише по-висок честотен ранг на думи, които имат значителен брой срещания, но в много ограничен брой документи. Както отбелязва П. Ханкс, за корпусната лексикография фреквентността е важна, но не е достатъчна, трябва да се взема предвид и дистрибуцията на явлението (Ханкс 2003: 58). Това може да се постигне чрез нормализиране на честотния ранг на думите в корпуса например чрез предложената от К. Симов и др. формула, в която се отчитат два параметъра: LexRank – вероятността за срещане на лексемата, изчислявана чрез разделяне на броя на срещанията на словоформите на лексемата на броя на всички словоформи в корпуса, и DocRank – броят на документите, съдържащи словоформи на лексемата (Симов и др. 2008: 413).

Корпусният подход не само улеснява началния етап от неографската работа – изграждането на словник, но осигурява и значително по-голяма пълнота и представителност в отразяването на новата лексика. Той позволява да се установи с по-висока достоверност неологичният статус на съответните единици и да се определи повече или по-малко точно хронологията на тяхното навлизане в езика. Ако се пренебрегне този

подход и се разчита единствено на други критерии (например субективното усещане на лексикографа за новост на думата, отсъствието на думата в съществуващите речници и пр.), може да се стигне до некоректни лексикографски решения. Само такива критерии са приложени например при съставянето на издадения през 2003 г. *Неологичен речник (за периода 1998–2003)* (*НеолР*), което е довело до включване в словника му на цяла редица думи и съчетания, неологичният статус на които може да бъде поставен под съмнение, ако се вземат предвид корпусните данни, срв.: *бурка* (със срещания в Българския национален корпус в текстове, датиращи от 1942 г. насам)[8], *смъртник* (1956 г.), *килт* (1957 г.), *публична тайна* (1959 г.), *аеромобил* (1966 г.), *прогнозист* (1971 г.), *полиграф* (1971 г.), *футуролог* (1972 г.), *снегоход* (1972 г.), *плаващ курс* (1976 г.), *спиричуъл* (1979 г.) и др.

Използването на корпусна информация би допринесло в голяма степен и за адекватното определяне на словника на правописните речници. По отношение на речниците от този жанр корпусноизвлечените честотни списъци от думи са важен и обективен ориентир за реално срещащите се в съвременната речева практика лексикални единици, за техните форми и фреквентност. Въз основа на такива списъци биха могли да се направят по-обосновани изводи за степента на узуализация на дадено явление, да се установят настъпили промени в речевата практика и да се прецени необходимостта от кодифицирането на тези промени. Пример за правописен речник, чийто словник е изработен при отчитане и на корпусни данни (в допълнение с използване на словника на предходни речници), е *ПРБЕ*. При изготвяне на словника на посочения речник са взети предвид статистически данни за честотата на словоформите в използвания корпус (BulTreeBank), като се отчитат и данните за броя на документи, в които са засвидетелствани употреби на съответните словоформи (Симов и др. 2008).

4.2. Вземане на по-точни и обективни лексикографски решения при трактовката на фонетичните и графичните варианти на думите

От корпуса може да се извлече информация за честотата на възможните фонетични и графични варианти на думата, което е особено полезно при изработването на някои жанрове речници като например неологичните. Известно е, че неологизмите са подчертано динамичен лексикален пласт, в който немалка част от единиците (особено заемките) са в процес на фонетична, графична, морфологична и друга адаптация и проявяват неустойчивост, изразяваща се в графично и/или фонетично вариране. Честотният критерий, приложен върху представителен текстов масив, може да бъде основание да се отдаде предпочитание на една или друга регистрирана форма като база за избор на вариант за заглавка в неологичния речник и в по-широк план – като насока за установяване на правописна норма. Тази практика е приложена в *РНДБЕ* въз основа на данните от Българския национален корпус. Така например информацията за преобладаващата честота на срещане в Българския национален корпус на графичната форма *плейър* в сравнение с *плейер* (в съчетание с данни за нейната преобладаваща честота и в интернет) мотивира обособяването на заглавка *плейър* в *РНДБЕ* и извеждането на *плейер* като неин вариант. Подобен е случаят и с графичните форми *мърчандайзинг*, *мърчандайзер*, предпочетени в *РНДБЕ* пред по-малко честотните *мърчъндайзинг*, *мърчъндайзер*, регистрирани например в *НеолР*.

Обективни корпусни данни от посочения тип следва да се вземат предвид и при изработването на правописни речници и въпреки че те не са основен критерий при вземането на решения в случаите на правописни проблеми, може да подпомогнат тяхното разрешаване.

4.3. По-точно и обективно отразяване на парадигматичните и синтагматичните особености на думите

Корпусът е източник на обективна информация относно употребата на отделни граматични форми на думите, а също и

на данни за синтактичното поведение на думите. Тази информация е особено важна за големи тълковни речници като *РБЕ*, които представят детайлна и разностранна информация за тълкуваните лексикални единици с отразяване на редица особености в тяхната семантика, парадигматика и употреба.

Така например въз основата на получени от Българския национален корпус данни се установява по обективен начин отсъствието на форми за множествено число при някои съществителни имена от среден род (например *разностилие*) и наличието на такива форми при други имена от същия род (например *разобличаване*, *разномислие*).

Въз основа на емпиричните данни от корпуса може да бъде мотивиран и изборът на граматични бележки като *Само мн. / Обикн. мн. / Само ед. / Обикн. ед.*, чрез които се отбелязва наличието на определени ограничения във функционирането на някои граматични форми на имената. Така например наблюденията върху употребите на прилагателното *разнокалибрен* в Българския национален корпус (с помощта на системата Sketch Engine)[9] показват категорично, че това прилагателно функционира преди всичко във форма за множествено число, което е достатъчно обективно основание в неговата речникова статия в *РБЕ* да бъде включена ограничителната бележка *Обикн. мн.*

Въз основа на корпусни данни може да се направят обосновани заключения и по отношение на ограниченията в синтактичното функциониране на някои думи. Така например информацията, извлечена от Българския национален корпус, показва, че съществителното име *разноглед* в значението ‘човек с определен недостатък в зрението’ функционира обикновено като сказуемно определение и поради това би трябвало към съответното значение в речниковата му статия в *РБЕ* да се добави бележка *Обикн. като сказ. опред*.

4.4. Мотивиране на решения при определянето на стилистичните квалификатори на заглавната дума

Възможността за извършване на наблюдения върху голям по обем текстов материал в корпусите улеснява определянето на някои квалификатори на заглавни думи в речниците (тъл-

ковни, неологични, синонимни) и същевременно мотивира вземането на лексикографски решения в такива случаи. Особено ценна е ролята на корпусния подход при уточняване на квалификаторите, които характеризират думите от гледище на историческата перспектива (например *остаряла дума*, *нова дума*) или според честотата на тяхната употреба (*рядка дума*, *рядко значение*).

В научната литература са изказвани възражения срещу употребата на квалификатора *рядка дума* в тълковните речници на българския език главно поради схващането, че е трудно да се определят ясни критерии за неговото приложение (Венедиктов 2000). Възможността за извличане на обективни корпусни данни за честотата на срещанията на думите и техните форми обаче са аргумент срещу подобни възражения и мотивират основателността на използването на посочената бележка при някои заглавни думи, фразеологични съчетания или отделни значения и нюанси.

Така например корпусните данни показват основателността за оформяне в *РБЕ* на заглавна дума *посаждения* и квалифициране на формата за единствено число *посаждение* като рядка: **ПОСАЖДѐНИЯ** *мн*., *ед*. (рядко) **посаждѐние** *ср*. *Спец*. … (РБЕ 2008, т. 13: 675). В БНК са регистрирани 22 употреби на множествената форма на това съществително срещу едно срещане на формата за единствено число. Също така основателно е приписването на стилистична бележка *Рядко* на глагола *поусещам / поусетя* (РБЕ 2008, т. 13: 891), който има само 8 срещания в 0,02% от общия брой текстове в корпуса.

Наблюденията върху големи масиви от текстове са необходими и в случаите, когато трябва да се прецени дали дадена дума или форма е индивидуализъм, т.е. среща се само в идиолекта на определен автор, или не. Дори най-богатите лексикални картотеки включват сравнително ограничен материал (в сравнение с корпусите) и в повечето случаи не са достатъчно надеждна основа за заключения в тази насока. Така например въз основа на данни от лексикалните картотеки на *РБЕ* прилагателното *еховит* е определено като индивидуализъм: **ЕХОВѝТ** ... *Индив*. Който предизвиква ехтене, екот, ехтеж. *По-*

късно щях да усетя хвойневия дим, щях да дочуя как към селцето слиза еховита странна песен. Н. Инджов, ПП, 91. (РБЕ, т. 4, 2. преработено и допълнено издание, под печат). В Българския национален корпус обаче се откриват и употреби на това прилагателно в произведения на други автори като Георги Марков, Траян Първанов и пр., което поставя под съмнение неговото квалифициране като индивидуализъм.

За разлика от разгледания по-горе пример корпусните данни потвърждават отразената в *РБЕ* информация, че съществителното *потай* ‘покой, спокойствие, тишина’ и множествената форма на съществителното *полъх – полъхи*, имат характер на индивидуализми: първото се среща само в произведенията на Ст. Чилингиров, а второто – само у Блага Димитрова (РБЕ, 2008, т. 13: 356, 822).

Въз основа на корпусни данни би могло да се прецени също така дали дадена дума или значение са в активна употреба в съвременната речева практика, или са остарели. Така например в *СТРБЕ* значението ‘човек, чиято работа е да кара каруца’ на съществителното *каруцар* е маркирано като остаряло (СТРБЕ: 340). Наличието на близо 450 срещания на това съществително в посоченото значение в съвременни текстове от Българския национален корпус обаче говори, че то далеч не е излязло от употреба.

4.5. Наблюдения, подпомагащи семантичната интерпретация на лексикалните единици

Този аспект на приложение на корпусите се отнася главно за общите тълковни и неологичните речници, при работата върху които най-важният и труден етап от лексикографското описание е разграничаването на отделните значения на думата и тяхното дефиниране. Семантичната интерпретация на лексикалните единици в най-голяма степен е зависима от субективните виждания и езиковия усет на лексикографа и съответно най-често може да предизвика възражения. Използването на корпусни данни несъмнено допринася за това, определянето на семантичната структура на тълкуваните думи да бъде поставено на относително по-обективна основа, което е предпоставка за

извеждане на по-прецизни речникови дефиниции.

Корпусният подход не може да бъде полезен единствено при дефинирането на термините в специалните и тълковните речници и това се дължи на спецификата на термините, които, както е известно, имат контекстно независими значения. Терминологичните значения отразяват създадените от експерти научни понятия и поради това тяхното дефиниране се извършва чрез описание на най-съществените признаци, които формират тези научни понятия (Попова 1990: 14–15).

Използването на общ, значителен по обем, представителен корпус като БНК дава възможност за достатъчно широки наблюдения върху употребата на думите в голямо множество реални контексти и извличане на по-богати данни за отделните значения на думите и затова корпусният подход е особено ползотворен при изработването на многотомния *РБЕ* във връзка с неговото предназначение да отразява по-детайлно и на по-конкретно равнище семантичната структура на заглавните думи в сравнение с по-малките по обем тълковни речници. Специфична особеност на *РБЕ* е и представянето на думи, значения и употреби, които не са типични и се срещат сравнително рядко (като в някои случаи имат характер на индивидуализми), поради което не се отразяват например в еднотомните речници. В това отношение традиционните методи за лексикографска работа като интроспекцията и използването на данни от лексикални картотеки са със значително по-малка ефективност, отколкото корпусният подход.

Така например данните, извлечени от БНК, позволяват да се направи по-детайлна семантична интерпретация на прилагателното име *разнополов* в *РБЕ* в сравнение с останалите речници на българския език, където това прилагателно присъства обикновено само с едно значение: ‘които са от различен пол’ (реализиращо се само при множественото число на думата). Наблюденията върху употребата на *разнополов* в БНК обаче показват, че семантичната структура на прилагателното е по-богата и включва още значения, нерегистрирани в съществуващите речници: ‘който е свързан с лица от различен пол’ (например *разнополов брак*) и ‘който се състои от лица от различен

пол’ (например *разнополова двойка*).

Важно предимство на корпусите е възможността с помощта на софтуерните средства за работа с тях да се извеждат редове от конкорданси на търсените думи, което улеснява наблюденията върху различните употреби и по този начин подпомага работата върху разграничаването на отделните значения. Става възможно извличането на семантична информация по един несъмнено по-лесен и достъпен начин в сравнение с традиционните методи, като при това употребата на лексикалните единици може да бъде проследена в рамките на цели текстове (или на сравнително големи извадки от текстове), а не само в откъслечни текстови сегменти (както е при работата с фишове от лексикални картотеки), което осигурява пълнота и изчерпателност на наблюденията.

Възможностите, които съществуващите софтуерни средства за работа с корпуси предлагат за сортиране на контекста на търсената дума, допълнително подпомагат идентифицирането на отделните лексикални значения, тъй като в редица случаи различното контекстно обкръжение е индикатор за наличие на различия в семантиката. Така например в изведените с помощта на системата Sketch Engine срещания на глагола *разменям* в извадка от БНК (при извършена сортировка на съседните му думи) лесно се откроява значението ‘давам някаква валута, срещу която получавам друга на същата стойност; обменям, сменям’, реализиращо се обикновено в страдателни конструкции (например *Един долар се разменя за 1.70 лв.*) (вж. Фигура 1 по-долу).

Сортировката на контекста на посочения глагол улеснява идентифицирането и на друго липсващо в съществуващите речници значение: ‘(в съчетание със съществителни като *поглед*, *удар* и под.) извършвам определено действие (уточнено от съществителното) по отношение на някого и бивам обект на същото действие от негова страна’ (например *разменям погледи*, *разменям удари*, *разменям усмивки*) (вж. Фигура 2).

Както посочват М. Рындъл и А. Килгариф: „Все повече се приема, че значението на дадена дума в даден контекст е тясно свързано със специфичния модел, в който се появява думата –

Фигура 1. Срещания на глагола разменям в БНК, изведени с помощта на Sketch Engine

Corpus: **BulgarianNC**
Hits: **85**

First | Previous Page 2 of 5 Go Next | Last

1	селски свещеник или на	**разменящите**	си непристойни за висши
36	военен конфликт , а реплики ,	**разменяни**	между част от членовете
29	стоманата стара . Омраза	**разменям**	по долар за лев , а всичко
826	седмицата щатската валута се	**разменяше**	на нива от 1 . 8075 85 марки
826	пазара в петък една марка се	**разменяше**	срещу 996 . 00 50 лева . Дилъри
138	инвестиционните посредници са	**разменяли**	акции извън борсата .
279	седмицата една марка се	**разменяше**	за 997 . 30 - 997 . 50 лева срещу
156	Първият курс - за най-активно	**разменящите**	се акции , се определя
826	седмицата щатската валута се	**разменяше**	на нива от 1 . 8075 85 марки
826	пазара в петък една марка се	**разменяше**	срещу 996 . 00 50 лева . Дилъри
138	инвестиционните посредници са	**разменяли**	акции извън борсата .
279	седмицата една марка се	**разменяше**	за 997 . 30 - 997 . 50 лева срещу
55	чужбина . Един долар се	**разменяше**	за 1 , 40 марки вчера , което
39	или . долара . Марката се	**разменяше**	за 3257 рубли . Балъков учуден
39	чужбина . Един долар се	**разменяше**	за 1 , 4110 марки вчера . БАЛКАНБАНК
76	чужбина . Един долар се	**разменяше**	за 1 , 411 марки вчера . 8
448	днес . 1 долар вчера се	**разменяше**	за 1 , 476 германски марки
44	валутна борса . 1 долар се	**разменяше**	за 4583 рубли в понеделник
65	чужбина . Един долар се	**разменяше**	за 1 , 419 марки вчера , което
136	чужбина . Един долар се	**разменяше**	за 1 , 43 марки вчера , което

First | Previous Page 2 of 5 Go Next | Last

Lexical Computing Ltd.
Sketch Engine (ver:5KE-2.34-2.63.2)

Фигура 2. Срещания на глагола разменям в БНК, изведени с помощта на Sketch Engine

Corpus: **BulgarianNC**
Hits: **85**

First | Previous Page 4 of 5 Go Next | Last

9	по за маса , а и намеците	**разменяни**	помежду ни (естествено
2	следващия ден в училище ще си	**разменяме**	потайни погледи като
20	кораба . Д . Ж . ги последва ,	**разменяйки**	пътем няколко реплики
2	отношенията . Да не си	**разменяме**	реплики пред микрофони
2	отношенията . Да не си	**разменяме**	реплики пред микрофони
94	анализирани систематично и	**разменяни**	с останалите държави
29	но големите флотилии си	**разменяли**	само дистанцирани удари
2	Със сестра си обикновено	**разменяше**	само по няколко общи фрази
156	Първият курс - за най-активно	**разменящите**	се акции , се определя
22	такава дума в езика си .	**Разменяли**	си жените за удобство
1	селски свещеник или на	**разменящите**	си непристойни за висши
0	селски свещеник или на	**разменящите**	си непристойни за висши
50	полковникът , уловени под ръка и	**разменяйки**	си усмивки , тръгнаха
94	напълно ясно , докато си	**разменяше**	сладникави безочливости
54	пазари , където 1 долар се	**разменяше**	срещу 1 , 408 марки . КНИЖАТА
826	пазара в петък една марка се	**разменяше**	срещу 996 . 00 50 лева . Дилъри
826	пазара в петък една марка се	**разменяше**	срещу 996 . 00 50 лева . Дилъри
1	Обезумели двамата бойци си	**разменяли**	страховити удари , от
45	да бъде прехвърлян или	**разменян**	съгласно определени формалности
4	се небрежно покрай тях ,	**разменяйки**	твърд поглед с моя дневален

First | Previous Page 4 of 5 Go Next | Last

там, където моделът обхваща различни синтактични и колокационни свойства, както и информация за тематичната

област на употреба” (Рындъл, Килгариф 2011: 30). Във връзка с това идентифицирането и систематизирането на клъстерите от модели на употреба (Ханкс 2004) на думите в корпус с помощта на компютърни методи е от значителна полза за лексикографа. Въпреки това обаче автоматизирането на сложната и по същество творческа дейност за идентифициране и формулиране на различните значения на думите е все още въпрос на по-далечно бъдеще. Както с основание отбелязват М. Рындъл и А. Килгариф: „Автоматизираната лексикография все още не е действителност. Далеч е моментът, когато съставянето на дефинициите (най-трудното от всичко), автоматичното определяне на значението на думата ще се извършва от компютри” (Рындъл, Килгариф 2011: 30).

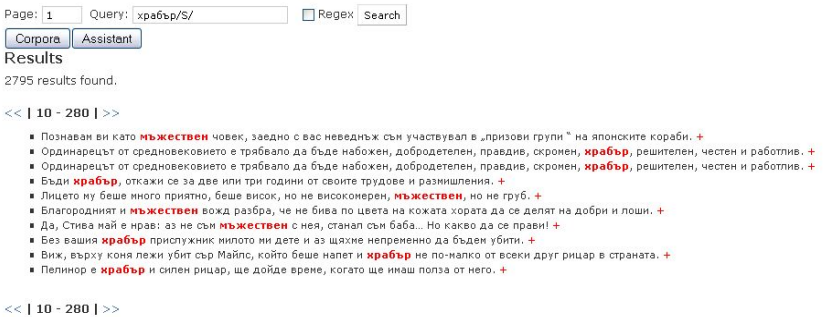
4.6. Улесняване на определянето на системните отношения, в които влиза заглавната дума

Корпусът, заедно с разработените софтуерни средства за работа с него, може да бъде източник и на извънредно ценна информация за системните връзки на тълкувана дума – синоними, антоними, хипероними и под. Тази информация е много полезна при работата както върху общи тълковни и неологични речници, така и върху синонимни речници и бази от данни.

Българският национален корпус е ценен източник на информация за системните отношения на тълкуваните думи, като достъпът до тази информация е възможен благодарение на разработената в Секцията по компютърна лингвистика в Института за български език система за търсене в корпуси, чрез която са възможни търсения по различни параметри и критерии: по форма, по лема или чрез регулярни изрази (Тинчев и др. 2007, Коева и др. 2011). Търсенето на синонимни релации на тълкувана дума се извършва чрез използване на Българския WordNet (http://dcl.bas.bg/en/wordnet_en.html), също разработен в Секцията по компютърна лингвистика (Коева 2007, Коева 2010). Така например при търсене в БНК на прилагателното име *храбър* освен конкордансни редове с негови срещания се извеждат и конкордансни редове със срещания на негови синоними като например *мъжествен* (вж. първия конкор-

дансен ред по-долу във Фигура 3). Тази информация улеснява лексикографа в установяването и по-пълното и точно отразяване на синонимните връзки на тълкувана дума с други лексикални единици.

Фигура 3. Срещания на прилагателното име храбър и негови синоними в БНК, извлечени чрез системата за търсене в БНК



При търсене в БНК на синоними на многозначни думи изведените конкордансни редове е възможно да съдържат синоними, отнасящи се към различни значения на търсената дума. Например в показаните по-долу във Фигура 4 конкордансни редове със срещания на прилагателното *тежък* и негови синоними присъства съчетанието *труден живот*, в което прилагателното функционира в значение, синонимизиращо се с едно от преносните значения на *тежък.* В случаи като този корпусната информация за синонимни релации помага на лексикографа полесно да идентифицира значенията, в които търсените лексикални единици влизат в синонимни връзки.

Корпусът е източник и на друг вид ценна информация за синонимните отношения на заглавните думи, свързана с фреквентността на отделните синоними. Тази обективна информация (в съчетание и с други критерии) подпомага лексикографа в определянето на това, кой е основният синоним в даден синонимен ред, и поради тази причина е особено полезна при работата върху синонимни речници и синонимни бази от данни. Сведенията за честотата на отделните лексикални или фразеологични единици, влизащи в отношения на синонимия, са много

Фигура 4. Срещания на прилагателното име тежък и негови синоними в БНК, извлечени чрез системата за търсене в БНК

полезни и при работата върху общи тълковни и неологични речници, като подпомагат вземането на мотивирани решения относно това, кой от синонимите е основен и изисква описателна дефиниция и кой би могъл да се тълкува чрез доминантния синоним. Такива обективни данни са важни особено по отношение на неологичните речници, представящи един много динамичен и неустойчив лексикален пласт, при работата с който понякога е трудно да се определи само по езиков усет кой е основният синоним в даден синонимен ред или основният фразеологичен израз от група равнозначни изрази. Така например информацията за честотата на лексикалните единици *сайт* и *уеб страница* в БНК, чиито срещания са в пропорция 65:1, бе обективен критерий при вземане на решение относно това, коя от двете единици да бъде смятана за основна и да се тълкува описателно в *РНДБЕ*. Корпусно-извлечена информация за фреквентността се използва и при представянето на равнозначните фразеологизми, които са често явление в българска неология. Въз основа на данните за фреквентността на равнозначните фразеологични единици *легнал полицай*, *спящ полицай* и *полегнал полицай* в БНК бе установено, че основният фразеологизъм е *легнал полицай* и той бе тълкуван описателно в *РНДБЕ*, а останали два бяха препратени към него.

4.7. Извличане на информация за лексикалната съчетаемост на думите

Данните за лексикалната съчетаемост на търсена дума и честотата на отделните нейни съчетания е друг тип лексикографски релевантна информация, която може да се изведе от корпуси. Тази информация трудно може да бъде получена по друг път и е извънредно полезна за лексикографа, като му помага да отрази значително по-пълно и адекватно съчетаемостта на съответната дума в речниковата статия.

Тук трябва да се уточни, че когато става дума за лексикална съчетаемост, в корпусната лингвистика обикновено се използва терминът колокация, въведен от Дж. Фърт (Фърт 1957) и впоследствие наложен от Дж. Синклер (Синклер 1991) във връзка със създаването и приложението на електронни корпуси. Сега този термин се използва много широко в англоезичната и немската лингвистика и лексикография (вж. Стоичков 2010 и посочената там литература), а отчасти и в славянската лексикография (Чермак 2007, Кошкова 2010). Без да коментираме различните значения, с които този термин се използва, а също и неговата неопределеност, ще уточним само, че тук както Фр. Чермак ще влагаме в него по-широк смисъл и ще го използваме за означаване на всички типични, редовно срещащи се словосъчетания на една дума – както с устойчив, така и със свободен характер (Чермак 2007: 193, цит. по Кошкова 2010: 178).

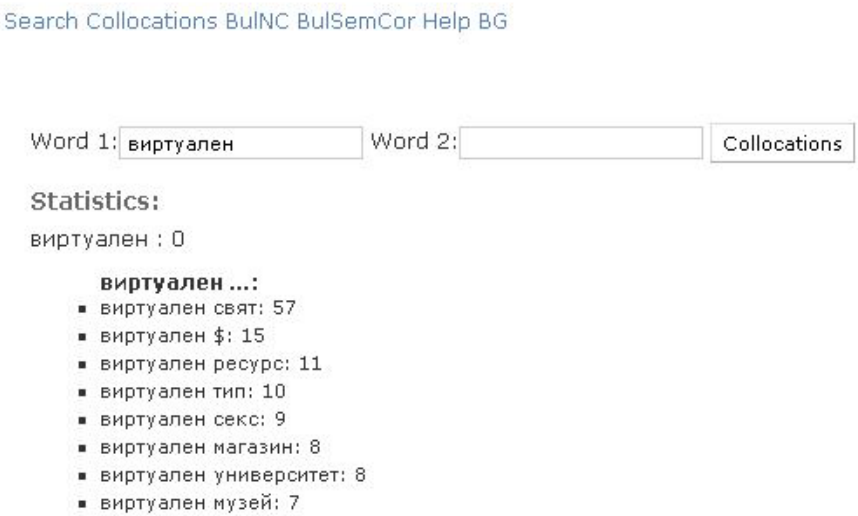
За изчисляване на важността (significance, salience) на колокациите се използват различни статистически методи: MI, MI3, T-score, log likelihood, minimum sensitivity, log dice, MI.log_f и пр. (вж. например Килгариф, Тъгуел 2001 и цитираната там литература). Системите за търсене в корпуси (corpus query systems) обикновено включват една или повече от тези функции. Така лексикографът има възможност да използва статистически подкрепена информация за обосноваване на своите решения, свързани с описание на лексикалната съчетаемост на тълкуваните единици.

Информацията за колокациите на дадена дума е ценна на първо място за общите тълковни и за неологичните речници, в които илюстрирането на значението (или значенията) на тъл-

куваната дума е немислимо без посочване на типичните словосъчетания, в които думата участва. Корпусните данни улесняват извънредно много лексикографа при установяването на тези типични съчетания.

Така например наблюденията показват, че сред изведените колокации на прилагателното име *виртуален* с помощта на системата за търсене в Българския национален корпус и на Sketch Engine като най-честотни се очертават следните съчетания: *виртуална реалност*, *виртуален секс*, *виртуална библиотека*, *виртуално пространство*, *виртуален магазин*, *виртуален свят* (вж. Фигура 5 и Фигура 6 по-долу). Някои от тях са устойчиви (например *виртуална реалност*), а други са свободни, но често срещани, типични (например *виртуален магазин*). Двата типа съчетания са представени в речниковата статия на прилагателното име *виртуален* в *РНДБЕ* съответно като част от фразеологичния блок на статията (устойчивите съчетания) и като част от илюстративния материал за съчетаемостта на прилагателното при отделните негови значения (съчетанията със свободен характер) (РНДБЕ: 102–103).

Фигура 5. Колокации на прилагателното име виртуален в БНК, изведени със системата за търсене в БНК



Фигура 6. Част от „кандидатите за колокати" на прилагателното име виртуален в БНК, изведени с помощта на Sketch Engine

Collocation candidates

Page 1 Go Next >

		Freq	T-score
p/n	вакуума	3	1.732
p/n	диапазон	4	2.000
p/n	частици	8	2.828
p/n	музикални	4	2.000
p/n	реалност	9	3.000
p/n	функционират	3	1.732
p/n	библиотека	5	2.236
p/n	двойки	4	1.999
p/n	мрежи	6	2.449
p/n	пространство	13	3.604
p/n	широк	4	1.999
p/n	страници	4	1.999
p/n	собствено	3	1.731
p/n	Интернет	3	1.731
p/n	мрежа	7	2.644
p/n	програми	4	1.997
p/n	свят	5	2.230
p/n	във	21	4.565

Особено ценна е предоставяната от системата Sketch Engine възможност за сортиране на непосредственото контекстуално обкръжение (отляво и отдясно) на търсената дума по морфосинтактични характеристики като например глаголни фрази с обект или субект, именни определителни фрази, предложни фрази и т.н.[11] По този начин се спестяват много усилия за проучване и обобщаване на множеството конкордансни редове, в които участва търсената дума.

На Фигура 7 по-долу са представени сортираните по синтактична характеристика думи от левия и десния контекст на съществителното име *момиче* в БНК, въз основа на които като най-честотни фрази с компонент *момиче* може да се изведат именните фрази *хубаво момиче*, *младо момиче*, *добро момиче*, вербалната фраза *прегръщам момиче* и др. Посочените колокации би трябвало да присъстват в речниковата статия на съществителното име *момиче* в един по-голям тълковен речник, както и сред съчетанията на съответната заглавка в двуезични речници с изходен език български.

Фигура 7. Колокати на съществителното име момиче в БНК, сортирани по синтактична характеристика с помощта на Sketch Engine

момиче () BulgarianNC freq = 3325

object_of	357	2.4
отвличам	4	8.34
поругая	3	8.05
отвърна	11	7.99
прегръщам	3	7.94
попитам	12	7.89
вълнувам	3	7.61
прошепна	3	7.51
извикам	5	7.49
заведа	3	7.4
обичам	5	7.03
кажа	22	7.01
работя	3	6.87
пусна	4	6.66
отбележа	3	6.64
погледна	4	6.55

modifies	842	2.2
хубав	54	9.71
млад	87	9.48
-годишно	18	9.32
умен	17	9.04
драконов	12	8.76
красив	18	8.55
змейски	9	8.39
леден	12	8.33
беден	11	8.18
непознат	11	8.12
любим	11	8.02
приятен	11	7.93
червенокос	6	7.82
чаровен	6	7.79
симпатичен	6	7.71

and/or	231	1.3
момче	73	10.4
дъщеря	3	6.58
жена	17	6.34
смърт	8	6.29
момиче	4	5.97
младеж	3	5.69
мъж	5	5.06
дете	3	3.98

pp_зад	7	11.7
щангистка	3	9.52

pp_с	58	1.9
дънки	4	9.26
коса	5	6.27

pp_в	62	1.0
рим	5	8.47
село	3	5.31
дом	3	4.67

pp_по	15	0.7
улица	4	5.55

pp_на	57	0.3
димитър	3	7.22
13	3	5.33
име	7	4.96
маса	3	4.44

pp_за	6	0.1
ръка	3	3.03

Детайлната синтезирано представена информация за колокациите на търсена дума (включваща и статистически данни) дава полезни насоки за разграничаването на отделните значения на думата, тъй като нерядко различните типове колокации са свързани с различни лексикални значения (Байбър и др. 1998: 35–36). Като отбелязва А. Килгариф, групирането на колокациите може да се използва като метод, подпомагащ автоматичното разграничаване на отделните значения на лексемите (Килгариф 2005).

Изведените с помощта на Sketch Engine колокации на прилагателното име *верен* ни насочват към три различни негови значения (цитирани според тълкуванията им в РБЕ, 2002, т. 2, 2. преработено и допълнено изд.: 118): ‘който съответства на истината’ (срв. на Фигура 8 по-долу колокациите *верен отговор*, *вярно твърдение*, *вярна констатация*), ‘който проявява постоянство в отношението, в чувствата си към някого или към нещо’ (срв. колокациите *верен приятел*, *верен съюзник*, *верен под-*

дръжник) и ‘който е последователен и твърд в отстояване на нещо’ (срв. *верен на традицията*, *верен на принципа*, *верен на идеята*, *верен на думата си*).

Фигура 8. Колокати на прилагателното име верен в БНК, сортирани по синтактична характеристика с помощта на Sketch Engine

верен () BulgarianNC freq = 1960

a_modifier	519	3.7
съюзник	14	9.14
отговор	52	8.62
слуга	6	8.04
приятел	21	7.9
страж	4	7.68
твърдение	6	7.56
другар	5	7.46
посока	14	7.42
поддръжник	3	7.38
последовател	3	7.33
спътник	4	7.29
поданик	3	7.29
усет	3	7.22
тълкуване	4	7.17
слух	4	6.97
клисурец	2	6.96
резервираност	2	6.95
гълъбче	2	6.92
констатация	3	6.91
теорема	2	6.89
равенство	4	6.86

pp_до	7	2.1
доказване	2	8.54
смърт	2	4.39

pp_по	10	1.3
2	7	4.24

pp_на	72	1.0
жорж	2	8.83
кремъл	2	8.64
крал	5	7.32
ярослава	2	7.24
обещание	2	6.92
традиция	4	6.51
вот	2	6.13
съпруг	2	5.72
момиче	2	5.09
принцип	4	4.8
идея	2	4.41
власт	3	3.23
дума	3	2.84
държава	2	2.24

Получаването на попълни данни за колокациите е особено ценно за неологичните речници. Това е свързано с обстоятелството, че в редица случаи появата на нови колокации на позната дума сигнализира за оформянето на ново значение в нейната семантична структура. Възможността за извеждане не само на най-честотните колокации, но и на по-малко фреквентните и по-нетипичните съчетания е важна за лексикографа, тъй като по думите на Фр. Чермак в някои случаи тези съчетания свидетелстват не за изолирани или фигуративни употреби, а за нововъзникнали типове на стандартното значение (Чермак 2010). В този смисъл данните за колокациите подпомагат извънредно трудоемката лексикографска задача по установяването на семантичния развой на думите. Така например откриването в БНК на колокации като *гореща новина*, *гореща информация* и под. подпомогнаха идентифицирането на съответното ново преносно значение на прилагателното

горещ, включено в *РНДБЕ* с дефиниция: ‘за новина, информация и под. – който е много актуален, току-що огласен, току-що станал известен’ (РНДБЕ: 119).

Корпусноизвлечената информация за колокациите е полезна за лексикографа и при идентифицирането на устойчивите съчетания, посочвани в тълковните речници в отделен блок (раздел) на речниковата статия или по друг специфичен начин. Така например сред извлечените с помощта на Sketch Engine колокации на прилагателното име *етнически* (вж. Фигура 9) се очертават две устойчиви съчетания с участие на това прилагателно: *етническо напрежение* и *етническо прочистване*, които намериха своето отражение в *РНДБЕ* (РНДБЕ: 175).

Фигура 9. „Кандидати за колокати” на прилагателното име етнически в БНК, изведени с помощта на Sketch Engine

Collocation candidates

Page 1 Go Next >

		Freq	T-score	MI	logDice
p/n	**принадлежност**	78	8.831	12.804	11.057
p/n	**малцинства**	63	7.937	13.720	10.999
p/n	**мигрантите**	55	7.416	14.494	10.906
p/n	**взаимоотношения**	60	7.744	12.048	10.563
p/n	**религиозни**	38	6.163	11.926	10.062
p/n	**прочистване**	24	4.899	13.933	9.750
p/n	**религиозна**	26	5.098	12.393	9.715
p/n	**общности**	36	5.995	10.230	9.369
p/n	**доброволността**	18	4.242	14.072	9.359
p/n	**учениците**	23	4.793	10.910	9.257
p/n	**албанци**	18	4.242	11.985	9.201
p/n	**групи**	73	8.531	9.344	9.118
p/n	**конфликти**	20	4.469	10.697	9.052
p/n	**религиозните**	15	3.872	12.095	8.984
p/n	**пол**	19	4.356	10.547	8.954
p/n	**турци**	17	4.121	10.962	8.949
p/n	**расова**	13	3.605	12.875	8.857
p/n	**произход**	34	5.821	9.250	8.757
p/n	**народност**	12	3.464	12.787	8.743
p/n	**напрежение**	19	4.353	9.611	8.595
p/n	**въпросите**	25	4.992	9.219	8.573

Информацията за контекстуалното поведение на думите е особено важна за двуезичните и многоезичните речници, в които отразяването на основните колокации на дадена заглавна единица е съществена част от речниковата статия. Дори според някои автори ползващите преводен речник се нуждаят повече от информация за колокациите, отколкото от информация за самите думи (Богаардс 1996, цит. по Верлинд, Селва 2001). Независимо дали подобно мнение е основателно, или не, със сигурност може да се каже, че изборът на типичните колокации на дадена заглавка е много важен етап при създаването на преводни речници. Това се дължи най-вече на съществуващата в редица случаи асиметрия на съчетанията на съотносителни лексикални единици в два езика (вж. Кошкова 2010 и цитираната там литература) и свързаната с нея националноезикова специфика. Посочването на колокации на заглавните думи се препоръчва и при двуезичните терминологични речници. Както отбелязва Б. Алексиев, предварително планираното включване на колокации на ключовия термин в речниковата му статия е много важно, тъй като така се подпомага преводаческата и учебната работа (Алексиев 2006). Въз основа на корпусноизвлечени данни за колокациите на заглавни думи може чувствително да се улесни и автоматизира и работата по изработване на новия жанр двуезични речници: речници на колокациите.

Честотността на колокациите на многозначна заглавна дума може да ориентира лексикографа и в подредбата на отделните значения на думата в тълковния речник, без обаче според нас това да има определящо значение. По принцип подредбата на значенията на думите в тълковните речници се обуславя от фактори от семантичен характер, свързани със семантичното развитие на думата, връзките между отделните семеми и др. В това отношение традиционно изработваните речници се различават от корпусноизведените, в които честотата на колокациите на думата в нейните отделни значения се възприема като критерий за тяхната подредба в речниковата статия, като значението, представено от най-честотните колокации, трябва да бъде посочено на първо място (Синклер 1991, Чермак 2010). По отношение на българската лексикография в посока към корпусноизведения подход е например предложението на М. Стамболиева за определяне на семантич-

ната структура на тълкуваните думи изцяло въз основа на данни от електронни архиви (Стамболиева 2003).

4.8. Извличане и подбор на илюстративен материал

Това е най-достъпното и лесно осъществимо направление в лексикографското приложение на корпусите (Рындъл, Килгариф 2011), като при тълковните и неологичните речници корпусният подход дава възможност значително да се оптимизира и ускори търсенето и систематизирането на езиковия материал, а също и да се извърши по-качествен подбор на този материал.

Така например с помощта на търсещата програма към Българския национален корпус може да се направи подбор на илюстративен материал за омоформи (например съвпадащи по форма прилагателно име и наречие). Чрез търсене с регулярни изрази може да се редуцират максимално нерелевантните случаи, като се елиминират тези примери, които се отнасят към заглавката от другата част на речта. Така например при търсенето за прилагателното име *практически* се използва регулярен израз <практически[0,0]*{POS=N}>, включващ лемата *практически*, следвана от съществително име, който отразява типичната позиция на прилагателното име пред съществително в изречението (вж. Фигура 10). При търсене в БНК за наречието *практически* се използва регулярен израз <*{POS=V}[0,0] практически>, включващ лемата *практически*, предшествана от глагол. Този регулярен израз отразява типичната за наречието позиция в изречението след глагол (вж. Фигура 11).

Много активно са използвани корпусноизвлечени примери при работата върху *РНДБЕ*, в който илюстративният материал е получен в преобладаващата си част от Лексикографския корпус като част от Българския национален корпус. БНК се използва много активно и при работата върху *РБЕ*, в който по принцип се представя богат илюстративен материал. Особено удобна за лексикографа е възможността за ограничено търсене в БНК – в рамките на отделен субкорпус или отделни субкорпуси, формирани на базата на хронологичен, стилов, тематичен или друг признак. Така лексикографът би могъл да концентрира наблюденията си върху по-тесен кръг от източници съобразно с конкретните си задачи.

Фигура 10. Срещания на прилагателното име практически в БНК – търсене с регулярни изрази

Page: 1 Query: <практически[0,0]*{POS=r Regex Search

Corpora Assistant

Results

2606 results found.

<< | 4 - 261 | >>

- Не всички от знака ще имат този късчет, нека да анализират задачите, с които се залавят, а не да се заблуждават за **практически неща.** +
- Посочва, че" учебното съдържание, учебните планове и програми се нуждаят от осъвременяване с оглед създаване на предпоставки за насърчаване на самостоятелното мислене и изграждане на **практически умения** у учениците. +
- Според основните ну критици липсвал ну **практически опит,** бил некушав, нямал снелост, образно казано, да сече и да реже престъпници в държава, намираща се в състояние на преход. +
- В делови и **практически план** усилията ви днес ще бъдат възнаградени - било с похвала, било с издигане в йерархията, било морално или материално. +
- В района на Солун и Катерини ние имаме **практически занятия** и вече повече от 1000 наши студенти са били тан на практика. +
- Уж всички се кълнат в народната воля и в мъдростта на избирателя, но стигне ли се дотам тези клетви да намерят някакви **практически изражения,** моментално се явяват изключително смислени контраргументи. +
- Опитът му стои в основата на този наръчник за **практически упражнения,** които ще ви помогнат да се отърсите от страховете си и да ги трансформирате в духовен мир, щастие и удовлетворение. +
- Ако вземем предвид достигнатото по-високо ниво на научните изследвания на Запад (особено в **практически аспект)**), нещата добиват още по-тъмни краски. +
- Следваше да развием инициативата на Рейгън за съвместно с руснаците разработване на противоракетните технологии в **практически план** – защо не задействахме в изследователската работа някои от техните учени? +
- Участвал е в **практически разработки** по борбата с криминалната и организираната престъпност със службите от Белгия, Германия, Франция, Англия и САЩ. +

<< | 4 - 261 | >>

Фигура 11. Срещания на наречието практически в БНК – търсене с регулярни изрази

Page: 1 Query: <*{POS=V}[0,0]практически Regex Search

Corpora Assistant

Results

1498 results found.

<< | 1 - 150 | >>

- Срокът на ефективност вероятно ще е значително по-дълъг, но по време на публикуването на статията само тази продължителност е **била практически** доказана. +
- Източници, близки до телекома, заявиха, че **е практически** невъзможно като член на борда на директорите той да има ограничени функции, тоест да отговаря само за прехвърлянето на т. нар. специални обекти. +
- Очевидно обратният вариант - да се издири всеки от милионите автори, техни наследници или издателства - **е практически** невъзможен за реализиране. +
- За много изолирани места това **е практически** единствен вариант за производство на ток например острови, които нямат връзка с електропреносна мрежа, или високопланински райони. +
- И въпреки всичко рано или късно човечеството ще бъде принудено да се възползва от соларните технологии все пак енергията, излъчвана от нашето слънце, си **остава практически** неизчерпаема. +
- При положение че насовата технология **е практически** една LCD, единствено липсата на големи количества монитори с диагонали над 26 инча спира залеза на дисплеите от едно време. +
- Всички останали кашания, препятания или придвижвания се смятат за уредени с предишни двустранни или многостранни споразумения, т. **е. практически** не се изисква разрешението на федералното правителство. +
- Облекчи се визовият режим, което ни позволява лесно да **пътуваме практически** по целия свят ^, изброява плюсовете на бизнес средата изпълнителният директор на „Дейзи техноподжи ^. +
- Това обаче **е практически** невъзможно, тъй като не може да се предположи кои ще бъдат купувачите, преди книжата да се появят на пазара. +
- Построена е през 1906 - 1907 г. по проект на Фридрих Грюнангер - австрийски архитект, **отдал практически** целия си творчески живот на България. +

<< | 1 - 150 | >>

Нови възможности пред лексикографите разкриват такива функции като алгоритмите за извличане на подходящи за речника примери (GDEX – Good Dictionary EXamples) и т.нар. „лексикография с отметки” (Tickbox Lexicography) (вж. по-подробно Рындъл, Килгариф 2011 и цитираната там литература), които тепърва предстои да бъдат изпробвани и въведени в родната лексикографска практика.

5. Заключение

В направеното изложение очертахме основните насоки, в които корпусният подход намира или би могъл да намери приложение в работата върху различни жанрове речници – тълковни, неологични, синонимни, терминологични, двуезични и др., както и някои конкретни наблюдения от собствения ни практически опит, свързан с използването на корпуси в лексикографската дейност. Въвеждането на корпусния подход в по-широка практика е изключително перспективно, тъй като ще постави българската лексикография на съвременни основи в съответствие с постиженията на модерните езикови технологии.

БЕЛЕЖКИ

[1] При корпусноизведения подход се изхожда от контекстуалната теория за лексикалното значение (Фърт 1957, цит. по Тонини-Бонели 2001: 87).

[2] За типологията на корпусите вж. например Синклер 1995, Макенъри, Уилсън 2001, Захаров 2005 и цитираната там литература.

[3] Разбира се, стандартите за обема на корпуса варират в отделните етапи от развитието на корпусната лингвистика. Първият в света корпус (създаден в началото на 60-те години на миналия век) – американският Brown Corpus, със своя обем от 1 милион думи е бил във всяко отношение революционен за своето време. По-късно, през 90-те години като еталон (включително по отношение на обема) се налага Британският национален корпус (British National Corpus, http://www.natcorp.ox.ac.uk/), който обхваща 100 милиона думи. На съвременния етап вече не са рядкост корпуси с обем от над милиард думи. С подобен обем в настоящия момент са например Чешкият национален корпус (Český národní korpus, http://ucnk.ff.cuni.cz/), Манхаймските корпуси на немския език (IDS-Korpora, http://www.ids-mannheim.de/kl/corpora.html), някои корпуси на китайския език и пр.

[4] Както отбелязват С. Аткинс и М. Рындъл, според закона на Зипф за отношението между ранг и честотност на думите в корпуса е необходимо да се борави с много голям обем от текстове, за да се получи адекватна информация за редките и много редките думи и значения (Аткинс, Рындъл 2008: 61).

[5] Така например приетото в Британския национален корпус съотношение между писмени и устни текстове (съответно 90% срещу

10%) е до голяма степен условно (и преди всичко прагматично мотивирано), като се има предвид, че устната текстова продукция всъщност преобладава над писмената и нейните реални количествени измерения (дори в много малък хронологичен отрязък) трудно биха могли да бъдат точно изчислени. Подобни затруднения са причина в някои случаи принципът за представителност да се пренебрегва, както е например при Манхаймските корпуси на немския език (Шимкова 2005).

[6] За постигнатите резултати и бъдещата работа по проекта *Български национален корпус* вж. по-подробно Коева, Стоянова 2009, Коева и др. 2010а, Коева и др. 2010б, Коева и др. 2011.

[7] Наблюдения в това отношение са правени например за двуезичните речници на френския език. Сравнявайки интуитивно подбран словник на двуезичен речник с изходен език френски с извлечени от корпус данни, С. Верлинд и Т. Селва установят, че 12% от заглавките в интуитивно подбрания словник въобще нямат срещане в корпуса (Верлинд, Селва 2001) и във връзка с това тяхната типичност и обичайност може да бъде поставена под съмнение.

[8] Цифрата в скобите означава годината на публикуване на текста, в който се открива най-ранната регистрация на съответната дума в корпуса, а не времето на навлизането на думата в езика.

[9] За системата Sketch Engine (софтуер за лексикално профилиране), разработена в Университета в Брайтън от А. Килгариф (http://the.sketchengine.co.uk/), вж. по-подробно Килгариф, Рындъл 2002, Килгариф 2003 и др.

[10] Повече информация за тези възможности на системата Sketch Engine вж. в Килгариф, Рындъл 2002.

ЛИТЕРАТУРА

Алексиев 2006: Aleksiev. B. Terminology structuring for learner's glossaries. – *Knowledge Organization*. Vol. 33, № 2, pp. 96–118.

Аткинс 2002: Atkins, B. T. S. Then and now: competence and performance in 35 years of lexicography. – In: *Proceedings EURALEX*. Vol. 1. Copenhagen, pp. 1–28.

Аткинс, Рындъл 2008: Atkins, B. T. S., M. Rundell. *The Oxford Guide to Practical Lexicography*. Oxford, Oxford University Press.

Байбър 1993: Biber, D. Representativeness in corpus design. – *Literary and Linguistic Computing* 8, pp. 243–257.

Байбър и др. 1998: Biber, D., S. Conrad, R. Reppen. *Corpus Linguistics: Investigating Language Structure and Use.* Cambridge, Cambridge University Press.

Благоева 2008а: Благоева, Д. Проблеми на изграждането на електронен корпус за лексикографски цели. – In: *Lexikografie v kontextu informační společnosti*. Praha, UJČ AV ČR, pp. 169–179.

Благоева 2008б: Благоева, Д. Съвременни подходи в българската неография (проблеми и перспективи). – *Български език*, № 1, с. 5–14.

Благоева 2009: Blagoeva, D. Electronic Corpora and Bulgarian New-Word Lexicography. – In: *Études Cognitives*. Vol. 9. Warszawa, SOW, pp. 143–150.

Благоева и др. 2009: Благоева, Д., С. Колковска, В. Сумрова, М. Чоролеева. *Електронизация на лексикографските ресурси*. – Лексикографски преглед, № 10, с. 5–56.

Благоева, Колковска 2006: Благоева, Д., С. Колковска. Някои проблеми на корпусната лексикография. – В: *Светът на речника. Светът в речника. Юбилеен сборник, посветен на 70-годишнината на чл.-кор. д.ф.н. Емилия Пернишка*. Велико Търново, ИК „Знак'94", с. 113–121.

Благоева, Колковска 2007: Благоева, Д., С. Колковска. Електронен корпус за целите на „Речник на българския език" – състояние и перспективи. – В: *Лексикографията и лексикологията в съвременния свят. Материали от IV национална конференция с международно участие на Българското лексикографско дружество в чест на проф. Кристалина Чолакова, София, 21-22 октомври 2006 г*. Велико Търново, ИК „Знак'94", с. 277–286.

Богаардс 1996: Bogaards, P. Dictionaries for learners of English. – *International Journal of Lexicography*, № 9 (4), pp. 277–320.

Венедиктов 2000: Венедиктов, Гр. О помете *редко(е)* в толковых словарях болгарского языка. – In: *Folia slavistica: Рале Михайловне Цейтлин*. Москва, Институт славяноведения, с. 9–13.

Верлинд, Селва 2001: Verlinde, S., Th. Selva. Coprus-based versus intuition-based Lexicography: defining a word-list for a French Learners' Dictionary. – In: *Proceedings of the Corpus Linguistics conference*. Lancaster, pp. 594–598.

Генадиева и др. 1966: Генадиева, З., К. Ничева, С. Спасова, Кр. Чолакова. *Ръководство за съставяне на Речник на българския език.* София, Изд. на БАН.

Захаров 2005: Захаров, В. П. *Корпусная лингвистика. Учебно-методическое пособие*. Санкт-Петербург, Издательство С.-Петербургского университета.

Карлсон 2000: Karlsson, F. Lexikografie a korpusova lingvistika. – In: Čermák, F., J. Klímová, V. Petkevič (eds.). *Studie z korpusové lingvistiky*. Praha, Karolinum, 2000, s. 427–454.

Карпова 2004: Карпова, О. М. Современный лексикографический портфель английской лексикографии (на материале словарей издательства HarperCollins). – В: *Межкафедральный словарный кабинет имени проф. Б. А. Ларина. XL. Сборник статей*. Санкт-Петербург, Издательство С.-Петербургского университета, с. 159–169.

Кенеди 1998: Kennedy, G. *An Introduction to Corpus Linguistics*. London – New-York, Longman.

Килгариф 2000: Kilgarriff, A. Business Models for Dictionaries and NLP. – *International Journal of Lexicography*, № 13, pp. 107–118.

Килгариф 2003: Kilgarriff, A. What computers can and cannot do for lexicography or Us precision, them recall. – In: *Proceedings of 3rd Congress of ASIALEX*, Tokyo. (http://www.kilgarriff.co.uk/publications.htm)

Килгариф 2005: Kilgarriff, A. Putting the corpus into the dictionary. – In: Ooi, V. B. Y., A. Pakir, I. S. Talib, P. K. W. Tan (eds.). *Perspectives in Lexicography: Asia and beyond*. Tel Aviv, KDictionaries, pp. 239–248.

Килгариф, Рындъл 2002: Kilgarriff, A., M. Rundel. Lexical profiling software and its lexicographic applications – a case study. – In: *Proceedings of Xth Congress of EURALEX.* Copenhagen, pp. 807–818.

Килгариф, Тъгуел 2001: Kilgarriff, A., D. Tugwell. WORD SKETCH: Extraction and Display of Significant Collocations for Lexicography. – In: *Proceedings of Collocations Workshop, ACL 2001*. Toulouse, pp. 32–38.

Коева 2000: Коева, Св. Автоматично създаване на лексемни честотни речници. – В: *За думите и речниците. Лексикографски и лексиколожки четения'98*. София, „Диос", с. 109–115.

Коева 2007: Коева, Св. БулНет (лексикално-семантична мрежа на българския език) – част от световната лексикално-семантична мрежа. – *Български език*, № 1, с. 34–50.

Коева 2010: Bulgarian Wordnet – current state, applications and prospects. – In: *Bulgarian-American Dialogues*. Sofia, Prof. M. Drinov Academic Publishing House, pp. 120–132.

Коева и др. 2010а: Коева, Св., Д. Благоева, С. Колковска. Българският национален корпус – състояние и перспективи. – В: *Лексикографията в европейското културно пространство. Материали от Петата национална конференция с международно участие по лексикография и лексикология, София, 19-20 октомври 2009 г.* Велико Търново, ИК „Знак'94", с. 289–300.

Коева и др. 2010б: Koeva Sv., D. Blagoeva and S. Kolkovska. Bulgarian National Corpus Project. – In: Calzolari, N., K. Choukri, B. Maegaard, J. Mariani, J. Odjik, S. Piperidis, M. Rosner, D. Tapias (eds.). *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10).* Valletta, European Language

Resources Association (ELRA), pp. 3678–3684.

Коева и др. 2011: Коева, Св., Д. Благоева, С. Колковска. Проектът *Български национален корпус* – резултати и перспективи. – *Български език*, 2011, № 3, с. 34–53.

Коева, Стоянова 2009: Коева, Св., И. Стоянова. Български национален корпус. – *Български език*, № 3, с. 137–145.

Косем, Кришнамурти 2007: Kosem, I., R. Krishnamurthy. A New Venture in Corpus-Based Lexicography: Towards a Dictionary of Academic English. – In: *Proceedings from Corpus Linguistics Conference*. (www.acorn.aston.ac.uk/RK-publications/2007-CL-proceedings40_Paper.pdf)

Кошкова 2010: Кошкова, М. Лексикалната съчетаемост в двуезичен речник (върху материал от български и словашки език). – В: *Лексикографията в европейското културно пространство. Материали от Петата национална конференция с международно участие по лексикография и лексикология, София, 19–20 октомври 2009 г.* Велико Търново, ИК „Знак'94", с. 175–184.

Кришнамурти 2008: Krishnamurthy, R. Corpus-driven Lexicography. – *International Journal of Lexicography*, Vol. 21, pp. 231–242.

Лийч 1992: Leech, G. Corpora and theories of linguistic performance – In: Svartvik, J. (ed.). *Directions in Corpus Linguistics*. Berlin and New York, pp. 105–122.

Макенъри, Уилсън 2001: McEnery, T., A. Wilson. *Corpus Linguistics*. Edinburgh, Edinburgh University Press.

Мейер 2004: Meyer, Ch. F. *English Corpus Linguistics. An Introduction*. Cambridge, Cambridge University Press.

Павловски 2003: Pawłowski, A. Uwagi na temat korpusu języka polskiego (reprezentatywność, aktualność, nazwa). – In: *Językoznawstwo w Polsce. Stan i perspektywy.* Opole, Polska Akademia Nauk – Uniwersytet Opolski, s. 162–180.

Попова 1990: Попова, М. *Типология на терминологичната номинация*. София, Изд. на БАН.

Попова 2010: Попова, М. Въпроси на макроструктурата на терминологичния речник. – В: *Слово и словесност. Сборник в чест на доц. д-р Ю. Балтова*. София, ИК „ЕМАС", с. 166–174.

Рындъл, Килгариф 2011: Рындъл, М., А. Килгариф. Автоматично създаване на речници: къде е границата? – *Български език*, № 3, с. 7–33.

Савицки, Хлавачова 2003: Savický, N., J. Hlaváčová. Measures of Word Commonness. – *Journal of Quantitative Linguistics*, Vol. 9, № 3, pp. 215–231.

Симов и др. 2008: Симов, К., П. Осенова, С. Видинска. Избор на словник и изработване на правописен речник на българския език на

базата на корпуса BulTreeBank. – В: *Изследвания по фразеология, лексикология и лексикография (в памет на проф. дфн Кети Анкова-Ничева).* София, АИ „Проф. Марин Дринов”, с. 407–414.

Синклер 1987: Sinclair, J. (ed.). *Looking Up: An Account of the COBUILD Project*. London, Collins.

Синклер 1991: Sinclair, J. *Corpus, Concordance, Collocation.* Oxford, Oxford University Press.

Синклер 1995: Sinclair, J. Corpus typology: a framework for classification. – In: Melchers, G., B. Warren (eds.). *Studies in Anglistics*. Stockholm, Almqvist & Wiksell.

Стамболиева 2003: Стамболиева, М. С програмите на електронния архив. За архитектурата и архитектите. – *Български език*, № 1, с. 99–102.

Стоичков 2010: Стоичков, Р. Структурно-семантично моделиране на глаголно-именните колокации (Върху материал от български и английски език). – В: *Лексикографията в европейското културно пространство. Материали от Петата национална конференция с международно участие по лексикография и лексикология, София, 19–20 октомври 2009 г.* Велико Търново, ИК „Знак’94”, с. 345–362.

Тинчев и др. 2007: Тинчев, Т., Св. Коева, Б. Ризов, Н. Обрешков. Система за разширено търсене в корпуси. – *Литературата*, № 2, с. 98–116.

Тонини-Бонели 2001: Tognini-Bonelli, E. *Corpus Linguistics at Work*. Amsterdam and Philadelphia, Benjamins.

Уи 1998: Ooi, V. B. Y. *Computer Corpus Lexicography*. Edinburgh, Edinburgh University Press.

Филмор 1992: Fillmore, Ch. „Corpus Linguistics” vs. „Computer Aided Armchair Linguistics”. – In: Svartvik, J. (ed.). *Directions in Corpus Linguistics*. Berlin and New York, 1992, pp. 35–60.

Филмор, Аткинс 1994: Fillmore, Ch. J., B. T. S. Atkins. Starting Where the Dictionaries Stop: the Challenge of Corpus Lexicography. – In: Atkins, B. T. S., A. Zampolli (eds.). *Computational Approaches to the Lexicon*. Oxford, Clarendon Press, pp. 349–393.

Фърт 1957: Firth, J. R. Modes of Meaning – In: Firth, J. R. *Papers in Linguistics 1934 – 1951*. Oxford, Oxford University Press.

Ханкс 2003: Hanks, P. Lexicography. – In: Mitkov, R. (ed.). *The Oxford handbook of computational linguistics*. Oxford, Oxford University Press, pp. 48–69.

Ханкс 2004: Hanks, P. Corpus Pattern Analysis. – In: Williams, G., S. Vessier (eds.). *Proceedings of the Eleventh Euralex Congress*. Lorient, UBS, pp. 87–97.

Хербст 1996: Herbst, T. On the way to the perfect learners' dictionary: a first comparison of OALD5, LDOCE3, COBUILD2 and CIDE. – *International Journal of Lexicography*, № 9, pp. 321–357.

Чермак 2007: Čermák, F. *Jazyk a jazykověda*. Praha, Karolinum.

Чермак 2010: Čermák, F. Notes on compiling a corpus-based dictionary. – In: *Lexikos 20 (AFRILEX 20: 2010)*, pp. 559–579.

Шаров 2003: Шаров, С. А. Представительный корпус русского языка в контексте мирового опыта. – *Научно-техническая информация*, сер. 2, № 6, с. 9–18.

Шимкова 2005: Шимкова, М. Репрезентативность корпуса как лингвистическая проблема. – В: *MegaLing'2005. Прикладная лингвистика в поиске новых путей. Материалы международной конференции*. Санкт-Петербург, „Осипов", с. 130–139.

ЦИТИРАНИ РЕЧНИЦИ

НеолР: Бонджолова, В. *Неологичен речник (за периода 1998–2003).* Велико Търново, „GABEROFF", 2003.

ПРБЕ: Попов, Д., К. Симов, П. Осенова, Св. Видинска. *Правописен речник на българския език.* София, „Наука и изкуство", 2003.

РБЕ: *Речник на българския език*. София, БАН, „ЕМАС", т. 1 (1977) – т. 13 (2008) –

РНДБЕ: Пернишка, Е., Д. Благоева, С. Колковска. *Речник на новите думи в българския език (от края на XX в. и първото десетилетие на XXI в.).* София, „Наука и изкуство", 2010.

СТРБЕ: *Съвременен тълковен речник на българския език с приложения*. 3. изд. Велико Търново, „GABEROFF", 2000.

CCELD: *Collins Cobuild English Language Dictionary*. Worthing, HarperCollins, 1987.

Индра Светла
(Прага)

ИЗГРАЖДАНЕ НА СЛОВНИКА НА ЛЕКСИКАЛНАТА БАЗА ДАННИ *ПРАЛЕКС*[*]

The index of the Pralex LDB was first built on the basis of the *Frequency Dictionary of Czech*; it was later complemented according to the *Dictionary of the Czech Literary Language* and during the basic treatment of the database items it is also complemented from the SYN Corpus (selectively so far). In order to serve further as an inventory of lexical units of today's Czech and as a base for a new monolingual dictionary of medium length, or for various specialist dictionaries, the index will continued to be expanded on an ongoing basis, namely both based on the selection of the relevant lexical units from the latest reference (dictionary) sources and also from material sources, e.g. further corpuses of written texts of the Institute of the Czech National Corpus or from our own material collections (the neological excerption database and field excerption database of the Institute for the Czech Language).

Ключови думи: лексикална база данни, словник, тълковен речник

Keywords: lexical database, index, monolingual dictionary

1. Увод

В лексикографската практика се очертават три основни възможности за изграждане на нов словник. Словникът може да се базира върху списък от словникови единици, регистрирани чрез самостоятелно извършена ексцерпция, да се създава чрез сравняване на наличните лексикографски справочници, т.е. на вече съществуващите тълковни и други видове речници, или да се изгражда, като се комбинират двете възможности.

При чешките тълковни речници събирането на лексикалния

---

[*] Изследването е проведено в рамките на изследователския проект „*Създаване на база данни на лексикалния фонд на чешкия език от началото на XX век*" (AV0Z90610521) на Института за чешки език при Чешката академия на науките, с ръководител доц. д-р Карел Олива.

материал посредством самостоятелна ексцерпция намира приложение най-вече при изготвянето на „Основен речник на чешкия език" (по-нататък ОРЧЕ[1]). За тази цел от началото на XX в. в Института за чешки език (нататък – ИЧЕ) се създава богат архив с фишове (за ОРЧЕ е използван т.нар. първи материал на лексикалния архив[2]), като при изграждането на словника на този речник са правени също така и справки с по-стари речници (като например речника на Йозеф Юнгман и на Франтишек Щепан Кот). Следващият лексикографски труд „Речник на книжовния чешки език" (по-нататък РКЧЕ[3]) се базира върху словника на ОРЧЕ, който е допълнен с единици, регистрирани чрез самостоятелно извършена ексцерпция (т.нар. втори материал). При съставянето на „Речник на книжовния чешки за училищата и широката общественост" (по-нататък РКЧ[4]) е взет под внимание най-вече словникът на РКЧЕ, към който е добавен нов ексцерпиран материал (съответно трети, четвърти и пети материал от архива с фишове на ИЧЕ). Следователно в чешката лексикографска практика се установява следният подход: с оглед на предвиждания обем и целите на съответния речник най-напред се оценява и модифицира словникът на предходния речник, тъй като всеки следващ тълковен речник на чешкия език е с по-малък обем от предходния, и едва след това се допълват нови единици в съответствие с предназначението на новия речник.

От края на 60-те години на XX в. ръчно извършваната ексцерпция започва да се прилага все по-рядко (ядрото на речниковия запас е ексцерпирано още при събирането на първия материал от архива, който е и основен лексикален материал), през 70-те години се извършва селективна ексцерпция, а през 80-те години в определен период ексцерпцията е преустановена напълно (през този период се прави ексцерпция само за целите на „Академичния речник на чуждите думи"). От началото на 90-те години ексцерпцията, извършвана в ИЧЕ към ЧАН, е насочена най-вече към събирането на неологизми, така че общоупотребявани думи и изрази се извличат преди всичко с оглед на установени промени в семантиката им.

Откриването на самостоятелно научно направление „*Съз-*

даване на база данни на лексикалния фонд на чешкия език от началото на XX век" през 2005 г. (Светла 2007, Светла 2008, Рангелова, Светла 2010) постави отново въпроса кой от методите да изберем и приложим за изграждането на новата база данни (дали да се ограничим с използването на наличните лексикографски справочници, дали да предприемем нова ексцерпция, или да потърсим други източници за извличане на материал). Освен това трябваше да решим дали словникът на лексикалната база данни (нататък ЛБД) ще бъде идентичен със словника на предвиждания речник, или не.

Анализът на достъпните източници на лексикален материал показа, че независимо от богатите лексикални картотеки в ИЧЕ в най-ново време не е била извършвана целенасочена и цялостна ексцерпция на общоупотребяваните думи и изрази от съвременни текстове (като изключим допълването на гореспоменатия архив с фишове и събирането на неологизми). Ето защо се насочихме към достъпните текстове от електронния корпус на чешкия език, които на сегашния етап изпълняват функцията на т.нар. пълна (тотална) ексцерпция. Намерението ни да се използват текстовете от корпуса беше продиктувано не само от възможността да се работи с изреченски контексти, фиксирани в електронен вид, но и поради факта, че създаването на актуален собствен материал би било прекалено трудоемко и би отнело много време. Така се придържаме и към тенденцията, която се наблюдава в съвременната лексикография на редица други страни (например Англия, Полша, Словакия и др.).[5]

2. Възможности и критерии за подбор на лексикалните единици в словника на ЛБД *Пралекс*

2.1. Подбор според честотата: ЧРЧЕ (*Честотен речник на чешкия език*)

Определянето на критериите за подбор на лексикалните единици в словника е важна част от концепцията на лексикографския труд. Базата данни *Пралекс* е изграждана като емпирична база за бъдещо лексикографско описание, т.е. преди да е изготвена цялостна концепция за съставянето на новия

тълковен речник, поради което нашата задача е особено трудна. Имайки предвид чисто описателната функция на тази фундаментална база данни с лексикален материал, подборът на заглавките в словника беше осъществен с оглед на тяхната фреквентност, така че на този първоначален етап работата беше концентрирана върху набор от лексеми с най-висока честота на употреба в съвременния чешки език. Подборът беше подчинен и на изискването всяка обработена заглавка от базата данни да има достатъчен брой срещания в източниците на материала. С други думи – основен критерий беше засвидетелстването на съответната лексикална единица в корпуса SYN2000 (респективно FSC2000) с достатъчно висока честота (в началото на нашата работа това беше единственият синхронен, лематизиран и морфологично тагиран корпус от писмени текстове, създаден в Института за чешкия национален корпус (ИЧНК).

Поради тази причина при изграждането на първоначалния словник на базата данни *Пралекс* използвахме заглавките с означена статистическа характеристика в ЧРЧЕ[6], базиран върху корпуса SYN2000, като изхождахме от предпоставката, че по брой включени единици масивът на този речник най-много се доближава до словника на еднотомния РКЧ, т.е. обхваща *ядрото* на съвременния речников запас[7] – основния лексикален фонд, който остава стабилен в езика. Допълването на словника за достигане на определения в настоящия изследователски проект общ брой заглавки на базата данни (100–120 000 единици) се предвиждаше да се осъществи на един следващ етап от работата ни, след като бъдат уточнени критериите за подбор на заглавни единици. Първоначалното ни намерение (за да бъде достигнат необходимият брой заглавки) беше да допълваме словника с поликомпонентни единици, но тъй като схващането за тях не бе достатъчно изяснено, обработката им на този първи етап беше твърде проблематична.

В основата на първоначалния вариант на словника на *Пралекс* залегнаха словниците, включени в ЧРЧЕ. Този фреквентен речник обхваща три масива: на прилагателните имена (50 000 заглавки), на съществителните собствени имена (2 000 единици) и на съкращенията (1 000 единици), и е съставен въз основа на

един-единствен критерий – честотата на думите, включени в първия корпус SYN2000 (за нуждите на ЧРЧЕ FSC2000), разработен от ИЧНК. Основното преимущество на ЧРЧЕ за нашите цели бе, че можехме да го ползваме в електронен вид, което изключително улесни работата ни по съставянето на словника. От друга страна обаче, ЧРЧЕ включва само списък от леми (лексикални форми), придружени от статистически данни, без каквато и да било друга лексикографска информация. Липсва дори характеризиране на лемите като части на речта. Освен това критериите за подбора на текстовете, върху които е базиран посоченият корпус, са твърде различни от критериите на лексикографите[8], което създаваше допълнителни затруднения. Последвалите анализи показаха, че с оглед на актуалните задачи на сегашния етап от работата ни този словник е крайно недостатъчен като основен словник за база данни с посочените цели и затова ще бъдат необходими съществени корекции и допълнения, тъй като посочената честота, извлечена от един-единствен корпус, не е достатъчно обективен критерий за съставянето на словник на съвременния чешки език в нужния обхват.

Не се потвърди и първоначалното ни предположение, че *ядрото*, формирано въз основа на честотната характеристика, ще се припокрива в по-голямата си част със словника на РКЧ[9]. Беше очевидно, че за създаването на нашия словник ще бъде необходимо да работим и с предходните тълковни речници[10].

2.2. Подбор според източниците на материал или според лексикографските справочници: корпус или тълковни речници?

Другата възможност беше да използваме корпус за съставянето на словника. Основният проблем при ползването на корпус за създаването на нашия словник се състоеше в това, че когато започнахме да работим по изследователския проект (в началото на 2005 г.), съществуваше само един корпус – SYN2000, който не можеше да ни предложи сравнително пълен списък с думи (освен 50 000 леми, включени в ЧРЧЕ, останалите единици бяха застъпени твърде слабо). Поради това се наложи да работим и с наличните тълковни речници. Въпросът беше дали да се

опрем върху словника на по-новия еднотомен речник РКЧ и впоследствие да се насочим към по-новата лексика (от 80-те години на XX в.), включително поликомпонентните единици и термините, или да се базираме върху по-стар речник със среден обем (РКЧЕ[11]), който обхваща по-широк спектър лексика – включва както основни, така и производни думи, лексикални единици както от активния, така и от пасивния речников запас, включително и от терминологията, след което да допълваме създадения словник с неологизми от различни лексикални пластове.

След като станаха достъпни останалите корпуси на ИЧНК (SYN2005 и SYN2006PUB), стана ясно, че от тях може да се извлече информация за много други думи, които вече са регистрирани в РКЧЕ. Поради това беше решено при понататъшното изграждане на словника на *Пралекс* да бъде използван РКЧЕ като основен лексикографски справочник, тъй като е много по-богат от РКЧ. За допълването на словника с лексикални единици използвахме също лемите от словообразователните гнезда в РКЧЕ, тъй като в *Пралекс* не се прилага гнездовият принцип. Иначе казано, трябваше да бъдат включени като самостоятелни заглавки и онези лексеми от РКЧЕ и РКЧ, които или изобщо не са обработени, или са добавени към съответното словообразувателно гнездо с частична информация, или са обработени селективно (например отглаголни съществителни, абстрактни съществителни на *-ost*, относителни и притежателни прилагателни и т.н.). Като самостоятелни заглавки се извеждат и всички собствени имена, редица словообразувателни компоненти, абревиатури, които в РКЧЕ са приведени към други заглавни думи, както и поликомпонентни единици, които в досегашните речници са били винаги част от речниковата статия.

За да бъде словникът на нашата база данни сравнително пълен, на настоящия етап от работата си го допълваме с почти всички лексеми, регистрирани в РКЧЕ като леми, т.е. не само със самостоятелно изведените заглавни думи, но и с всички варианти, посочени към заглавките, с производни думи, дадени в края на разработваната статия, и с омоними, присъединени към

основната словникова статия, а при съществителните и с формите за женски род, образувани от мъжки род, които в РКЧЕ са посочени в скоби след изходната лема. Не добавяме само архаизмите (т.е. думите, които в РКЧЕ са означени като остарели), както и онези най-редки думи, които не са регистрирани в корпуса.

Тъй като изгражданата база данни е замислена като подготовка, т.е. емпирична база за бъдещ лексикографски труд (евентуално и за други лингвистични изследвания), много по-целесъобразно е, ако при създаването на нейния словник бъде отразен целият процес за вземане на решения и извършване на справки, вместо да бъде представен само крайният резултат, т.е. окончателният подбор за бъдещия речник.

Въз основа на събраните сведения в *Пралекс* ще бъде възможно да се уточнят критериите за подбор на заглавните единици в подготвяния тълковен речник съобразно с неговото предназначение, тип и обхват. Смятаме, че събраните до момента сведения по отношение на словника, който е на етап извършване на справки, ще бъдат използвани за по-нататъшна работа, като част от резултатите може да бъдат използвани за изследване на съвременното състояние на речниковия състав на чешкия език спрямо времевата ос (по този начин ще получим обща представа както за остарелите думи, така и за актуалното развитие на неологизмите в съвременния чешки език).

Макар че нямаме за цел да представим изцяло лексикалния състав на съвременния чешки език, стремежът ни е да извлечем и да характеризираме потенциалните деривати, регистрирани в РКЧЕ, за да се изясни актуалното им състояние с оглед на тяхната употреба в съвременния писмен чешки език. В крайна сметка словникът на *Пралекс* вероятно ще обхване почти всички леми, включени в РКЧЕ, така словникът на този речник ще послужи като отправна точка и долна времева граница при оценката на съвременния узус и честотата на употреба на масива от леми в текстовия корпус.

От друга страна, необходимо е да се вземе под внимание, че някои остарели думи, които са били актуални по времето на подбора на заглавките и тяхната обработка за РКЧЕ, не са били включени в речника по една или друга причина (например ня-

кои некнижовни, тясноспециализирани, сленгови думи, думи табу и др.), така че границите, очертани от словника на РКЧЕ, също не може да се приемат като решаващ и единствен критерий. Тук може да бъдат отнесени и някои вече регистрирани названия на остарели реалии, които на практика са изчезнали от узуса, но все още можем да се натъкнем на тях (например при посещение на етнографски музей, при четене на по-стари текстове), или такива лексикални единици, чиято употреба е възобновена, така че преминаването на времевите граници може да се осъществява в различна посока.

След като бъде направена преценка на словника на РКЧЕ, ще се насочим към по-новата лексика, регистрирана в издадените след него речници (тълковни и специализирани, например НАРД, РН1, РН2). Предполагаме, че нашият словник значително ще надхвърли предвиждания брой заглавки (100 000 единици), но при по-нататъшната работа това значително ще облекчи окончателния подбор за бъдещия речник, тъй като обработените заглавки в базата данни ще бъдат допълнени с примери или поне ще бъдат коментирани с оглед на тяхното засвидетелстване в корпуса, с който работим. Даваме си сметка, че част от периферната лексика, която е била включена в словника на базата данни (например някои стилистично маркирани единици, принадлежащи към субстандартния пласт или към периферни сфери на комуникацията и др.) по една или друга причина ще остане обработена само в *Пралекс* и няма да бъде включена в бъдещия речник поради неизбежното ограничение за обема му. Именно затова се налага словникът на *Пралекс* да е по-обемен от този на бъдещия речник. От една страна, така ще бъде извършен по-прецизен подбор, от друга страна, базата данни ще изпълнява основната си функция – да бъде източник на разнообразна информация за съвременното състояние на лексикалната система, включително и за нейните периферни части.

3. Представяне на отделните части на речта и на видовете заглавни единици

В словника на *Пралекс* ще бъдат обхванати всички части на речта. Засега обаче обработката на базата данни протича после-

дователно по отделните части на речта, поради което попълването им се намира на различен етап – първоначално (от 2007 г.) започнахме с обработката на съществителните имена[12]; в края на 2008 г. пристъпихме към извеждането в самостоятелни заглавки на сложните съществителни имена от словообразувателните редове, засвидетелствани в РКЧЕ[13]; започна обработката на някои по-малобройни части на речта (местоимения, числителни и междуметия), които бяха извлечени изцяло от РКЧЕ; през 2009 г. преминахме към системно допълване с прилагателни имена и наречия; до края на 2010 г. се предвиждаше в словника на *Пралекс* да бъдат застъпени всички глаголи от РКЧЕ.

Що се отнася до общата типология на заглавните единици, в словника на *Пралекс* различаваме еднокомпонентни и поликомпонентни единици, части от думи, абревиатури, така че наред с еднокомпонентните единици словникът бе попълван постепенно и с други видове заглавки въз основа на РКЧЕ: найнапред през 2008 г. в словника бяха включени всички фразеологични единици, които в РКЧЕ са добавени в края на речниковата статия на съответната еднокомпонентна заглавка след знака ромб (♦)[14]. През 2009 г. в словника бяха добавени почти всички съставни части на думи[15], през 2010 г. се съсредоточихме предимно върху попълването на масива на съкращенията[16] и на масива на съществителните собствени имена[17] (посочените два вида заглавни единици бяха извличани и от други източници).

За разлика от тълковните речници засега не сме се придържали към традиционното изискване, според което в словника трябва да бъдат застъпени пропорционално както различните компоненти на речниковия състав[18] (като се отчита тяхното функционално разслоение), така и различните видове заглавни единици (думи, съставни части на думи, словосъчетания). Поскоро стремежът ни е да обхванем в пълнота отделни подсистеми, поради което прибягваме и до други източници.

Всички леми в словника на *Пралекс* се преценяват и при необходимост се коментират съобразно с наличните сведения за тях в корпуса SYN, обединяващ SYN2000, SYN2005 и SYN2006PUB.

4. Използване на други източници за допълване на словника

За систематичното протичане на работата ни се предвижда, след като изготвим нашия списък от лексикални единици и техните варианти (извлечени от РКЧЕ), той да бъде съпоставен със словника на *„Академичния речник на чуждите думи*" (АРЧД, респективно ГАРЧД), за да го допълним с подбрани чужди думи; за по-новия лексикален пласт ще използваме допълнението към РКЧЕ: *„Какво няма да открием в речниците*" (Сохова, Пощолкова 1994), както и речниците *„Нови думи в чешкия език. Речник на неологизмите*" 1, 2 (РН 1, РН 2); за специализираната лексика ще бъдат правени справки и допълнения относно различните специалности и сфери на човешката дейност, като ще се опрем върху специализирани речници и енциклопедии. По аналогичен начин ще бъдат използвани и други източници на материал и бази данни.

Що се отнася до източниците на материал, засега словникът на базата данни е допълван системно най-вече от корпуса SYN[19] – в случаите, когато при обработката на някои единици от базата данни се натъкнем на други единици, които са свързани с обработваната единица по един или друг начин. Впоследствие смятаме да разширим емпиричната база[20] – в скоро време се предвижда корпусът SYN да утрои обема си, след което ще можем да правим справки и да отчитаме дистрибуцията на лемата в други корпуси, както и в други електронни източници (като базите данни на фирмата „Нютон Медиа АД"[21]), или в интернет посредством различни търсачки.

Не можем да не споменем и двете бази данни, изграждани в ИЧЕ, в които се съхранява значителен обем лексикален материал, натрупван и архивиран с помощта на компютърната техника.

За по-нататъшното допълване на словника с нови думи ще бъде използвана базата данни с неологизми *Неомат* (програмата за ексцерпиране се нарича *EDA*, от означението ***E**xcerpční **da**tabáze*), която се изгражда от началото на 90-те години на XX в. Основното предназначение на тази база данни е да отразява динамиката в речниковия състав и навлизането на нови думи.

Нейната първата част, наречена *Архив 1*, съдържа около 202 000 записа, извадки от текстове в електронен вид[22]. За нуждите на бъдещата лексикографска работа ексцерпцията продължава в масива, наречен *Архив 2*.

За извършването на справки и допълването на словника със специализирана лексика ще се опираме и на специализираната база данни *ODE*[23], чието предназначение е да извлича материал от учебниците за началните и средните училища, от различни специализирани речници и енциклопедии, както и от други научнопопулярни текстове, така че да допълва материала от корпусите. На първо място изхождаме най-вече от печатните източници, които имаме на разположение в библиотеката на лексикографския отдел, като се стремим да обхванем и онези източници, които не са достъпни в електронен вид. Тази по същество ръчна ексцерпция, която обаче се архивира в електронен вид, се извършва с цел да се съберат не само тясноспециализирани изрази от съответните области, но и данни за производни прилагателни имена и поликомпонентни единици с компонент съществително, характерни за специализираните и учебните текстове (най-вече отглаголни съществителни и абстрактни названия, т.е. предимно названия за действия, състояния, процеси, методи и свойства).

Специализираната база данни ще ни даде възможност да се постигне нужният баланс както по отношение на несъразмерното представяне на отделните специализирани области в използваните от нас корпуси, така и по отношение на неравномерния подбор спрямо степента на специализираност на отделните текстове от корпусите; евентуално биха се очертали и други сфери на комуникация. Тази допълнителна ексцерпция ще ни помогне да приведем и подходящ материал, илюстриращ употребата на някои общоупотребявани лексеми като термини, и евентуално да засвидетелстваме дистрибуцията им в други области. Не на последно място ексцерпцията има за цел да открие подходящи контексти, с чиято помощ да се уточни семантичното съдържание на различни термини.

Така въз основа на извършената от нас ексцерпция в специализираната база данни ОДЕ постепенно ще се оформят

отделни специализирани словници. В базата данни *Пралекс* ще влезе само подбрана част от всяка специализирана област, тъй като селекцията е задължителна за един общ тълковен речник, а главното предназначение на *Пралекс* е, както вече беше казано, да бъде емпирична основа за бъдещ тълковен речник.

При подбора на словниковите единици за бъдещия тълковен речник можем да използваме като допълнителен източник и дигитализирания лексикален архив с фишове на ИЧЕ, който, както вече беше посочено, отразява едно по-старо състояние на лексикалната система на чешкия език. Сканираните фишове с ексцерпиран материал може да се приемат като по-стара добавка към съвременните езикови корпуси и да послужат както при допълването на словника на лексикалната база данни с релевантни единици, така и при обогатяването на някои нерегистрирани в корпусите единици с важни сведения за употребата им във времето. В рамките на настоящия изследователски проект не се предвижда да допълваме лексикалната база данни с единици от по-старите лексикални архиви.

БЕЛЕЖКИ

[1] Речникът излиза в периода 1935–1957 г.

[2] Този материал е дигитализиран и публикуван в интернет на адрес http://bara.ujc.cas.cz/psjc/, достъпен е също така на уеб страницата LEXIKO (http://lexiko.ujc.cas.cz, рубрика Databáze heslářů).

[3] Първото издание на речника излиза през 1960–1971 г., второто – през 1989 г.

[4] Първо издание 1978 г., второ издание (преработено и допълнено) 1994 г., трето издание (преработено) 2003 г., четвърто издание (идентично с третото) излиза през 2004 г., 2005 г. и 2009 г.

[5] Срв. ОРАЕ 2005, РССЕ 2006, ГРПЕ и др.

[6] ЧРЧЕ беше издаден в книжен и електронен вариант през 2004 г.

[7] „Речниковият запас на всеки език се състои от ядро и периферия. Схващането за основния речников запас е различно: например може да става въпрос за основния лексикален фонд, съставен от думи, присъстващи отдавна и трайно в езика, или да се приеме структурната теория за център и периферия, според която центърът се състои от думи, които са непроизводни, основни по отношение на семантиката си, стилистично неутрални, продуктивни и високочестотни” (ЕРЧЕ 2002: 244).

[8] Както посочват авторите на ЧРЧЕ, „SYN2000 се базира на прецизно балансирана и пропорционална извадка от всички писмени жанрове на чешкия език, като обхваща съвременния език на вестниците, художествената литература и специализираните публикации". Това обаче не означава, че е извършен систематичен подбор на лексиката от различни специализирани сфери и области на живота.

[9] Анализът на ЧРЧЕ показа, че въз основа на честотната им характеристика в някои случаи са представени цели словообразователни редове, например: *rod*, *rodák*, *rodič*, *rodička*, *rodičovství*, *rodičovský*, *rodina*, *rodinný*, *rodilý*, *rodiště*, *rodit*, *rodný*, *rodový*, *rodokmen*; *relace*, *relační*, *relativismus*, *relativita*, *relativnost*. В подобни редове обаче понякога най-фреквентната дума е съществително име, в други случаи – глагол или прилагателно име. В ЧРЧЕ например е включено прилагателното име *radiologický*, докато съществителните имена *radiologie* и *radiolog* не присъстват.

[10] Тъй като е необходимо материалът от корпуса да бъде съотнесен към съответните значения на всяка дума, в илюстративната част към заглавките на базата данни е наложително да се опираме върху дефинициите в някой тълковен речник.

[11] В предговора към РКЧЕ се посочва, че речникът съдържа 192 000 речникови статии, освен това обаче към съответните словообразователни гнезда са включени още лексикални единици, така че реалният брой е по-висок.

[12] В съответствие с нашата концепция в базата данни са изведени като самостоятелни заглавни думи всички омоними, варианти, а при съществителните имена и формите за женски род, образувани от названия за лица от мъжки род, както и умалителните форми.

[13] Става дума за над 6 000 словникови единици.

[14] В началото този масив наброяваше около 9000 словникови единици, които впоследствие бяха обработени.

[15] Общо наброяват около 1 000. Това са предимно сегменти, състоящи се от първата част на сложни думи; в РКЧЕ са обработени самостоятелно около 650 такива единици.

[16] Първоначалният масив от ЧРЧЕ съдържаше около 1 000 единици.

[17] От ЧРЧЕ бяха извлечени 2 000 единици.

[18] Срв. Филипец 1995: 16.

[19] От 2011 година този корпус беше преименуван на ORIG_SYN.

[20] В *Пралекс* под буква **а** са включени 4221 заглавни единици (включително съкращения и поликомпонентни единици, засега без глаголите). В първия корпус SYN2000 списъкът с леми под буква **а** наброява 2953 единици, в общия корпус SYN те са вече 6014. Това оз-

начава, че ако подборът за словника не е подчинен на строга селекция, броят на единиците под всяка буква би могъл най-малко да се удвои – достатъчно е да се съберат лемите от корпуса SYN и от отделните списъци (използвани, примерно, за допълване със съществителни собствени имена и съкращения).

[21] Фирмата Нютон Медиа АД (NEWTON Media, a. s.) създава и поддържа комерсиални бази данни, съдържащи текстовете на всички мас-медии с национално разпространение, включително преписи в текстов вид на предаванията, излъчвани по националната телевизия и националното радио, както и съдържанието на редица интернет страници.

[22] Тази част от неологичната ексцерпция, която послужи за изработването на РН 1 и РН 2, вече е достъпна за широката общественост в интернет на адрес www.neologismy.cz.

[23] Названието е създадено като съкращение на израза ***O**borová **da**tabáze **e**xcerpčné* (ODE), т.е. база данни, която съдържа материал от специализирани текстове от различни области на човешката дейност.

ЛИТЕРАТУРА

Рангелова, Светла 2010: Рангелова, А., И. Светла. За създаването и изграждането на лексикалната база данни Пралекс. – В: *Лексикографията в европейското културно пространство. Материали от Петата национална конференция с международно участие по лексикография и лексикология, София, 19–20 октомври 2009 г.* Велико Търново, ИК „Знак ’94”, с. 287–300.

Светла 2007: Světlá, J. The possibilities and limits of lexikographical description of the Czech lexicon in database form. – In: *Computer Treatment of Slavic and East European Languages. Slovanské a východoeurópske jazyky v počítačovom spracovaní. Fourth International Seminar (Bratislava 25–27 October 2007).* Bratislava, Tribun, s. 244–248.

Светла 2008: Světlá, J. K návrhu, vývoji a funkcím lexikální databáze češtiny. – In: *Lexikografie v kontextu informační společnosti*. Praha, ÚJČ AV ČR, s. 19–32.

Филипец 1995: Filipec, J. Teorie a praxe jednojazyčného slovníku výkladového. – In: *Manuál lexikografie*. Jinočany, H & H, s. 14–49.

ЦИТИРАНИ РЕЧНИЦИ

АРЧД (Академичен речник на чуждите думи)**:** *Akademický slovník cizích slov*. 1. vyd. Praha, Academia, 1995, 1. díl A–K, 2. díl L–Ž, 834 s.; 1998, A–Ž, 834 s.

ГАРЧД (Голям академичен речник на чуждите думи)**:** *Velký slovník*

cizích slov. Verze slovníku 2.0. Voznice, Leda spol. s r. o., 2005. (CD-ROM).

ГРПЕ (Голям речник на полския език)**:** *Wielki słownik języka polskiego*, Warszawa, 2007-2012, http://www.wsjp.pl/.

ЕРЧЕ (Енциклопедичен речник на чешкия език)**:** *Encyklopedický slovník češtiny*. Praha, Nakl. Lidové noviny, 2002. 604 s.

РН 1 (Речник на неологизмите 1)**:** Martincová, O. a kol. *Nová slova v češtině. Slovník neologizmů 1*. Praha, Academia, 1998. 356 s.

РН 2 (Речник на неологизмите 2)**:** Martincová, O. a kol. *Nová slova v češtině. Slovník neologizmů 2.* Praha, Academia, 2004. 568 s.

НАРЧД (Нов академичен речник на чуждите думи)**:** *Nový akademický slovník cizích slov A–Ž*. Praha, Academia, 2005. 880 s.

ОРАЕ (Оксфордски речник на английския език)**:** *Oxford Dictionary of English*. Oxford, Oxford University Press, 2003, 2005.

ОРЧЕ (Основен речник на чешкия език)**:** *Příruční slovník jazyka českého*. Praha, Státní nakladatelství – SPN, 1935–1957.

РКЧ (Речник на книжовния чешки)**:** *Slovník spisovné češtiny pro školu a veřejnost*. Praha, Academia, 2003. 647 s.

РКЧЕ (Речник на книжовния чешки език)**:** *Slovník spisovného jazyka českého*. 2., nezměněné vyd. Praha, Academia, 1989.

РССЕ (Речник на съвременния словашки език)**:** *Slovník súčasného slovenského jazyka*, т. 1, A–G. Bratislava, Veda, 2006. 1133 s.

Сохова, Пощолкова 1994: Sochová, Z., B. Poštolková. *Co v slovnících nenajdete*. *Novinky v české slovní zásobě.* Praha, Portál. 204 s.

ЧРЧЕ (Честотен речник на чешкия език)**:** *Frekvenční slovník češtiny*. Praha, Nakl. Lidové noviny, 2004. 596 s.

ЕЗИКОВИ КОРПУСИ И ЕЛЕКТРОННИ ИЗТОЧНИЦИ

Databáze heslářů [online]. Verze 4.5. Praha. Ústav pro jazyk český AV ČR, v. v. i., 2008. <http://lexiko.ujc.cas.cz/heslare/>.

Český národní korpus – FSC2000. Ústav Českého národního korpusu FF UK, Praha 2004. <http://ucnk.ff.cuni.cz/english/fsc2000.php>.

Český národní korpus – SYN2000. Ústav Českého národního korpusu FF UK, Praha 2000. <http://ucnk.ff.cuni.cz>.

Český národní korpus – SYN2005. Ústav Českého národního korpusu FF UK, Praha 2005. <http://ucnk.ff.cuni.cz>.

Český národní korpus – SYN2006PUB. Ústav Českého národního korpusu FF UK, Praha 2006. <http://ucnk.ff.cuni.cz>.

LEXIKO – webové hnízdo o novodobé české slovní zásobě a výkladových slovnících [online]. Verze: 1.8.2. Praha. Ústav pro jazyk český AV ČR, v. v. i., 2009. <http://lexiko.ujc.cas.cz/index.php>.

Textový archiv Newton Information Technology, s.r.o., <www.newtonit.cz>.

Индра Светла, Албена Рангелова
(Прага)

ЛЕКСИКАЛНАТА БАЗА ДАННИ *ПРАЛЕКС* – СТРУКТУРА И ОБРАБОТКА НА ЗАГЛАВНИТЕ ЕДИНИЦИ*

The article deals with the programme and content parts of the treatment of the Czech lexis in the form of a lexical database. It briefly outlines the development of the Praled software, describes the overall structure of the Pralex LDB in terms of its macrostructure and microstructure and presents the main principles of the treatment of database items (data filled in compulsorily with each entry). The conclusion is devoted to the possible uses of this lexical database not only for a future monolingual dictionary and specialist dictionaries but also for various linguistic research projects.

Ключови думи: лексикална база данни, словник, тълковен речник, методология

Keywords: lexical database, index, monolingual dictionary, methodology

Увод

Основната цел на изследователския проект *„Създаване на база данни на лексикалния фонд на чешкия език от началото на XXI век”*[1], разработван в Института за чешки език при Чешката академия на науките (ИЧЕ ЧАН), е да се подготви разработването на модерен тълковен речник на чешкия език, като се изследва речниковият запас и най-новото му състояние както в теоретичен аспект, така и с оглед на езиковия материал. Научният контекст на проекта е определен от развитието на корпусната лингвистика и навлизането на новите информационни технологии в научната работа. Обща тенденция в развитието на модерната лексикография е създаването на обширни лексиког-

* Изследването е проведено в рамките на изследователския проект *„Създаване на база данни на лексикалния фонд на чешкия език от началото на XX век”* (AV0Z90610521) на Института за чешки език при Чешката академия на науките, с ръководител доц. д-р Карел Олива.

рафски трудове въз основа на езикови корпуси, срв. например ОРАЕ (2003, 2005), РССЕ 2006. Друга важна тенденция е изграждането на бази данни с възможност за различно лексикографско приложение, които се финансират от издателствата (Аткинс, Рындъл 2008: 263–379).

Методологично базата данни *Пралекс* беше замислена така, че нейното съдържание, т.е. събраната лингвистично релевантна информация, да даде основата **за изготвянето на бъдещ тълковен речник на съвременния чешки език**, който да служи еднакво добре както на широк кръг потребители, така и за целите на информационните технологии, насочени към обработване на естествени езици (Рангелова 2008, Рангелова, Светла 2010). Ето защо отделните заглавки в базата данни ще бъдат **основата, върху която ще се разработват съответните речникови статии**. Оттук произтича и предимно описателната функция на базата данни – основната ни задача е да съберем репрезентативен езиков материал, който да представя състоянието на съвременния узус, а не да извършим цялостен анализ на съвременната лексикална норма – това ще бъде следващата ни задача, към която ще пристъпим, след като натрупаме достатъчно лексикален материал в базата данни. Предвижда се в базата данни да бъдат обхванати 100 000 – 120 000 лексикални единици. За разлика от речниците под **самостоятелна единица в базата данни** разбираме всяка еднокомпонентна лексикална единица и нейните възможни вариантни проявления, омоними, съкращения и словообразователни елементи, както и поликомпонентни единици, изведени самостоятелно в словника.

Работата по това научно направление започна през 2005 г., когато след анализ на достъпната информация за подобни проекти (*Canoo.net*, *DWDS*, *Celex*, *TLFi*)[2] се пристъпи към разработването на лексикална база данни, като за целта се обърнахме към Центъра за обработка на естествения език към Факултета по информатика на университета „Т. Г. Масарик" в гр. Бърно (относно нейната структура и функциониране вж. Светла 2007, Светла 2008, Рангелова, Светла 2010). В резултат на това сътрудничество беше създаден лексикографски софтуер за чешкия език, наречен ***Пралед*** (***Praled***). Названието възникна в

Центъра за обработка на естествения език на Масариковия университет в Бърно от означението ***Pra**žská **le**xikální **d**atabáze* („Пражка лексикална база данни”). *Пралед* представлява **система** за обработка на лексикографски релевантни данни[3], която съдържа основното ядро на базата данни, организирано на принципа клиент – сървър, формуляри с полета за записване на информацията, програми за тяхното изобразяване и сортиране на данните („програмният” компонент от нашата задача беше представен на конференцията *Граматика и корпус* 2005[4]). Вторият важен компонент е създаването на съдържанието на базата данни. За да го разграничим от софтуера, този компонент беше наречен ***Пралекс*** (***Pralex* –** съкращението възникна от израза ***Pra**žský **lex**ikon,* Пражки лексикон) – става въпрос за **база данни от думи, словоформи и словосъчетания**).[5]

През 2007 г., след като софтуерът беше разработен и тестван, започна въвеждането на лексикографски данни, като концепцията за съдържанието на лексикалната база данни (понататък ЛБД) беше актуализирана (първоначално тя беше замислена като речникова база данни с название Лексикон 21), а концепцията за самата база данни беше променена, т.е. беше пренасочена към създаването на широка емпирична и методическа основа за бъдещия лексикографски труд. Ето защо беше необходимо да се създаде **теоретична концепция** за представянето на чешката лексика в база данни от този тип; тъй като концепцията не беше оформена окончателно, беше необходимо да се доуточнява в процеса на работата ни в зависимост от актуалните задачи, които решавахме. При разработката на програмата беше изяснена концепцията относно цялостната структура на базата данни, т.нар. формуляри за записване на данните за отделните части на речта, включващи всички менюта, от които се избират различни стойности, и полетата, предвидени за вписване на текст. На следващия етап се съсредоточихме върху съдържанието, т.е. бяха уточнени принципите за обработка на съществителните имена, а в последствие и на останалите части на речта, както и на отделните типове заглавни единици.

Основното ни усилие бе насочено към съществителните имена, от останалите части на речта вече са обработени между-

метията, числителните и местоименията (работата ни протичаше в посочената последователност, като върху всеки от посочените неголеми масиви работеше един от членовете на работния екип). Що се отнася до видовете заглавни единици, наред с еднокомпонентни заглавки се обработват и поликомпонентни (фраземи и словосъчетания). В словника бяха включени и заглавки от *Речника на книжовния чешки език* (РКЧЕ), представляващи части от думи (словообразователни елементи).

1. Структура на ЛБД *Пралекс*

В **макроструктурата** на базата данни се различават еднокомпонентни и поликомпонентни заглавни единици и съкращения; като самостоятелни единици се извеждат и някои части от думи – засега те само се ексцерпират от наличните речници, без да се обработват допълнително. Заглавките се подреждат по азбучен ред; единиците, които се изписват с главна или съответно с малка буква, се извеждат самостоятелно в словника, така че на практика са налице два обединени / преплетени азбучни реда. Макар че е предвидена възможност за едновременно търсене, първоначалните настройки различават малки и главни букви, така че се работи или с масив от единици, започващи с малка буква (това са най-вече съществителни нарицателни имена, както и някои съкращения и означения), или с масив от единици, започващи с главна буква (предимно съществителни собствени имена и абревиатури).

Според предлагания подход всяка лема е изведена като самостоятелна единица в базата данни, което на практика означава, че всеки вариант, било то формален (дума – словосъчетание), било то само графичен (дублети), е представен самостоятелно в словника. Така че се изключва използването на словообразователните гнезда като организиращ принцип на микроструктурата. Този тип системни отношения са представени чрез активни хипертекстови връзки.

На равнище **микроструктура** към всяка единица от базата данни се въвежда конкретна информация за заглавната дума и нейните проявления. Самостоятелно изведената единица в словника на ЛБД *Пралекс* наричаме заглавка, или заглавна

единица, макар че не се припокрива напълно с речниковата заглавка; обработката на нашите заглавни единици като словникови единици от лексикалната база данни в много отношения се различава от обработката на заглавките в един традиционен тълковен речник, независимо че отделни части от тях формално наподобяват обичайна речникова статия (отразяват различните аспекти на съществуването и функционирането на лексикалните единици, например дефинирането на значението). Структурата на „статията" в базата данни – в съответствие с речниковото разбиране за структура на лексемата – е йерархична. Различаваме, разбира се, еднозначни и многозначни единици, но в базата данни дори еднозначните единици са представени като йерархична структура. Поради това към всяка единица от ЛБД е възможно да се въведат данни, имащи отношение към заглавката като цяло (в т.нар. **карта Заглавка**), а впоследствие и данни, отнасящи се към различните значения (т.нар. **карта Значение**).

От самото начало сме се стремили да създадем не само формуляр за записване на данни, но и прегледно и удобно от гледна точка на потребителя изобразяване на цялата информация за заглавната единица, което максимално да наподобява структурата на речниковата статия в традиционен тълковен речник. Използваме и всички възможности на изчислителната техника, с които разполагаме благодарение на електронната обработка на данните, включително и хипертекстова връзка между отделните заглавки за проследяване на системните отношения. Освен с изображение на една или повече заглавки можем да работим и с изображението на т.нар. **Словник** – това е началното изображение, върху което в табличен вид са представени подредените една под друга заглавки с най-важните данни за тях в съответната колона, включително данни от административен характер.

2. Обработка на заглавните единици в ЛБД *Пралекс*

Основната обработка на всяка единица (с работно означение Заглавка) от базата данни е многостепенен процес, който започва с подбор на заглавката, включването ѝ в словника и по-

сочване на източника, от който е ексцерпирана. След това се определя формата, съответно и видът на заглавката, каква част на речта е, ако е необходимо, се посочват омоними и варианти, докато се стигне до създаването на дефиниция и привеждането на илюстративен материал.

Всяка заглавна единица („статия”) в ЛБД съдържа:

• **лема**: лемата съответства на основната форма в речниковите статии, в базата данни обаче в полето, предвидено за лемата, се посочва една-единствена форма, т.е. без вариантите (в ЛБД те се извеждат самостоятелно) и без подзаглавки, които в някои речници биват добавяни към изходната речникова статия – например омоними в една обща речникова статия (изграждането на словника е разгледано по-подробно в статията „Изграждане на словника на лексикалната база данни *Пралекс*” на И. Светла в настоящия сборник);

• **знак за омонимия**: омонимите се означават с римски цифри и на този етап ги подреждаме според частите на речта;

• **тип на заглавната единица**: вж. по-горе, прокараното разграничение намира отражение и в макроструктурата на ЛБД *Пралекс*;

• **част на речта / тип**: определянето на заглавката като част на речта се извършва в съответствие с РКЧЕ, в някои случаи се коментират различията или съвременните схващания, например в съпоставка с РКЧ. При определянето на типа лексикална единица, освен класификацията за частите на речта, се взема предвид дали става въпрос за словосъчетание, фразема, цитат, собствено име, съкращение; техническото изпълнение наложи видовете заглавки да бъдат включени в един списък с частите на речта (като опция в менюто), тъй като в базата данни е желателно данните от различните опции да се комбинират помежду си (най-вече с оглед на еднокомпонентния и поликомпонентния състав);

• **варианти**: под варианти разбираме онези формални прояви на лексемите, при които се наблюдават фонетични или графични разлики (Филипец 1995: 21). В ЛБД регистрираме като самостоятелни единици всички варианти, посочени в РКЧЕ и в други източници, независимо дали са кодифицирани, или не,

като засега само посочваме дали са засвидетелствани в корпуса. При допълването на заглавките от РКЧЕ не включваме вариантите, които са означени като остарели (например думите, завършващи на *-ism* в опозиция с *-ismus*, *-izmus*);

• **препратките** между отделните заглавни единици се осъществяват чрез т.нар. **Списъци**, които са разположени във всяка една от тях и дават възможност за хипертекстови връзки, а следователно и за лесни преходи между корелиращите заглавки, за отваряне на няколко заглавки върху екрана и за едновременната им обработка, оценка и евентуална корекция. По този начин – посредством хипертекстовите препратки и отделните списъци – са обхванати и представени различни системни и други отношения, предимно омонимия, антонимия, вариантност, както и връзката между еднокомпонентни и поликомпонентни единици (между фраземи или словосъчетания и техните компоненти), между съществителни собствени и съществителни нарицателни имена, между съкращения и техните пълни съответствия, обработени като самостоятелни заглавки;

• **източник за словника**: тези данни са работна информация, насочваща към източника, от който дадена лема (с оглед на формата ѝ) е извлечена за словника. И тъй като се извършва съпоставка между корпуси и определени лексикографски източници, тя се състои от две части. При лексикографските източници се придържаме към следния принцип: **наличие на дадена форма в тълковен речник**, като се започне от РКЧЕ; от съществено значение е също дали дадена форма на лексикалната единица е регистрирана в *Честотния речник на чешкия език* (ЧРЧЕ), или не, както и дали дадена лема се открива в корпуса.

Това са необходимите данни за всяка заглавна единица като цяло, включително и за многозначната. По-нататък се работи върху отделните значения и се въвеждат работни дефиниции, които се базират върху РКЧЕ, евентуално и върху други речници, след което при необходимост дефинициите се преработват или коментират (разполагаме с богат апарат от бележки за отделните равнища на обработка).

Апаратът от бележки е предназначен най-вече за нуждите на лексикографската обработка: в бележките се въвежда

различна информация, която впоследствие ще бъде използвана при работата върху речника. Някои от бележките обаче може да бъдат използвани и от широк кръг потребители (например енциклопедичната бележка към личните имена, в която се дават сведения за името).

Всички единици от нашия словник, т.е. отделните заглавки, съдържат коментар (съобразен с частите на речта и типовете заглавки) относно наличието на съответната лема и нейните варианти в използваните източници, най-вече в тълковните речници. Разработката на заглавката като цяло и на нейните отделни значения в предходните речници се прави в **Бележки към тълкуването**. За разлика от DEBDict тук привеждаме повече информация и в по-експлицитен вид, а именно:

– първа регистрация в тълковен речник (започвайки от РКЧЕ)[6];

– как е разработена лексикалната единица – самостоятелно или в гнездо към друга лексикална единица;

– дадено ли е тълкуване, или не;

– към коя речникова единица е отнесена;

– върху какво можем да се опрем по-нататък (например тълкуване при синоним, тълкуване в друг речник) и т.н.

- **тълкуване на значението:**

Както беше споменато, засега основната ни задача е към всяка единица от базата данни да бъде събран и класифициран достатъчно илюстративен материал от езиковия корпус, въз основа на който да се проведе по-нататъшната лексикографска обработка. Преценихме, че ще е по-уместно да отнасяме езиковия материал директно към съответните значения, затова и заглавките са прецизно структурирани. За целта беше необходимо да изходим от някоя вече разработена семантична структура, като в това отношение се опираме най-вече върху тълкуването и семантичното членение на заглавните единици в РКЧЕ.

В случай че тълкуването в РКЧЕ по една или друга причина не ни удовлетворява, ползваме дефиниции от други речници, модифицираме тълкуванието от речниците или въвеждаме свое предложение. Тъй като методологичната база на тълкуванията в *Пралекс* не е единна, дефинициите в ЛБД в никакъв случай не

бива да се смятат за окончателни. За нуждите на бъдещия тълковен речник дефинициите трябва да се редактират, голяма част от тях да се доработят и да се унифицират с оглед на бъдещата концепция за новото лексикографско описание, която засега не е изготвена и трябва да се изработи, преди да пристъпим към следващата задача – системно лексикографско описание. По-нататъшната ни работа с езиковия материал от корпуса и другите източници ще се съсредоточи върху т.нар. дефиниращ контекст, който би ни помогнал да формулираме съответната дефиниция и да доуточним значението.

Както беше казано, когато е необходимо, заимстваните дефиниции се редактират и понастоящем, а въз основа на анализирания материал от корпуса добавяме и други значения или предлагаме нови решения за тълкуването. Необходимо е да споменем също така, че работата върху значенията произтича от изискването за класифициране на материала, с който разполагаме, но в никакъв случай не отменя необходимостта от цялостен анализ на семантичната структура на всяка лексикална единица. Този анализ неминуемо ще бъде извършен при цялостната обработка на заглавките в бъдещия тълковен речник (очевидно е, че при описанието на сферата на употреба на лексикалната единица се прилагат по-различни подходи в сравнение с описанието на нейната семантична структура).

Както беше споменато по-горе, ЛБД *Пралекс* е създадена като модел за комплексно лексикографско описание. В настоящия изследователски проект обаче не се описват всички аспекти и категории, които традиционно са застъпени в един тълковен речник, макар и да са предвидени рубрики за тях в работния формуляр. Редица аспекти остават задача на бъдещата лексикографска разработка, например прагматичните аспекти на значението, стилистичната принадлежност на единиците, специализираността им според различни сфери на употреба. Специално внимание изисква отношението им към езиковата норма, наличието на варианти (книжовни и некнижовни). Ето защо актуално създаваната база данни има преди всичко дескриптивна, а не прескриптивна функция. Поради това не бива да се очаква, че ще изпълнява всички функции, присъщи на

тълковния речник, от друга страна – дори в случай че въз основа на нея не бъде разработен нов лексикографски труд, ЛБД *Пралекс* представлява база с богато структуриран и коментиран езиков материал и разнообразна метаезикова информация, която може да послужи за различни лингвистични изследвания. В никакъв случай ЛБД не бива да се смята за готов речник, независимо че и в този си вид отговаря на някои от параметрите за цялостно лексикографско описание.

- **илюстративен материал към заглавната единица и нейните значения:**

Съществена част от работата на нашия лексикографски екип е извличането на материал от големи езикови масиви – работим и с трите корпуса на Института за Чешкия национален корпус (ИЧНК): SYN2000, SYN2005 и SYN2006PUB, обединени в електронен вид за нуждите на лексикографската практика в корпуса SYN (включващ трите посочени корпуса), като събраният материал следва да послужи както за създаването на бъдещия тълковен речник, така и за по-нататъшни проучвания върху съвременното състояние на чешкия речников състав и неговата динамика, а също и да намери приложение в сферата на информационните технологии. Благодарение на това, че от хронологична гледна точка материалът е сравнително хомогенен (включва текстове от последните 20–30 години), можем да илюстрираме как се използват думите в съвременния чешки език.

Тъй като разполагаме с изключителната (но днес вече стандартна) възможност да работим с огромен обем лематизирани и морфологично характеризирани езикови данни, съчетавайки класическите лексикографски методи с модерни софтуерни продукти за тяхното прилагане, преди всичко за сортиране и класифициране на материала в корпуса (Bonito, Word Sketch Engine), днес сме в състояние да съберем богат илюстративен материал за бъдещия речник от публицистични, художествени и специализирани текстове от последните 20–30 години, които дават ясна представа за съвременната употреба на лексикалните единици и тяхното функциониране в езика, а следователно и за динамиката в съвременната лексикална система на чешкия език. Във връзка с това става дума не за ексцерпция, а за

анализ, подбор и класифициране на контекстовите реализации на лексикалните единици в корпуса. Така се постига предварителна обработка на много по-широк спектър от реален езиков материал в сравнение с досегашните тълковни речници, които са създадени въз основа на лексикалните архиви на ИЧЕ ЧАН, съдържащи материал от класическата ексцерпция.[7]

За разлика от класическата ексцерпция, която се е извършвала върху фишове (протичала е в ИЧЕ от края на XIX в. до 70-те години на XX в.) и е била ориентирана предимно към отделни изреченски цялости, т.е. към регистрирането на думата в по-широк контекст, необходим за дефиниране на значението, понастоящем за целите на ЛБД *Пралекс* се работи с езиков корпус, който представлява модерен аналог на т.нар. тотална ексцерпция от зората на лексикографската практика. Затова в илюстративната част на заглавката можем да представим не само отделни изреченски цялости (с посочена дата, източник и тип текст), но и да преценим всеки контекст с оглед на съчетаемостта на заглавните думи.

При съответната част на речта, най-вече при съществителните, се вземат предвид т.нар. обработени примери, които включват синтагми или части от изречения, свидетелстващи за семантичния обхват, в който се употребяват заглавните думи, и за отделните им значения. Структурата на илюстративните блокове (при съществителните означени като E1 – E14) следва синтактичната съчетаемост на единиците и е максимално формализирана, например за съществителните имена блок E1 (Adj+SUBST), E2 (SUBST+Adj), E3 (SUBST+Subst-gen), E4 (Subst+SUBST-gen) и т.н. В рамките на всеки блок е отразена семантичната съчетаемост на имената (Чермак 2006, Ярошова 2007).

Наред с това се използват, макар и по-ограничено, т.нар. примери цитати, които са извлечени от корпуса и се въвеждат без каквито и да било корекции. Това означава, че при обработката на материала за нашата база данни се работи с езиковия материал на следващото равнище на абстракция, което се надгражда над ръчната (днес вече и електронна) ексцерпция от определени текстове. Доказателственият материал се въвежда директно към съответните значения, като така въз основа на

материала от корпуса, събран в електронен вид, и на възможностите, които ни дават структурираната лексикална база и наличните тълковни речници, можем да работим и върху семантичната структура на заглавните единици.

3. Предназначение на ЛБД *Пралекс*

В заключение можем да констатираме, че благодарение на богатия основен словник и гъвкавия модел на описание в ЛБД *Пралекс* постепенно се оформя масив от почти готови заглавки за бъдещия тълковен речник, като отделните части на речта са в различен стадий на обработка. Най-общо можем да кажем, че в рамките на настоящия изследователски проект усилията ни бяха насочени към изграждането на словника[8] и събирането на богат и вътрешно структуриран илюстративен материал към заглавките (Опавска, Щепанкова 2007), като едновременно с това придобихме ценен опит за работа със семантичната структура на лексикалните единици и с лексикографските техники за нейното представяне.

Извършеното досега е предпоставка за преминаването към следващия етап от работата ни. Изхождайки от направените анализи и натрупаните резултати, можем да твърдим, че създаването на бъдещия речник ще бъде облекчено в няколко насоки: целият словник на ЛБД *Пралекс* ще бъде снабден с коментари с оглед на извършените справки в корпуса; към редица заглавки са отбелязани нови значения, които са били установени при справките; илюстративният материал представя изцяло възможностите за съчетаемост на заглавните думи; повечето заглавки са придружени от богат апарат с бележки относно използваните източници и т.н.

След като бъде разработена концепцията на бъдещия речник и бъдат уточнени неговите цели, обхватът и практическата му насоченост, по-нататъшната обработка на словника ще бъде подчинена на избраните критерии за подбор на единиците, а работата по микроструктурата ще бъде ориентирана най-вече към семантиката на заглавните единици, т.е. към дефинирането на значението и семантичната структура на полисемните единици, като тази структура в ЛБД *Пралекс* отразява най-вече

РКЧЕ, но също и новия лексикален материал в езиковия корпус. Коригирането на дефинициите цели актуализиране на тълкуванията, така че да отговарят на съвременната лексикална норма, а метаезикът да бъде изчистен от идеологически наслоения. Впоследствие ще се работи и върху други задължителни компоненти от лексикографското описание, които в рамките на настоящата изследователска задача не са били отразени.

Съвсем естествено е, че при разработката на бъдещия речник ще бъдат максимално използвани отделните заглавки в ЛБД *Пралекс*, например вариантите. В речника те ще бъдат представени в рамките на една речникова статия, докато в базата данни са налице две (или повече) самостоятелно обработени единици, тъй като целта е била да се съберат отделно всички налични данни, които имат отношение към всяка графична форма на лексикалната единица, след което те да бъдат характеризирани с оглед на правописната норма, стилистичната им принадлежност и под.

Онези единици от базата данни, при които не се налагат корекции и допълнения, ще останат в настоящия си вид. Наред с това готовите заглавки ще бъдат използвани като класифициран материал за извършването на други лингвистични изследвания. Те могат също така да служат като основа за други лексикографски трудове – например за специализирани речници. Периодът, предвиден за разработка на лексикалната база данни *Пралекс*, приемаме като подготвителен етап за създаването на модерен тълковен речник също и поради факта, че в лексикографския отдел на ИЧЕ ЧАН през последното десетилетие постъпиха много млади хора, така че понастоящем нашият работен екип се състои предимно от млади, перспективни кадри, които благодарение на проекта добиват ценен опит за понататъшната си лексикографска работа.

БЕЛЕЖКИ

[1] Ръководител на разработваната база данни е д-р Индра Светла.

[2] *Canoo.net* (швейцарски проект на фирмата Canoo и университетите в Базел, Амстердам и Лугано), *DWDS* (Das digitale Wörterbuch der deutschen Sprache des 20. Jahrhunderts), *CELEX* (the

Dutch Centre for Lexical Information), TLFi (Trésor de la langue francaise informatisé) и др.

[3] „*Пралед* е предназначен за събиране и обработка на лингвистични данни от различни източници като например: корпуси, електронни речници, скици на употребата на думите (word sketches), морфологичен анализ и други подобни" (вж. информацията, посочена на уеб страницата на Центъра за обработка на естествения език при Факултета по информатика на университета „Т. Г. Масарик" в гр. Бърно за клиента *Пралед* (Praled, http://deb.fi.muni.cz/clients-praled-cs.php).

[4] Вж. Пала и др. 2007, срв. също Светла 2007, Светла 2008.

[5] Така базата данни се разграничава и от бъдещия речник, чието название е „*Лексикон 21*" (*Lexikon 21*, *L21*, т.е. „Лексиконът на XXI век") – това название използвахме в началото на работата ни, когато създавахме концепцията и модела на лексикографското описание на съвременния чешки език; вж. също Светла 2007, Рангелова, Кралик 2007.

[6] Следва се хронологичният ред на издаването на речниците РКЧЕ, РКЧ, при лемите от чужд произход се взема предвид и ГАРЧД, при съкращенията се работи със специализирани справочници.

[7] Лексикалните сбирки на ИЧЕ ЧАН са електронизирани и са достъпни в интернет на уеб портала *Лексико* (*LEXIKO*, http://lexiko.ujc.cas.cz, Databáze heslářů), за директен достъп – http://bara.ujc.cas.cz.

[8] Относно изграждането на словника на ЛБД *Пралекс* вж. статията на И. Светла в настоящия сборник.

ЛИТЕРАТУРА

Аткинс, Рындъл 2008: Atkins, S. B. T., M. Rundell. *The Oxford Guide to Practical Lexicography*. Oxford, Oxford University Press. 540 p.

Опавска, Щепанкова 2007: Opavská, Z., B. Štěpánková. Tools for Working with Corpus Evidence in the Lexical Database LEXIKON 21 (Program PRAMAT and the Exemplification Tool). – In: *Computer Treatment of Slavic and East European Languages. Slovanské a východoeurópske jazyky v počítačovom spracovaní. Fourth International Seminar (Bratislava 25–27 October 2007). Proceedings*. Bratislava, Tribun, s. 190–197.

Пала и др. 2007: Pala, K., A. Horák, A. Rambousek, A. Rangelova. Nové nástroje pro českou lexikografii – DEB2. – In: *Gramatika a korpus*. Praha, ÚJČ AV ČR, s. 197–204.

Рангелова 2008: Rangelova, A. K úkolům výzkumného záměru Vytvoření databáze lexikální zásoby českého jazyka počátku 21. století. – In: *Lexikografie v kontextu informační společnosti*. Praha, ÚJČ AV ČR, s. 11–18.

Рангелова, Кралик 2007: Rangelova, A., J. Králík. Wider Framework of the Research Plan Creation of a Lexical Database of the Czech Language of the Beginning of the 21st Century. – In: *Computer Treatment of Slavic and East European Languages. Slovanské a východoeurópske jazyky v počítačovom spracovaní. Fourth International Seminar (Bratislava 25–27 October 2007)*. *Proceedings*. Bratislava, Tribun, s. 209–217.

Рангелова, Светла 2010: Рангелова, А., И. Светла. За създаването и изграждането на лексикалната база данни Пралекс. – В: *Лексикографията в европейското културно пространство. Материали от Петата национална конференция с международно участие по лексикография и лексикология*, *София, 19–20 октомври 2009 г.* Велико Търново, ИК „Знак ’94”, с. 287–300.

Светла 2007: Světlá, J. The possibilities and limits of lexikographical description of the Czech lexicon in database form. – In: *Computer Treatment of Slavic and East European Languages. Slovanské a východoeurópske jazyky v počítačovom spracovaní. Fourth International Seminar (Bratislava 25–27 October 2007)*. Bratislava, Tribun, s. 244–248.

Светла 2008: Světlá, J. K návrhu, vývoji a funkcím lexikální databáze češtiny. – In: *Lexikografie v kontextu informační společnosti*. Praha, ÚJČ AV ČR, 2008, s. 19–32.

Филипец 1995: Filipec, J. Teorie a praxe jednojazyčného slovníku výkladového. – In: *Manuál lexikografie*. Jinočany, H & H, s. 14–49.

Чермак 2006: Čermák, F. Kolokace v lingvistice. – In: *Studie z korpusové lingvistiky 2. Kolokace*. Praha, Nakl. Lidové noviny, 2006, s. 9–16.

Ярошова 2007: Jarošová, A. Problém vymedzenia kolokácií. – *Jazykovedný časopis*, № 2, s. 81–102.

ЦИТИРАНИ РЕЧНИЦИ

ГАРЧД (Голям академичен речник на чуждите думи)**:** *Velký slovník cizích slov*. Verze slovníku 2.0. Voznice, Leda spol. s r. o., 2005. (CD-ROM).

ОРАЕ (Оксфордски речник на английския език)**:** *Oxford Dictionary of English*. Oxford, Oxford University Press, 2003, 2005.

РКЧ (Речник на книжовния чешки)**:** *Slovník spisovné češtiny pro školu a veřejnost*. Voznice, Leda spol. s r. o., 2005. (CD-ROM).

РКЧЕ (Речник на книжовния чешки език)**:** *Slovník spisovného jazyka českého*. Praha, Academia, 1989.

РССЕ (Речник на съвременния словашки език)**:** *Slovník súčasného slovenského jazyka*, т. 1, A–G. Bratislava, Veda, 2006. 1133 с.

ЧРЧЕ (Честотен речник на чешкия език)**:** *Frekvenční slovník češtiny*. Praha, NLN, 2004. 596 s.

ЕЗИКОВИ КОРПУСИ И ЕЛЕКТРОННИ ИЗТОЧНИЦИ

Databáze heslářů [online]. Verze 4.5. Praha, Ústav pro jazyk český AV ČR, v. v. i., 2008. <http://lexiko.ujc.cas.cz/heslare/>.

Český národní korpus – SYN2000. Ústav Českého národního korpusu FF UK, Praha, 2000. <http://ucnk.ff.cuni.cz>.

Český národní korpus – SYN2005. Ústav Českého národního korpusu FF UK, Praha, 2005. <http://ucnk.ff.cuni.cz>.

Český národní korpus – SYN2006PUB. Ústav Českého národního korpusu FF UK, Praha, 2006. <http://ucnk.ff.cuni.cz>.

LEXIKO – webové hnízdo o novodobé české slovní zásobě a výkladových slovnících [online]. Verze: 1.8.2. Praha, Ústav pro jazyk český AV ČR, v. v. i., 2009. <http://lexiko.ujc.cas.cz/index.php>.

Textový archiv Newton Information Technology, s.r.o., <www.newtonit.cz>.

Надежда Костова
(София)

ПРОБЛЕМИ НА ЛЕКСИКОГРАФСКОТО ПРЕДСТАВЯНЕ НА НЯКОИ ГЛАГОЛИ ЗА РЕЧЕВА ДЕЙНОСТ В БЪЛГАРСКИЯ ЕЗИК*

The article presents the results of a corpus-based study of the verbs *speak* and *say* denoting speech activity. The study aims to show how the use of corpora can lead to solving many lexicographical problems. The results of the corpus-based study were compared with corresponding vocabulary items in the multivolume explanatory „Dictionary of Bulgarian Language". Options are available to optimize the lexicographical presentation of these verbs. A conclusion is drown, that corpus-based approach allows a comprehensive and detailed presentation of the lexical semantics in the dictionary.

Ключови думи: лексикография, корпуси, глаголи, български език
Key words: lexicography, corpora, verbs, Bulgarian language

Въведение

За съвременната лексикография основен ресурс са електронните корпуси. Измествайки традиционната лексикална картотека, електронните корпуси предоставят на лексикографите богат материал от различни по стил и жанр цели текстове, който превъзхожда фишовете с примери, ексцерпирани от художествена и научна литература, периодични издания и други източници. Корпусите предлагат и разнообразни възможности за значително ускоряване и улесняване на изработването на речници като: автоматично генериране на азбучни честотни списъци на лексиката, които се използват за изготвяне на словник и подбор на думи за включване в речника; лесно и бързо преглеждане и сортиране на материалите; извличане на обективна инфор-

* Изследването е проведено в рамките на научноизследователския проект „Информационна среда за комплексно лексикографско описание на словното богатство на българския език (от началото на XIX в. до наши дни)", финансиран от Фонд „Научни изследвания", договор ДТК 02/53 от 2009 г.

мация за употребата и съчетаемостта на лексикалните единици; систематизиране на данните от търсенето и статистическата им обработка.

Корпусният подход осигурява изчерпателност и точност на лексикографските решения. Той дава възможност за оптимизиране на изискващата много труд и време лексикографска работа и подпомага по-обективното и по-прецизното описание на лексикалните единици (Благоева, Колковска 2007). Корпусният подход се налага като по-подходящ и перспективен от традиционните методи и подходи в лексикографията. Както посочват Верлинд и Селва: „При сравняване на двата лексикографски подхода – корпусно базирания и интуитивния – се налага заключението, че корпусно базираната лексикография дава убедителни и необходими емпирични доказателства за личната интуиция на лексикографа" (Верлинд, Селва 2001).

От 2005 г. корпусният подход се прилага успешно при създаването на най-мащабния тълковен речник на българския език – многотомния академичен „Речник на българския език" (РБЕ), заедно с традиционните лексикографски методи и лексикални източници. Тук ще се спрем върху перспективите, които разкрива корпусният подход с оглед на усъвършенстване и прецизиране на лексикографското представяне на глаголите за речева дейност. Ще представим резултатите от корпусно базирано проучване на съчетаемостта на някои глаголи за речева дейност и ще сравним тези резултати с речниковите статии на съответните глаголи в „Речник на българския език". Чрез това сравнение ще очертаем някои насоки за постигане на по-адекватно лексикографско описание на семантиката на разглежданите лексикални единици, на тяхната употреба, както и на лексикалната и синтактичната им съчетаемост в голям тълковен речник.

Езиковите ресурси, които използваме, са: **Българският национален корпус** (БНК), създаден в Института за български език „Проф. Любомир Андрейчин" от сътрудници от Секцията по компютърна лингвистика и Секцията за българска лексикология и лексикография (http://www.ibl.bas.bg/BGNC_bg.htm), с обем в сегашния момент от над 400 милиона графични думи (Коева и др. 2010а, Коева и др. 2010б), както и **специализираният Лексикографски корпус**, дело на Секцията за българска лексикология и

лексикография (Благоева, Колковска 2007, Благоева и др. 2009), който през 2009 г. беше включен като интегрална част в БНК. И двата корпуса отговарят на изискванията за представителност и балансираност по отношение на стиловата и жанровата характеристика на текстовете, което ги прави изключително полезни ресурси за българската лексикография.

Софтуерните средства, използвани при изследването, са следните:

- системата за търсене в корпуси на БНК (http://search.dcl.bas.bg/) (Тинчев и др. 2007);
- WordSmith Tools (http://www.lexically.net/wordsmith/index.html), софтуер за лексикален анализ, с който ще извлечем статистически данни от специализирания лексикографски корпус;
- Sketch Engine (http://www.sketchengine.co.uk/), софтуер за лексикално профилиране, който в сегашния момент работи с български корпус от 26,5 млн. графични думи (токъни), част от БНК (подробно за Sketch Engine вж. Килгариф, Рындъл 2002).

1. Корпусно проучване на съчетаемостта на глаголите за речева дейност *говоря* и *казвам*

Резултатите от направеното корпусно проучване на съчетаемостта на основните глаголи за речева дейност *говоря* и *казвам* нагледно показват сходствата и различията в семантиката на изследваните глаголи. От големия по обем материал бяха извлечени статистически данни, които предлагат обективен критерий за идентифициране, дефиниране и подреждане на значенията на тези глаголи в тълковен речник.

Броят на конкорданските редове (съответно срещанията) на глаголите за речева дейност *говоря* и *казвам* и техните словоформи в разглежданите източници е следният:

- БНК: *говоря* – 20638 срещания (точно съвпадение на словоформа), *казвам* – 27656 срещания (точно съвпадение на словоформа); търсене на лема: *говоря/F/* – 220621 срещания, *казвам/F/* 804847 – срещания;
- Лексикографски корпус (срещанията са извлечени с

помощта на WordSmith Tools): *говоря* – 7603 срещания, *казвам* 9576 срещания; търсене на всички форми: *говор** – 229068 срещания, *каз** – 769327 срещания;

- Извадка от БНК (срещанията са извлечени с помощта на Sketch Engine) – конкорданси на лемите *говоря* и *казвам*: *говоря* – 19668 срещания, *казвам* – 26722 срещания.

Статистическите данни за броя на срещанията на изследваните глаголи показват повече употреби на глагола *казвам* в сравнение с тези на глагола *говоря*.

Освен чрез извличане на конкордансни редове изследваме двата глагола чрез функцията Collocations на системата за търсене в БНК и функцията Collocates на WordSmith Tools. С термина „колокация", който се използва в програмите за обработка на корпуси, най-общо се означава всяко съчетание на търсената дума с най-близко намиращата се дума в ляв или десен контекст. Използваме и функцията Word Sketch на Sketch Engine, която също дава най-често срещаните съчетания на глаголите с обекти, обстоятелства, субекти, както и с предложни фрази. С функцията Cluster на WordSmith Tools намираме най-често срещаните съчетания на глаголите с няколко думи в ляв и десен контекст, а с функцията Sketch-Diff на Sketch Engine сравняваме най-често срещаните съчетания на два глагола.

1.1. Говоря

Резултатите от търсене на най-често срещаните съчетания на глагола *говоря* в БНК и в специализирания Лексикографски корпус чрез функцията Collocations на системата за търсене в БНК и функцията Collocates на WordSmith Tools са представени съответно във Фигура 1 и Фигура 2 по-долу. Получените данни за съчетаемостта и честотата на употреба на глагола *говоря* показват значително сходство, което произтича и от факта, че лексикографският корпус е интегрална част от БНК.

При глагола *говоря* с най-голяма честота на употреба са: конструкции на глагола с частицата *да – да говоря,* формите *не говоря, ще говоря* (срв. Фигура 1 и Фигура 2). Близки резултати показват срещанията на глагола в края и началото на изречението: *говоря $* (2565 срещания) – *$ говоря* (2012 срещания) (вж. Фигура 1).

Search Collocations BulNC BulSemCor Help BG

Word 1: говоря Word 2: Collocations

Statistics:

говоря : 0

говоря ...:

- говоря за: 4649
- говоря с: 4286
- говоря $: 2565
- говоря на: 451
- говоря и: 405
- говоря сериозно: 392
- говоря само: 344
- говоря по: 307
- говоря от: 301
- говоря така: 240

Search Collocations BulNC BulSemCor Help BG

Word 1: Word 2: говоря Collocations

Statistics:

говоря: 0

...:

... говоря:

- да говоря: 8543
- $ говоря: 2012
- не говоря: 1699
- ще говоря: 1394
- ти говоря: 827
- какво говоря: 707
- аз говоря: 635
- ви говоря: 555
- че говоря: 468
- си говоря: 292

Фигура 1. Съчетания на глагола говоря, установени чрез системата за търсене в БНК

С почти еднаква фреквентност са конструкциите *говоря за (нещо)* и *говоря с (някого)* (срв. Фигура 1 и Фигура 2), които представят различни елементи от комуникативната ситуация. Логично на първо място е темата на разговора, но не по-малко значим е

фактът кой е събеседникът, както показва статистиката в БНК.

За разлика от фразите с комитативно значение *говоря с (някого)*, конструкциите *говоря на (някого)*, *ти говоря, ви говоря* имат значително по-ниска фреквентност. Следователно при глагола *говоря* е по-силно изразена „солидарническата" семантика, отколкото „властническата" (термини на Б. Георгиев, вж. Георгиев 2006).

Съчетанията на глагола *говоря* с именни фрази с предлога *по* експлицират канала, контакта при комуникацията *(говоря по телефона, по радиото)*, а чрез именни фрази с предлога *от* може да се представи както мястото на говорещия (*говоря от трибуната, от мястото си*), така и отправителят на съобщението, който съвпада или не с говорещия (*говоря от свое име, говоря от името на*), а също и контекстът на съобщението (*говоря от личен опит, говоря от сърце, не мога да говоря от радост*).

От съчетанията на глагола с наречия най-често се срещат *говоря сериозно, говоря само, говоря така*; освен с лични местоимения глаголът *говоря* се среща често и с въпросителното местоимение *какво*; среща се и с подчинени изречения, въведени със съюзите *и* и *че*.

WordSmith Tools дава по-богата информация за съчетанията на глагола *говоря* с думи в ляв и десен контекст, както и за тяхната фреквентност. Близки резултати имат *мога (да) говоря* и *трябва (да) говоря* (вж. Фигура 2), *искам (да) говоря* е с малко по-голяма честота на употреба.

При заявка *говор** в WordSmith Tools (вж. Фигура 3) получаваме сведения за фреквентността на личните и неличните форми на глагола – най-често се срещат *говори, говореше, говорим, говоря, говорят, говориш, говорите, говорейки* (1084 срещания); както и за отглаголното съществително и за сродни думи (и техните форми): *говорене* (1790 срещания), *говорител* (3046 срещания), *говорителят* (3044 срещания) и т.н. Освен това се извличат статистически данни за съчетаването на формите на глагола с думи в ляв и десен контекст. Най-често срещаните контекстуални партньори на глагола *говоря* са предлозите *за*, *с, на, по, в, от,* частиците *да*, *не, се, си, ще, ли*, съюзите *че, и, като, но, а*, местоименията *какво, ми, му, когато, той, които* и т.н.

говоря.cnc

File Edit View Compute Settings Window Help

N	Word	With	elation	Texts	Total	tal Left	al Right	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	ГОВОРЯ	говоря	0,000	3 804	19 862	97	97	16	37	20	15	9	19 668	9	15	20	37	16
2	ДА	говоря	0,000	2 752	11 208	10 076	1 132	331	377	289	1 036	8 043	0	55	191	244	347	295
3	ЗА	говоря	0,000	2 749	7 751	1 731	6 030	253	230	389	857	2	0	4 353	708	422	303	244
4	С	говоря	0,000	1 653	5 231	414	4 817	96	123	104	90	1	0	4 075	311	180	130	121
5	НЕ	говоря	0,000	2 185	5 093	4 182	911	408	594	1 402	198	1 580	0	145	110	225	228	203
6	И	говоря	0,000	1 657	3 349	1 553	1 796	324	359	359	383	128	0	404	201	522	362	307
7	НА	говоря	0,000	1 336	2 252	520	1 732	148	148	94	130	0	0	429	116	619	299	269
8	АЗ	говоря	0,000	1 209	2 251	1 980	271	161	292	339	632	556	0	109	46	42	35	39
9	ЩЕ	говоря	0,000	1 080	2 166	1 792	374	91	137	143	124	1 297	0	40	56	88	110	80
10	ЧЕ	говоря	0,000	1 087	1 862	1 464	398	209	293	273	248	441	0	69	45	97	86	101
11	СЕ	говоря	0,000	1 055	1 661	1 019	642	174	241	390	190	24	0	11	116	112	197	206
12	ТОВА	говоря	0,000	928	1 414	423	991	56	89	103	115	60	0	72	662	103	76	78
13	ТИ	говоря	0,000	749	1 409	922	487	50	53	22	9	788	0	185	61	127	60	54
14	СИ	говоря	0,000	922	1 371	648	723	79	131	111	44	283	0	40	56	386	123	118
15	ИСКАМ	говоря	0,000	791	1 312	1 224	88	12	20	100	1 089	3	0	6	20	27	16	19
16	КАКВО	говоря	0,000	746	1 225	1 143	82	18	35	65	356	669	0	19	15	19	14	15
17	МОГА	говоря	0,000	823	1 211	1 143	68	27	33	232	851	0	0	2	19	11	17	19
18	НО	говоря	0,000	845	1 201	652	549	139	213	131	132	37	0	93	85	159	127	85
19	Е	говоря	0,000	816	1 094	647	447	95	149	259	138	6	0	68	52	75	124	128
20	ТРЯБВА	говоря	0,000	603	1 024	949	75	31	28	90	799	1	0	6	16	21	20	12
21	МИ	говоря	0,000	708	1 015	741	274	141	267	262	67	4	0	5	38	73	81	77
22	ОТ	говоря	0,000	752	970	249	721	85	75	68	20	1	0	265	72	146	118	120
23	В	говоря	0,000	736	953	357	596	118	88	105	46	0	0	142	54	146	133	121
24	НЕГО	говоря	0,000	590	900	76	824	13	12	19	18	14	0	0	746	49	17	12
25	ВИ	говоря	0,000	499	877	615	262	26	29	20	15	525	0	91	18	65	43	45
26	КАТО	говоря	0,000	621	868	344	524	41	63	45	51	144	0	194	84	93	72	81
27	СЕРИОЗНО	говоря	0,000	503	806	220	586	2	2	1	101	114	0	383	168	29	2	4
28	ТЕБ	говоря	0,000	470	801	73	728	7	4	6	15	41	0	0	648	36	25	19
29	ЛИ	говоря	0,000	518	765	659	106	31	90	197	277	64	0	38	15	20	21	12
30	МЕ	говоря	0,000	553	752	590	162	70	141	295	82	2	0	4	21	31	57	49
31	А	говоря	0,000	556	722	238	484	54	69	69	33	13	0	89	34	139	131	91
32	ПО	говоря	0,000	493	632	128	504	38	52	19	19	0	0	260	48	115	54	27
33	САМО	говоря	0,000	487	594	123	471	28	26	26	37	6	0	329	50	34	27	31

concordance collocates plot patterns clusters filenames follow up source text notes

1 788 Type-in

Фигура 2. Съчетаия на глагола говоря, установени чрез WordSmith Tools

Concord

File Edit View Compute Settings Window Help

N	Word	With	elation	Texts	Total	tal Left	al Right	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	ЗА	говор*	0,000	7 130	89 867	21 280	68 587	3 132	4 036	5 565	8 209	338	0	48 305	8 581	4 655	3 678	3 368
2	ДА	говор*	0,000	6 782	84 198	71 122	13 076	3 447	3 948	2 712	19 879	41 136	0	763	2 024	2 756	3 740	3 793
3	ГОВОРИ	говор*	0,000	6 953	74 643	613	591	141	127	125	119	101	73 439	105	115	116	124	131
4	И	говор*	0,000	6 574	57 399	27 208	30 191	4 858	5 485	6 074	5 886	4 905	0	6 955	4 580	7 649	5 907	5 100
5	НА	говор*	0,000	6 721	55 357	18 596	36 761	4 183	4 450	4 683	3 477	1 793	0	13 034	2 766	9 737	5 974	5 250
6	НЕ	говор*	0,000	6 179	45 627	35 863	9 764	3 196	5 133	6 493	4 501	16 540	0	1 431	1 537	2 018	2 499	2 279
7	СЕ	говор*	0,000	6 431	42 473	28 793	13 680	2 609	2 681	3 555	1 530	18 418	0	3 470	1 925	2 576	2 845	2 864
8	С	говор*	0,000	5 135	33 295	6 275	27 020	1 421	1 636	1 643	1 317	258	0	18 795	2 653	2 190	1 820	1 562
9	ЧЕ	говор*	0,000	5 615	28 973	12 735	16 238	2 282	2 521	2 552	2 826	2 554	0	8 060	3 075	1 994	1 616	1 493
10	ГОВОРЕШЕ	говор*	0,000	3 213	26 534	143	147	26	18	28	43	28	26 244	31	45	29	24	18
11	В	говор*	0,000	5 626	24 019	11 433	12 586	2 769	3 006	3 345	1 717	596	0	3 224	1 485	2 704	2 668	2 505
12	СИ	говор*	0,000	4 733	22 832	13 081	9 751	1 315	1 471	1 587	1 500	7 208	0	1 009	1 099	3 961	1 960	1 722
13	ГОВОРИМ	говор*	0,000	5 470	22 426	98	103	28	34	19	10	7	22 225	8	11	18	32	34
14	ГОВОРЯ	говор*	0,000	3 803	19 951	144	140	32	58	26	17	11	19 667	11	16	30	54	29
15	ГОВОРЯТ	говор*	0,000	5 301	19 471	123	142	28	21	25	21	28	19 206	26	19	22	34	41
16	Е	говор*	0,000	5 102	18 935	10 507	8 428	1 974	2 589	2 452	1 072	2 420	0	878	1 524	1 972	2 026	2 028
17	ОТ	говор*	0,000	4 838	15 966	7 019	8 947	1 824	2 059	1 660	961	515	0	1 662	1 093	2 254	2 029	1 909
18	ТОВА	говор*	0,000	4 326	13 810	6 364	7 446	873	1 044	1 428	1 535	1 484	0	496	4 269	1 225	789	667
19	ГОВОРИШ	говор*	0,000	2 199	12 238	65	55	15	17	11	12	10	12 118	4	9	12	18	12
20	КАКВО	говор*	0,000	2 969	11 950	10 780	1 170	238	418	1 132	2 896	6 096	0	288	242	222	207	211
21	НО	говор*	0,000	3 682	11 380	5 677	5 703	1 287	1 361	1 366	1 273	390	0	1 291	1 048	1 377	1 157	830
22	КАТО	говор*	0,000	3 673	11 370	4 646	6 724	686	857	728	740	1 635	0	2 284	1 096	1 308	1 144	892
23	МУ	говор*	0,000	2 846	10 338	6 763	3 575	673	850	816	1 316	3 108	0	586	506	992	729	762
24	ТОЙ	говор*	0,000	2 966	10 321	6 476	3 845	731	909	1 083	1 055	2 698	0	1 147	726	640	735	597
25	ЩЕ	говор*	0,000	3 351	9 802	6 796	3 006	581	693	758	847	3 917	0	233	655	702	758	658
26	ПО	говор*	0,000	3 711	9 714	2 742	6 972	629	736	667	582	128	0	3 618	784	1 176	786	608
27	А	говор*	0,000	3 580	9 617	4 650	4 967	937	1 211	1 401	827	274	0	1 030	616	1 311	1 139	871
28	МИ	говор*	0,000	2 394	8 882	6 379	2 503	509	685	630	856	3 699	0	439	457	658	457	492
29	КОГАТО	говор*	0,000	3 255	7 732	5 772	1 960	431	568	601	1 343	2 829	0	379	390	506	386	299
30	ЛИ	говор*	0,000	2 444	7 479	4 723	2 756	418	607	968	1 205	1 525	0	1 528	262	343	327	296
31	КОИТО	говор*	0,000	3 323	7 176	4 713	2 463	412	500	609	1 251	1 941	0	81	207	861	683	631
32	ГОВОРИТЕ	говор*	0,000	2 189	7 097	44	52	6	8	8	13	9	7 001	7	19	8	10	8
33	ТАКА	говор*	0,000	2 405	6 856	2 154	4 702	332	308	499	598	417	0	3 614	306	328	250	204

concordance collocates plot patterns clusters filenames follow up source text notes

21 734 Type-in ЗА

Фигура 3. Съчетания на лемата говоря, установени чрез WordSmith Tools

При сравнение на Фигури 2 и 3 се вижда, че основната форма на глагола се среща най-често в конструкции с частицата *да*, а конструкциите *говоря за (някого)* и *говоря с (някого)* имат почти еднаква фреквентност, докато различните форми на глагола

говоря се съчетават най-често с именни или местоименни фрази с предлога *за*, след тях по честота се нареждат *да*-конструкциите, а конструкциите с предлога *с* са два пъти по-малко от тези с предлога *за*. Следователно може да приемем, че глаголът *говоря* преимуществено изисква уточняване на темата на разговора, докато събеседникът и обстоятелствата на комуникацията са с по-малка важност. Както основната, така и всички останали форми на глагола най-често образуват конструкции с частицата *да*, а след това с частицата *не*, откъдето се налага изводът, че глаголът *говоря* по-често влиза в модални конструкции, отколкото в отрицателни.

Чрез WordSmith Tools може да се извлекат и клъстери на глагола *говоря*, които дават информация за комбинации на изследвания глагол с по-голям брой думи в ляв или десен контекст (вж. Фигура 4). Данните от клъстерите дават възможност за детайлизиране на резултатите. Например *да говоря* се съдържа в следните клъстери: *да говоря с, да говоря за, искам да говоря, мога да говоря, трябва да говоря, няма да говоря* и др.; *говоря за – да говоря за, не говоря за, говоря за това* и т.н. От клъстерите получаваме и нова информация за типични употреби на глагола: *не мога да говоря, не искам да говоря, за какво говоря, няма да говоря, да не говоря* и др.

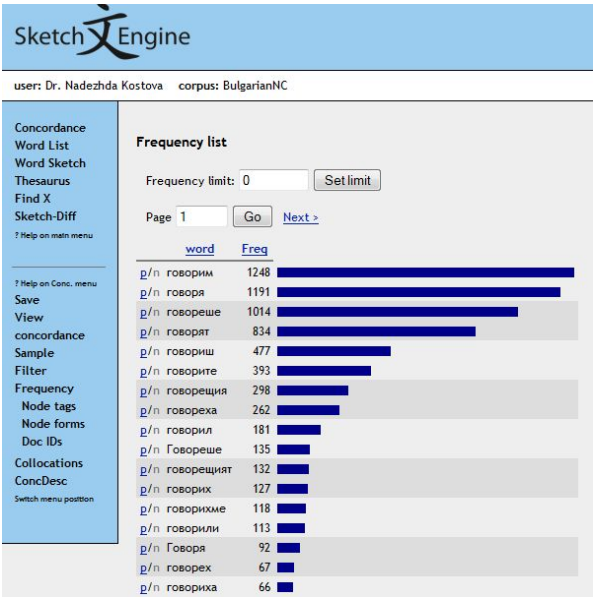
Concord

File Edit View Compute Settings Window Help

N	Cluster	Freq.	Length
1	ДА ГОВОРЯ С	2 992	3
2	ДА ГОВОРЯ ЗА	1 359	3
3	ИСКАМ ДА ГОВОРЯ	1 085	3
4	НЕ ГОВОРЯ ЗА	863	3
5	МОГА ДА ГОВОРЯ	851	3
6	ТРЯБВА ДА ГОВОРЯ	799	3
7	ГОВОРЯ С НЕГО	614	3
8	ГОВОРЯ ЗА ТОВА	607	3
9	ГОВОРЯ С ТЕБ	598	3
10	ЩЕ ГОВОРЯ С	553	3
11	НЕ МОГА ДА	514	3
12	ГОВОРЯ С ВАС	388	3
13	НЕ ИСКАМ ДА	359	3
14	ЗА КАКВО ГОВОРЯ	314	3
15	НЯМА ДА ГОВОРЯ	298	3
16	ГОВОРЯ С НЕЯ	276	3
17	ДА НЕ ГОВОРЯ	267	3
18	ЛИ ДА ГОВОРЯ	244	3
19	ДА ТИ ГОВОРЯ	193	3
20	ДА ГОВОРЯ И	193	3

Фигура 4. Клъстери на глагола говоря, извлечени с WordSmith Tools

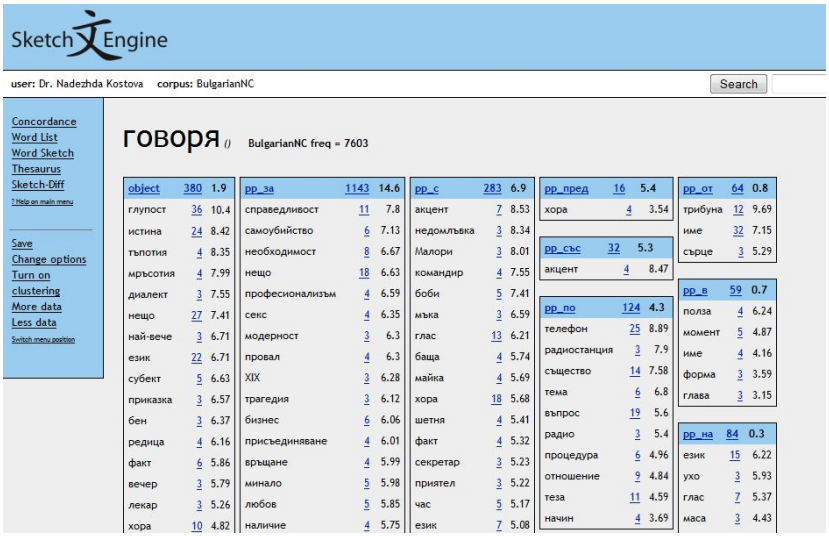
Sketch Engine разкрива по-разнообразна информация за контекстуалното поведение на глагола *говоря*, но поради това, че използва по-малък корпус с български текстове, данните за фреквентността са относителни. Това се вижда при сравняване на Фигури 3 и 5. Въпреки че липсва във фреквентния списък на формите на глагола (вж. Фигура 5), за формата *говори*, която е с най-много срещания в Лексикографския корпус, Sketch Engine показва 3499 конкордансни редове (тук влизат и безличните форми *говори се*), докато за *говоря* – 1285 редове.



Фигура 5. Данни за фреквентността на формите на говоря, извлечени чрез Sketch Engine

Функцията Word Sketch на Sketch Engine позволява да намерим типичните комбинации на лемата *говоря* (вж. Фигура 6). От получените данни виждаме, че най-често срещаните обекти на глагола *говоря* са думите: *глупости, истината, (някакъв) език, нещо*; с по-малка честота на употреба са съчетанията *говоря (празни) приказки, (някакви) думи*. С голяма фреквентност са следните конструкции с предложни фрази с обектна или обстоя-

телствена функция: *говоря за (някакви) неща, говоря за справедливост, говоря с хора, говоря с (някакъв) глас, говоря с господин, говоря по телефон, говоря по (някакъв) въпрос, говоря по същество, говоря от трибуната, говоря от името на, говоря на (някакъв) език.* С по-малка фреквентност са конструкциите: *говоря с акцент, говоря с недомлъвки, говоря с мъка, говоря с часове, говоря пред хората, говоря със (странен) акцент, говоря от сърце, говоря в полза на, говоря в момента, говоря на глас, говоря на ухо, говоря на масата, говоря на (някакви) хора, говоря зад гърба (на някого), говоря като депутат, говоря като човек*.



Фигура 6. Типични комбинации на лемата говоря, извлечени с функцията Word Sketch на Sketch Engine

Повечето данни, систематизирани чрез Word Sketch (т.нар. скици на употребата на думите), са лексикографски релевантни. В някои случаи обаче се получава „информационен шум”, като се извежда нерелевантна или неточна информация. Така например при словоредна инверсия (субект в следходна позиция спрямо глагола – позиция, типична за прекия обект), фразата на

субекта (напр. *Бен, лекар, хората, фактите*) се интерпретира неточно от програмата като обект. В полето за обекта понякога неправилно попадат и обстоятелствени фрази в конструкции като *говоря вечер, говоря (половин) час*.

За да проверим данните от Word Sketch, ще потърсим подобни данни чрез WordSmith Tools и системата за търсене в корпуси на БНК.

При конкордансите в БНК може да се наблюдават обектите на глагола *говоря*: *Говоря езика сравнително добре; Вярвате ли, че говоря истината; Вие ме обвинихте, че говоря неистини; Ти мислиш, че говоря лъжи; Говоря глупости може би; Не ме карай да говоря неща, които не съм обмислил добре*.

Обекти на глагола *говоря* в Лексикографския корпус, извлечени чрез WordSmith Tools: *истината* (148 срещания), *глупости (67), езика (33),* често се срещат и комбинациите *говоря (някакви) приказки (30), говоря празни (приказки) (24), говоря английски (50), френски (39), български (24)*. Там откриваме и често срещани съчетания на глагола с наречия: *говоря толкова (41), говоря малко (29), говоря направо (68), говоря откровено (46), говоря открито (44), говоря свободно (38), говоря спокойно (35), говоря тихо (14), говоря високо (14)*.

За разлика от Sketch Engine, където формата *глупости* е по-често срещан обект на глагола *говоря* (във всичките му форми при търсене по лема) от формата *истината*, в данните от Лексикографския корпус, получени с WordSmith Tools, се среща два пъти по-често съчетанието *говоря истината* от *говоря глупости*, но при търсене на всички форми на глагола WordSmith Tools дава приблизително равни резултати: *истината* (886 срещания), *глупости* (733 срещания).

Данни за конструкции на глагола *говоря* с различни предлози:

• *говоря за*: в БНК – 4572 срещания, в Лексикографския корпус – 4329 срещания; често срещани конструкции с предложна фраза с обектна функция, извлечени чрез WordSmith Tools: със съществителни имена – *за времето* (19 срещания), *за любов* (11), *за пари* (13), *за политика* (13), *за миналото* (13), *за работа* (9), *за живота* (8), *за бъдещето* (9), *за проблемите* (6),

за секс (5), *за бизнес или за (някакъв) бизнес* (7), *за война или за (някаква) война* (7) и т.н.; с местоимения – *говоря за това* (602 срещания), *за него* (112), *за себе си* (94), *за нея* (64), *за тях* (53), *за нещо* (37), *за всичко* (25); често срещани клъстери (данни от WordSmith Tools): *няма да говоря за* (89), *да не говоря за* (89), *не говоря за това* (88), *аз не говоря за* (78), *обичам да говоря за* (45), *говоря за това, че* (35), *не говоря за себе си* (22), *говоря за неща, които* (14) и т.н. Конструкциите *говоря за справедливост, за самоубийство, за необходимост*, които са изведени чрез Sketch Engine като често срещани, се откриват в БНК при различни форми на глагола.

• *говоря с*: в БНК – 4276 срещания, в Лексикографския корпус – 4073 срещания; често срещани конструкции с предложна фраза с обектна функция, установени чрез WordSmith Tools: с местоимения – *говоря с него* (612 срещания), *с теб* (599), *с тебе* (100), *с някого* (34), *с никого* (32), *с никой* (23), *с всички* (10), *с всеки* (8); със съществителни имена – *говоря с господин (X)* (55 срещания), *с баща (си)* (39), *с хората* (26), *с доктор (X)* (25), *с шефа* (13), *с управителя* (10), *с полицията* (6); други конструкции с предложни фрази с обектна или обстоятелствена функция: *говоря с непознати*, *с приятели*, *с раздразнение*, *с тънък глас*, *с пълна уста*, *с часове* и т.н.

• *и говоря със*: в БНК – 179 срещания, в Лексикографския корпус – 180 срещания; често срещани конструкции, установени чрез WordSmith Tools: *говоря със съпруга (си)* (7), *със сестра (си, ти)* (6), *със съдията* (6); клъстери (данни от WordSmith Tools): *трябва да говоря със* (26), *искам да говоря със* (18) и др.; други конструкции с предложни фрази с обектна или обстоятелствена функция: *говоря със сина си, със семейството си, със служителите, със съквартиранта, със студенти, със секретарката, със спокоен глас, със заповеден тон* и т.н.

В БНК намираме различни форми на глагола *говоря* в съчетания *с акцент, с недомлъвки*, посочени като резултати с голяма фреквентност от Sketch Engine.

• *говоря на:* в БНК – 443 срещания, в Лексикографския корпус – 423 срещания; често срещани конструкции с предложна фраза с обектна функция, установени чрез WordSmith Tools: *на*

вас (14), *на теб* (8), *на себе си* (10), *на хората* (7); конструкции с предложна фраза с обстоятелствена функция: *говоря на ти* (24), *на вие* (10), *на английски* (17), *на български* (8), *на френски* (8), *говоря на (някакъв) език* (10), *говоря на (някаква) тема* (26), *говоря на глас* (11) и др.; клъстери (данни, извлечени чрез WordSmith Tools): *говоря на тази тема* (24), *говоря на себе си* (10), *не говоря на вас (9), ви говоря на ти (9)* и др.; други примери: *говоря на висок глас, искам да говоря на четири очи с вас, говоря на правилен английски, говоря на диалект* и др.

Конструкциите *говоря на ухо, на масата*, изведени от Sketch Engine, намираме и в БНК при различни форми на глагола, въпреки че при основната форма не се срещат.

• *говоря по*: в БНК – 306 срещания, в Лексикографския корпус – 261 срещания; често срещани конструкции с предложна фраза с обстоятелствена функция, установени чрез WordSmith Tools: *говоря по телефона* (71), *по (някакъв) въпрос* (37), *по (някакъв) начин* (22), *по принцип* (16), *по същество* (10), *по работа* (6), *по (някаква) тема* (4), *по време (на)* (3) и др.; клъстери (данни от WordSmith Tools): *говоря по телефона с* (12), *говоря по този начин* (12) и др.; други примери: *говоря по съвест, говоря по първа точка, да говоря по радиото, когато говоря по телевизията, говоря по адрес на, да говоря по улиците, не говоря по цели дни с никого* и др.

• *говоря от*: в БНК – 294 срещания, в Лексикографския корпус – 265 срещания; клъстери (данни от WordSmith Tools): *говоря от името на* (108), *от свое име* (14), *от тяхно име* (13), *от личен опит* (10), *от собствен опит* (5) и др.; други конструкции с предложна фраза с обстоятелствена функция: *говоря от трибуната, ще говоря от мястото си, говоря от Женева, говоря от телефонна кабина, отидох да говоря от спалнята, говоря от сърце, не мога да говоря от радост* и др.

• *говоря в*: в БНК – 147 срещания, в Лексикографския корпус – 162 срещания; често срещани конструкции с предложна фраза с обстоятелствена функция, установени чрез WordSmith Tools: *говоря в момента* (10), *в минало (време)* (5); клъстери (данни от WordSmith Tools): *какво говоря в* (9), *говоря в множествено число* (6) и др.; други примери: *в полза на, в*

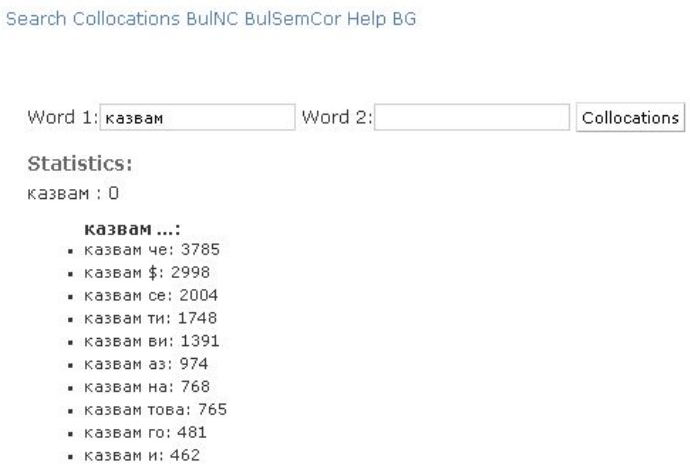
името на, в защита на, в трето лице, в преносен смисъл, в буквалния смисъл, в слушалката, в присъствието на адвокат, в стихове, в писмото си, говоря в парламента и т.н.

• *говоря пред*: в БНК – 111 срещания, в Лексикографския корпус – 104 срещания; често срещани конструкции с предложна фраза с обстоятелствена функция, установени чрез WordSmith Tools: *говоря пред вас* (7), *пред него* (7), *пред публика* (7), *пред хора* (3); клъстери (данни от WordSmith Tools): *не мога да говоря пред* (9), *да говоря пред публика* (7) и др.; други примери: *говоря пред журналисти, децата, студенти, медиите, всички, приятелките си, баща му* и т.н.

• *говоря против*: в БНК – 20 срещания, в Лексикографския корпус – 16 срещания; конструкции с предложна фраза с обектна функция: *говоря против тебе, против тях, против мъжа ѝ, против предателите, против убежденията си* и др.

1.2. Казвам

Данните за глагола *казвам*, извлечени от БНК и от специализирания Лексикографски корпус, са много близки с тези, представени при *говоря*. Получените резултати за съчетаемостта и честотата на употреба на глагола са представени съответно във Фигура 7 по-долу и Фигура 8 на стр. 92.



Search Collocations BulNC BulSemCor Help BG

Word 1: Word 2: казвам Collocations

Statistics:

казвам: 0

...:

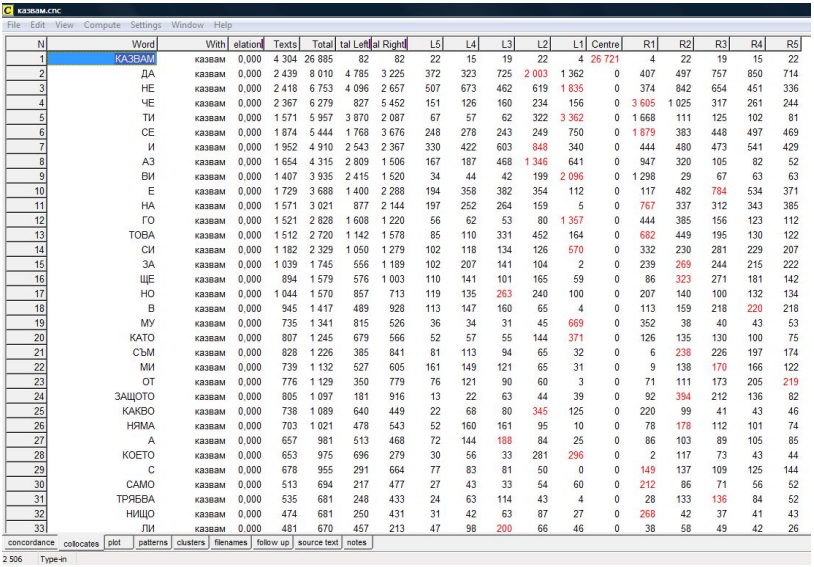
... казвам:

- $ казвам: 5432
- ти казвам: 3518
- ви казвам: 2314
- не казвам: 1932
- го казвам: 1462
- да казвам: 1391
- се казвам: 811
- аз казвам: 677
- му казвам: 675
- си казвам: 597

Фигура 7. Съчетания на глагола казвам, установени чрез системата за търсене в БНК

В БНК глаголът *казвам* се среща най-често в началото на изречение (*$ казвам* - 5432 срещания), но може да бъде и в края (*казвам $* - 2998 срещания). И в БНК, и в Лексикографския корпус е голяма фреквентността на глагола *казвам* във функцията му да въвежда подчинено изречение със съюза *че* (срв. Фигура 7 и Фигура 8). Голяма е фреквентността и на съчетанията на глагола с личните местоимения *ти, ви, му, аз*; с много по-малко употреби са конструкциите, включващи предлога *на*. Непреходните глаголи *казвам се* и *казвам си* се срещат често в използваните корпуси, широко представени са отрицателни и модални конструкции като *не казвам*, *да казвам*. Съчетанията *казвам това, казвам го, казвам нищо* представят обекта на глагола и също са едни от най-често срещаните в двата корпуса.

Чрез WordSmith Tools се извличат статистически сведения и за други съчетания на глагола *казвам* с думи в ляв и десен контекст (вж. Фигура 8): *казвам за* (някого или нещо) и *казвам* (нещо) *за* (някого или нещо), *като казвам, казвам само, (не) казвам нищо* и т.н.



казвам.cnc

File Edit View Compute Settings Window Help

N	Word	With	elation	Texts	Total	tal Left	al Right	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	КАЗВАМ	казвам	0,000	4 304	26 885	82	82	22	15	19	22	4	26 721	4	22	19	15	22
2	ДА	казвам	0,000	2 439	8 010	4 785	3 225	372	323	725	2 003	1 362	0	407	497	757	850	714
3	НЕ	казвам	0,000	2 418	6 753	4 096	2 657	507	673	462	619	1 835	0	374	842	654	451	336
4	ЧЕ	казвам	0,000	2 367	6 279	827	5 452	151	126	160	234	156	0	3 605	1 025	317	261	244
5	ТИ	казвам	0,000	1 571	5 957	3 870	2 087	67	57	62	322	3 362	0	1 668	111	125	102	81
6	СЕ	казвам	0,000	1 874	5 444	1 768	3 676	248	278	243	249	750	0	1 879	383	448	497	469
7	И	казвам	0,000	1 952	4 910	2 543	2 367	330	422	603	848	340	0	444	480	473	541	429
8	АЗ	казвам	0,000	1 654	4 315	2 809	1 506	167	187	468	1 346	641	0	947	320	105	82	52
9	ВИ	казвам	0,000	1 407	3 935	2 415	1 520	34	44	42	199	2 096	0	1 298	29	67	63	63
10	Е	казвам	0,000	1 729	3 688	1 400	2 288	194	358	382	354	112	0	117	482	784	534	371
11	НА	казвам	0,000	1 571	3 021	877	2 144	197	252	264	159	5	0	767	337	312	343	385
12	ГО	казвам	0,000	1 521	2 828	1 608	1 220	56	62	53	80	1 357	0	444	385	156	123	112
13	ТОВА	казвам	0,000	1 512	2 720	1 142	1 578	85	110	331	452	164	0	682	449	195	130	122
14	СИ	казвам	0,000	1 182	2 329	1 050	1 279	102	118	134	126	570	0	332	230	281	229	207
15	ЗА	казвам	0,000	1 039	1 745	556	1 189	102	207	141	104	2	0	239	269	244	215	222
16	ЩЕ	казвам	0,000	894	1 579	576	1 003	110	141	101	165	59	0	86	323	271	181	142
17	НО	казвам	0,000	1 044	1 570	857	713	119	135	263	240	100	0	207	140	100	132	134
18	В	казвам	0,000	945	1 417	489	928	113	147	160	65	4	0	113	159	218	220	218
19	МУ	казвам	0,000	735	1 341	815	526	36	34	31	45	669	0	352	38	40	43	53
20	КАТО	казвам	0,000	807	1 245	679	566	52	57	55	144	371	0	126	135	130	100	75
21	СЪМ	казвам	0,000	828	1 226	385	841	81	113	94	65	32	0	6	238	226	197	174
22	МИ	казвам	0,000	739	1 132	527	605	161	149	121	65	31	0	9	138	170	166	122
23	ОТ	казвам	0,000	776	1 129	350	779	76	121	90	60	3	0	71	111	173	205	219
24	ЗАЩОТО	казвам	0,000	805	1 097	181	916	13	22	63	44	39	0	92	394	212	136	82
25	КАКВО	казвам	0,000	738	1 089	640	449	22	68	80	345	125	0	220	99	41	43	46
26	НЯМА	казвам	0,000	703	1 021	478	543	52	160	161	95	10	0	78	178	112	101	74
27	А	казвам	0,000	657	981	513	468	72	144	188	84	25	0	86	103	89	105	85
28	КОЕТО	казвам	0,000	653	975	696	279	30	56	33	281	296	0	2	117	73	43	44
29	С	казвам	0,000	678	955	291	664	77	83	81	50	0	0	149	137	109	125	144
30	САМО	казвам	0,000	513	694	217	477	27	43	33	54	60	0	212	86	71	56	52
31	ТРЯБВА	казвам	0,000	535	681	248	433	24	63	114	43	4	0	28	133	136	84	52
32	НИЩО	казвам	0,000	474	681	250	431	31	42	63	87	27	0	268	42	37	41	43
33	ЛИ	казвам	0,000	481	670	457	213	47	98	200	66	46	0	38	58	49	42	26

concordance collocates plot patterns clusters filenames follow up source text notes

2 506 Type-in

Фигура 8. Съчетания на глагола казвам, извлечени с WordSmith Tools

И при глагола *казвам* търсенето по лема дава различни резултати (вж. Фигура 9). Различните форми на глагола, както и основната му форма, най-често се свързват с подчинено изречение със съюза *че* (срв. Фигура 8 и Фигура 9). С голяма фреквентност при различните форми на глагола *казвам* са и съчетанията с местоимения (вж. Фигура 8) – *той, ми, му, ти, тя, аз, ви* и т.н. Съчетаването с местоимения зависи от глаголната форма, но при основната форма се срещат повече местоимения във функция на обект, докато при другите форми на глагола с по-голяма фреквентност са местоименията в позицията на субект. При конструкциите с предлози има малка разлика (срв. Фигура 8 и Фигура 9) – и основната, и другите форми на глагола най-често се съчетават с именни фрази с предлог *на*. Основната форма *казвам* влиза в множество конструкции с предлога *за* (239 срещания), а другите форми на глагола – с предлога *в* (8564 срещания). Фреквентността на другите конструкции с предлози е следната: на основната форма с предлога *с* (149 срещания), с предлога *в* (113 срещания), с предлога *от* (71 срещания); на

другите форми на глагола с предлога *за* (7738 срещания), с предлога *с* (6045 срещания), с предлога *от* (4841 срещания).

Различните форми на глагола *казвам* се съчетават със съюзи *че, и, но*, с местоимения *той, ми, тя, му, ти, аз, го, ви, това, както, какво*, с частици *да, не, се, си, ще*, с предлози *на, в, за, с, от* (вж. Фигура 9).

Concord

File Edit View Compute Settings Window Help

N	Word	With	elation	Texts	Total	tal Left	al Right	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	КАЗА	каз*	0,000	6 608	84 467	1 031	1 180	278	265	253	141	94	382 256	179	240	282	250	229
2	И	каз*	0,000	7 238	66 877	92 944	73 933	11 561	13 206	16 165	23 414	28 598	0	9 461	27 061	13 520	11 850	12 041
3	ЧЕ	каз*	0,000	7 212	63 425	20 411	43 014	4 978	5 111	4 931	3 974	1 417	0	101 464	17 664	10 422	7 675	5 789
4	ДА	каз*	0,000	6 992	53 716	79 005	74 711	15 051	16 411	19 699	16 219	11 625	0	10 035	9 731	18 350	19 404	17 191
5	НА	каз*	0,000	7 324	47 095	60 845	86 250	11 412	14 151	16 440	15 432	3 410	0	21 630	18 634	14 953	15 656	15 377
6	СЕ	каз*	0,000	7 091	30 586	74 483	56 103	10 969	11 829	10 838	10 636	30 211	0	6 322	9 016	17 132	13 097	10 536
7	НЕ	каз*	0,000	6 678	17 100	71 778	45 322	9 472	10 528	14 607	18 242	18 929	0	6 844	12 997	10 975	8 271	6 235
8	Е	каз*	0,000	6 985	12 058	56 402	55 656	7 076	8 916	10 607	13 404	16 399	0	3 170	14 287	16 380	12 486	9 333
9	КАЗВА	каз*	0,000	7 172	85 857	490	484	126	112	104	111	37	84 883	26	72	119	132	135
10	ТОЙ	каз*	0,000	5 559	83 071	22 040	61 031	2 405	2 648	2 591	5 765	8 631	0	47 704	8 737	2 042	1 411	1 137
11	СИ	каз*	0,000	5 958	76 574	39 440	37 134	4 499	5 637	6 933	8 711	13 660	0	9 701	6 039	8 062	6 726	6 606
12	В	каз*	0,000	6 958	75 501	38 699	36 802	6 953	7 873	10 877	8 303	4 693	0	8 564	6 578	6 170	7 560	7 930
13	ЗА	каз*	0,000	6 601	61 287	29 445	31 842	5 864	6 906	7 725	8 041	909	0	7 738	5 627	5 879	5 994	6 604
14	МИ	каз*	0,000	4 811	59 947	42 491	17 456	2 269	2 705	3 320	6 518	27 679	0	6 180	3 438	3 279	2 479	2 080
15	ОТ	каз*	0,000	6 594	56 839	27 604	29 235	5 420	6 309	7 308	6 638	1 929	0	4 841	4 648	6 138	6 784	6 824
16	КАЗАХ	каз*	0,000	4 122	55 925	329	283	84	101	54	60	30	55 313	32	54	57	69	71
17	ТОВА	каз*	0,000	5 853	47 769	28 103	19 666	3 867	4 973	6 337	6 764	6 162	0	6 710	5 584	2 977	2 327	2 068
18	С	каз*	0,000	5 873	47 640	19 522	28 118	3 929	4 492	5 545	4 962	594	0	6 045	8 471	4 065	4 457	5 080
19	МУ	каз*	0,000	4 668	45 977	27 929	18 048	1 592	1 798	1 981	5 263	17 295	0	7 735	2 415	2 987	2 511	2 400
20	ТИ	каз*	0,000	3 179	42 758	27 102	15 656	1 768	2 354	3 351	5 481	14 148	0	6 199	3 700	2 434	1 805	1 518
21	ЩЕ	каз*	0,000	5 432	41 762	17 228	24 534	4 287	4 447	5 130	2 947	417	0	2 102	8 511	5 686	4 524	3 711
22	ТЯ	каз*	0,000	3 790	41 641	11 042	30 599	1 332	1 469	1 454	3 043	3 744	0	23 367	4 362	1 358	822	690
23	КАЗАЛ	каз*	0,000	5 178	38 601	217	209	52	67	58	25	15	38 175	9	13	37	77	73
24	ГО	каз*	0,000	4 506	37 064	23 065	13 999	1 935	2 430	2 768	5 520	10 412	0	2 348	3 189	3 472	2 759	2 231
25	КАТО	каз*	0,000	4 811	33 668	13 897	19 771	2 218	2 231	2 255	3 229	3 964	0	1 424	9 307	4 594	2 410	2 036
26	АЗ	каз*	0,000	3 814	33 512	14 206	19 306	2 027	2 389	2 842	4 383	2 565	0	11 988	3 383	1 775	1 214	946
27	НО	каз*	0,000	4 781	30 823	17 645	13 178	3 003	3 329	4 880	5 150	1 283	0	2 100	4 644	2 390	1 953	2 091
28	КАЗВАМ	каз*	0,000	4 304	27 275	292	263	67	61	82	64	18	26 720	29	44	55	76	59
29	КАКТО	каз*	0,000	5 077	23 895	22 212	1 683	591	936	2 479	9 138	9 068	0	329	369	357	323	305
30	КАКВО	каз*	0,000	3 790	23 335	15 920	7 415	780	1 268	2 470	5 679	5 723	0	2 801	1 545	1 252	959	858
31	КАЗВАТ	каз*	0,000	5 518	22 592	153	128	34	38	43	22	16	22 311	12	20	31	41	24
32	СЪМ	каз*	0,000	3 846	22 586	12 553	10 033	1 402	1 677	2 073	3 411	3 990	0	498	3 933	2 537	1 794	1 271
33	ВИ	каз*	0,000	3 003	21 934	14 120	7 814	934	1 205	1 533	2 878	7 570	0	2 616	1 550	1 366	1 206	1 076

concordance | collocates | plot | patterns | clusters | filenames | follow up | source text | notes

52 993 | Type-in | 382 256

Фигура 9. Съчетания на лемата казвам, извлечени с WordSmith Tools

Клъстерите на глагола *казвам* (вж. Фигура 10), извлечени с WordSmith Tools, детайлизират информацията. Най-често срещаната употреба на глагола е в следните клъстери: *не казвам, че*; *ти казвам, че*; *ви казвам, че*; *казвам, че не*; *казвам ти, че*; *да казвам, че*; *казвам ви, че* и др. Други често срещани клъстери: *каквото ти казвам*; *истина ви казвам*; *не трябваше да казвам*; *прави каквото ти казвам*; *излишно е да казвам*; *казвам само, че*; *казвам всичко това*; *нали ви казвам* и т.н.

Резултатите от търсенето на съчетанията на свършения глагол *кажа* (вж. Фигура 11) се различават от тези на несвършения *казвам* (вж. Фигура 8 и Фигура 9).

Concord

File Edit View Compute Settings Window Help

N	Cluster	Freq.	Length
1	НЕ КАЗВАМ ЧЕ	824	3
2	ДА ТИ КАЗВАМ	610	3
3	ТИ КАЗВАМ ЧЕ	577	3
4	ВИ КАЗВАМ ЧЕ	447	3
5	ДА ВИ КАЗВАМ	377	3
6	КАЗВАМ ЧЕ НЕ	328	3
7	КАЗВАМ ТИ ЧЕ	317	3
8	ТИ ГО КАЗВАМ	292	3
9	ДА НЕ КАЗВАМ	289	3
10	ДА КАЗВАМ ЧЕ	268	3
11	АЗ ТИ КАЗВАМ	256	3
12	КАЗВАМ ВИ ЧЕ	253	3
13	КАКВОТО ТИ КАЗВАМ	238	3
14	ДА ГО КАЗВАМ	230	3
15	АЗ СЕ КАЗВАМ	221	3
16	АЗ ВИ КАЗВАМ	220	3
17	ИСТИНА ВИ КАЗВАМ	204	3
18	КАЗВАМ ЧЕ Е	197	3
19	КАКВО ТИ КАЗВАМ	191	3
20	ТОВА КОЕТО КАЗВАМ	189	3

Фигура 10. Клъстери на глагола казвам, извлечени с WordSmith Tools

Concord

File Edit View Compute Settings Window Help

N	Word	With	elation	Texts	Total	tal Left	al Right	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	КАЖА	кажа	0,000	5 732	67 401	221	221	58	58	52	53	0	66 959	0	53	52	58	58
2	ДА	кажа	0,000	5 459	60 495	52 407	8 088	1 215	874	1 711	21 463	27 144	0	1 118	1 121	1 992	2 090	1 767
3	ЧЕ	кажа	0,000	4 339	21 700	1 575	20 125	551	574	349	94	7	0	15 839	1 617	1 228	844	597
4	ЩЕ	кажа	0,000	3 312	17 151	14 610	2 541	573	518	708	9 106	3 705	0	234	861	534	503	409
5	ТИ	кажа	0,000	2 107	16 315	14 918	1 397	152	100	43	822	13 801	0	100	403	382	274	238
6	НЕ	кажа	0,000	3 729	15 788	9 367	6 421	1 220	3 055	4 090	322	680	0	926	1 985	1 512	1 073	925
7	ВИ	кажа	0,000	2 730	12 122	11 366	756	106	95	41	455	10 669	0	2	156	221	184	193
8	ИСКАМ	кажа	0,000	2 334	10 416	10 064	352	46	172	1 476	8 366	4	0	26	140	94	49	43
9	И	кажа	0,000	3 004	10 360	5 466	4 894	1 167	1 538	2 086	598	77	0	1 181	870	886	940	1 017
10	МОГА	кажа	0,000	3 458	9 838	9 519	319	72	426	3 551	5 464	6	0	16	86	114	58	45
11	Е	кажа	0,000	2 910	8 359	2 309	6 050	617	719	642	330	1	0	439	1 532	1 711	1 387	981
12	НА	кажа	0,000	2 844	8 088	1 479	6 609	716	546	208	6	3	0	2 725	813	1 009	1 020	1 042
13	СЕ	кажа	0,000	2 434	7 681	2 796	4 885	704	1 002	814	265	11	0	40	1 257	1 313	1 206	1 069
14	КАКВО	кажа	0,000	2 082	7 295	5 110	2 185	227	678	2 990	1 214	1	0	1 339	390	183	162	111
15	ЗА	кажа	0,000	2 369	5 969	1 872	4 097	561	492	647	171	1	0	1 042	790	886	741	638
16	ТОВА	кажа	0,000	2 109	5 172	2 377	2 795	572	979	722	103	1	0	885	1 001	324	313	272
17	АЗ	кажа	0,000	1 926	4 953	3 303	1 650	578	939	1 270	429	87	0	534	512	274	187	143
18	НО	кажа	0,000	2 039	4 597	2 780	1 817	718	1 137	837	85	3	0	632	301	304	291	289
19	СИ	кажа	0,000	1 877	4 521	1 885	2 636	314	345	281	98	847	0	10	763	742	576	545
20	ТРЯБВА	кажа	0,000	1 783	4 329	3 668	661	63	191	2 149	1 263	2	0	40	238	161	137	85
21	ГО	кажа	0,000	1 727	4 311	3 109	1 202	123	117	20	5	2 844	0	8	346	334	260	254
22	МУ	кажа	0,000	1 471	4 025	3 464	561	117	92	32	198	3 025	0	4	136	174	122	125
23	НЕЩО	кажа	0,000	1 621	3 890	493	3 397	166	126	163	38	0	0	2 385	544	216	131	121
24	МИ	кажа	0,000	1 746	3 873	1 845	2 028	498	651	584	93	19	0	15	515	668	487	343
25	В	кажа	0,000	2 042	3 697	870	2 827	426	310	129	4	1	0	380	635	530	606	676
26	АКО	кажа	0,000	1 644	3 681	2 405	1 276	437	252	263	891	562	0	403	467	175	128	103
27	СЪМ	кажа	0,000	1 747	3 421	768	2 653	186	257	264	61	0	0	11	1 353	606	415	268
28	КОЕТО	кажа	0,000	1 432	2 976	1 821	1 155	144	603	832	213	29	0	19	599	251	144	142
29	ОТ	кажа	0,000	1 598	2 781	557	2 224	288	215	53	0	1	0	180	316	506	607	615
30	НЯМА	кажа	0,000	1 488	2 657	1 634	1 023	157	191	663	621	2	0	151	340	214	179	139
31	САМО	кажа	0,000	1 461	2 604	1 182	1 422	151	206	397	424	4	0	964	146	111	99	102
32	С	кажа	0,000	1 460	2 468	693	1 775	350	271	71	1	0	0	364	225	326	413	447
33	ИСКАХ	кажа	0,000	1 054	2 403	2 328	75	17	136	707	1 468	0	0	10	20	18	14	13

concordance collocates plot patterns clusters filenames follow up source text notes

5 048 Type-in и и изсунтя. — Защо бяхте заинтригувана, мисис Мърдок? — попитах, колкото да кажа нещо. — Ако този човек беше търговец на монети, който поне малко разбира

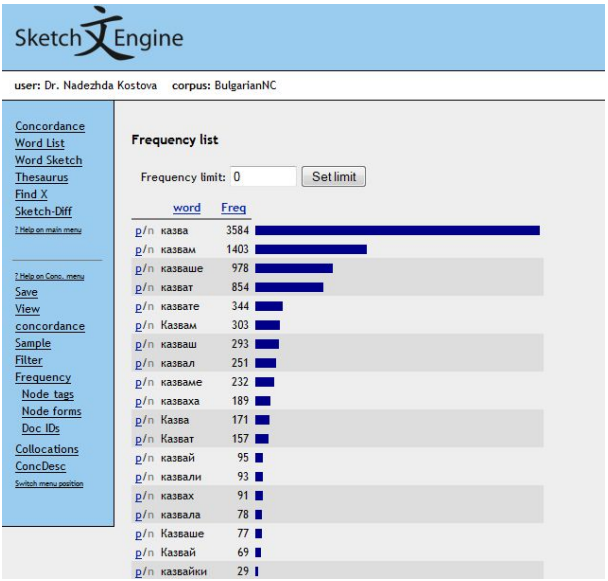
Фигура 11. Съчетания на глагола кажа, извлечени с WordSmith Tools

За разлика от глагола *казвам*, който и в основната, и в другите си форми най-често се свързва с подчинени изречения, въведени

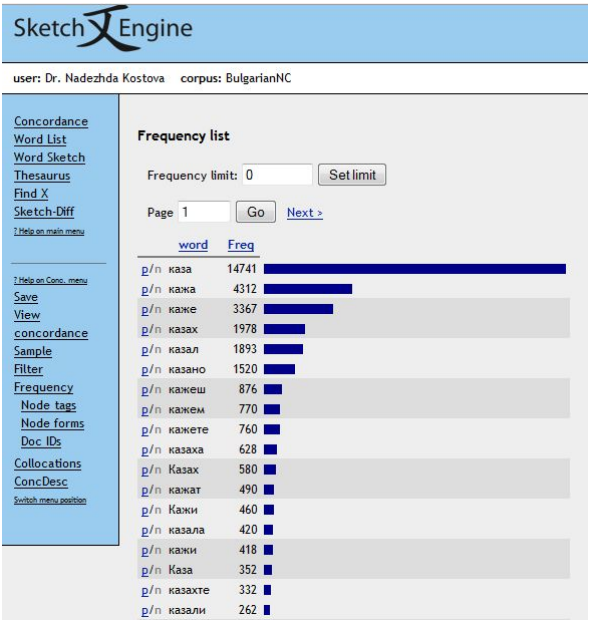
със съюза *че*, глаголът *кажа* най-често влиза в съставно сказуемо – *мога / трябва да кажа*. Освен съюза *че*, в обкръжението на глагола *кажа* се появява и съюзът *ако*: *ако (не) кажа, ако кажа*; подобни примери са по-редки при *казвам*. Интересно е също, че при най-често срещаните обектни съчетания се наблюдава противопоставяне: *(не) казвам нищо – (искам да) кажа нещо*.

Разбира се, двата глагола имат много повече прилики в съчетаемостта си, отколкото разлики. И несвършеният глагол *казвам*, и свършеният *кажа* се срещат често в съчетание с личните местоимения *ти, ви, му*, с предлога *на, за, в, от, с*. С голяма фреквентност са съчетанията *казвам това* и *кажа това*, *казвам само* и *кажа само* и т.н.

Фреквентността на формите на глаголите *казвам* и *кажа* в лексикографския корпус е следната: най-често се срещат *каза, казва, казах, казал, казвам, казват* (вж. Фигура 9). За фреквентността на формите на *казвам* и *кажа* получаваме идентични данни и от Word Sketch (вж. Фигури 12 и 13). Става ясно, че най-често срещаните форми в използваните корпуси са *каза, кажа* и *казва*.



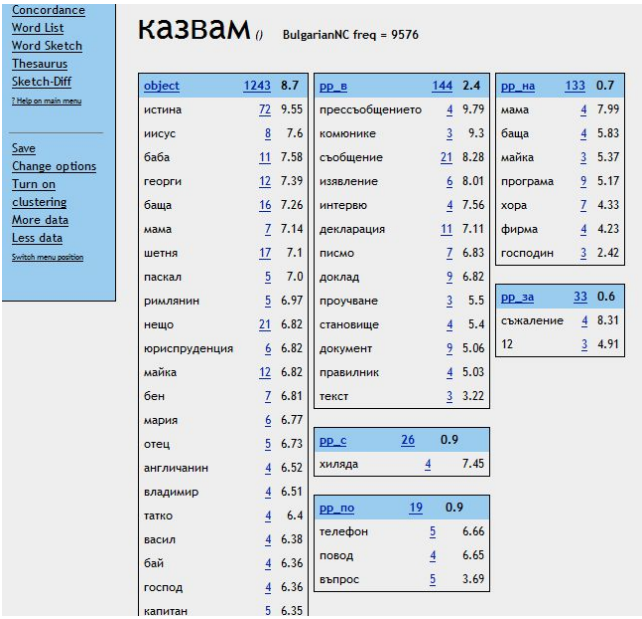
Фигура 12. Данни за фреквентността на формите на глагола казвам, извлечени чрез Sketch Engine



Фигура 13. Данни за фреквентността на формите на глагола кажа, извлечени чрез Sketch Engine

Sketch Engine отново дава по-подробни данни за типичните употреби на глагола *казвам* (вж. Фигура 14), както и при глагола *говоря*.

С най-голяма фреквентност при глагола *казвам* са обектите *истина* и *нещо* (вж. Фигура 14). И тук, както и при глагола *говоря*, наблюдаваме не само съчетания на глагола с именни фрази в обектна функция, но и с обстоятелствени пояснения. Например конструкциите с предлога *в* представят съчетаването на безлични или страдателни форми на глагола с предложна фраза с обстоятелствена функция: *се казва в съобщението, в декларацията, в писмото, в доклада*, които са много характерни за журналистическите текстове. Сред конкордансите в Sketch Engine намираме и пример с изявителна форма на глагола: *Флобер го казва в писмата си*.



Фигура 14. Типични комбинации на лемата казвам, извлечени чрез функцията Word Sketch на Sketch Engine

На трето място по честота на употреба Word Sketch извежда съчетанието на *казвам* с предлог *на – на мама, на баща (си, ти), на майка (си, ти), на хората*. При търсене на конкорданските редове чрез Sketch Engine откриваме и конструкции с предложна фраза с обстоятелствена функция като: *Приемам, че го казваш на шега*.

След това се нареждат съчетанията на глагола *казвам* с предложни фрази с предлога *за*. Word Sketch представя като най-често срещан пример израза *казвам за съжаление*. Подобен е случаят с предлога *с* – в примери като *Аз ви казвам с хиляди бе* (Ив. Вазов) глаголът *казвам* въвежда ремата от предишно изказване, за да бъде потвърдена или обяснена.

Сред конкорданските редове от Sketch Engine намираме примери за типични употреби на глагола:

– казвам нещо за някого: *Лекарите какво казват за състоянието ви? Той наистина вярва в това, което казва за*

Георги Марков;

– казвам нещо за пореден път: *Не ми го казвате за пръв път;*

– казвам нещо с някакво средство: *казвам с поглед, с думи;*

– казвам нещо по някакъв начин: *казвам с молба, с ирония, с дълбока убеденост, с цялата отговорност.*

Типични съчетания на глагола *казвам* с предложни фрази с предлога *по* са: *казвам по телефона* (нещо на някого), *казвам* (нещо) *по повод* (на нещо), *казвам* (нещо) *по въпроса* (*Казах всичко, каквото имах да казвам по въпроса*).

Типичните комбинации на свършения глагол *кажа*, извлечени с Word Sketch (вж. Фигура 15), се различават от тези на *казвам*.

Фигура 15. Типични комбинации на глагола кажа, извлечени чрез функцията Word Sketch на Sketch Engine

В полето за обекти при *кажа*, както и при *казвам*, програмата неточно извежда предимно субекти (причина за това е типичната и за двата глагола употреба след пряка реч). И *казвам*,

и *кажа* се съчетават най-често със съществителното *истина* във функцията на обект. И двата глагола образуват конструкции с предложни фрази във функцията на непряк обект, включващи предлога *на*, както и с предложни фрази с предлозите *в, по,* които показват сходна фреквентност. Съчетанията с предлозите *с*, *за*, които са характерни и за двата глагола, при глагола *кажа* са по-широко застъпени: конструкции с предложна фраза с обстоятелствена функция *(кажа с усмивка, кажа с (някакъв) глас, кажа със сигурност),* конструкции с предложна фраза с обектна функция *(кажа за правителството, кажа за хората)*. При *кажа* са представени и съчетанията с предлози *пред, след, през*: конструкции с предложна фраза с обектна или обстоятелствена функция *кажа пред журналистите, пред съда, кажа след срещата, след мача, кажа през седмицата*.

Чрез WordSmith Tools извличаме данни за съчетанията на глагола *казвам* в Лексикографския корпус, за да сравним с данните от Word Sketch. Използваме и резултати от търсенето в БНК.

Обекти при *казвам*: *истината* (297 срещания)*, всичко (152), нещо (150), същото (22), казвам благодаря (17), съжалявам (2), стига (13)* (данни от Лексикографския корпус, установени чрез WordSmith Tools).

Данни за конструкции на глагола *казвам* с различни предлози:

- *казвам на*: в БНК – 757 срещания, в Лексикографския корпус – 739 срещания; честотни съчетания с обектна или обстоятелствена функция в Лексикографския корпус: *казвам на никого* (65 срещания), *на хората* (33), *на всички* (19), *на вас* (12), *на някого* (7), *на мама* (7), *казвам на глас* (11) (данните са извлечени чрез WordSmith Tools); в БНК – *казвам на шега* (6), *на господин (X)* (3); често срещани клъстери (данните са извлечени чрез WordSmith Tools): *да не казвам на* (59)*, не казвам на никого* (36), *няма да казвам на* (15) и пр.; други примери: *казвам на Елена, на студентите, на продавачката, на леля, на тебе* и т.н.
- *казвам за*: в БНК – 261 срещания, в Лексикографския корпус – 152 срещания; честотни съчетания с обектна или обстоятелствена функция в Лексикографския корпус:

казвам за него (6), *за вас* (5), *за (пореден) път* (8), *за (нечие) добро* (5) (данните са извлечени чрез WordSmith Tools); други примери: *казвам за празненството, за нашия разговор, казвам за твое добро, казвам за втори път, за последен път, казвам за ваше сведение* и др.

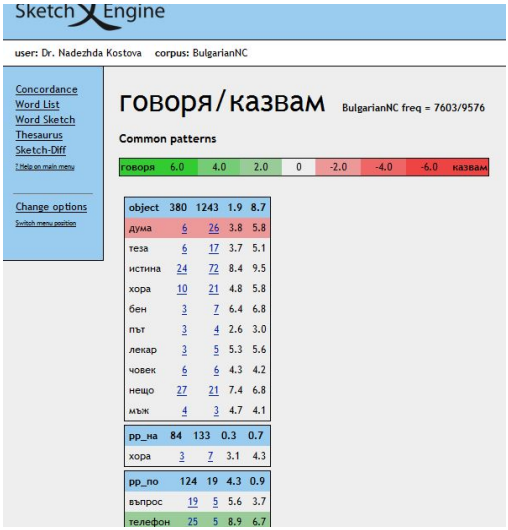
- *казвам с*: в БНК – 153 срещания, в Лексикографския корпус – 109 срещания; честотни съчетания в Лексикографския корпус, установени чрез WordSmith Tools: *казвам с (някакъв) глас* (8 срещания); примери: *казвам с ирония, с треперещ глас, с гордост, с тъга, с лоши чувства, с умисъл, с добро, със спокоен тон, с очи, с ясното съзнание* и др.
- *казвам в*: в БНК – 117 срещания, в Лексикографския корпус – 89 срещания; честотни съчетания в Лексикографския корпус, установени чрез WordSmith Tools: *казвам в очите* (6 срещания); примери: *казвам в буквален (положителен, отрицателен) смисъл, в прав текст; казвам в интервю, в писмо, в книга; казвам в гнева си; казвам в кръга на шегата, казвам в отговор, в заключение, казвам в свой интерес, казвам в лицето, казвам в слушалката, казвам в хола* и др.
- *казвам от*: в БНК – 83 срещания, в Лексикографския корпус – 57 срещания; примери: *казвам от опит, от уважение, от скромност, от учтивост, от сърце, от позицията на млад човек, казвам от трибуната.*
- *казвам по*: в БНК – 62 срещания, в Лексикографския корпус – 37 срещания; примери: *казвам по телефона, по високоговорителя, по навик, по този въпрос, по повод запитването, казвам по едно време*
- *казвам пред*: в БНК – 25 срещания, в Лексикографския корпус – 19 срещания; примери: *казвам пред него, пред всички, пред хората, пред нея, пред теб, пред съда, пред микрофона* и др.

1.3. Обобщение

Корпусните данни за съчетаемостта и контекстуалното поведение на основните глаголи за речева дейност *говоря* и

казвам може да бъдат основа за открояване на сходствата и различията в семантиката на тези глаголи. В обкръжението на глагола *говоря* най-често се срещат предложни именни фрази, съдържащи предлога *за*, докато глаголът *казвам* най-често се свързва с подчинени допълнителни изречения, въведени с подчинителния съюз *че*. Иначе казано, глаголът *говоря* обикновено изисква уточняване на темата, докато при глагола *казвам* се експлицира информацията, заради която се осъществява комуникацията. Ето един пример (цитиран по Георгиев 2006), в който ясно личи разликата в значението между двата глагола: *Той много говори, но какво всъщност каза и каза ли всъщност нещо – не стана ясно*.

Приликите и разликите в съчетаемостта на двата глагола могат да се изследват и чрез функцията Sketch-Diff на Sketch Engine. Представят се съчетанията, които са регистрирани само за единия или за другия глагол, както и съчетанията, които са общи за двата.



Фигура 16. Резултати за съчетаемостта на говоря и казвам, получени с функцията Sketch-Diff на Sketch Engine

От резултатите, получени чрез функцията Sketch-Diff (вж. Фигура 16), ставя ясно, че при сравняване на глаголите *говоря* и *казвам* се установява съществена разлика в честотата на

срещане на съчетанията *казвам (някакви) думи*, *говоря по телефона*, докато *говоря истината* и *казвам истината*, *говоря (някакви) неща* и *казвам (някакви) неща*, *хората говорят* и *хората казват* се срещат почти еднакво в процентно съотношение. При *казвам* най-много са примерите с *казвам истината* (72 срещания), а при *говоря – говоря нещо* (27 срещания). Двата глагола са със сходна фреквентност в корпуса: *говоря* със *7603* срещания, *казвам* с *9576* срещания, но съвпаденията в съчетаемостта им не са много.

При сравняване на контекстуалното поведение на несвършения глагол *говоря* и свършения *кажа* (вж. Фигура 17) се виж-

Фигура 17. Резултати за съчетаемостта на говоря и кажа, получени с функцията Sketch-Diff на Sketch Engine

да, че съчетанията *кажа дума, кажа на хората, кажа по (някакъв) начин* са по-често срещани от съответните съчетания на глагола *говоря*, и обратното, *говоря за (някаква) теза, говоря от името на* са по-често срещани от съответните съчетания на глагола *кажа*. Почти еднаква е честотата на съчетанията *говоря истината* и *кажа истината*, *говоря нещо* и *кажа нещо*, *говоря за времето* и *кажа за времето, говоря на (някакъв) език* и *кажа на (някакъв) език, говоря на ухо* и *кажа на ухо, говоря с (някакъв) глас* и *кажа с (някакъв) глас, говоря в момента* и *кажа в момента, говоря по телефона* и *кажа по телефона* и т.н. При *кажа* най-много са примерите с *кажа истината* (104 срещания), а при *говоря – говоря нещо* (27 срещания)*, истината* (24 срещания). Двата глагола са с голяма разлика във фреквентността: *говоря със 7603* срещания – *кажа* с *35433* срещания, но със сходна честота на съчетанията.

И така, резултатите от корпусното проучване на *говоря* и *казвам / кажа* показват по-голяма прилика между *говоря* и *кажа*, отколкото между *говоря* и *казвам*. Получените данни може да послужат при лексикографското представяне на тези глаголи, като еднаквите съчетания се използват за доказване на синонимни отношения, а различните – на семантична разлика между тези глаголи.

2. Сравняване на резултатите от корпусното проучване на съчетаемостта на глаголите за речева дейност *говоря* и *казвам* с речниковите статии на тези глаголи в многотомния тълковен „Речник на българския език"

Подробно ще разгледаме значенията на глаголите *говоря* и *казвам,* представени в многотомния тълковен „Речник на българския език", като ги съотнасяме с корпусните данни, за да проучим възможностите за подобряване на лексикографското представяне чрез включване на най-често срещаните типични съчетания като най-подходящ и най-представителен илюстративен материал, чрез използване на фреквентността на съчетанията за по-обективно подреждане на значенията, както и за отстраняване на субективни или неточни лексикографски решения.

2.1. Говоря

Глаголът *говоря* има 18 значения в РБЕ; някои от значенията се реализират при функционирането на глагола като непреходен, а други – като преходен.

Първото значение на глагола *говоря* е ‘служа си с устна реч като средство за общуване’, глаголът е непреходен. С това свое значение глаголът влиза в съчетания с други глаголи като *научавам се, започвам*, които представят овладяването на тази дейност от човека. В РБЕ е посочен само един пример (*Децата започват да говорят през втората си година*), в БНК откриваме малко примери, които да експлицират това значение: – *Да, но не приказва много – каза Джон. – Но когато се научи да говори, той ще бъде нашият дипломат!* Също и за нюанса ‘умея, мога да си служа с устна реч’: *Детето вече говори* (в РБЕ); – *Ти излъга! – извика едно от децата, което малко стресна Саманта, тъй като не знаеше, че децата могат да говорят; Остана безкрайно учудена, че малкият вече може да говори* (в БНК).

Другият нюанс – ‘изричам, произнасям отделни думи в реч по някакъв начин’ – се реализира при съчетаването на глагола с наречия като *гръмко, високо, меко, носово*. В БНК е налице възможност да се търсят съчетания на глагола с наречия чрез регулярния израз <говори/F/ *{POS=ADV}>. От съчетанията, извлечени с WordSmith Tools от Лексикографския корпус, виждаме, че в контекстуалното обкръжение на глагола най-често се срещат наречията *спокойно* (35 срещания)*, тихо* (14 срещания)*, високо* (14 срещания). В РБЕ са посочени примери и съчетания: *Тя говореше гръмко. Говоря високо. Говоря меко*.

При следващия нюанс субектите са животни, птици – ‘издавам характерни звуци, които са сигнали за нещо, изразяват нещо’. За този нюанс се откриват много малко примери в корпусите: *Учени открили, че северните и южните птици говорят на съвършено различни езици* (в БНК). Обратно, за нюанса ‘владея и използвам, служа си устно с някакъв език’ има много примери и за представянето му ще са полезни статистически данни: *говоря английски* (50 срещания)*, френски* (39 срещания)*, български* (24 срещания) (данни от съчетанията,

извлечени от Лексикографския корпус чрез WordSmith Tools). Примерите в РБЕ изчерпват всички най-често срещани съчетания: *Вие сте много любезни, но жена ми далеко не говори български език, както се следва*. Вазов; *Говоря отлично английски и френски, а сега уча и руски* и др.

Второто значение ‘изговарям, изричам нещо’ се реализира при функционирането на глагола като преходен; в това значение *говоря* е синоним на *казвам*. При това значение на глагола в обектна функция са съществителните *думи, приказки,* както и местоимения като *нещо, нищо*: *Говорим думи, без да вникваме в същността им; Добри, но ненужни думи говориш ти, жено; Я гледай какви приказки говори; Не говори неща, за които утре ще съжаляваш* (примери от БНК).

Третото значение е непреходно – ‘изказвам с думи, чрез устна реч мислите си, изразявам се гласно в съответствие с особеностите на мисленето, характера, темперамента си и под.; приказвам’, то е синтактично ограничено, тъй като се реализира само в съчетания на глагола с обстоятелствено пояснение. Представя се от съчетанията *говоря малко* (*спокойно, свободно, гладко*), които лесно може да се намерят в корпуса. Нюансът конкретизира на какъв език се говори – ‘изказвам с думи, гласно мислите си, като си служа с някакъв език, диалект и под.’ и се представя от съчетания като *говоря на английски* (17 срещания), *на български* (8 срещания), *на френски* (8 срещания), *на (някакъв) език* (10 срещания) (данни от Лексикографския корпус, получени чрез WordSmith Tools); в РБЕ са посочени само *говоря на диалект, на влашки*.

Четвъртото значение ‘изразявам устно нещо, изказвам някаква мисъл, съобщавам някакъв факт и под.’ се реализира в повелителни изречения, в изречения с пряка реч. Тук корпусите ще са полезни с търсене на повелителни форми (*говори, говорете*) и съчетания като *говори той, говореха хората* и други. За нюанса ‘съобщавам, разкривам (обикн. при насилие, измъчване) нещо пред някого, което му е нужно’ са характерни примери като *ще го накарам да говори*; в корпусите ще търсим конструкции с глаголите *накарам, принудя, заставя* и подобни.

Петото значение ‘обръщам се към някого с думи, насочвам

изказването си към някого’ е непреходно и се разкрива в конструкции с предлога *на*, които имат голяма фреквентност в корпусите, а количествените данни посочват най-често срещаните съчетания: *говоря на вас* (14 срещания)*, на теб (8), на себе си (10), на хората (7),* както и клъстери – *не говоря на вас* (9 срещания)*, да говоря на хората (5), не говоря на теб (5)* (от лексикографския корпус). Също и при нюанса ‘обръщам се към някого, общувам с някого с думи по определен начин’: съчетания – *говоря на ти* (24 срещания)*, на вие (10)*; клъстери – *ви говоря на ти* (9 срещания)*, му говоря на вие (6)* (от лексикографския корпус, получени чрез WordSmith Tools). За нюанса ‘съобщавам или разяснявам нещо на някого’ най-характерни са примери като *той говореше нещо важно на приятеля си*, а за нюанса ‘провеждам разговор с някого, за да му съобщя, разясня нещо поверително или което се отнася специално до него’ – *имам нещо да ти говоря*.

Шестото значение ‘уговарям, съветвам някого за нещо, стремя се да въздействам на някого, като му обяснявам нещо’ се представя от съчетания с местоимения в дателна форма: *говоря му, говоря ѝ*, за които корпусите предоставят богат материал.

За седмото значение ‘изказвам мнението си, становището си за нещо, по някакъв въпрос или за някого (устно или писмено)’ може да се извлекат конструкции с предлозите *за, на, по*: *говоря за това* (602 срещания)*, за него (112), за себе си (94), за нея (64), за тях (53), за времето (19), за хората (15), за любов (11), за пари (13), за политика (13), за миналото (13), за работа (9), за живота (8), за бъдещето (9), за жените (6), на (някаква) тема (26), по въпроса (11)* (от Лексикографския корпус). Нюансът ‘изразявам се, изказвам се за някого или за нещо, като давам някаква оценка, като го преценявам по някакъв начин’ се открива в примери като *говоря добре за учителите си; на публични места говорят положително за държавната администрация* (от БНК).

При осмото значение ‘произнасям слово, реч, правя изказване’ глаголът има за субекти съществителните *министър, президент, премиер, депутат, председател* и т.н., характерни

предложни конструкции са: *говоря от трибуната, пред публика, пред хората, в парламента, в съдебната зала* и др., както и съчетания с числителни редни, наречия, прилагателни (като *говори първи, говори най-напред, говори последен*). За контекстуалните партньори на глагола в това значение корпусите предоставят богат материал.

Към деветото значение ‘съобщавам писмено, давам някакви сведения, обикн. за исторически факти, събития’ се отнасят както конструкции с предлога *за* (*говори за пленените, за битката, за войната и т.н.*), така и с предлога *в* (*в разказа, в книгата, в писмото* и др.), а при нюанса субекти са съществителните *предание, легенда, писмо.* Такива конструкции бързо и лесно се извличат от корпусите.

Десетото значение е ограничено в конструкции с предлога *с* – ‘обменям мисли, информация с някого, общувам с някого чрез реч; разговарям’; корпусите предлагат голям избор: *говоря с него* (612 срещания)*, с теб (599), с вас (385), с нея (275), с тях (121), с тебе (100), с някого (34), с никого (32), с никой (23)*; *с господин (X) (55), с баща (си) (39), с доктор (X) (25), с шефа (13), с адвоката (12), с управителя (10), с директора (10), с мама (9), с полицията (6)* и т.н. (данни от Лексикографския корпус, извлечени чрез WordSmith Tools). Нюансът ‘обсъждам, коментирам, разисквам нещо, някакъв въпрос с някого’ пък е ограничен само в множествено число на глагола, като за него са характерни конструкции с предлозите *за, на, по* (*говорихме за войната, по разни въпроси, на тая тема*); съответно търсенето в корпуси ще бъде съобразено с тези ограничения.

Единадесетото значение на глагола ‘общувам, поддържам връзки с някого’ се реализира обикновено в отрицателен контекст: *не говоря с нея, с онзи няма да говориш*, примери като *говоря с нея* не експлицират в достатъчна степен значението. В корпусите може да се търсят конструкции с предложна фраза с предлог *с*, а също и с отрицателни форми на глагола.

За дванадесетото значение ‘имам уговорка, споразумение с някого за нещо’ са характерни конструкции с предлозите *с* и *за* (*говорил съм с него, говорихме с него за парите*). Тринадесетото значение ‘разпространявам слух’ се реализира с форми за

множествено число, обикновено в конструкции с предлога *за* (*говорят за нея, хората говорят*). Четиринадесетото значение ‘одумвам някого’ се реализира в изрази като: *цяло село я говори, малко ни говорят хората*.

За петнадесетото значение ‘изразявам нещо (мисъл, чувство, настроение) по някакъв начин без реч, без думи (обикн. с поглед, жест)’ са характерни съчетания с предлога *с* (*говоря с очи, с поглед, с жестове*), а при нюанса субекти са съществителните *лице, очи, поглед, движение, жест* и под. При шестнадесетото значение ‘давам възможност да се съди, да се направят изводи за нещо, давам сведения за нещо, свидетелствам’ субектите са неодушевени (*фактите, дрехите, предметите, постъпките говорят*); в корпусите търсим тези конструкции, например при заявка *дрехите говорят* от Лексикографския корпус извличаме чрез WordSmith Tools четири конкордансни реда.

Седемнадесетото значение ‘въздействам, влияя върху съзнанието, чувствата на някого’ също е за неодушевени субекти; то се реализира с конструкции с предлога *на* (*всичко говори на душата ми*), а нюансът ‘въздействам на съзнанието на някого, като му напомням за нещо’ – с конструкции с кратко лично местоимение в дателна форма (*всяко дърво ми говори*). При осемнадесетото значение ‘надделявам, проявявам се ясно, открито, като влияя върху постъпките на някого’ функцията на субект се изпълнява от съществителните *мисъл, идея, чувство* и под. (*в нея говореше чувството на справедливост*); съответно в корпусите се търсят такива конструкции.

2.2. Казвам

Глаголът *казвам* е преходен и има 20 значения в РБЕ, които ще разгледаме подробно, за да преценим как прилагането на корпусния подход може да допринесе за усъвършенстване на представянето на лексикографската информация в тълковен речник.

За първото значение ‘изразявам гласно мислите си; говоря, изричам’ най-типични са примерите с пряка реч. Примери в РБЕ: *Бенковски ги изгледа със съжаление и каза с ироничен тон: – Пък аз си помислих, че бързаш да опиташ оръжието си; – Вът-*

решно страда нещо – му каза лекарят и др. Обикновено глаголът в авторовия израз е в трето лице единствено число минало свършено време; други форми на глагола в авторовата реч са *казва, каже, казваше, казвам* и т.н. Формата *каза* е с най-голяма фреквентност в корпусите. В БНК най-често срещани са съчетанията: *каза той* (44961 срещания) и *каза тя* (22502 срещания); от Лексикографския корпус чрез WordSmith Tools получаваме сходни резултати: *каза той* (41990 срещания) и *каза тя* (20943 срещания), а също и *каза мъжът* (661 срещания), *жената (506), господин (645), госпожа (445), министърът (585), председателят (536), шефът (541), президентът (488), съдията (464), момчето (439), татко (204), някой (285)* и т.н. В примерите, извлечени със Sketch Engine, намираме *каза баща, баба, майка, шефа, Георги, Мария* и др. Разбира се, всички посочени съчетания не може да бъдат изчерпателно отразени в речниковата статия в тълковния речник, но поне тези от тях, които показват най-голяма фреквентност, би трябвало да бъдат включени в зоната с илюстративен материал.

Другото типично съчетание, в което се реализира разглежданото значение, е *казвам на някого*. В РБЕ намираме примери със съчетанието *му каза* (примерът е посочен по-горе) и със съчетания с остарялата дателна форма: *малка мома права стои ... и юнаку тихом казва*; *Калинка каже Стояну*. В лексикографския корпус *му каза* се среща 5849 пъти, с по-голяма фреквентност е само *ми каза* (16769 срещания). В авторовата реч може да има и наречие – за време: *каза вчера* (1062 срещания), за начин: *каза тихо* (1985 срещания), *спокойно (607), бавно (329), замислено (322), сериозно (166)* (данни от Лексикографския корпус, извлечени чрез WordSmith Tools). В РБЕ са включени и примери с изразите *каза с ироничен тон, казвам на шега*; има и пример, в който глаголът е в пряката реч в повелителна форма: – *Кажи бе, момче?* (такива примери са по-малко от тези, при които глаголът е в авторовия израз).

За нюанса към първото значение ‘произнасям дума, израз, звук и под.; изговарям’ в РБЕ са приведени илюстративни примери: *Без да каже сбогом на дружината, тръгна към града*; *Помниш ли, като бях малък, как ми се смеехте, дето не можех*

да кажа „р”. В Лексикографския корпус чрез WordSmith Tools намираме *казвам „благодаря”* (17 срещания)*, казвам „съжалявам”* (2 срещания)*, казвам „стига”* (13 срещания); БНК също предлага подходящи примери: *Беше толкова мил ... когато каза „Благодаря ти”; Едно време щеше да се извини, да каже: „Съжалявам, че съм ви причинил неудобство”*.

Второто значение ‘разправям, разказвам нещо, което зная, което съм научил, някаква случка’ се реализира в съчетания като *казвам какво (кога, как* се случи нещо), *казвам нещо, всичко*. В РБЕ е включено съчетанието *казвам урока*, което е остаряло, в БНК намираме примери като: *Тара умираше от желание да разведри обстановката, да каже някоя смешка*. Нюансът ‘добавям към нещо казано’ се разкрива с типичното съчетание *казвам още нещо*, а нюансът ‘изпълнявам някакво литературно произведение пред публика’ – с *казвам стихотворение*.

Третото значение ‘говори се, носи се слух, мълви се, приказва се’ е граматически ограничено и в корпусите трябва да се търсят само форми на глагола в трето лице множествено число. Типични конструкции за това значение са: *казват, че* (8408 срещания); *както казват* (1313 срещания); *да казват* (1189 срещания, като тези конструкции имат най-голяма фреквентност в Лексикографския корпус. Обикновено в посоченото значение глаголът не се свързва със субект; във функцията на субекти може да се срещнат само следните думи: *хората* (261 срещания)*, всички* (267 срещания)*, някои* (303 срещания) (данни от Лексикографския корпус, получени чрез WordSmith Tools).

Четвъртото значение ‘съобщавам устно или писмено някакъв факт, събитие и др.; уведомявам, известявам’ се разпознава в по-широк контекст. В РБЕ намираме примерите: *А бе, забравих да ти кажа каква работа стана с даскалицата*; *Казаха ни, че в село Могила е дошъл един богат черкезин*; *Ако ти кажат, че азе паднал съм с куршум пронизан, и тогаз, майко, не плачи*; *В едно свое писмо Ботев казва следующето за Левски* и др. В корпусите конструкцията *казвам, че* има най-голяма фреквентност – 3785 срещания в БНК и 3605 срещания, извлечени с WordSmith Tools от Лексикографския корпус, и съот-

ветно големият брой конкордансни редове предлагат богат избор на материал за това значение. След *казвам, че* се нареждат съчетанията *ти казвам, ви казвам* – съответно с 3518 срещания и 2314 срещания в БНК и с 3362 срещания и 2096 срещания, извлечени с WordSmith Tools от Лексикографския корпус. Тези статистически данни дават основание посоченото значение да бъде отразено като второ по ред в речниковата статия на глагола *казвам.*

При нюанса ‘предавам нечии думи, нареждане, поздрав и под.’ глаголът се свързва с две или три фрази със значение ‘нещо’, ‘на някого’, ‘от някого’: *да му каже да дойде, да му кажете много здраве от мене* (РБЕ), при другия нюанс ‘споменавам’ фразите са ‘нещо’, ‘на някого’, ‘за нещо’: *не каза нито дума за срещата си с нея* (РБЕ). Съответно в корпусите ще се търсят конструкции от типа на *кажи от мене: Много здраве му кажи от мене* (примери от БНК); а също и *не казва нито дума* – 51 срещания в БНК. Третият нюанс ‘заявявам нещо (обикн. като уверение или предупреждение)’ също изисква по-широк контекст, в който да се проявят причинно-следствените отношения между две действия: *Казах ти да не ходиш там, че ще загазиш* (РБЕ).

При петото значение ‘споделям нещо с някого, признавам пред някого някакво чувство, преживяване или нещо, което съм извършил’ глаголът се свързва с именна фраза с обектна функция, която съдържа съществителни като *мъка, болка, скръб* и подобни, и с фраза за непряк обект ‘на някого’ или ‘пред някого’. В РБЕ са включени примерите: *Майко, само пред теб една мойта скръб да кажа мога.* Вапцаров*; Ако ви кажа болката си, ще ви изкажа срамо си.* П. Р. Славейков; в БНК намираме само няколко примера като този: *никому не мога да кажа мъката си*.

Шестото значение ‘соча, посочвам, отбелязвам или твърдя нещо като факт или поука’ се отнася както за одушевени субекти (*казах, че*), така и за неодушевени субекти (*историята казва, че*). Нюансът ‘показвам, предсказвам’ също се отнася както за одушевени субекти (*Кажи ми колко години ще живея?*), така и за неодушевени (*сънищата нищо не казват*). Субектите на другия нюанс ‘определям’ са само одушевени, като той се реализи-

ра в конструкции от типа *мога да кажа на колко*; пример от РБЕ: *Никой не можеше да каже на колко години е Месечката*.

Седмото значение ‘отговарям на зададен въпрос; обяснявам, изяснявам, разкривам’ се реализира във въпросителни изречения или изисква въпросителни и други местоимения: *Кажи де, ще ме закараш ли?; да каже кои са, как да ти кажа, какво да му кажа, всичко ще ти кажа, право да ми кажеш*. За нюанса ‘давам някакъв отговор, който съдържа и съвет’ е необходим по-широк контекст (*кажи ми как да постъпя; Аз съм решил да приема, ти какво ще кажеш?*). Осмото значение ‘на мнение съм по даден въпрос, имам определено становище; мисля, смятам’ също е свързано с въпросителните изречения (напр. *какво ще кажеш?*). Нюансът му ‘намислям, решавам’ се реализира извън такъв тип изречения (*каже ли нещо Асен, то и става*), докато нюансът ‘възразявам’ се открива във въпросителни или отрицателни изречения (*каза ли нещо?; татко нищо няма да каже*).

Деветото значение ‘помислям, допускам’ е ограничено във формите за условно и повелително наклонение, за бъдеще време с *нека* и *да*. В корпусите се реализира от конструкции като: *би казал, че* (546 срещания в БНК)*, ще кажеш, че* (448 срещания)*, нека кажем, че* (69 срещания)*, да кажем, че* (2649 срещания).

Десетото значение ‘поръчвам, заръчвам, нареждам нещо’ изисква съюз *да* (*казал да го хванат*) – търсенето в корпусите е насочено към такива конструкции; нюансът ‘пожелая, поискам, намеря за добре’ се реализира от съчетания *каквото кажеш* (150 срещания в БНК)*, както кажеш* (422 срещания)*, когато кажеш* (155 срещания), в които глаголът е в свършен вид.

При единадесетото значение глаголът има за субекти съществителните *поглед, очи, жест* – ‘изразявам нещо (мисъл, чувство, настроение) по някакъв начин’ (*очите ѝ нищо не казват*); в корпусите търсим съответните конструкции, например за *казват очите* от Лексикографския корпус извличаме чрез WordSmith Tools 7 конкордансни реда.

При граматически ограничени и конструктивно обусловени значения търсенето в корпусите е за съответните форми и съчетания. Дванадесетото значение е граматически ограничено – из-

разява се само от свършения вид на глагола в повелително наклонение или бъдеще време и в съчетание с лично местоимение в дателен падеж (*ще ти кажа аз на тебе, ще ви кажа аз на вас*), като служи за изразяване на закана към някого. При търсене на конструкцията *ще ти кажа аз на тебе* в лексикографския корпус намираме 5 примера.

И тринадесетото значение ‘наричам някого или нещо с някакво име; назовавам, именувам’ е ограничено – проявява се само в съчетание с лично местоимение във винителен падеж (пример от РБЕ: *В наше село имаше един човек, казваха го дядо Славчо*); при търсене на съчетанието *казват го* в Лексикографския корпус от 37 конкордансни редове, извлечени чрез WordSmith Tools, само един пример илюстрира това значение: *Аз зная това кученце. Казват го Меджи*. За разкриване на нюансите към тринадесето значение е необходим по-широк контекст – ‘наричам някого или нещо с някакво друго име или с някакъв прякор’ (пример от РБЕ: *Разказваше една попрестаряла жена, на която казваха Каломенчанката*) и ‘назовавам нещо по друг начин въз основа на сходство’ (пример от РБЕ: *Какъв живот живеем ний тук! Кажи го затвор, кажи го пустиня*).

Четиринадесето значение се реализира в съчетания с *така, тъй* – ‘извършвам определено движение с тялото си, с ръцете, краката и др. с дадена цел’ (*кажи така, кажи тъй*). Петнадесетото значение ‘показвам нещо чрез някакво движение, жест’ се представя от примери като: *Де, Станке, кажи сега как баба хоро играла* (РБЕ). Шестнадесетото значение ‘издава някакъв звук, шум, трясък и под.’ се разкрива в съчетания на глагола с междуметия (*нещо каза пуф-пуф*).

При седемнадесетото значение ‘значи, означава; ще рече’ ще се търси формата на свършения глагол за трето лице бъдеще време *ще каже*; при осемнадесетото значение – конструкции като *кажи го от радост, кажи го от скръб; кажи го бял, кажи го черен*.

Последните значения (19 и 20) също са конструктивно обусловени и представят употребите на глагола *казвам* като вметната дума (само свършен вид и само в трето лице единствено

число) – предположение за достоверността на някакъв факт (*вече, кажи, си дойдохме*) или приблизителност (*Имаше къща, кажи, цял конак*); при търсене на формата *кажи* от извлечените от корпусите конкордансни редове трябва да се намерят тези примери, които илюстрират разгледаните значения.

2.3. Обобщение

От направения анализ на речниковите статии на глаголите за речева дейност *говоря* и *казвам* в многотомния тълковен „Речник на българския език", сравнени с резултатите от корпусното проучване на съчетаемостта на тези глаголи, се вижда, че корпусният подход може да подобри лексикографското представяне не само като предостави подходящ илюстративен материал, но и чрез статистическите данни, които дават основание на лексикографа да подреди значенията на глаголите в определен ред, а също и да включи най-често срещаните, най-типичните примери.

В заключение можем да кажем, че корпусният подход предоставя на лексикографа изключително богат материал и различни възможности за неговото сортиране като търсене на форми на глаголите, на съчетания, на фрази, както и статистическа обработка на данните. Именно заради големия си обем и бързото и лесно търсене и подбор на примери корпусите изместват традиционните лексикографски източници – лексикалните картотеки. Статистическите данни, които се извличат от корпусите, предлагат обективен критерий за включване в речниковата статия на примери с най-често срещаните съчетания. Корпусният подход дава възможност за изчерпателно и детайлизирано представяне на семантиката на лексемите в тълковен речник.

ЛИТЕРАТУРА

Благоева и др. 2009: Благоева, Д., С. Колковска, В. Сумрова, М. Чоролеева. Електронизация на лексикографските ресурси. – *Лексикографски преглед*, № 10, стр. 5–56.

Благоева, Колковска 2007: Благоева, Д., С. Колковска. Електронен корпус за целите на „Речник на българския език" – състояние и перспективи. – В: *Лексикографията и лексикологията в съвременния*

свят. Материали от IV национална конференция с международно участие на Българското лексикографско дружество в чест на проф. Кристалина Чолакова. София, 21–22 октомври 2006 г. Велико Търново. Знак'94, с. 277–286.

Верлинд, Селва 2001: Verlinde, S., Th. Selva. Corpus-based versus intuition-based lexicography: defining a word list for a French learners' dictionary. – In: Rayson, P., A. Wilson, T. McEnery, A. Hardie, Sh. Khoja (eds.). *Proceedings of the Corpus Linguistics conference.* Lancaster.

Георгиев 2006: Георгиев, Б. *Значение и употреба на глаголи и имена за описание на общуването*. Второ издание. Електронно издателство LiterNet. <http://liternet.bg/publish4/b_georgiev/znachenie /index.html>.

Килгариф, Рандъл 2002: Kilgarriff, A., M. Rundell. Lexical Profiling Software and Its Lexicographic Applications: a Case Study. – In: *Proceedings of the 10th Euralex International Congress.* Copenhagen, pp. 807–818.

Коева и др. 2010а: Koeva Sv., D. Blagoeva and S. Kolkovska. Bulgarian National Corpus Project. – In: Calzolari, N., K. Choukri, B. Maegaard, J. Mariani, J. Odjik, S. Piperidis, M. Rosner, D. Tapias (eds.). *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10).* Valletta, European Language Resources Association (ELRA), pp. 3678–3684.

Коева и др. 2010б: Коева, Св., Д. Благоева, С. Колковска. Българският национален корпус – състояние и перспективи. – В: *Лексикографията в европейското културно пространство. Материали от V национална конференция с международно участие по лексикография и лексикология. София, 19–20 октомври 2009 г.* Велико Търново. „Знак'94", с. 289–300.

Тинчев и др. 2007: Тинчев, Т., Св. Коева, Б. Ризов, Н. Обрешков. Система за разширено търсене в корпуси. – *Литературата*, № 2, с. 98–116.

Йорданка Трифонова
(София)

СЛАВЯНСКО-БЪЛГАРСКАТА ДВУЕЗИЧНА ЛЕКСИКОГРАФИЯ С ОГЛЕД НА ИЗПОЛЗВАНЕТО Ѝ В Е-РЕЧНИЦИТЕ

Slavonic-Bulgarian dictionaries, published in the last 50 years, are confronted through a complex of macrostructure, microstructure and social-linguistic parameters: *kind of a dictionary*, *necessity of creating*, *purpose*, *extent*, *chose of entries and sources*, *structure of the dictionary and the entry*, *others special features (special lists*, *abbreviations*, *presence of proper names etc.).* In the report the results of comparing are generalized. These dictionaries, connected with the culmination in the development of the Slavonic studies in Bulgaria in the second half of 20. century, are oriented at the language practice, which is mostly comprehended like an understanding of the foreign language and translating of the native.

Ключови думи: двуезични речници, славянско-български речници, е-речници, лексикография, славистика

Key words: bilingual dictionaries, Slavonic-Bulgarian dictionaries, e-dictionaries, lexicography, Slavonic studies

Двуезичните речници стоят в основата на модерната българска лексикография (срв. Филипова-Байрова 1962, Кювлиева-Мишайкова 1997, Чоролеева 2008, Чоролеева 2009), съответно и на модерното българско езикознание, но българската езиковедска наука продължава да бъде в дълг[1] както към обобщаването на фактите, така и към осмислянето на резултатите в областта на двуезичната ни лексикография[2].

Причините за това са няколко. На първо място – отношението към този тип езиковедски съчинения само като към *помагала* за определена практическа дейност, отразено в разбирането, че двуезични речници не се пишат, а се съставят или изработват[3]. На второ място – въпреки че са най-обемистите, най-разпространените и най-често създаваните лингвистични съчинения, двуезичните речници никога не са „във”, а винаги са „между” две лингвистични традиции (Кошкова 2006: 905). Те стоят извън границите на „родното”, предполагат определена

степен на необходимост от „чуждото” и изискват съответната степен на интерес и толерантност към него. На трето място – двупосочната им ориентация в пространството е съпроводена с преходност и бързо остаряване във времето[4].

Общата ситуация в областта на българската двуезична лексикография има съответното отражение при двуезичните речници на „малките” славянски езици, които са обект на изследване в това изложение. Целта на работата е да се разгледат създадените през последния половин век двуезични „традиционни” българско-славянски и славянско-български речници като ресурси за изработване на двуезични електронни речници на тези езици.

*

За изпълнение на поставената задача славянските двуезични речници[5], свързани с българския език, се описват и съпоставят чрез следния комплекс от макроструктурни, микроструктурни и социолингвистични параметри: *тип речник*, *необходимост от създаването му*, *предназначение*, *обхват*, *избор на словник и източници*, *структура на речника и речниковата статия*, *други особености* (*специални списъци, съкращения, присъствие на проприи и под.*)[6]. Както проличава от избраните критерии, целта е особеностите на речниците да бъдат представени във връзка с изпълняваната от тези лексикографски трудове социална роля и с оглед възможностите за по-нататъшното им приложение.

Избраните параметри се основават на разбирането за двуезичния речник като вид „помагало с комплексен парадигматично-синтагматичен характер, отразяващо лексикално-семантичните корелации и всички други видове системни отношения между двата езика, в които се реализират тези отношения.” (Чермак 1995: 231).

Приема се, че ДР може да бъдат разделени според:

а) потребностите за създаването им на *общи, терминологични* и *учебни*;

б) общите функции (дихотомии) на езика, т.е. кодиране : декодиране, синтез : анализ, говорене : писане, слушане : четене

на *кодиращи* и *декодиращи*, от което следва, че идеалните лексикографски отношения между два езика изискват създаване на четири речника;

в) традиционната издателска практика на *малки (джобни), средни и големи*, въпреки че характерът и обемът на ДР се определят от предназначението му, типа ползватели и техните нужди.

Детайлната типология на ДР отчита повече показатели (срв. Чермак 1995: 233, Аткинс, Рындъл 2008: 24–35), но типът на разглежданите тук речници може да бъде определен чрез отбелязаните три делитбени признака.

В съответствие с Фр. Чермак приемаме, че ДР *се създават* поради „нуждата да се облекчи, ускори и стандартизира контактът между общности, които използват различни езици[7], следователно и за различни видове превод", а за цел на двуезичната лексикография като вид дейност се определя „търсенето на възможности и средства за взаимна субституция, по-точно намирането на такива лексикални единици от целевия език, които, поставени в съответния контекст, дават гладък превод" (Чермак 1995: 231). Създаването на ДР се разглежда като „практическо-теоретична дейност", извършвана от хора, чиито познания са ограничени във времето и пространството. Всеки ДР е „изкуство на възможното" (Чермак 1995: 231).

През последните петдесет години ДР са използвани за задоволяване главно на потребностите, отразени по-горе в точка а): 1. пътувания, търговия, различни видове културни и икономически връзки, заради които се създават *общите* речници (академични и „големи" речници, речници на мъртвите езици); 2. специални контакти, свързани с определени области на знанието, използващи специална терминология, заради които се създават *терминологичните* речници; 3. изучаване на чужди езици, особено без непосредствен контакт, поради което възникват *учебни* речници. Според Фр. Чермак основни ползватели на първия тип речници са администраторите и преводачите (устни и писмени); вторият тип речници се използват главно от писмени преводачи специалисти, а третият тип – от ученици, студенти, обучавани и лингвисти.

Параметърът *обхват на лексиката* зависи от типа речник и

предназначението му. Този параметър, както и следващият – *избор на словник и източници* – се намират в пряка връзка с равнището на съответната национална лексикография. В следващото изложение те се разглеждат като показатели за книжовноезиковата, тематичната и хронологичната съвместимост на двете лексикални системи, отразени в речниците.

При характеризиране на ДР в *структурно отношение* се изхожда от разбирането за лексиката като система с определена структура, проявяваща се в разделянето ѝ на ядро и периферия. ДР се схваща като „комплексна и структурирана система от заглавни думи, към които се търси еквивалентна заглавна дума в друга система, а не като списък от думи, чиито съответствия се търсят в друг списък от думи" (Чермак 1995: 232). Приема се, че в добрия ДР трябва да бъдат представени следните основни типове информация: а) точното място на съответната лексема в системата от взаимозависимости (категории), към които тя принадлежи; б) валентността, основната синтагматика и основните видове съчетаемост за всяка заглавна дума и за съответния ѝ еквивалент. Във всеки ДР водещ трябва да бъде принципът, че на заглавната дума като комплекс от информации съответства заглавна дума със същия комплекс от информации (Чермак 1995: 231)[8].

Параметрите, използвани за описание на разгледаните речници, се отнасят за добрия ДР от „предтехнологичния" период на двуезичната лексикография. С навлизането на компютрите в лексикографията възниква понятието *електронен речник* (е-речник), под което в настоящия момент се разбира: а) речник на електронен носител, за разлика от „хартиения" речник (Аткинс, Рындъл 2008: 24, Бергенхолц и др. 2009: 102); б) речник, който се изработва с помощта на новите технологии, т.е. компютри, бази данни, програми (Аткинс, Рындъл 2008: 3, Бергенхолц и др. 2009: 102), и в) речник, в който данните, получени чрез нови технологии, се представят и използват по начин, различен и по-ефективен в сравнение с начина на представянето им в традиционните ДР (Хартман, Джеймс 1998: 96, Бергенхолц 2009: 102, Липовска 2009: 166-189)[9]. Следващото изложение се основава на последното от посочените определения за двуезичен е-речник.

Изработването на добър електронен ДР изисква съществуването на: (1) две специални картотеки за двата езика и множество преводни ексцерпции, или (2) абсолютно паралелни компютърни корпуси на двата езика със съответния идентифициращ и търсещ софтуер (Чермак 1995: 232). Възможността традиционните славянски ДР да бъдат използвани за целите на електронната лексикография се преценява според тези две изисквания.

Фактите показват, че в настоящия момент електронните двуезични и многоезични речници на малките славянски езици са рядкост, тъй като електронната лексикография се развива преди всичко *от* и *към* езиците за международно общуване[10]. Лексикалното родство между българския и останалите славянски езици, съчетано с различната степен на граматични различия, изисква специфичен подход при изработване на българско-славянски двуезични е-речници, срв. Липовска 2009: 171–189. Славянският ДР речник включва определен вид опит – положителен или отрицателен, който не може да бъде пренебрегван. Освен това е показателен за степента на зрелост на националната ни двуезична лексикография и разкрива особеностите към достигане на тази зрелост.

*

Описанието на двуезичната славянска лексикографска продукция, свързана с българския език, бе направено с оглед на хронологията на речниците[11]. Този начин на подредба на материала бе предпочетен пред разделянето по подгрупи (*източни, южни* и *западни* славянски езици) или разделянето според посоката на речниците (*българско-славянски* или *славянско-български*) поради: а) желанието българско-славянските и славянско-българските лексикографски контакти да бъдат показани в тяхната цялост[12], б) все още съществуващото близко лексикално родство вътре в славянската езикова група, което позволява тя да бъде разглеждана като едно цяло, в) необходимостта да бъдат обединени речници на езици с различен брой ползватели и с различен социален престиж – на единия полюс стоят лужишките езици, които имат ограничено географско разпространение

и социално влияние, а на другия – руският, който изпълнява ролята на lingua franca за славянските народи през последните петдесет години и оказва влияние върху двуезичната лексикография на всеки един от тях.

Изследваният материал обхваща двуезичните общи средни речници на „малките" славянски езици и основните руско-български и българско-руски речници, създадени след 1960 г. В прегледа не са включени речниците от по-ранния период на славистиката, т.е. изданията преди 1960 г., както и специалните речници преди и след тази дата[13]. Поради споменатото положение на руския през разглеждания период, както и поради излизането през последните години на две сериозни изследвания върху руско-българската (Липовска 2009) и върху българско-чуждата лексикография (Павлова 2010), в приложението са отбелязани и в текста са разгледани само тези руско-български издания, които по преценка на авторката са оказали най-силно влияние върху развитието на двуезичната лексикография у нас.

Очертаните хронологически рамки маркират периода на най-ползотворното развитие на славистиката в България и по света през XX в.: „изключително благоприятен за развитието на славянската филология е периодът след края на Втората световна война, когато най-голямата славянска страна – Русия, се нарежда между победителите в нея и се превръща в една от двете световни суперсили, определящи политическия, икономическия, военния и културния развой на планетата" (Куцаров 2008б: 17).

Важен показател за активизиране на славянската двуезична лексикография през разглеждания период е количеството на изданията. Като изключим най-ранните лексикографски пособия на славянски езици и изданията до Втората световна война[14], разглежданите речници започват да се появяват в края на 50-те и началото на 60-те години на миналия век: през 1960 г. излиза първото издание на големия руско-български речник, а през 1961 г. – полско-българският речник на Ив. Леков и Фр. Славски, както и преработеният и разширен чешко-български речник на Цв. Романска (III издание); през 1966 г. излиза ново разширено издание на големия българско-руски речник на С. Б. Бернщейн,

през 1967 г. – първият българско-сърбохърватски речник на М. Младенов, а през 1970 г. – първото издание на словашко-българския речник на Ив. Леков и В. Бланар. През 1986 г. се появява големият руско-български речник на Г. Тагамлицка и С. Влахов, а през 1988 г. излиза първият българско-украински речник на Ив. Стоянов и Е. Чмир. Най-оскъдни откъм лексикографска продукция са 90-те години на миналия век. Началото на хилядолетието обаче е белязано с активизиране на двуезичните славянски издания: през 2000 г. излиза преработеният българско-сръбски речник на М. Младенов, през 2002 г. – готовият отпреди десетилетие чешко-български речник под редакцията на покойния Св. Иванчев, през 2003 – новият българско-руски речник на А. Кошелев, през 2004 г. – първата част на тритомния българско-словашки речник на М. Кошкова, а също новият като концепция руско-български речник на С. Влахов. Второто издание на същия речник през 2007 г., допечатката от 2009 г., както и допечатката на двутомния чешко-български речник от 2005 г.[15] свидетелстват за обществения интерес към подобни издания и за жизнеността на славистичната лексикография.

Съществен с оглед на насочеността на речниците и качеството на лексикографската продукция е въпросът за създателите им. Данните показват, че през последния половин век инициатори и редактори на средните и големите двуезични славянски речници по-често са изтъкнати преводачи и университетски преподаватели[16], отколкото езиковеди. В течение на времето активността на академичните институции нараства, но създаването на славянски речници у нас продължава да се ръководи повече от нуждите на образованието, отколкото от потребностите на науката и икономиката.

*

Разгледаният материал показа, че в началото на второто хилядолетие „бели петна" в двуезичната българско-славянска лексикография са *белоруският*, *хърватският* език и *лужишките* езици. През изминалия период лексикографските отношения между български, от една страна, и белоруски и лужишки, от друга, са поддържани чрез съответните руски, донякъде и

немски речници[17]; поради геополитическата ситуация в бивша Югославия лексикографските контакти между български и хърватски са осъществявани чрез сърбохърватско-българските и обратните им речници.

Количеството на славянско-българските речници (15) превишава това на българско-славянските (10). Данните съответстват на наблюденията, очертаващи посоката на лексикографските релации между руски и български: за период от 150 години броят на общите руско-български речници (33) е двойно по-голям от броя на общите българско-руски (16) (Липовска 2009: 49–140). Налага се заключението, че при благоприятните за развитието на славистиката условия през последния половин век потенциалът на двуезичната лексикография откъм българска страна не е достатъчно използван[18].

При езиците, с които са налице лексикографски отношения, се наблюдава асиметрия: не съществува нито среден, нито голям украинско-български речник; в словника на големия българско-чешкия речник [22] е отразено състоянието на българската лексикална норма отпреди повече от половин век; въпреки актуализациите през 80-те години на миналия век подобно е състоянието и при словниците на българско-полските и полско-българските речници ([15] и [16]).

Специализираните издания са няколко[19] – чешко-български и българско-чешки технически речник ([23], [24]), сръбско-български речник на идиоми и омоними ([8]), многоезичен речник в областта на туризма ([27]), както и съвсем новият речник на чешката некнижовна лексика ([28]). Тематичната насоченост на специалните речници показва, че възникването им е резултат колкото на социални потребности (технически и туристически термини), толкова и на научен интерес (идиоми, омоними, некнижовна лексика).

Поради екстралингвистични причини до 1989 г. тиражите са доста по-високи от съвременните – средните речници имат тираж около 10 000 бройки, а кратките дори и по-висок[20]. Данните за последните двайсет години обаче не може да бъдат смятани за достоверни, тъй като в новите речници тиражът или не се посочва, или е отбелязан като 1 000 бройки, без да може

да се установи размерът на допечатките.

По обем[21] преобладаващата част от разгледаните речници са „средни" – броят на заглавните им статии варира от 35 000 до 60 000 думи, най-често между 40 000 и 50 000 заглавки. Най-богат лексикално е руско-българският речник ([2]), публикуван през 1986 г., за който са посочени 75 000 заглавни думи.

Най-пълно, най-цялостно и на най-високо теоретично равнище през изминалия половин век се е развивала руско-българската двуезична лексикография[22], което е обяснимо с оглед значението и ролята на руския език в България[23] през разглеждания период. На второ място по брой, обем и поява във времето се намират полско- и чешко-българските лексикографски издания, съответно отношения. Възникването и поддържането на последните е следствие, от една страна, на културните връзки между България и тези страни преди и особено след Втората световна война[24], а от друга – на активния масов туризъм между 60-те и 90-те години на миналия век. За чешкия известна роля изиграва присъствието на голяма чешка диаспора в България още от следосвобожденската епоха.

Според отбелязаното в предговорите преобладаващата част от речниците са създадени поради общи, неконкретизирани потребности: укрепване и разширяване на културни отношения, усъвършенстване на езиковата комуникация, учредяване и развитие на университетски специалности (липса или незадоволително равнище на лексикографски пособия за изучаване на полски, чешки, сърбохърватски, словашки, срв. [7], [11], [12], [18], [26]). Два малки речника ([14], [25]), замислени и изпълнени като „огледални", са изработени поради практически нужди (активизиране на туризма). Краткият българско-украински речник ([6]) пък възниква поради актуализирания стремеж към запазване на езиците на малцинствата.

Наред с общия тип потребности, характерни за по-ранния етап от развитието на славянско-българската лексикография, в изданията след 1990 г. се забелязват и други тенденции. От една страна, като причина за създаване на нови речници се посочва остаряване на предишни издания поради естествения развой на езиците или геополитически промени, срв. [3], [4], [9],

[10], [18], [27]. От друга – наблюдава се стремеж към достигане на равнището на световната двуезична лексикография. В [4] този стремеж е изразен чрез новия тип речникова макро- и микроструктура, в [18] – чрез стриктно придържане към понятието преводна еквивалентност, в [27] – чрез декларацията, че се съпоставят не отделни думи, а лексикални системи, които трябва да бъдат представени според новите изисквания пред двуезичната лексикография. Съзнанието за теоретичната значимост на създаваните пособия присъства в почти всички разгледани средни и по-големи речници, но е изведено като основна цел само в един от тях [17].

В съответствие с общите потребности предназначението на разглежданите речници също е общо – предполага се, че те ще се използват за *изучаване* на езика и за *четене* на художествена и специална литература, както и периодика. Като потенциални техни ползватели са посочени на първо място студентите, преподавателите, преводачите, а след това изобщо лицата, които професионално или любителски имат контакт със съответния език. Мястото, където речниците ще се използват, е преди всичко „другата" страна, тъй като ползвателите в „родната", намиращи се в непосредствен контакт с носители на езика, порядко прибягват до допълнителни помагала[25]. Преобладаващата част издания са декодиращи; при специалните речници ([14], [25], [23], [24], [8], [29]) има няколко опита за по-ефективно, т.е. „огледално", използване на лексикографския материал, а в три средни речника – [3], [15], [16] – двупосочната ориентация е отбелязана като търсена цел[26]. Разглежданите издания са предназначени главно за пасивно ползване на съответния език; само в [27] целта е създаване на говорни умения.

Кръгът на потенциалните ползватели, както и фактът, че става дума за преводни речници на национални езици, обясняват обхвата на включената лексика. Предполага се, че разглежданите издания ще се ползват от образовани или желаещи да се образоват хора, за които основна форма на езика е националният книжовен език. Поради тази причина в тях присъства изобилно: *неактуална лексика* (архаична, остаряла, книжна), понякога непозната дори за редовите носители на съответните ези-

ци, но срещана в литературата; национално-характеризираща *диалектна лексика*, въпреки че става дума за речници на книжовни езици; *терминологична лексика*, вследствие все по-силната интелектуализация на съвременните славянски езици; *езикови реалии и фразеология*, чието присъствие може да обясни ред културологични особености на съответната нация. За сметка на очевидните предпочитания към писмената форма на езика ограничено са включени *разговорни думи* – според големината на речника, *неологизми* – според узуализирането им, а *вулгаризмите* и *жаргонизмите* отсъстват напълно. Интересна е степента, в която са представени така наречените *интернационализми,* обозначавани като заемки или чужди думи в зависимост от отношението към тях в съответната лексикална традиция. Присъствието им не се подчертава, тъй като става дума за речници на национални езици, но в някои случаи, напр. [9] и [10], споменаването е неизбежно поради изобилието им и вероятността от грешки; в други, напр. [29], те са основна група при търсенето на обща основа за изучаване на близкородствени езици; в трети, напр. [4], еднаквите в български и руски „чужди" думи са основната група „под чертата".

Данните от предговорите показват, че съставителите са определяли словника на съответния речник чрез: а) ползване на словниците на съществуващи тълковни речници, напр. в [1], [3], [10], [15], [16], [17]; б) използване на словник, специално създаден за целите на преводните речници, напр. в [2], [27], [29]; в) ползване на вече съществуващи двуезични помагала и словници (при почти всички речници); г) уточняване според картотеки на езикови институции, напр. [5] и [18]; д) позоваване на личен опит и наблюдения, напр. [1], [7а], [9], характерно като цяло за по-ранен исторически период. Най-разпространен е първият начин, при който обаче съществува опасност от хронологично несъответствие между изходния словник и еквивалента му. Възниква въпросът доколко адекватно разглежданите ДР отразяват лексикалната норма на съответния изходен език през съответния исторически период. По-горе вече бе посочено, че влияние върху обхвата на включената в разглежданите речници лексика оказват и екстралингвистични фактори (потребност,

предназначение, потенциални ползватели и др.). Поради това – противно на декларираното в предговорите – не може категорично да се твърди, че фреквентността[27] и мястото на дадена лексикална единица в националния лексикален фонд са главни критерии за нейното включване и в словника на преводния речник[28].

Структурата на речниците и речниковите статии отразява динамиката в развитието на славянската лексикографска теория и практика през последния половин век и позволява да се направят следните обобщения:

а) първа особеност на разглежданите лексикографски издания е *високата степен на структурно сходство* между тях. То се изразява в еднаква азбучна подредба според началния писмен знак на заглавките в изходния език, в двудялба на линейното графично представяне на материала и в дозиране на информацията във вид на дискретни единици (думи и фраземи) в основна (речникова) форма. На макроструктурно равнище разлики съществуват само при представянето на системните лексикални връзки (използване / неизползване на препратки и на гнездовия принцип; показване / непоказване на синонимни, антонимни и омонимни отношения). На равнище речникова статия структурните разлики се изразяват в приемането / неприемането на концепцията за главно значение и второстепенни значения и съответно в графичното ѝ представяне[29].

Новите възможности на електронните речници, т.е. организирането на материала не само по хоризонтален, но и по вертикален принцип, се отразяват в структурата на създадените през последното десетилетие речници: в [4] на макроструктурно равнище се използва не само линейна хоризонтална, но и вертикална двудялба, при която вместо един възникват два паралелни азбучни списъка; в [29] линейната структура се запазва, дори „удължава", но изходен може да бъде всеки език; в [4], в [29], както и в по-рано създадените специални речници ([8], [14], [23], [24], [25], [28]) за основна структурна единица е приета лексата, а не думата;

б) втора обща особеност на разглежданите речници е *системната им ориентираност*, с други думи, съпоставянето и успоредяването на лексикалните единици като части на лекси-

калните системи, а не като като части на текста. В някои речници, напр. [2], [4], [9], [17], [18], е декларирано, че са изградени на *принципа на превода*, но това твърдение не е коректно, тъй като се отнася само до разграничението тълковни vs преводни речници[30]. В повечето разглеждани издания се извършва превод от системата на единия език в системата на другия, а функционалният аспект на лексикалната единица се разкрива чрез регистриране на съчетаемостта ѝ и чрез повече или по-малко обилна екземплификация в зависимост от обема на речника. Затова е по-коректно да се посочи, че разглежданите ДР са изградени върху тезата за съществуване на *системна* лексикална (семантична, преводна) еквивалентност между лексикалните единици на два езика, и да се работи с термини като *заглавка* и *съответка*[31] (или *еквивалент*), без да се използва понятието *превод*.

В два от създадените през последното десетилетие речници ([4], [18]) разграничението между системната и функционалната еквивалентност на съпоставяните лексикални единици е ясно изразено: С. Влахов и А. Липовска „търсят такива похвати за лексикографиране, при които при представянето на руски думи и фрази с български езикови средства да се постигне максимално вярно разкриване на *съдържанието* им и *употребата* в речта.” ([4]: 7). Според авторите начинът за изпълнение на тази задача е „поставяне на дадената единица в естествената ѝ контекстуална среда” ([4]: 7). М. Кошкова, в съответствие с теорията за преводната еквивалентност, развивана в словашката двуезична лексикография (Секанинова 1993: 58–79), определя словашкия еквивалент на българската лексикална единица не чрез превод в рамките на словашката лексикална система, а чрез съпоставка на употребите и съчетаемостта ѝ в български и словашки текстове;

в) ограниченията, налагани от „хартиения” облик на ДР, определят следващата обща особеност, а именно – *съвместяването на различен вид информация* – семантична, граматична, стилистична. Тъй като основна структурна единица на разглежданите речници е думата като част на речта, необходимо е тя да бъде представена чрез различните си словоформи (граматична

информация), значения (отделни лекси) и употреби (така наречените стилистични ограничения). Поради това всички разглеждани речници изобилстват с допълнителни граматични бележки, а по-ранните от тях включват дори граматични очерци за съответните езици или описания на графичните им системи. Тези факти още веднъж показват, че декларираният в част от речниците *принцип на превода* означава превод от една лексикална система в друга, а не превод на функциониращa в текста лексикална единица. Представените различни видове информация рядко са равностойно разработени[32].

Разглежданите речници са преди всичко речници на пълнозначните думи, тъй като общ техен недостатък[33] е елементарното представяне на неизменяемите части на речта, дължащо се на неразработеността на славянския синтаксис и славянската прагматика през този период.

Отразените в лексиката културни различия между славянските народи и опасността от погрешни интерпретации поради езиково родство налагат в речниците да бъде включена допълнителна информация под формата на обяснения. Вид допълнителна информация представляват и списъците на единици, стоящи в периферията на националните лексикални системи (топоними, антропоними, съкращения, термини).

Поради изброените факти разглежданите речници може да бъдат определени, от една страна, като синкретични трудове, намиращи се на границата между речника и граматиката, а от друга страна – като хибридни съчинения, включващи елементи на ДР, тълковния речник и енциклопедията;

г) общият характер, средният обем и търсената двупосочна ориентация[34] се отразяват в *различния баланс* между ДР като практическо помагало и ДР като научно съчинение. В разглежданите издания се наблюдава изразена тенденция към засилване на научния характер за сметка на практическото предназначение на речника. Стройната и премислена организация на лексикалния материал, към която се стремят създателите на речниците в съответствие с постиженията на лексикографската теория, обаче не винаги съвпада с интуитивното разбиране, възможностите и изискванията на потенциалните ползвате-

ли. Опитвайки се да обяснят усложнената структура и обхвата на произведенията си, авторите добавят все по-обширни предговори, показващи къде се разминават лексикалните системи на двата съпоставяни езика[35], какво съдържа или не съдържа речникът, как трябва да се ползва и т.н., срв. напр. предговорите в [3], [4], [18], [27]. Читатели на тези предговори обаче са главно създателите на други подобни речници, тъй като колкото по-съвършен и икономичен е речникът от научна гледна точка, толкова по-тесен е кръгът на лицата, които могат да използват пълноценно цялата събрана в него информация.

Посочените особености на славянско-българските и българско-славянските речници от последния половин век се намират в пряка зависимост както от равнището на българската и съответната национална лексикография, така и от фиксираната („хартиената") форма на ДР в предкомпютърния период. Свързани с кулминацията в развитието на славистиката в България през втората половина на XX в. (срв. Славистиката 2003, Славистика и общество 2006, В търсене на смисъла и варианта 2008, Куцаров 2008а), тези речници са синкретични и хибридни лексикографски съчинения, ориентирани към езиковата практика, схващана преди всичко като разбиране на „чуждия" език и превеждане на „родния".

За да преценим значимостта на разгледаните традиционни ДР като ресурси за електронната лексикография, ще сравним обобщено състоянието на двуезичната ни славянска лексикография с основните изисквания към двуезичните е-речници:

Състояние	**Изисквания**
- лексикографска асиметрия, изразяваща се както в „празни клетки" за отделни езици, така и в различно качество и различна хронологична отнесеност на материала *от* и *към* българския език;	- за симетрия, паралелизъм и равнопоставеност на лексикографския материал;
- фиксиран от „хартиения" формат тип на речника – общ, терминологичен, учебен, малък,	- да не налага върху ползвателя мярката на речника, а да позволява всеки да работи с е-

Състояние	Изисквания
голям и т.н.	речник по своя мярка;
- насочени главно към писмената форма на езика, към четене, разбиране и превод;	- да обхваща възможно най-пълно различните форми на съществуване на езика;
- снабдени със сложен обяснителен апарат речници за образовани ползватели;	- опростени откъм обяснения речници, съобразени с ползвателя;
- с опростена макроструктура вследствие на линейното разположение на материала, стремежа да се улесни ползвателят и недостатъчно проучената еквивалентност между лексикалните системи на двата езика;	- да създаде сложна „мрежова” макроструктура, даваща достъп до максимално количество включена информация;
- речници на лексикалните системи с думата в основна форма като речникова единица, основаващи се на понятието преводна еквивалентност;	- да представя функционирането на думата в контекста въз основа на превода;
- събрана на едно място (около основната речникова единица) информация с различна йерархия и различен вид	- да представя информацията модулно и да позволява различни видове използване

Съпоставката показва съществени и както изглежда на пръв поглед – непреодолими различия между традиционните и (засега бъдещите) славянски електронни ДР. Изброените разлики обаче засягат главно технологията на речниците; вследствие на благоприятните за славистиката условия през изминалите петдесет години лексикалната система на българския език е повече или по-малко разработена в семантичен аспект с оглед на останалите славянски езици. Ако не желаем да се окажем в ситуацията на хора, затворени в родната си езикова компетентност и употребяващи родния си език само за домашно общуване, трябва да намерим адекватен начин за „износ” на българския сред

славянските езици и чрез средствата и методите на модерната двуезична лексикография. През последните десетилетия унификацията в лексикографската теория и практика, стабилизираната[36] норма на „малките" славянски езици, сходните тенденции в развитието на националните им лексикални системи и все по-активното развитие на корпусната лингвистика създават благоприятни предпоставки за създаване на ДР, съобразени както с тълковните речници на славянските книжовни езици, така с единния европейски стандарт за ежедневна и професионална комуникация.

*

В заключение ще обобщим значението на разгледаните славянски ДР за развитието на българската двуезична лексикография и лексикология.

Безспорно най-съществен е приносът на изследваните речници за лексикографската практика, за изграждането и поддържането на лексикографските контакти между българския и останалите славянски езици, за натрупването на солиден материал, който – макар и зависим от странични фактори като степен на развитие на националната лексикография, близост между езиците, обществен престиж на езика, фиксирана „книжна" форма – дава надежди за създаване на двуезични славянски речници от ново поколение.

Реализираното в славянските ДР съпоставяне между лексикалните системи на българския и останалите славянски езици допринася за разкриване на спецификата на българската лексикална система.

Славянските ДР имат принос и за развитието на българската лексикографска теория – както чрез идеите, породени при успоредяването ѝ с лексикографските традиции на другите славянски езици, така и чрез възникналите на родна почва лексикографски иновации вследствие на натрупаната солидна и стабилна материална основа.

ПРИЛОЖЕНИЕ

Източнославянски езици

Руско-български и българско-руски речници[37]:

[1] Болгарско-русский словарь. Составил С. Б. Бернштейн. Издательство „Советская энциклопедия”. Москва 1966. *(Около 58 000 слов) (Тираж 35 000*).

[2] Русско-болгарский словарь. Т. I. А–О. Т. II. П–Я. Второе переработанное издание. Под редакцией С. Влахова и Г. А. Тагамлицкой. Наука и искусство. София 1986. – Руско-български речник. Т. I. А–О. Т. II. П–Я. Второ преработено издание. Под редакцията на С. Влахов и Г. А. Тагамлицка. Наука и изкуство. София 1986. *(Около 75 000 думи) (Тираж 13 106).*

[3] Доц. д-р А. К. Кошелев. Болгарско-русский словарь. – Доц. д-р А. К. Кошелев. Българско-руски речник. Наука и изкуство. София 2003. *(Около 45 000 думи) (Тираж непосочен).*

[4] Сергей Влахов. Нов руско-български речник. Под редакцията на Анна Липовска. – Сергей Влахов. Новый русско-болгарский словарь. Под редакцией Анны Липовской. Парадигма. София 2004. (II издание 2007) (III издание 2009). *(Около 50 000 думи) (Тираж непосочен).*

Украинско-български и българско-украински:

[5] И. А. Стоянов. Е. Р. Чмир. Българско-украински речник – Болгарсько-український словник. Академия на науките на Украинската ССР. Институт за езикознание „А. А. Потебня”. Наукова думка. Киев 1988. *(Около 43 000 думи) (Тираж 3 000)*.

[6] Българско-украински речник. Около 20 000 думи. Съставителка Катерина Потапенко-Калоянова. Киев 2004. – Болгарсько-український словник. Близько 20000 слів. Укладач Катерина Потапенко-Калоянова. Київ 2004*. (Тираж непосочен).*

Южнославянски езици

Сърбохърватско-български и българско-сърбохърватски:

[7a] Д-р Марин Младенов. Българско-сърбохърватски речник. 1967 г.

[7] Ангел Игов. Српскохрватско-бугарски речник. Друго допуњено издање. София 1983. Издавачко предузеће „Наука и изкуство”. – Сърбохърватско-български речник. Второ допълнено издание. Издателство „Наука и изкуство”. София 1983 (III издание 1987). *(Около 45 000 заглавни думи) (Тираж 4 603).*

[8] Ценка Иванова, Маријана Алексић. Српско-бугарски тематски

речник и Лексикон међујезичких хомонима и полисема. Издателство ПИК. Велико Търново 1999. *(Непосочени брой думи и тираж).*

[9] Д-р Марин Младенов. Бугарско-српски речник. Българско-сръбски речник. Завод за уџбенике и наставна средства, Београд, 2000. *(Около 50 000 думи) (Тираж 1000)*.

[10] А. Зарков, Пл. Трифонов. Сръбско-български речник. ИК ПАЛ. София 2004. *(Около 50 000 думи). (Тираж 1000)*.

Западнославянски езици

Полско-български и българско-полски:

[11] Иван Леков. Полско-български речник. Университетска печатница. София 1944. *(Тираж 1 050).*

[12] Słownik polsko-bułgarski. Pod redakcią czł.-koresp. BAN prof. Iw. Lekowa i prof. Dra Fr. Sławskiego. Bułgarska akademia nauk. Institut języka bułgarskiego. Sofia, 1961. Wydawnictwo Bułgarskiej Akademii nauk – Полско-български речник. Под редакцията на чл.-кор. проф. Ив. Леков и проф. д-р Фр. Славски. БАН – ИБЕ. Издателство на БАН. София 1961. *(Непосочен брой думи) (Тираж 3 000)*.

[13] Franciszek Sławski. Podręczny słownik bułgarsko-polski. Warszawa 1963.

[14] Францишек Славски, Сабина Радева. Джобен българско-полски и полско-български речник. Ведза повшехна. Варшава 1965. – Franciszek Sławski, Sabina Radewa. Kieszonkowy słownik bułgarsko-polski i polsko-bułgarski. Wiedza powszechna. Warsawa 1965. *(Около 15 500 думи) (Тираж 20 233)*.

[15] Franciszek Sławski. Podręczny słownik bułgarsko-polski z suplementem. T. I, II. Państwowe wydawnictwo Wiedza powszechna. Warszawa 1987. – Фр. Славски. Наръчен българско-полски речник с допълнение. Т. I, II. ДИ „Наука и изкуство”. София 1987. *(Около 50 000 думи) (Тираж 9 700 + 300).*

[16] Sabina Radewa. Podręczny słownik polsko-bułgarski z suplementem. T. I-II. Państwowe wydawnictwo Wiedza Powszechna. Warszawa, 1988. – Сабина Радева. Наръчен полско-български речник с допълнение. Том I-II. ДИ „Наука и изкуство”. София 1988. *(Около 45 000 думи) (Тираж 9 700 + 300).*

Словашко-български и българско-словашки:

[17] Slovensko-bulharský slovník. Pod redakciou člena korešpondenta prof. Ivana Lekova a doc. dr. Vincenta Blanára. Vydavateľstvo Bulharskej akademie vied. Sofia 1970. – Словашко-български речник. Под редакцията на чл.-кор. проф. Иван Леков и доц. д-р Винцент Бланар.

Съставили Стефан Знеполски, доц. Светомир Иванчев, д-р Карел Мара, проф. д-р Цветана Романска. Издателство на Българската академия на науките. София, I издание 1970, II издание 1979. *(Непосочен брой думи) (Тираж на II-то издание 3 000).*

[18] Bulharsko-slovenský slovník. I. (A–K). Maria Košková a kolektiv. Slavistický kabinet SAV. Bratislava 2004. *(Непосочен брой думи)* (*Тираж непосочен*).

Чешко-български и българско-чешки

[19] Vl. Šak. Bulharsko-český slovník. J. Otto. Praha 1920. *(Непосочени брой думи и тираж).*

[20] Vl. Šak. Česko-bulharský slovník. J. Otto. Praha 1920. *(Непосочени брой думи и тираж).*

[21] Vl. Sís, B. Ivanov. Nový kapesní slovník jazyka českého a bulharského. Díl I: česko-bulharský, Díl II: bulharsko-český. Nakladatel J. Lorenz. Třebíč 1922. *(Непосочени брой думи и тираж).*

[22] K. Hora. Bulharsko-český slovník. Nakladatelství ČSAV. Praha 1959. *(Около 60 000 думи) (Тираж 5 000)*.

[23] Bulharsko-český a česko-bulharský technický slovník (SNTL 1958).

[24] Българско-чешки и чешко-български технически речник. Bulharsko-český a česko-bulharský technický slovník. SNTL – Nakladatelství technické literatury Statní nakladatelství Technika. Praha – Sofie 1988. *(Около 24 000 заглавни думи) (Тираж 2 600)*.

[25] Bohdan Prošek. Bulharsko-český a česko-bulharský kapesní slovník. Státní pedagogické nakladatelství. 1. vydání Praha 1964, 2. vydání Praha 1969. (*Съответно около 13 000 и около 14 000 думи*). (*Тираж 15 000*).

[26] Д-р Цветана Романска. Чешко-български речник. I издание 1947, II издание 1949, III преработено издание. Държавно издателство „Наука и изкуство”. София 1961. (*Около 35 000 заглавни думи*) (*Тираж 8 079*).

[27] Чешко-български речник в два тома под редакцията на Светомир Иванчев. Колектив. Т. I. A–O, т. II. P–Ž. Труд & Прозорец, София 2002. Допечатка Труд & Прозорец, София 2005. Допечатка Труд & Прозорец, София 2007. (*Т. I около 26 000 заглавни думи, том II около 35 400 заглавни думи*) (*Тираж 1 000 бройки без допечатките)*.

[28] Я. Бъчваров и кол. Кратък чешко-български речник на некнижовната лексика. София, Издателство ЕТО, 2011, XXXII + 244 стр. (*4 100 заглавни думи*) (*Тираж непосочен*).

Многоезични

[29] Language in Tourism. Многоезичен учебен речник за целите на туризма. Večjezični slovar za potrebe turistične dejavnosti. Viacjazyčný

študijný slovník pre potreby cestovného ruchu. Vícejazyčný studijní slovník pro potřeby cestovního ruchu. Language in Tourism. Multilingual Learner´s Dictionary (Bulgarian-Slovenian-Slovak-Czech). Rangelova A. and Ticha Zd. (eds). „Valentin Trajanov" Publishing House 2008. (*3 230 думи и изрази*) (*Тираж 600*)

БЕЛЕЖКИ

[1] Едва през последните две години се появиха две обстойни изследвания върху българската двуезична лексикография, срв. Липовска 2009, Павлова 2010.

[2] Почти половин век дели първия опит за обобщение на М. Филипова-Байрова от прегледа на историята и съвременното състояние на българската лексикография, в който двуезичните речници заемат скромно място (Чоролеева 2008: 44–49). Освен С. Павлова с проблеми на двуезичната лексикография през този период се занимава и Ст. Калдиева-Захариева (Калдиева-Захариева 2003).

[3] Студията на М. Филипова-Байрова носи показателното заглавие „Изработване на двуезични речници" (Филипова-Байрова, 1962: 223–262).

[4] Както отбелязва М. Филипова-Байрова, „от гледище на пълното отражение на лексиката, характерна за определен исторически момент, следва, че двуезичните речници от 25–50 хиляди думи не могат да служат дълги години. Това е работа на един голям, солиден речник. В случай на преиздаване на такъв (среден), той трябва непременно да се преработи, като се попълва и съкращава непрекъснато." (Филипова-Байрова 1962: 236).

[5] По-нататък се използва съкращението ДР – двуезичен речник.

[6] Количеството и разнообразието на българско-руските лексикографски трудове определят много по-детайлизирания набор от параметри, с които работи А. Липовска (Липовска 2009: 45).

[7] Въпреки че в славянската лексикография този въпрос рядко се коментира, за реалното разграничаване на видовете двуезични речници решаваща роля имат пазарът и числеността на дадена езикова общност. При малките общности обхватът на ползвателите и възвръщаемостта на инвестициите са много ограничени, от което следва, че ако не съществува специална причина, преводните речници на два „малки" езика нямат перспектива. Затова е необходимо да се отговори на въпроса кои видове речници може да се интегрират в един от гледна точка на разходите (с оглед на ползвателите им) и кой е оптималният подход за постигане на това (Чермак 1995: 232).

[8] Тези изисквания отразяват идеалния ДР. В действителност това,

което се представя за ДР, представлява „полуфабрикат”, изработен от националните тълковни речници (Чермак 1995: 232).

[9] Най-общо може да се каже, че електронните речници представляват „сбор от електронно структурирани данни, които може да бъдат достъпни с много средства, да бъдат обогатени с много видове функции и използвани при различни обстоятелства” (Бергенхолц 2009: 102).

[10] Поради това, че през изминалия половин век машинният превод се разработва с оглед на превода от английски на руски и обратно, срв. Щаудингер, Смит 1998.

[11] Статията е съкратена версия на по-широко изследване и не включва анализирания материал.

[12] Отговорността на славистичната наука за филологиите на отделните славянски езици се подчертава и днес, срв. Радева 2006: 16.

[13] Присъстват в Приложението към статията.

[14] За първите славянско-български речници вж. Кювлиева-Мишайкова 1997: 177–183. Хронология на изданията за отделни езици е представена в Рангелова, Младенова 2000: 127–130, Липовска 2009: 15–45, Павлова 2010, а също в предговорите на част от речниците, срв. [7]: 5; [3]: 11–12; [18]: 3; [4]: 7. Цифрите в квадратни скоби показват пореден номер в Приложението.

[15] Във втория том като година на издаване погрешно е посочена 2002.

[16] Първоначално изключенията са само няколко: В. Бланар и Ив. Леков като редактори на словашко-българския речник и Ив. Стоянов и Е. Чмир като създатели на първия българско-украински речник. Това се отнася не само за славянските речници, срв. напр. Тодор Нейков при испанско-български речник, проф. Марко Минков при английско-български речник, Сергей Влахов при руско-български речник, проф. Светомир Иванчев при чешко-български речник, Ангел Игов при сърбохърватско-български речник и др.

[17] Срв. библиографията към отделните раздели в Славянски езици 1994.

[18] Една от причините за това е дългогодишното отсъствие на основен типов български словник за българско-чуждите речници, срв. Павлова 1986, Павлова 2002.

[19] Изключваме руско-българските и българско-руските специализирани речници − фразеологични и специални терминологогични (железопътен, електротехнически, машиностроителен, минно-геоложки, технически и др.), срв. [4]: 7, Липовска 2009: 15–44.

[20] Джобният чешко-български речник има тираж 15 000 бройки, а същият тип полско-български речник излиза в тираж 20 000 бройки.

[21] В съответствие с издателското разделение в Чермак 1995: 233.

[22] Става дума за обща особеност на славянската двуезична лексикография, срв. Секанинова 1993: 45-58; Филипец 1973: 279-295; Филипец, Чермак 1986: 150-165.

[23] Според С. Влахов „от речничето на Сава Филаретов, приложено към разговорника „Карманная книга для русских воинов, находящихся в походах по болгарским землям" (1854) до второто преработено издание на двутомния „Руско-български речник" (1986) са издадени около 150 речника, различни по обем, съдържание, предназначение, качество" ([4]: 7). В течение на почти половин век руският има за България водеща „образователна" роля, която трябва да се прецени внимателно, обективно и многостранно – от „прозорец към света" до „спирачка за вътрешноезиковите развойни процеси" на българския език.

[24] Образованието на едни от най-изявените български езиковеди и литературоведи е свързано с Полша и Чехия, а след оформянето на Източния блок много български студенти от различни специалности се подготвяха в тези страни.

[25] Т.е. страната само на единия от езиците.

[26] Всеки ДР може да бъде използван двустранно, но в случая разглеждаме намеренията и целите на създателите.

[27] Недостатъчното използване на количествения критерий се дължи и на неравномерното развитие на квантитативната лингвистика в отделните страни през последния половин век и на различното равнище на корпусните изследвания през последните петнайсет години. Показателно например е състоянието в български и чешки – в български съществуват честотни речници само на определени стилове, докато в чешки – на целия език.

[28] Ясно е, че типът речник (общ или със специално предназначение) определя словника, но в случая се разглеждат средни и големи общи речници.

[29] Така наречените отделни „значения" (лекси) са дадени с номерация или без номерация.

[30] Названието двуезични речници се разглежда обикновено като синоним на преводни речници, срв. напр. Секанинова 1993: 45–49; Чермак 1995: 231 и др.

[31] Термин на С. Влахов, вж. [4]: 9.

[32] Вж. по-горе коментарите за стилистичните бележки в българските тълковни речници.

[33] До известна степен този недостатък е предолян в [4] и [18].

[34] Да служат едновременно за декодиране и кодиране.

[35] М. Кошкова определя два главни и няколко второстепенни типа

асиметрия между езиковедската теория и приложението ѝ в практическата лексикография (Кошкова 2006: 907–212).

[36] С изключение на македонски и хърватски по геополитически причини.

[37] Отбелязани са само разгледаните тук речници. Преглед на руско-българската лексикография в течение на 150 години е даден в [4], както и в Липовска 2009.

ЛИТЕРАТУРА

Аткинс, Рындъл 2008: Atkins, B. T. S., M. Rundell. *The Oxford Guide to Practical Lexicography*. Oxford, Oxford University Press.

Бергенхолц и др. 2009: Bergenholtz, H., S. Nielsen, S. Tarp (eds). *Lexicography at a Crossroads: Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*. Bern.

В търсене на смисъла и инварианта 2008: *В търсене на смисъла и инварианта. Сборник в чест на 80-годишнината на проф. Д. С. Станишева.* Съставители А. Бъчваров, М. Виларова, Й. Трифонова. София, УИ „Св. Климент Охридски".

За думите и речниците 2000: *За думите и речниците. Лексикографски и лексиколожки четения '98*. София, „Диос".

Калдиева-Захариева 2003: Калдиева-Захариева, Ст. За някои проблеми на двуезичната фразеография. – В: *Проблеми на българската лексикология, фразеология и лексикография. Българско езикознание*. Том 3. Съставител В. Кювлиева-Мишайкова. София, АИ „Проф. Марин Дринов", с. 438–456.

Кошкова 2006: Košková, M. Asymetria jazykovednej teorie a lexikografického spracovania v dvojjazyčnom slovníku (na príklade nového bulharsko-slovenského slovníka). – In: *Studia Balcanica Bohemo-Slovaca*. Svazek 2. Brno, Matice moravska, s. 905–912.

Куцаров 2008а: Куцаров, Ив. Европейската славистика между двете световни войни. – В: *В търсене на смисъла и инварианта. Сборник в чест на 80-годишнината на проф. Д. С. Станишева.* Съставители Я. Бъчваров, М. Виларова, Й. Трифонова. София. УИ „Св. Климент Охридски", с. 225–232.

Куцаров 2008б: Куцаров, Ив. Славистиката – забележително минало, скромно настояще, съмнително бъдеще. – В: *Паисиеви четения. Международна славистична конференция „Интеркултурният диалог – традиции и перспективи"*. Научни трудове. Том 461, кн. 1, сб. А, 2008. Филология, с. 11–20.

Кювлиева-Мишайкова 1997: Кювлиева-Мишайкова, В. *Българското речниково дело през Възраждането.* София. АИ „Проф.

Марин Дринов”.

Липовска 2009: Липовска, А. *Русско-болгарская лексикография: традиции и тенденции развития.* София, Издательство Софийского университета.

Младенова 2011: Младенова, М. Рец. за А. Липовска. *Русско-болгарская лексикография: традиции и тенденции развития*. София, Издательство Софийского университета, 2009. (под печат)

Павлова 1986: Павлова, С. Необходимостта от изработване на основен български словник за българо-чуждите речници. – В: *Въпроси на съвременната българска лексикология и лексикография.* София, БАН, с. 169–175.

Павлова 2002: Павлова, С. Модел за изграждане на общ типов словник на родно-чужд речник. – В: *Проблеми на българската лексикология, фразеология и лексикография. Българско езикознание*. Том 3. Съставител В. Кювлиева-Мишайкова. София. АИ „Проф. М. Дринов”, с. 399–438.

Павлова 2010: Павлова, С. *Българската двуезична лексикография от Освобождението до края на XX в.* София, „ЕМАС”.

Радева 2006: Радева, В. Славянските езици в обединена Европа. – В: *Славистика и общество*. Съставители М. Младенова, И. Монова. София. „Херон Прес”, с. 12–16.

Рангелова, Младенова 2000: Rangelová, A., M. Mladenova. Česko-bulharské a bulharsko-české slovníky – stav a perspektivy. – *Acta Universitatis Nikolai Copernici. Studia Slavica 5*. Zeszyt 343, 2000, s. 127–133.

Секанинова 1993: Sekaninova, E. *Dvojjazyčna lexikografia v teorii a praxi*. Bratislava, Veda.

Славистика и общество 2006: *Славистика и общество*. Съставители М. Младенова, И. Монова. София. „Херон Прес”.

Славянски езици 1994: *Славянските езици. Граматични очерци*. София, Издателство на БАН.

Славистиката 2003: *Славистиката в началото на XXI век. Традиции и очаквания*. Съставители В. Панайотов, Я. Бъчваров. София, „СемаРШ”.

Филипова-Байрова 1962: Филипова-Байрова, М. Изработване на двуезични речници. – В: *Известия на Института за български език*, кн. VI, с. 223–262.

Филипец 1973: Filipec, J. Ke konfrontaci dílčích sémantických systému v slovné zásobě dvou různých jazyků. – In: *Československé přednášky pro VII. Mezinárodní sjezd slavistů (Varšava 1973)*. Praha, Nakladatelství ČSAV, s. 279–295.

Филипец, Чермак 1985: Filipec J., F. Čermák. *Česká lexikologie*. Praha, Academia.

Хартман, Джеймс 1998: *Dictionary of Lexicography*. Comp. by R. R. K. Hartmann & Gregory James. London, Routledge.

Чермак 1995: Čermák, F. Překladová lexikografie. – In: *Manuál Lexikografie*. Praha, H&H, s. 230–248.

Чоролеева 2008: Чоролеева, М. История и современное состояние болгарской лексикографии. – В: *Теория и история славянской лексикографии. Научные материалы к XIV съезду славистов*. Отв. редактор М. И. Чернышева. Москва, с. 44–69.

Чоролеева 2009: Чоролеева, М. Периодизация на развитието на българската лексикография. – В: *Езиковедски изследвания в чест на чл.-кор. проф. д-р Тодор Бояджиев, проф. д-р Венче Попова и проф. Петър Пашов*. София, УИ „Св. Климент Охридски”, с. 351–358.

Щаудингер, Смит 1998: Staudinger, B., N. Smith. Machine translation and Slavic languages: a brief history and state of art. – In: *Jazyk a kultura vyjadřování. Milanu Jelínkovi k pětasedmdesátinám.* Eds. M. Krčmová, P. Karlík. Brno, s. 167–180.

Божана Нишева
(София)

НЕОГРАФСКОТО ПРЕДСТАВЯНЕ НА ПОСТПОЗИТИВНИТЕ СЛОВООБРАЗУВАТЕЛНИ КОМПОНЕНТИ В БЪЛГАРСКИЯ И ЧЕШКИЯ ЕЗИК (С ОГЛЕД НА СЪВРЕМЕННАТА ЛЕКСИКАЛНА ДИНАМИКА)

This article presents the post-position word-formation components included in the dictionaries of the new words of the Bulgarian and Czech languages under three distinct groups: 1) the so called suffixoids (units of functional and semantic character falling between the root and the morpheme); 2) meaningful post-position units common to whole series of new compound words; 3) internationalisms active in the formation of new words both in pre- and in post-position. The Bulgarian and Czech neologism groups using these are thus the object of a functional and semantic, as well as a quantitative analysis, thereby proving the comparison of materials from a relatively new type of dictionary – that of neologisms – is not only conductive of valuable knowledge regarding the dynamics in the individual Slavic lexical systems, but is also directly related to the employment of contemporary methods and approaches in the bilingual lexicographic practice.

Ключови думи: неография, словообразователни компоненти, лексикална динамика

Key words: neography, word-formation components, lexical dynamics

1. Тенденцията към разширяване на сферата на употреба и функциониране на словообразователни модели с участието на афиксоиди, респ. на неологизми от афиксоиден тип, се определя като една от най-съществените прояви на лексикална и словообразователна динамика в съвременните славянски езици (Аврамова 2003, Земска и др. 1999, Смирнов и др. 1998 и пр.). За източник на иновации в славянските езикови системи се смята също така повишената активност на композитообразуването, свързана с навлизане на нови интернационални типове за композиционно оформяне и с увеличаване на продуктивността на някои домашни словообразователни средства и модели (Боздехова 1994, Онхайзер 2003, Колковска 2003, Мейстржик 2000, Мурдаров 1983 и др.). Посочените динамични явления водят до

нарастване на ролята на аналогията в словообразователния процес, до серийното образуване на неологизми (Мартинцова, Вачкова 1992, Вачкова, Вачков 1998) и в частност са свързани с появата на цели редове нови единици с общи постпозитивни (основни) компоненти[1]. В езиковедската литература по въпроса – с оглед на функционалните и семантичните им черти – тези постпозитивни словообразователни компоненти най-често се разделят на два основни типа: *суфиксоиди*[2] и *суфиксоидни лексеми* (вж. по-подробно Яновец, Рангелова 2005 и др.).

2. Динамиката, съпровождаща обогатяването на репертоара от постпозитивни словообразователни средства с различни генетични, функционални и семантични характеристики, както и на групите имена, образувани с тяхно участие, е отразена и в речниците на новите думи във всички славянски езици[3]. С оглед на този факт основна цел на изследването ще бъде съпоставката на неографското представяне на следните типове български и чешки постпозитивни словообразователни средства и на новите думи, образувани с тях:

1) суфиксоиди, т. е. морфеми с формални и функционални особености, отреждащи им място на границата между корените и суфиксите, които не се употребяват самостоятелно и чиято семантика „кристализира" само в рамките на новосъздадената дума, напр.: б. **-воз**, **-ман**, **-троника** и ч. -**fil**, **-log**, **-kracie**;

2) пълнозначни и самостоятелно употребявани лексеми, налагащи се като обща втора съставна част на цели редове новообразувани композитуми, напр. **-център/-centrum**, **-терапия/-terapie**, **-туризъм/-turismus/-turizmus**, **-turistika**;

3) интернационализми като **бизнес**, **хит**, **шоу**, отличаващи се от останалите постпозитивни лексеми, тъй като проявяват активност при образуването на нови думи по два различни словообразователни модела: „афиксоиден/лексикален детерминиращ член + **-бизнес**, -**хит**, -**шоу** с функция на постпозитивна опорна дума" и „**бизнес-**, **хит-**, **шоу-** с функция на препозитивен модифициращ член + друга основа (най-често съществително име)".

3. По-нататък групите неологизми с общ постпозитивен компонент се подлагат на функционално-структурен анализ, комбиниран с квантитативен метод на съпоставка. С други думи, функционалната и структурната характеристика, както и класификацията на неологизмите, се основават на количествените

данни[4], съдържащи се в следните два български и два чешки речника на новите думи: *Речник на новите думи и значения в българския език* (по-нататък РНДЗБЕ), *Речник на новите думи в българския език* (по-нататък РНДБЕ); *Nová slova v češtině. Slovník neologizmů* (по-нататък SN 1), *Nová slova v češtině. Slovník neologizmů 2* (по-нататък SN 2)[5].

3.1. Неографските подходи при представянето на репертоара от нови постпозитивни компоненти се интерпретират с оглед на съвременната българска и чешка лексикална динамика, като за целта речниковият материал от нови или употребявани в нова функция, ново значение или с повишена словообразувателна активност постпозитивни компоненти се допълва с ексцерпции от езика на печатните медии[6]. Избраният метод на изследване е свързан с обстоятелството, че изчерпателността на неографското представяне често пъти е повлияна от редица ограничаващи фактори, например личното мнение на лексикографа, концепцията на речника и под. (Яновец 2003).

4. В българските и чешките неологични речници намираме следните суфиксоиди, представени или в отделна речникова статия с приложени към нея примери с общ постпозитивен член, или в рамките на различни речникови статии за неологизми с еднаква високопродуктивна втора съставка[7] (вж. Таблици 1 и 2).

Таблица 1

Честота на новите суфиксоиди и на неологизмите, образувани с тях, според РНДЗБЕ и РНДБЕ		
-воз 1.	12	*тежкотоварен автомобил за превозване на някакъв товар*
-воз 2.	2	*танкер за превозване на някакъв товар*
-лог	12	
-логия	12	
-ман	12	
-гейт	9	
-тека	6	
-холик	6	
-фоб	5	
-мен	5	
-холизъм	3	
-троника	3	

Честота на новите суфиксоиди и на неологизмите, образувани с тях, според РНДЗБЕ и РНДБЕ (продължение)		
-иада	2	
-фил	2	
-крат	1	
-крация	1	

Таблица 2

Честота на новите суфиксоиди и на неологизмите, образувани с тях, според SN 1 и SN 2		
-log	41	
-logie	37	
-fil	23	
-man 2.	13	*kdo je (nadměrně až chorobně) něčím zaujat*
-man 1.	9	*kdo se něčím zabývá, něco koná, představitel něčeho*
-kracie	21	
-téka	13	
-mil	11	
-filie	8	
-holik	8	
-krat	7	
-milec	6	
-holismus	4	
-gate	4	
-fob/-fobik	3	

4.1. Резултатите от таблиците най-общо показват, че в български числен превес имат постпозитивните афиксоиди за образуване на названия на абстрактни признаци и понятия, като с по-висока словообразувателна активност (в сравнение с чешкия) се отличават суфиксоидите, „отсечени" от англицизми. Съществен е и фактът, че на фона на силното влияние на англо-американизацията, в РНДЗБЕ и РНДБЕ на първо място по брой неологизми, образувани с него, се оказва постпозитивният компонент -**воз** 1., 2., с произход от славянската терминологична лексика[8].

4.2. За разлика от българските речници, SN 1 и SN 2 свидетелстват, че в чешки са значително по-активни суфиксои-

дите с източник интернационалната (терминологична) лексика от т.нар. класически езици, като с тях възникват предимно нови названия за лица. На първо място по продуктивност при този тип постпозитивни афиксоиди е елементът **-log**. Тъй като традиционно се приема, че словообразователните стъпки са в посока от номинацията на понятието към номинацията на лицето, свързано с него, тази ситуация в чешкия на първо място показва, че динамиката в езиковите системи предизвиква асиметрия в отделните подсистеми, а на второ илюстрира „преходния" характер на афиксоидния репертоар и на имената, оформени с него.

4.3. От гледна точка на генетичните и комуникативно-стиловите специфики на отделните постпозитивни компоненти, суфиксоидните неологизми от РНДЗБЕ, РНДБЕ и SN 1, SN 2 може да бъдат разпределени в следните няколко основни групи.

4.3.1. В българския и чешкия език съществуват немалко на брой терминологични и абстрактни названия, образувани по интернационални словообразователни модели. С други думи, по-голямата част от суфиксоидите от таблици 1 и 2, които с различна активност започват да се налагат като продуктивни членове на словообразователния инвентар, респ. да се употребяват в по-широки сфери на съвременната комуникация, са генетично свързани с типове названия, традиционно причислявани към т.нар. културна лексика в двата езика. Това са компонентите: **-иада/-iáda**, **-крат/-krat**, **-крация/-kracie**, **-лог/-log, -логия/-logie**, **-ман/-man 1.**[9], **-тека/-téka**, **-фил/-fil**, **-филия/-filie**, **-фоб/-fob**. Всеки от тях първоначално прониква в езика приемник като част от словообразователната структура на название, отнасящо се към определена научна, професионална или интелектуална област, срв. напр. терминологичните названия, образуваните в един и същ език източник и по общи словообразователни модели: б. *етнология – етнолог*, *орнитология – орнитолог*; ч. *antropologie – antropolog*, *kardiologie – kardiolog* и мн. др. Под влияние на все по-засилващите се след 1989 г. тенденции към демократизация и интернационализация (или повторна интернационализация[10]), единиците от този тип започват да функционират паралелно с

по-нови образувания, оформени със същия задпоставен елемент, б. *миксология – миксолог*, *нумерология – нумеролог (номерология – номеролог)*; ч. *počítačologie – počítačolog*, *simpsonologie – simsponolog*. Процесът става импулс за функционалното и семантичното трансформиране на разглежданите постпозитивни компоненти, като води до разширяване на съчетаемостта им с домашни или одомашнени мотивиращи основи, б. *пазаролог*, *циганолог*; ч. *havlolog*, *záhadolog*. Съществен принос за задълбочаването на явлението има по-конкретно детерминологизацията, т.е. навлизането на терминологична лексика в нетерминологични сфери на употреба. Под влияние на всички изброени дотук фактори постпозитивните морфеми с чужд произход все повече се приближават до статута на останалите продуктивни суфикси, като особено активно участват в процеса на парадигматична деривация[11]. Иначе казано, *въз основа на формалните, семантичните и функционалните специфики на посочените в таблици 1 и 2 втори съставки на сложни думи, по-нататък формациите, образувани с тях, ще бъдат разглеждани в контекста на „преходния" им характер, т.е. като лексикални единици, постепенно доближаващи се до българските и чешките имена, образувани чрез деривация.*

4.3.1.1. Друга характерна иновация, отразена в РНДЗБЕ, РНДБЕ и SN 1, SN 2 е постепенното разширяване на терминологичното значение на суфиксоидите. Така напр. деконкретизацията на **-лог/-log** протича паралелно с детерминологизацията на елементите -**логия/**-**logie** и всеки от тези компоненти се използва активно при адаптацията на множество англицизми, напр.: а. *scientology – scientologist* → б. *сциентология – сциентолог*; ч. *scientologie – scientolog* и др.

4.3.1.1.1. Новите български имена с краен член **-логия** са значително по-малко от формациите с **-logie** в чешки. В българския материал също така отсъстват национални единици със структура: „проприална основа + **-лог/-log** и **-логия/-logie**", широко застъпени в чешкия: *beckettolog*, *shakespearolog*, *tolkienolog*. В двата езика навлизат и названия на нови области от човешкото познание и на специалисти, работещи в тях, напр. б. *медиология – медиолог*; ч. *aromatologie – aromatolog*, *kos-*

metologie – *kosmetolog* и др.

4.3.1.1.2. Неографските източници открояват и още една разлика между българския и чешкия, в който се обособява немалка група експресивни/пейоративни единици, напр. *komunistolog* (име, оформено с **-logie,** не е регистрирано), *pojmologie* – *pojmolog*, *pornolog* (име, оформено с **-logie,** не е регистрирано), *trendologie* – *trendolog*, *upírologie* – *upírolog* и др. От примерите личи, че мотивационните връзки невинаги са линейни, т.е. че част от неологизмите се образуват по аналогия. За контекстуалната обусловеност при създаването на български суфиксоидни неологизми са показателни и изречения от медийния език като: *„Покрай внимателното фокусиране на избирателните списъци се роди и нова наука. Фантомология.”*

4.3.1.2. С по-слаба продуктивност се отличават суфиксоидите -**крат/-krat** и **-крация/-kracie**. Неологичните речници и на двата езика регистрират предимно групи от нови абстрактни названия: б. *еврокрация*, *медиокрация*, *наркокрация*; ч. *bankokracie*, *partitokracie*, които и днес се разширяват с нови елементи. В SN 1, SN 2 са посочени и някои експресивни названия: *eurokracie*, *stupidokracie*, а в българския медиен език се срещат оказионални образувания като *бандитокрация* и *чалгокрация*.

4.3.1.2.1. Новите имена за лица с -**крат/-krat** са значително по-малко от съществителните с -**крация/-kracie**, напр. б. *еврократ*; ч. *etnokrat*, те нерядко имат експресивен семантичен оттенък: б. *партократ*; ч. *telekrat*. Формации с проприална основа, разкриващи още веднъж словообразувателния потенциал на компонента, засега се срещат само в езика на медиите: б. *доганокрация*, *костокрация*; ч. *havlokracie*, *klausokracie*.

4.3.1.3. На елемента **-filie** със значение ‘obliba něčeho, záliba v něčem’ е посветена отделна речникова статия само в SN 2. Неографският подход в този случай вероятно се основава на факта, че на този етап с **-филия/-filie** се оформят предимно чужди основи, при това по-активно в чешкия. В медийните текстове обаче се очертава и още една немалобройна група от експресивни неологизми, възникнали чрез комбинацията на словообразувателни елементи с различна комуникативно-стилистична характеристика, срв. б. *еврофилия*, *хомофилия*, *циганофилия* и

ч. *keltofilie*, *mobilofilie*, *technofilie*, *úřadofilie* и др.

4.3.1.3.1. По-ниската продуктивност на -**filie** в чешки контрастира с активното парадигматично образуване на неологизми с **-fil**[12], напр. *gramofil*, *mobilofil*. В SN 1 и SN 2 е налице и голяма група, образувана по модела: „проприална основа + **-fil**", напр. *cimrmannofil*, *hrabalofil* и др.

4.3.1.3.2. Цитати от медийните текстове като: *„В България има германо**фили**, има американо**фили**, има русо**фили**, има турко**фили**, имало е и гърко**фили** и съответно **фоби**. Това е част от пейзажа. Въпросът е дали зад всичките тези **фили** и **фоби** успяваме да прокараме българо**филията**."* потвърждават функционирането на българските **-филия** и **-фил** най-вече като средства за образуване на нови думи по аналогия и свидетелстват за възникването на антонимни отношения между придобилите самостоятелност съставки **-фил** и **-фоб**.

4.3.1.4. Интернационалният тип nomina actionis, оформени със суфиксоид -**иада/**-**iáda** за назоваване на различни мероприятия, игри и състезания, е представен в най-новия лексикален пласт с множество единици: б. *азиада*, *бондиада*, *потериада*, *фолклориада*; ч. *bondiáda*, *oskariáda*, *puzzliáda*, *ramboiáda* и др. Посочените примери, повечето от които от речниците на новите думи, също разкриват тенденцията, обща за всички разгледани дотук постпозитивни афиксоиди, а именно – нарастването на продуктивността на проприалните названия в суфиксоидното словообразуване.

4.3.1.5. Суфиксоидът с произход от гръцки език с форма **-тека/-téka** се налага при образуването на неологизми по следните словообразователни модели: а) „радиксоид **-тека/-téka**": б. *видеотека*, *порнотека*; ч. *fototéka*, *infoteka* (в българския понякога и със съединителна гласна: *артотека*); б) „съществително нарицателно + **-тека/-téka**": б. *медиатека*; ч. *hobbytéka*; в) почесто в чешки – „абревиатурна основа + **-тека/-téka**": б. *DVD-тека*; ч. *CDtéka*, *DVDtéka*; г) само в чешки – „абревиатурна дума + **-тека/-téka**": *cédétéka*. Вероятно поради сравнително ограничените си сфери на употреба последните два типа имена не са поместени в нито един от разглежданите речници. За намаляващата продуктивност на **-téka** в чешкия говори и включва-

нето му единствено в SN 1.

4.3.1.5.1. Въпреки слабата си продуктивност суфиксоидът **-тека/-téka** развива полисемност и в двата езика, което до голяма степен се дължи на по-ранното заемане на интернационализми като *diskotéka/дискотека*. В резултат на това се обособява четвърта група имена с **-téka/-тека** в постпозиция, назоваващи места за забавление или забавна вечер, ориентирана към слушането на музика от определен жанр, напр. б. *латинотека*, *ретротека*, *фолкотека*; ч. *punkotéka*, *retrotéka*, *salsatéka* и др.

4.3.2. Втората значително по-малочислена група суфиксоиди от таблици 1 и 2 представляват единиците, навлезли заедно със заемки от английски език: **-мен/-man** 2., **-гейт/-gate**, **-холик/-holik**[13], -**холизъм/-holizmus**.

4.3.2.1. Тенденцията към интернационализация и/или англоамериканизация, характеризираща тази група, води и до семантичното профилиране на някои от посочените постпозитивни компоненти[14]. Така напр. Л. Яновец и А. Рангелова (Яновец, Рангелова 2005: 87) обръщат внимание на промяната в значението на **-man** и изтъкват, че -**man** освен в думи със семантика, сродна на терминологичната[15] и спадаща към следните два подтипа: 1. ‘който е (прекалено, дори болестно) пристрастен към нещо’: *agentoman*; и 2. ‘който е болестно зависим от нещо’: *internetoman*; започва да се употребява и в трето (нетерминологично) значение: ‘който се интересува от нещо, който прави нещо, представител е на нещо’[16], срв. *salesman*, *technoman* и др.

За разлика от чешкия в българския със същото значение се установяват цели редове неологизми с крайна част **-мен**, тъй като тя е фонетично по-близка до произношението на [-mæn] в английски, напр. *фронтмен*, *яхтсмен*, *шоумен*. Процесът на смислова, звукова и формална диференциация на **-ман/-мен** протича независимо от наличието на по-отдавна заети англицизми, означаващи професионална принадлежност и регистрирани в БТР, напр. *барман*, *ватман*. Поради тази причина формата на суфиксоида **-ман** със семантика „болестно зависим” и при по-старите заемки, напр. *меломан*, и при по-новите едини-

ци, напр. *интернетоман*, *кафеман*, *киберман*, *митоман* и др., съвпада.

4.3.2.2. Във всички славянски езици суфиксоидът **-гейт/ -gate** навлиза заедно с англицизма *Watergate*[17]. В българския той постепенно започва да се употребява като словообразувателно средство със значение ‘афера, скандал’[18], а статистиката от речниците показва сравнително високата му продуктивност, особено в сравнение с чешкия.

4.3.2.2.1. В езика на българските медии се очертават три основни групи с -**гейт** със следните словообразувателни структури: а) „съществително нарицателно + -**гейт**“: *адвокатгейт*, *досиегейт*, *парногейт*, *петролгейт*, *посланикгейт*, *прокуроргейт*, *телефонгейт*; б) „проприално название (лично/географско/прозвище) + -**гейт**“: *Батетогейт*, *Бригогейт*, *Ивановгейт*, *Ловечгейт*, *Овчаровгейт*, *Симеонгейт*; в) „инициално съкращение + -**гейт**“: *МВР-гейт*, *ДДС-гейт.*

4.3.2.2.2. Членовете на чешките неологични групи с **-gate** преобладаващо запазват характера си на цитатни изрази. Неясната семантика на **-gate** в чешки се преодолява например чрез създаването на словосъчетания със съществителните: *aféra*, *kauza*, срв. aféra *Čunekgate*, aféra *Grossgate*, kauza *Langergate*, aféra *Orlengate*, но и в тези случаи **-gate** се комбинира предимно с лични имена. За по-слабата продуктивност на компонента в чешки има значение и несклоняемостта на единиците с **-gate**.

4.3.2.3. Сравнението на данните в таблици и 1 и 2 показва, че и в двата езика суфиксоидът **-холик/-holik** е словообразувателно по-активен от суфиксоида -**холизъм/-holizmus**, напр. б. *трудохолик*, *шопохолик*; ч. *jídloholik*. От съпоставката също така личи, че компонентите -**holik** и **-holizmus** са с обща по-висока продуктивност в чешкия. В SN 1 и SN 2 са посочени например единиците *netholik*, *surfholik*, *workoholik/workholik/workaholik*, а в РНДЗБЕ и РНДБЕ с **-холик** е регистрирана само малка, затворена и вече доста време съществуваща група: транслитерираният англицизъм *уъркохолик*, хибридните имена, образувани чрез калкиране на първия компонент на същата заемка: *работохолик*, *трудохолик*, неологизмите *магазинохолик* и *сексхолик*, както и женсколичното име *козметохоличка*[19].

4.3.3. В третата и най-малка група суфиксоиди от българските и чешките речници на новите думи попадат елементите от славянски произход. В българския това е девербалният компонент -**воз**, който започва да функционира самостоятелно, след като се отделя от композиционно-суфиксни формации, образувани чрез нулева суфиксация, напр. *бетоновоз*, *панеловоз*[20] и др. Съгласно констатацията на Л. Яновец и А. Рангелова (Яновец, Рангелова 2005: 88), че процесът на функционално и семантично преосмисляне в чешки е особено напреднал при суфиксоидите от групата: **-fil**, **-man** и **-mil**, **-milec**, което води до синонимната им употреба в структурата на съществителни, изразяващи степен на „положително отношение към обекта (предмета или лицето) по скалата „възхищение – въодушевление – любов – обсебеност", към тази група спадат и елементите **-mil** и **-milec**[21], напр. *filmomil – filmomilec*, *čajomil – čajomilec*.

4.4. Обособяването на инвентар от постпозитивни компоненти с различни функционални и семантични черти, изразяващи широк диапазон от семантични отношения към означеното с мотивиращата основа – от симпатизиране, възхищение, интерес, ангажираност и привързаност до болезнена зависимост от определени явления или предмети – активизира и образуването на цели редове нови названия за лица. В РНДЗБЕ, РНДБЕ и SN 1, SN 2 най-многобройна е именно групата суфиксоидни имена за лица според идейната им ориентация, сферата им на интереси и под., напр. б. *глобофоб*, *медиократ*; ч. *homofobik*, *myšoman*. По активност в словообразувателния процес при тези названия в българския на първо място се нарежда суфиксоидът **-лог**, следван от **-ман**, **-фоб**, **-фил,** а на последно място остава **-крат**. В чешкия на първо място е компонентът **-fil**, следван от **-man** 1., 2., по-слабо застъпени в неологичния материал са **-mil**, **-krat**, **-milec**, а накрая остава суфиксоидът с варианта на форма **-fob/-fobik**.

4.5. Разширяването на съчетаемостта на суфиксоидите е свързано също така с появата на нови семантични (респективно парадигматични) отношения между новите думи, образувани с тях. Така например въз основа на материала от речниците може да се съди за още една характерна за суфиксоидните неологизми черта – способността им да изразяват различна степен на

привързаност/зависимост на лицето от предмета или явлението, означени от мотивиращата основа, срв. б. *технофил – техноман*[22] *– техноманиак*[23]; *нетофил – нетохолик – нетоман – нетоманиак*; ч.: *keltofil – keltomil – keltomilec*; *mobilofil – mobilomil – mobiloman*; *nedvědofil – nedvědomilec* и др. Този начин за семантична интензификация, постигана чрез прибавяне на постпозитивен компонент, може да се сравни с начините за степенуване на прилагателните имена в синтетичните езици. Налага се заключението, че РНДЗБЕ и РНДБЕ регистрират един нов тип изразяване на семантични отношения в рамките на лексемата, неприсъщ за аналитизма на книжовния български език.

4.5.1. За напредналата фаза на семантична трансформация при някои суфиксоиди може да се съди и по парадигматичните отношения между формациите, завършващи на **-ман/-man**, **-холик/-holik** и **-фил/-fil**, със значение ‘любител на нещо, почитател; фанатичен, дори болестно зависим от нещо’, които се включват в редове много близки по значение двойки/тройки, напр. б. *кинофил – кинохолик – киноман*; ч. *čokofil – čokoholik – čokoman*.

4.6. Процесите, регистрирани в българските и чешките речници на новите думи, кореспондират с активизирането на някои други нови суфиксоиди, непосочени в тях, но присъстващи в езика на медиите.

4.6.1. Повишава се например словообразователната активност на компонента **-вед**[24]. В българския с него се оформят две групи нови думи със следните значения: ‘специалист в дадена област’: *рекламовед*, *религиовед*, *филмовед* (срв. с по-старите единици като *изкуствовед*) и по-общото ‘който разбира от нещо’ – по аналогия с вече съществуващите в езика думи като *сърцевед*, напр. *звездовед*, *социовед* (успоредно със *социолог*), нерядко имащи експресивен оттенък[25]. Третата и най-малка част неологизми с **-вед** отново са свързани с тенденцията за повишаване на съчетаемостта на проприалните основи, напр. *костовед*, *бойковед*.

4.6.2. Появата на нови стопански отрасли влияе на словообразователната продуктивност на моделите със задпоставен компонент **-въдство/-въд**. В РНДБЕ са регистрирани например нео-

логизмите *охлювъдство – охлювъд*, *щраусовъдство – щраусовъд*, възникнали по аналогия с по-стари единици, например *пчеловъдство – пчеловъд*, които в БТР са характеризирани като заемки от руски език[26].

4.6.3. В групата на суфиксоидните неологизми, очертаващи границата между неологичното и оказионалното словообразуване, попадат и някои лексикално и семантично несамостоятелни единици, образувани със следните постпозитивни компоненти.

4.6.3.1. Сравнително широко застъпен в най-новия пласт от експресивна лексика, със сфера на употреба предимно българския медиен език, е словообразувателният модел, оформен с компонент **-ландия**, напр. *Абсурдландия*, *Евроландия*, *Простоландия*, *Първулландия*, *Шоплания* (по аналогия с названията на държави като *Холандия*, *Шотландия* и др.). Макар и по-рядко, моделът се използва и при образуването на хрематоними, напр. *Коколандия*. За разлика от българския, в чешкия е регистриран единствено интернационализмът *euroland/Euroland*, разпространен предимно в политическата и икономическата публицистика (вж. и SN 2).

4.6.3.2. При имената със силно експресивен, разговорен и неодобрителен оттенък и в двата езика се налага компонентът **-стан/-stán** (по аналогия с названията на държави като б. *Таджикистан*; ч. *Turkmenistán* и под.). Неографските източници съдържат неологизмите публицизми б. *Абсурдистан*; ч. *Absurdistán/absurdistán*, а ексцерпираният материал показва, че компонентът **-стан** има по-висока словообразувателна активност в българския, срв. *мангалистан*, *циганистан*, *чалгастан*; ч. *banánistán*, като между единиците, представители на последните два разгледани словообразователни типа, се установяват дори и вариантни/синонимни отношения, напр. *бананландия – бананистан.*

5. Заемането на редица терминологични названия с общ пълнозначен задпоставен компонент, който в български и чешки (и в останалите славянски езици) функционира и като самостоятелна лексема, напр. б. **(-)терапия**: *вокалотерапия*, *рефлексотерапия*; ч. **(-)fobie**: *agyrofobie*, *lachanofobie*, става импулс

за аналогичното образуване на нетерминологични единици по същите словообразователни модели. Паралелно с това към редовете думи със зает компонент възникват единици с домашен постпозитивен член, напр. б. -**лечение** (към **-терапия**); ч. **-milec** (към **-fil**). Процесът кореспондира с тенденцията към одомашняване на чуждите термини „чрез заместване на части от чуждиците с домашни лексикални основи" (Мартинцова, Савицки 1987: 131–132, преводът мой – Б. Н.).

5.1. Според РНДЗБЕ, РНДБЕ и SN 1, SN 2 в структурата на новите композитуми с общ втори член най-активно участват постпозитивните компоненти, представени в таблици 3 и 4 (вж. понататък). Както при суфиксоидите не всички посочени в честотните таблици съставки на сложни думи са разгледани в самостоятелни речникови статии, такъв е например случаят със задпоставените части -**терапевт/-terapeut, -турист/-turista, -фобия/-fobie**. Също така само в българските речници са отделени компонентите **-бос** и **-фест**, а само в чешките – компонентите **-centrismus/-centrizmus**, **-centrum**, -**festival**, **-maniak**, **-turismus/ -turizmus**, **-turistika** 1., 2.

Таблица 3

Честота на словообразувателно активните постпозитивни лексикални основи и на неологизмите, образувани с тях, според РНДЗБЕ и РНДБЕ	
-терапия	32
-мания	20
-туризъм	11
-фест	11
-**маниак**	10
-център	9
-терапевт	8
-бос	8
-лечител	6
-турист	4
-фобия	4
-центризъм	2

Таблица 4

Честота на словообразувателно активните постпозитивни лексикални основи и на неологизмите, образувани с тях, според SN 1 и SN 2		
-manie	72	
-centrum	32	
-terapie	22	
-maniak	16	
-centrismus/-centrizmus	12	
-terapeut	12	
-fobie	9	
-turistika 1.	6	*organizované, speciálně zaměřené cestování po trasách s charakteristickým obsahem, mající charakteristický účel, cíl*
-turistika 2.	3	*ježdění, cestování za jiným cílem než turistickým*
-festival	6	
-turista	6	
-turismus/-turizmus	4	
-léčitel	1	

5.2. Данните от РНДЗБЕ и РНДБЕ са красноречиви по отношение на това, кои са най-високочестотните пълнозначни компоненти. Новите сложни думи с **-терапия** и **-мания** неколкократно превишават останалите формации с общи постпозитивни лексемни основи. За разлика от ситуацията при суфиксоидния инвентар сред продуктивните втори съставки при композитумите няма елементи от домашен произход. С други думи, най-широко разпространените неологизми от разглеждания тип възникват предимно по аналогия с интернационални словообразувателни модели. От друга страна, за съществената роля на англоамериканизацията като катализатор на словообразувателна динамика в приемащата езикова система свидетелства участието в процеса на композитообразуване на заемки като **-бос**: *кримибос*, *радиобос*, *тотобос* и -**фест**: *велофест*, *магфест*, *фотофест*[27].

5.3. SN 1 и SN 2 регистрират явления и процеси, които са сходни с тези в българския. По-съществени различия се проявя-

ват в честотата на употреба на конкретните компоненти. Прави впечатление особено високата продуктивност на втората съставка **-manie**[28]. В сравнение с българския е значителна и продуктивността на вариантния компонент **-centrismus/-centrizmus**, напр. *bruselocentris(z)mus*, *pragocentris(z)mus* и др., която очевидно е следствие от силно изразената (във всички славянски езици) тенденция към образуване на абстрактни имена за означаване на идейни платформи и течения. Не на последно място в силен контраст със ситуацията при суфиксоидните неологизми е много по-слабата словообразувателна активност на задпоставените лексикални компоненти за образуване на имена за лица като **-турист/-turista**, напр. б. *арттурист*, *велотурист*; ч. *agroturista*, *ufoturista* и др. За разлика от **-туризъм** в българския, в чешкия неологичен материал елементът **-turismus/-turizmus** е слабо застъпен, тъй като той е силно конкуриран от единствения задпоставен компонент от тази група в процес на полисемантизация: **-turistika** 1.: *gastroturistika* и **-turistika** 2.: *etnoturistika*, срв. също и *sexturis(z)mus* – *sexturistika.*

5.4. Българският и чешкият неографски материал от разглеждания тип може да бъде разделен в две основни подгрупи, които имат следната словообразувателна структура: 1) „първа произвеждаща дума – самостоятелно функционираща в езика лексема + серийно употребявана задпоставена (опорна) дума”, напр. б. *бебецентър*; ч. *sportcentrum*, понякога свързани и чрез съединителна гласна: б. *цвето**о**терапия*; ч. *diet**o**terapie*; 2) „първа съставка препозитивен афиксоид (в различен етап на функционална и семантична трансформация) + пълнозначна втора основа”, напр. б. *авиофестивал*; ч. *cyklocentrum.*

С оглед на казаното дотук обект на по-нататъшното изложение ще бъдат отделните пълнозначни основи в постпозиция и съпоставката на новите думи, образувани с тях.

5.4.1. В РНДЗБЕ и РНДБЕ компонентът **(-)център/(-)centrum** не е представен самостоятелно, но в отделни речникови статии се срещат неологизми и нови словосъчетания, в които той функционира като опорен компонент, напр. *експоцентър*, *таласоцентър*; *бизнес център/бизнесцентър*, *мол център*,

офис център, *тунинг център* и др.[29] Най-значителна по големина е групата, в която **(-)център/(-)centrum** се комбинира с препозитивни радиксоиди, някои от тях дори полисемантични[30], напр. б. *артцентър*, *инфоцентър*, *автоцентър*, *демоцентър*; ч. *ekocentrum*, *videocentrum*. В езика на българските медии се срещат също така терминологични названия като *медиацентър* ‘преносимо мултимедийно устройство’ и вторични композитуми като *евроинфоцентър*. Със значение ‘център на столицата’ се употребяват формациите *суперцентър*, *топцентър*, а в чешкия *minicentrum* означава ‘малък център за услуги’, като всичко това показва, че съчетаемостта на компонента -**център/-centrum** се разширява.

5.4.2. Функционално и смислово лексемата **фест/fest** се отнася към полисемантичния интернационализъм *фестивал/festival*[31], от който е образувана чрез механично отделяне на първата му част. Българските речници разглеждат **(-)фест** както самостоятелно, така и като втора съставка на нови сложни думи. SN 2 посочва само лексемата **fest**, но в чешкия материал от медиите се срещат и нови названия на различни мероприятия, напр. *pivofest Trutnov*, *punkfest Proti Proudu*, *Sázavafest*, отразяващи известна промяна в словообразувателната активност на суфиксоида.

Българските неологизми с -**фест** спадат към няколко основни групи: 1) „интернационална лексемна основа + -**фест**”: *джазфест*, *кинофест*, *филмфест*; 2) „съкратена основа + -**фест**”: *бирфест*, *винфест*; 3) „афиксоид + -**фест**”: *авиофест*, *екофест* и др. Съчетаването на -**фест/-fest** с афиксоид нерядко се осъществява паралелно с образуването на нови имена по модела „афиксоид + **-фестивал/-festival**”, което от своя страна поражда конкуренция и вариантност между формациите с идентичен препозитивен компонент, срв. б. *етнофест – етнофестивал*; ч. *fotofest – fotofestival*.

5.4.3 Както в българския, така и в чешкия език лексемите от гръцки **фобия/fobie** и **мания/mánie** се употребяват в извънтерминологичен контекст значително по-дълго време, отколкото разгледаните дотук компоненти. Конвергенцията в тази област се засилва и от следните едновременно проявяващи се

иновационни процеси. От една страна, допълват се липсващите членове на редовете с постпозитивни части **-филия/- filie**, **-мания/-manie** и -**фобия/-fobie**, между които се установяват синонимни и антонимни релации, точно копиращи прагматично-семантичните връзки при терминологичните названия, срв. напр. б. *еврофилия*, *евромания* и *еврофобия*; *киберофилия*, *кибермания* и *киберфобия*; *технофилия*, *техномания* и *технофобия*; ч. *etnofilie*, *etnomanie* и *etnofobie*; *technofilie*, *technomanie* и *technofobie*; *xenofilie*, *xenomanie* и *xenofobie*. Отделните членове на подобни групи се различават по сферата си на образуване и употреба, респективно по стилистично-комуникативните си специфики. От друга страна, в езика на медиите чрез неологизмите с препозитивен компонент съществително нарицателно име или проприално название, комбинирано с **-фобия/-fobie//-мания/-manie**, често се изразява лично отношение към дадено явление или лице, напр. б. *диетомания*, *звездомания*, *кафемания*, *потърмания*, *пъзелмания*; ч. *agentomanie*, *blondiemanie*, *bondomanie*, *sluncemanie*, *titulomanie*. Експресивните единици с оказионален характер възникват и в зависимост от актуалността на дадена тема в обществото, напр. б. *бойкомания*, *гербомания*; ч. *jágrmanie*, *jelcinomanie.* Единиците от неографския и медийния неологичен материал, оформени с втори компонент **-фобия/-fobie**: б. *ислямофобия*, *наркофобия*, *юдеофобия*; ч. *elitofobie*, *jugofobie*, *komunistofobie*, са значително по-малко от неологизмите, завършващи на **-мания/ -manie**.

5.4.4. Новите имена със съставка -**маниак/-maniak**, напр. б. *бингоманиак*, *киберманиак*, *кечманиак*; ч. *keltomaniak*, *mobilomaniak*, *potteromaniak*, изграждат семантични релации с неологизмите със суфиксоид **-фил/-fil** и -**ман/-man** + идентична препозитивна част, срв. б. *аудиоманиак – аудиофил*, *сексоман – сексманиак*, а това води до поява на вариантност както по отношение на степента на книжовност/некнижовност на новите думи, така и с оглед на употребата им в рамките на един текст.

5.4.4.1. РНДЗБЕ и РНДБЕ не посочват неологизми с втора част **-любител**, напр. *астролюбител*, *еколюбител*, *метеолюбител*, *фитнеслюбител/фитнес любител* и др., която днес съ-

що е активна при допълването на разглежданите вариантни редове[32], срв. *видеофил – видеолюбител* или *кафелюбител – кафеман* (изразяващ по-силна степен на пристрастеност към кафето). Близки до значението на българските неологизми с **-любител** са експресивните чешки единици с **-milovník**, срв. шестчленната група: *filmomilovník – filmomil – filmomilec – filmofil – filmoman – filmožrout*[33]. Съседните членове на подобни неологични редове функционират не само като вариантни лексикални средства, но и като синонимни единици в рамките на един и същи текст.

5.4.5. В съпоставителен чешко-български план участието на интернационалния словообразувателен компонент **-terapie/ -терапия** в словообразувателния процес е подробно разгледано от З. Тиха (Тиха 2003). Авторката достига до заключението, че първият член на формации с крайна част **-терапия/-terapie** обикновено означава ‘това, с чиято помощ лечението/терапията се осъществява’ и подчертава, че основната разлика между двата езика е силната конкуренция между постпозитивните лексикални основи -**лечение** и **-терапия**[34] в българския (Тиха 2003: 216), напр. *медолечение – медотерапия*, *парафинолечение – парафинотерапия*, *плодолечение – плодотерапия* и др.

5.4.5.1. Цитираната констатация като цяло съвпада със състоянието, отразено в речниците на новите думи, въпреки че в РНДЗБЕ и РНДБЕ е самостоятелно представен единствено компонентът -**терапия**, докато В SN 2 освен по-многочислената група с втора част **-terapie**, напр. *dietoterapie*, *muzikoterapie*, *vitaminoterapie*, са регистрирани и следните единици с компонент **-léčení/-léčba**: *samoléčba/samoléčení* (към англицизма *self-medication*), *domoléčba*.

5.4.5.2. Неографските източници сочат, че към новите думи с **-терапия/-terapie** се образуват и имена с втора част **-терапевт/-terapeut**, напр. б. *арттерапевт*, *мезотерапевт*, *музикотерапевт* (срв. б. *енерголечител*, *природолечител)*; ч. *etikoterapeut*, *urinoterapeut*, *zooterapeut*, а единици като б. *паралечител*; ч. *kvaziterapeut* по словообразувателния модел „префиксоид + **-лечител/-léčitel**” се появяват доста по-рядко. Лексемите и от двата разглеждани типа могат да се интер-

претират като имена с двойна мотивация: *ароматерапия → ароматерапевт* или *арома + терапевт → ароматерапевт.*

РНДЗБЕ и РНДБЕ отразяват и новата само за българския език[35] тенденция към засилено образуване на женсколични имена от мъжколични неологизми, напр. *шокотерапевт → шокотерапевтка*; *гладолечител → гладолечителка.*

6. РНЗДБЕ разглежда англицизмите **бизнес**, **хит** и **шоу** самостоятелно и като продуктивни членове на словообразователния инвентар – **бизнес** в препозиция и постпозиция, а **хит** и **шоу** само като първи членове на нови думи. Иначе казано, според РНДЗБЕ в българския език „фонологичният родов модел функционира почти без изключение" (Вачков 1997: 21–22), което позволява заемки като **бизнес**, **хит** и **шоу** лесно да се включат в субстантивните парадигми за мъжки/среден род. Този факт улеснява създаването на аналогични на англоамериканските български конструкции, води до повишаване „функционалната натовареност"[36] на **бизнес**, **хит** и **шоу** и предизвиква тяхното функционално преосмисляне, респ. постепенното им преминаване към групите на словообразувателно активни препозитивни/постпозитивни компоненти на сложни думи[37].

За разлика от РНДЗБЕ в РНДБЕ е отделен единствено компонентът **-бизнес**, а нови думи с **хит-** и **шоу-**, напр. *видеохит*, *стрийтшоу* присъстват само в илюстративния материал към други речникови статии. Обстоятелството, че активността на **бизнес**, **хит** и **шоу** в последно време е насочена предимно към образуването на терминологични словосъчетания, в които те са неизменяеми апозитивни членове, личи от множеството включени в речника словосъчетания от типа *риалити шоу*, *хит майстор*, *бизнес администрация* и др.

Чешките речници на неологизмите не характеризират нито един от англицизмите **byznys**, **hit** и **show** като пълноценен член на словообразователния инвентар, но регистрират отделни заети композитуми с тяхно участие, напр. *byznysplán*, *cyklobyznys*, *retrohit*, *showroom*. Във връзка с това И. Мравинацова отбелязва, че в чешкия компонентите на англоамерикански конструкции от типа *business class – show business*; *hit parade – super hit*, независимо от това, дали са изписани слято, полуслято

или разделно, се възприемат по-скоро като „заемки без морфологичен признак” (Мравинацова 2005: 198, преводът мой – Б. Н.).

6.1. Казаното дотук показва, че е необходимо типовете български и чешки неологизми, възникнали с **бизнес/byznys**, **хит/hit** и **шоу/show** и регистрирани в речниците на новите думи, да бъдат разгледани по-подробно, тъй като те разкриват някои актуални и особено интересни прояви на съвременната словообразувателна и лексикална динамика.

6.1.1. В българския език композитумите с втори основен компонент **-бизнес** са от два основни словообразувателни типа: 1) „пълнозначна лексема + **-бизнес**”: *интернет-бизнес*, *радиобизнес*, *сексбизнес*, *шоубизнес* и под.; 2) „радиксоид + -**бизнес**”: *авиобизнес*, *агробизнес*, *автобизнес*, *наркобизнес*, като и в двата случая първата съставка има семантика, уточняваща предмета на дейността. В чешкия еквиваленти на разгледаните неологизми с компонент **-byznys** се срещат по-рядко, напр. *sexbyznys*, *agrobyznys*, *narkobyznys*.

6.1.1.1. Като първи компонент на български композитуми **бизнес-** обикновено има широко значение – ‘отнасящ се към бизнеса’ – и е особено продуктивен с домашни основи, напр. *бизнесдела*, *бизнесминало*, *бизнесобединение*, *бизнесопит*, *бизнеспечалба*, *бизнессдружение*, *бизнессреда* и др. С негово участие се оформя и група по модела „**бизнес**- + субстантивирано прилагателно, назоваващо чужд език”: *бизнеснемски*, *бизнесруски* със значение ‘уроци по чужд език, насочени към нуждите на комуникацията в областта на бизнеса’, напр. *Фирмата предлага курсове по бизнесанглийски*. В сферата на търговията и услугите неологизми с компонент **бизнес-** като *бизнесабонат*, *бизнесклиент*, *бизнеспотребител*, *бизнестарифа* се образуват с цел реализацията на семантичната опозиция ‘предназначен за частен (единичен) потребител’ – ‘предназначен за бизнес (групов, фирмен) потребител’.

6.1.1.2. Паралелно с интернационализма *бизнесмен* (конкуриращ имена като *предприемач*, евентуално *частник*) в българския медиен език се употребяват и неологизмите *бизнесдама*, *бизнесжена.* Явлението отново е свързано с тенденцията за различаване на половия признак при назоваването на лица.

6.1.1.3. На фона на ниската словообразувателна активност на чешката лексикална единица **byznys** прави впечатление, че с **(-)бизнес(-)** в български се образуват и вторични композитуми, напр. *агробизнесфорум*, *бизнесвсекидневник* и др.

6.1.2. Групата нови български думи с втора опорна съставка -**шоу** се обособява в контекста на навлизането на редове транслитерирани англоамериканизми като *геймшоу*, *роудшоу*, *сноуборд-шоу*, *травестит-шоу* и др. Опорната задпоставена част -**шоу** се комбинира предимно с радиксоиди в различен етап на функционална трансформация, напр. *авиошоу*, *автошоу*, *мотошоу*, *радиошоу*. В чешкия англицизмът **show/šou** остава словообразувателно непродуктивен и това се дължи на несклоняемостта на заемката със специфичен завършек, която се употребява както в женски, така и в среден род. SN 1 регистрира само заетия английски израз *strip show*.

6.1.2.1. По аналогия с употребата на *show* в англоамерикански изрази субстантивната част **шоу(-)** със значение **‘**който се отнася до забавен спектакъл, представление’ започва да се налага и като първи компонент в състава на хибридни български неологизми, напр. *шоуводещ*, *шоугрупа*, *шоуновини*, *шоупредаване*.

6.1.3. Англицизмът **хит** се установява в български със значение ‘особено популярно, търсено, любимо произведение, изделие и др.’. Той запазва семантиката си и като постпозитивен лексикален компонент на нови сложни думи, възникнали основно при свързването му с радиксоид, напр. *ретрохит*, *фолкхит*, по-рядко с коренна морфема, напр. *кинохит*, *радиохит*, *попхит.* Подобно на българските и чешките неологични речници регистрират предимно единици с компонента **-hit** в постпозиция и афиксоид в препозиция, напр. *gigahit*, *velehit* и *radiohit*, *videohit*. За неотдавнашното активизиране на препозитивния член **hit-** свидетелства фактът, че неологизми като *hitrádio* и неосемантизми като *hitparáda* са включени едва в SN 2.

6.1.3.1. В български език между сложните думи с първа съставка **хит-** със значение ‘който се отнася до хитове’ и словосъчетанията със съгласувано определение *хитов* нерядко възникват конкурентни отношения, срв. *хитова марка – хитмарка*,

хитово предаване – хитпредаване, *хитов сингъл – хитсингъл*, *хитова стока – хитстока*. Не всички калкирани изрази с **хит-** обаче имат свои двусловни еквиваленти, напр. *хитлиста*, *хитпарад.*

7. Въз основа на проведеното изследване може да бъдат направени следните изводи:

7.1. Според данните от неографските източници словообразувателно активните български и чешки постпозитивни единици са значително по-малко от препозитивните словообразувателни компоненти, регистрирани в неологичните речници. Тази констатация е обяснима с факта, че както показа анализът, *постпозитивните компоненти не само участват в словообразувателната и семантичната структура на новата лексема, но изпълняват и функцията на неин морфологичен (граматичен) маркер*. От друга страна, проявите на динамика в областта на суфиксоидното словообразуване са значително по-силно изразени в чешкия, което произтича от флексивния му характер. В аналитичния български – напротив – между новите сложни думи със суфиксоидна лексема в постпозиция се изграждат разнообразни семантични отношения, подобни на семантичните релации, установяващи се между чешките членове на суфиксоидни групи.

7.2. *Процесът на функционално и семантично преосмисляне на редица пълнозначни постпозитивни компоненти (от чужд, но и от домашен произход) е обусловен от заемането на цели редове терминологични и абстрактни названия, т.е. създаването на нови думи с постпозитивен компонент се осъществява предимно по аналогия с интернационални словообразувателни модели.* Словообразувателните средства и механизмите за морфологична (словообразувателна) адаптация на интернационализмите, които активизира всеки отделен език, нерядко обаче се оказват различни.

7.3. *В българския и чешкия неологичен материал най-голям е делът на новите думи с постпозитивен словообразувателен компонент, назоваващи имена на лица в зависимост от тяхната идейна, политическа, професионална, любителска и под. ориентация.* Освен конституирането на богати

словообразувателно-семантични скàли за изразяване на заинтересованост, привързаност и дори болестна зависимост от определени предмети или явления, разгледаните функционални и семантични специфики на суфиксоидните формации от РНДЗБЕ, РНДБЕ и SN 1, SN 1 показват, че *и в двата езика се установяват групи от постпозитивни словообразувателни средства за изразяване на лично/оценъчно отношение към назоваваната от неологизма субстанция и за интензифициране на семантиката*. С други думи, към инвентара от постпозитивни компоненти с „преходен" характер по-бързо преминават елементите, активно участващи в изграждането на подобен тип лексикалносемантични (парадигматични) отношения.

7.4. В сравнение с предишния етап от развитието на българския и чешкия книжовен език активизирането на разгледаните в статията словообразувателни средства и модели се оказва една от най-съществените прояви на съвременната лексикална динамика. От тази гледна точка детайлната съпоставка на материала в неологичните речници на различни славянски езици с цел описанието на функционално-семантичните особености на отделните членове на неологичния репертоар и на парадигматичните връзки, възникващи между тях, е тясно свързана с прилагането на съвременни методи и подходи в двуезичната славянска лексикография.

СЪКРАЩЕНИЯ

а. – английски език
б. – български език
ч. – чешки език

БЕЛЕЖКИ

[1] Терминът *постпозитивен словообразувателен компонент* се използва като надреден, т.е. с него се означават както суфиксоидите, така и суфиксоидните лексеми.

[2] Тъй като въпросът за статута на т.нар. суфиксоиди (като значителна част от афиксоидния репертоар) е един от най-сложните въпроси на съвременното славянско словообразуване, е необходимо да се уточни, че в това изследване с термина *суфиксоид* се означават

афиксоидите, употребявани в постпозиция, които като несамостоятелни морфеми с преходен (междинен) характер много приличат на традиционните наставки, но значително се различават от тях по своите формални, семантични и функционални характеристики. Вж. също Аврамова 2003, Колковска 2003, Лотко 2000, Яновец 2003 и др.

[3] За неографското отразяване на иновационните процеси и тенденции вж. Мартинцова, Светла 2000, Благоева 2008, Мартинцова 2003 и др.

[4] Степента на продуктивност на даден компонент нерядко се оказва решаваща за самостоятелното или несамостоятелното му представяне в речниците на новите думи. В това изследване обаче критерият „степен на продуктивност" няма да е от такова значение за изводите от функционално-семантичния анализ.

[5] През последните двадесет години в България и Чехия излязоха и някои други неографски справочници, но за целите на това изследване ще бъдат използвани само посочените четири речника, тъй като те са сходни по обем и изчерпателност, по място на изготвяне – Лексикографската секция на Института за български език към БАН и Лексикографската секция на Института за чешки език към ЧАН – както и по отношение на приблизително еднаквия период от време, който обхващат.

[6] За източник на допълнителен неологичен материал се използва именно езикът на медиите поради това, че в съвременната лингвистика той се определя като „невралгичната точка на днешната езикова ситуация" (Гладкова, Ликоманова 2002: 106).

[7] Последното се отнася само до българския лексикографски материал, затова е необходимо да се уточни, че при обработката на честотните данни в таблици 1 и 2 (и във всички останали таблици) се изхожда от неографските методи и подходи, приложени в българските и съответно в чешките речници на новите думи, макар това понякога да поражда съществени разлики.

[8] По въпроса за активността на **-воз** в съвременния български и руски език вж. също Аврамова 2003: 183.

[9] Има се предвид само значението на **-ман/-man**, свързано със семантиката на по-стари терминологични заемки като напр.: б. *пироман*; ч. *kleptoman*. За чешки по-подробно вж. таблица 2 и **4.5.1.**, **5.3.4**.

[10] По въпроса за т.нар. повторна интернационализация вж. Бузашиова 1991: 89–104.

[11] За процесите на парадигматично словообразуване/парадигматична деривация вж. по-подробно Мартинцова 1983, Яновец, Рангелова 2005 и др.

[12] За разлика от **-filie** компонентът **-fil** е посочен в самостоятелна речникова статия още в SN 1.

[13] Въпреки че компонентът **-холик** се среща и в някои по-отдавна заети термини, напр. *алкохолик/alkoholik*, той става словообразувателно активен значително по-скоро, отколкото например **-fil/-фил**, **-man/-ман** и др., а това води до заключението, че вероятно става дума за влияние на повторната интернационализация. За суфиксоидите **-холик/-holik** и -**холизъм/-holizmus** вж. също Аврамова 2003: 182.

[14] Това е и един от малкото случаи на поява на полисемизация при суфиксоидите, в български същия процес наблюдаваме при **-воз**.

[15] В SSJČ: 1169 суфиксоидът **-man** е характеризиран като заемка от гръцки: „řec. mania, šílenství".

[16] По този въпрос вж. SN 1 SN 2. В РНДЗБЕ не намираме нито един от двата суфиксоида, а в РНДБЕ тълкуването на **-ман** е: 'човек, който проявява силно пристрастие, мания към това, което е изразено в първата част', синонимно на компонента **-маниак**.

[17] За **-гейт/-gate** вж. също Ухлиржова 1998: 218–219, Рангелова 1999: 34, Аврамова 2003: 181.

[18] Срв. с РНДЗБЕ.

[19] Примерите са посочени още у Рангелова 1996, Аврамова 2003, Благоева 2006.

[20] Примерите са от Радева 1991, вж. там по-подробно и с. 48, 69, 95.

[21] Според Яновец, Рангелова 2005: 91 **-mil** и **-milec** обаче също така се отличават и с черти на архаичност, което показва, че те все още пазят предметността и експресивността на значението си, следователно може да се заключи, че степента им на трансформация е значително по-ниска от тази на **-fil/-фил**, **-man//-ман/-мен**.

[22] Компонентът **техно-** се отнася както към техниката, така и към стила техно в музиката, точното му значение може да бъде изведено само от контекста.

[23] За задпоставения компонент **-маниак** вж. подробно **5.3.4**.

[24] Радева 1991: 62 определя компонента като зает лексикален елемент, който показва висока активност при участието си в деривационния процес.

[25] Вероятно под влияние на този процес към вече регистрираната лексема *душеведец* възниква и неологизма *душевед.*

[26] Докато обаче към *пчеловъдство – пчеловъд* съществуват и домашни двойки: *пчеларство – пчелар*, то неологизми като **щраусарство – *щраусар* не се образуват.

[27] Например в LDCE 2009 компонентът -**fest** е характеризиран като суфикс за образуване на сложни думи от разговорната лексика, по-

сочени са и примери като *talkfest*, *snoozefest*. За разлика от това **boss** се разглежда като съществително име, участващо в апозитивни конструкции, напр. *He will probably be none other than General Jaruzelski, the **party boss**.*

[28] За **-manie** вж. по-подробно **5.3.3**.

[29] По този модел в българския език по-рано е образувано например названието *киноцентър*, с което е наречена Националната филмова студия в Бояна и целият квартал в същия район.

[30] Яновец, Рангелова 2005: 96 посочват, че многозначността на този тип лексеми „отразява полисемията на компонента -**centrum** в нейното взаимодействие с възможната синонимия на препозитивната част на композитума" (преводът мой – Б. Н.).

[31] Вж. напр. представянето на *фестивал/festival* в РЧД 2007: 795, ASCS 2001: 227.

[32] Компонентът се установява в по-ранен етап, регистриран е напр. от Радева 1987: 82–83, най-вероятно думите, образувани с него, са свързани с влиянието на заемки от руски с втора част **-любитель** *(кинолюбитель)*.

[33] Чешкият постпозитивен словообразувателен компонент **-žrout** се използва сравнително рядко и то само при образуването на силно експресивни неологизми. В речниците на неологизмите с него са посочени и неологизми като *reklamožrout*.

[34] Някои имена с втора съставка **-лечение** разглежда и Аврамова 2003: 115, като ги включва към групата на новите nomina actionis.

[35] Т.нар. přechylování (образуване на имена, назоваващи лица от женски пол от имена за лица от мъжки пол) в чешкия не се разглежда като динамично явление, тъй като то е характерно за него.

[36] Мартинцова 2005: 223 отбелязва по-голямата функционална натовареност на лексикални средства със сравнително широка семантика, напр. *centrum*, *studio*, *management*, участващи като опорен член в редица чешки словосъчетания.

[37] На двата модела бе обърнато детайлно внимание в **2.**, подточка 3.

ЛИТЕРАТУРА

Аврамова 2003: Аврамова, Ц. *Словообразувателни тенденции при съществителните имена в българския и чешкия език в края на XX век.* София, „Херон прес". 259 с.

Благоева 2006: Благоева, Д. Неологизмите в съвременния български език. – *Български език и литература*, № 6, с. 16–25.

Благоева 2008: Благоева, Д. Съвременни подходи в българската неография (проблеми и перспективи). – *Български език*, № 1, с. 5–14.

Боздехова 1994: Bozděchová, I. *Tvoření slov skládáním.* Praha, Institut sociálních vztahů. 202 s.

БТР 2008: *Български тълковен речник.* Четвърто издание. Допълнено и преработено от Димитър Попов. София, „Наука и изкуство”. 1093 с.

Бузашиова 1991: Buzássyová, K. Opakovaná internacionalizácia a problem identifikácie morfologických a lexikálnych jednotek. – *Jazykovědný časopis*, № 2, s. 89–104.

Вачков 1997: Вачков, В. Формални модели на родова адаптация на английските заемки в българския и чешкия език. – *Съпоставително езикознание*, № 2, с. 19–25.

Вачкова, Вачков 1998: Вачкова, К., В. Вачков. Словообразуване и аналитизъм. Типологична обусловеност на някои словообразователни процеси в българския език. – В: *Славянска филология*. Т. 22. *Доклади за XII международен конгрес на славистите*, *Краков 1998*. София, с. 95–101.

Гладкова, Ликоманова 2002: Гладкова, Г., И. Ликоманова. *Языковая ситуация. Истоки и перспективы (болгаро-чешские параллели).* Praha, Karolinum. 452 s.

Земска и др. 1999: Zemskaja, E. A., O. P. Ermakova, Z. Rudnik-Karwatowa. Tendencje rozwojowe w słowotwórstwie języka polskiego i rosyjskiego końca XX stulecia. – *Slavia*, № 1, s. 9–18.

Колковска 2003: Колковска, С. Тенденции в композитообразуването с афиксоиди в българския език (в съпоставка с чешкия език) през 90-те год. на XX в. – In: *Dynamika a inovace v češtině a bulharštině (90. léta 20. století). Sborník příspěvků z pracovního setkání. Sofie 7. října 2002*. Praha, UJČ AV ČR, s. 60–78.

Лотко 2000: Lotko, E. *Slovník lingvistických termínů pro filology.* Olomouc, Univerzita Palackého, 128 s.

Мартинцова 1983: Martincová, O. *Problematika neologismů v současné spisovné češtině.* Praha, Univerzita Karlova. 160 s.

Мартинцова, Савицки 1987: Martincová, O., N. Savický. Hybridní slova a některé obecné otázky neologie. – *Slovo a slovesnost*, № 2, s. 124–139.

Мартинцова, Вачкова 1992: Мартинцова, О., К. Вачкова. Из чешката и българската неология. – *Съпоставително езикознание*, № 2, с. 11–18.

Мартинцова, Светла 2000: Martincová, O., J. Světlá (eds.). *Nová slovní zásoba ve výkladových slovnících.* Praha, UJČ AV ČR, 2000. 113 s.

Мартинцова 2003: Martincová, O. Obecná charakteristika lexikografického přístupu k neologizmům. – In: *Komparacja systemów i funkcjonowania współczesnych języków słowiańskich. Słowotwórstwo / Nominacja*. Opole, Uniwersytet Opolski, s. 34–39.

Мартинцова 2005: Martincová, O. Nová víceslovná pojmenování. – In: *Neologizmy v dnešní češtině*. Praha, UJČ AV ČR, s. 212–231.

Мейстржик 2000: Mejstřík, V. K zpracování nových kompozit ve výkladovém slovníku. – In: *Nová slovní zásoba ve výkladových slovnících*. Praha, UJČ AV ČR, s. 28–34.

Мравинацова 2005: Mravinacová, J. Přejímání cizích lexémů. – In: *Neologizmy v dnešní češtině.* Praha, UJČ AV ČR, s. 187–211.

Мурдаров 1983: Мурдаров, В. *Съвременни словообразователни процеси.* София, „Наука и изкуство”. 137 с.

Онхайзер 2003: Ohnheiser, I. Системно-структурные тенденции. – In: *Komparacja systemów i funkcjonowania współczesnych języków słowiańskich. Słowotwórstwo / Nominacja*. Opole, Uniwersytet Opolski, s. 317–338.

Радева 1987: Радева, В. *Българското словообразуване*. София, УИ „Св. Климент Охридски”. 165 с.

Радева 1991: Радева, В. *Словообразуването в българския книжовен език.* София, УИ „Св. Климент Охридски”. 226 с.

Рангелова 1996: Rangelova, A. Workholismus a to ostatní. – *Naše řeč*, № 2, s. 110–112.

Рангелова 1999: Rangelova, A. K dvěma česko-bulharským paralelám. – *Slavia*, № 1, s. 33–36.

РЧД 2007: *Речник на чуждите думи в българския език.* Пето издание. Допълнено и преработено от Емилия Пернишка. София, „Наука и изкуство”. 904 с.

Смирнов и др. 1998: Смирнов, Л. Н., Г. К. Венедиктов, М. И. Ермакова, Ю. Е. Стемковская. Общее и специфическое в процессах интернационализации в славянских литературных языках на современном этапе их развития. – В: *Славянское языкознание*. *XII. международный съезд славистов, Краков 1998 г. Доклады российской делегации*. Москва, „Наука”, с. 521–535.

Тиха 2003: Tichá, Z. *TERAPIE (K jednomu internacionálnímu slovotvornému komponentu). – In: *Internacionalizmy v nové slovní zásobě.* Praha, UJČ AV ČR, s. 215–219.

Ухлиржова 1998: Uhlířová, L. Zálivgate. – *Naše řeč*, № 4, s. 218–219.

Яновец, Рангелова 2005: Janovec, L., A. Rangelova. Sufixoidy a sufixoidní lexémy u substantivních kompozit. – In: *Neologizmy v dnešní češtině*. Praha, UJČ AV ČR, s. 85–98.

Яновец 2003: Janovec, L. Produktivnost některých slovotvorných komponentů v češtině a bulharštině. – In: *Dynamika a inovace v češtině a bulharštině (90. léta 20. století). Sborník příspěvků z pracovního setkání. Sofie 7. října 2002.* Praha, UJČ AV ČR, s. 79–84.

ASCS 2001: *Akademický slovník cizích slov.* Kolektiv autorů pod vedením V. Pětráčkové a J. Krause. Praha, Academia. 834 s.

LDCE 2009: *Longman Dictionary of Contemporary English.* DVD-ROM.

SSJČ 1960: *Slovník spisovného jazyka českého.* I. díl. A–M. Praha, Academia. 1311 s.

ИЗТОЧНИЦИ

РНЗДБЕ: Пернишка, Е., Д. Благоева, С. Колковска. *Речник на новите думи и значения в българския език*. София, „Наука и изкуство”, 2001. 309 с.

РНДБЕ: Пернишка, Е., Д. Благоева, С. Колковска. *Речник на новите думи в българския език (от последното десетилетие на XX и началото на XXI век).* София, „Наука и изкуство”, 2010. 515 с.

SN 1: 1998: *Nová slova v češtině. Slovník neologizmů*. Praha, Akademia, 1998. 356 s.

SN 2: *Nová slova v češtině. Slovník neologizmů 2*. Praha, Akademia, 2004. 568 s.

Český národní korpus – SYN2006pub. Ústav Českého národního korpusu FF UK. Praha 2006 – http://ucnk.ff.cuni.cz/

HPSG-based Syntactic Treebank of Bulgarian – http://www.bultreebank.org/

Neologická databáze oddělení lexikograficko-terminologického – ÚJČ AV ČR, v. v. i.

СЪВРЕМЕННИ МЕТОДИ И ПОДХОДИ
В ЛЕКСИКОГРАФСКАТА ПРАКТИКА

СБОРНИК СТУДИИ И СТАТИИ

Българска
Първо издание

Редактори и съставители на сборника:
доц. д-р Диана Благоева
доц. д-р Сия Колковска

Рецензент:
проф. дфн Лилия Крумова-Цветкова

Автори:
Диана Благоева, Сия Колковска, Индра Светла,
Албена Рангелова, Надежда Костова,
Йорданка Трифонова, Божана Нишева

ISBN 978-954-323-923-8
Авангард Прима
София, 2011