

ИНСТИТУТ ЗА БЪЛГАРСКИ ЕЗИК
БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ

БЪЛГАРСКИЯТ СЕМАНТИЧНО АНОТИРАН КОРПУС

The background of the cover is a deep blue gradient. It features a complex network of glowing blue lines and nodes, resembling a circuit board or a data network. Several overlapping, semi-transparent squares and rectangles are scattered across the lower half of the page, some containing faint grid patterns. The overall aesthetic is futuristic and technological.

Семантично
анотиран корпус
на
българския език

София 2010

Издава: Институт за български език
Българска академия на науките

Научната разработка е частично финансирана от МОМН, ФНИ,
проект ДИД 02 /19, и от ОП Развитие на човешките ресурси 2007-
2013, проект № BG051PO001-3.3.04/27

Редактор и съставител *доц. д-р Светла Пенева Коева*

- © *Светла Пенева Коева, Борислав Красимиров Ризов, Екатерина Николова Търпоманова, Мария Андреева Тодорова, Светлозара Илиева Лесева, Христина Николова Кукова, Цветана Иванова Димитрова – автори, 2010 г.*
- © *Божидар Тодоров Чемширов – художник на корицата, 2010 г.*
- © *Ангелина Владиславова Аврамова – графичен дизайн, 2010 г.*

ISBN 978-954-779-124-4 София, 2010

СЪДЪРЖАНИЕ

I. Част: Българският семантично анотиран корпус: основни принципи	5
1. Българският семантично анотиран корпус – теоретични постановки	7
<i>Светла Коева</i>	
2. Система за анотиране Chooser.....	43
<i>Борислав Ризов</i>	
3. Общи лингвистични конвенции при анотацията....	51
<i>Светлозара Лесева</i>	
II. Част: Българският семантично анотиран корпус: лингвистични конвенции	65
4. Лингвистични конвенции при анотация на съществителните	66
<i>Екатерина Търпоманова</i>	
5. Лингвистични конвенции при анотация на глаголите	82
<i>Светлозара Лесева</i>	
6. Лингвистични конвенции при анотация на прилагателните	108
<i>Христина Кукова</i>	
7. Лингвистични конвенции при анотация на наречията	119
<i>Христина Кукова</i>	
8. Лингвистични конвенции при анотация на числителните	126
<i>Светлозара Лесева</i>	
9. Лингвистични конвенции при анотация на местоименията	133
<i>Екатерина Търпоманова</i>	
10. Лингвистични конвенции при анотация на затворените класове.....	141
<i>Цветана Димитрова</i>	
11. Съставни единици в Българския семантично анотиран корпус.....	166
<i>Мария Тодорова</i>	
ПРИЛОЖЕНИЕ	186

I. Част

Българският семантично аотиран корпус: основни принципи

Българският семантично аотиран корпус е част от Българския Браун корпус, където на всяка лексикална единица е приписано коректното значение посредством съотнасяне с подходящото синонимно множество от Българския wordnet. Създаването на семантично аотиран корпус изисква решаването на множество теоретични и практически проблеми, свързани с интерпретацията на избраното множество от лексикални единици и съществуващите общотеоретични постановки и частноезикови решения за български език.

В първа част са описани някои от основните принципи при съставяне на анотационната схема, изборът на езиков ресурс за семантична анотация, базовите лингвистични конвенции и програмата за семантична анотация.

Създаването на Българския семантично аотиран корпус се съпътства с проверка, корекция и обогатяване на информацията в Българския wordnet и изисква приложение и преосмисляне на лингвистичните конвенции за анотация, следователно и на теоретични постановки в българската граматична традиция.

Предназначението на Българския семантично аотиран корпус е автоматично отстраняване на семантична многозначност – асоциирането на произволна лексикална единица от текста с най-коректното за дадения контекст значение. Наред с това Българският семантично аотиран корпус представлява източник на езикови данни за теоретични изследвания в областта на граматиката, стилистиката, лексикографията, корпусната лингвистика.

1. Българският семантично аотиран корпус – теоретични постановки

Светла Коева

1.1. Въведение

Най-трудните за компютърна обработка свойства на езика произтичат от неговата многозначност и неопределеност на различни равнища. Буквата ‘з’ в различен контекст може да означава различни звукове – пред гласна и пред звучна съгласна – ‘з’, а в краесловие и пред беззвучна съгласна – ‘с’ (така звуковата форма на думата *коз* има поне две омонимни значения). Ще има и много варианти на неопределен изговор – между ‘з’ и ‘с’. Думата *под* може да бъде предлог или съществително (в словосъчетанието *кутията е под леглото* е предлог, но в изречението *Не стъпвай по прясно боядисания под* е съществително, а в словосъчетанието *под огледало* може да бъде както предлог, така и съществително, като неопределеността може да се разреши в по-голям контекст. В изречението *Купих книга от автора на детски книжки* може да се различи синтактична структура – именна група към глагол, със съответно значение ‘*купих книга, написана от автора на детски книжки*’ или структура – именна и предложна група към глагол, със съответно значение ‘*купих книга, която ми продаде авторът на детски книжки*’.

Проблемът става още по-сложен, когато се намесят добре известните феномени омонимия – множество несвързани помежду си значения на дума, която се пише и/или произнася еднакво, и полисемия – множество „свързани” помежду си значения на дадена дума. В словосъчетанието *голям град* не е ясно дали думата *град* означава ‘*голямо селище*’ или ‘*градушка*’, но омонимите се различават ясно в изреченията *Построиха голям град* и *Падна голям град*. В словосъчетанието *тънка глазура* също не е ясно кое от значенията на многозначната дума *глазура* е употребено – ‘*Тънък стъкловиден слой, с който се покриват керамични, порцеланови и др. изделия*’ или ‘*Покритие от захарен сироп, шоколад и др. върху сладкиши или плодове за украса и вкус*’, за разлика от употребата ѝ в словосъчетанията *водоустойчива глазура* и *захарна глазура*.

Многозначността и неопределеността на езика не затрудняват езиковото общуване – те са негови свойства. При компютърната обработка на естествения език обаче разрешаването на многозначността и неопределеността представлява сериозен научен проблем, поради факта, че не съществува пълен и непротиворечив модел на езиковото общуване.

Процесът на (автоматично) разрешаване на многозначност може да бъде описан най-общо като избор в определен контекст на най-коректното значение (в широк смисъл) за даден езиков израз от множеството възможни значения. Отстраняването на семантична многозначност, съответно, е асоциация на дадена дума от текста (единична или съставна) със значение, различимо сред множеството от значения, потенциално свързани с тази дума.

Така дефиниран процесът на отстраняване на семантична многозначност поставя няколко въпроса: 1) какво се приема за значение на думата и дали и по какъв начин значението може да се дефинира извън контекста; 2) кои са значенията на дадена дума (и как се формулират) и 3) кое е правилното значение, което се асоциира с дадена дума в даден контекст. Първият въпрос е не само лингвистичен и към момента е извън обекта на изследването – приемаме, че значението на дадена дума се изразява комплексно посредством уникалното множество от семантични релации с други думи в езика. Проблемът за определянето на множеството от значения, които могат да се асоциират с дадена дума, също не е част от изследването – ако е необходимо, се добавят нови значения, без стремеж за пълна изчерпателност. Следователно задачата, чието решение се търси в рамките на създаването на Семантично аотирания корпус на българския език, е *кое е правилното значение, което се реализира при употребата на дадена лексикална единица*. Под лексикална единица се разбира както дума (графична дума, обозначаваща уникално и константно понятие), така и съставна дума (две или повече думи, не задължително последователни, обозначаващи уникално и константно понятие, в релация на еквивалентност с (графична) дума от същия език или друг език) (Коева 2006б).

Българският семантично аотиран корпус (БулСемКор) представлява извадка от Българския Браун корпус, в която всяка лексикална единица е аотирана с подходящо за употребата си значение от Българския wordnet (БулНет). В контекста на формулирането на теоретичните постановки на изследването накратко се разглеждат равнищата на лингвистична аотация, прави се сравнение с подобни проекти за други езици, описва се методологията за съставяне на изходния корпус, разглеждат се основните подходи за представяне на значение с техните предимства и недостатъци, коментират се основанията за избор на Българския wordnet като източник на значения за семантична аотация, излагат се основните принципи, които са приети при съставянето на аотационната схема, и се описват получените резултати и някои от възможностите за приложение.

1.2. Равнища на лингвистична анотация

Една от дефинициите за корпус е „*съвкупност от автентични езикови данни, които могат да се използват за лингвистични изследвания*” (Лиич 1997:1). Подобно е и определението за корпус като „*колекция от езикови примери, които са избрани и подредени според експлицитни лингвистични критерии, за да се използват като модел на езика*” (ЕГСКОЕ 1996а:4). С развитието на компютърната лингвистика областите на приложение на корпусната лингвистика се увеличават, затова по-скоро е приложима дефиницията: „*съвкупност от езиков материал в електронна форма, подходящ за компютърна обработка с приложение в лингвистичните изследвания или езиковите технологии*” (Лиич 1997:1). Или „*компютърен корпус е корпус, кодиран по стандартизиран и хомогенен начин, даващ възможност за извличане на информация от различен характер*” (ЕГСКОЕ 1996а:5). Като компилация на многобройните дефиниции за корпус може да се предложи следното определение:

Корпус е голяма колекция от езикови примери, представени по начин, даващ възможност за компютърна обработка, и избрани по определени (лингвистични) критерии, така че да представляват адекватен езиков модел.

Както се отбелязва, аотирианият корпус съдържа допълнителна „*тълкувателна, най-вече лингвистична информация*” (ЕГСКОЕ 1996б:3). Могат да се отделят различни равнища на лингвистична анотация (Лиич 1997:8-15), например: морфологично, морфо-синтактично, синтактично, семантично и дискурсно (ЕГСКОЕ 1996б:3), като аотирианият корпус обикновено се асоциират с повече от едно равнище (такъв е и случаят със Семантично аотириания корпус на българския език).

На морфологично равнище всяка лексикална единица в текста може да се асоциира със своята основна форма – лема (лематизация). Анотацията на лемата за флективни езици като български, където преходен глагол от несвършен вид има петдесет и две синтетични форми, е важна част от компютърната обработка, за да може се осъществява връзката при употребата на различните форми на дадена дума в различни контексти. В Българския семантично аотириан корпус посредством автоматична лематизация всяка лексикална единица (независимо в каква форма е употребена) се асоциира с подходящите синонимни множества от БулНет, където литералите (лексикални единици, свързани с релация на еквивалентност) също са съотнесени с основната си форма, например <LITERAL>продумвам<LEMMA>продумвам</LEMMA></LITERAL>.

Морфо-синтактичната анотация въвежда информация за граматичния клас на дадена лексикална единица и (възможно) съответните

стойности на граматичните категории, характеризиращи единицата. Морфо-синтактичната анотация отстранява граматичната многозначност на омографите и се използва в голям брой приложения за компютърна обработка на езика: за определяне на коректната фонетична стойност на дадена форма, за съотнасяне на формите към правилната парадигма, за идентификация на съставните лексикални единици, за определяне на строежа на фразите, за съотнасяне с допустимо лексикално значение и т.н. Семантично аотирианият корпус на българския език съдържа анотация за съответната част на речта, към която принадлежи дадена лексикална единица, представена като част от идентификационния номер на избраното синонимно множество от wordnet, например ENG20-01543395-n.

Синтактичната анотация най-общо може да представя граматичните зависимости между конституентите на словосъчетание или фразовата структура, в която конституентите се комбинират помежду си. В Българския семантично аотириан корпус синтактична анотация няма, освен частична информация за групирането на компонентите на съставните думи в една лексикална единица и определяне на главната ѝ част, например аналогичната информация в БулНет <SYNONYM><LITERAL> *министър* /на/ /правосъдието/<SENSE>1</SENSE><LEMMA>*министър* на *правосъдието*</LEMMA></LITERAL><LITERAL>/*правосъден*/ *министър*<SENSE>1</SENSE><LEMMA>*правосъден* *министър*</LEMMA></LITERAL></SYNONYM>.

При семантичната анотация могат да се разграничат два типа (МакЕнери и Уилсън 2001:61-62) – представяне на семантичните релации между думите в текста (например анотация на семантичните роли на аргументите, на валентността на лексикалните единици, на реализацията на семантичните фреймове) или на семантичните свойства на думите (например анотация на значението на думите). Представянето на семантичните свойства на думите може да асоциира дадена дума с нейното значение, но може да има и по-комплексен подход, при който думите се съотнасят със „семантичното поле”, към което принадлежат (Уилсън и Томас 1997:54) – множество от думи с общи семантични свойства, например непосредствените хипоними. В Семантично аотириания корпус на българския език се представя вторият тип семантична анотация – всяка лексикална единица е свързана със синонимното множество от Българския wordnet, към което принадлежи (по този начин и с експлицитно и имплицитно изразяваното от него значение). Може да се твърди, че в Семантично аотириания корпус на българския език има семантична анотация и към семантично поле – посредством връзката със съответното синонимно множество може да се идентифицират връзките със семантично свързаните надредни думи – хипероними (а чрез тях и

с хипонимите от същия или по-нисък ред), както и екстралингвистични релации – например *тематична област* и т.н.

Дискурсната анотация може да включва определяне на референцията на местоименията за отстраняване на многозначност при анафора, маркиране на изрази за изразяване на определена конвенция – извинение, поздравление и др.

Както се вижда, равнищата на анотация съответстват най-общо на равнищата на многозначност, от което следва, че една от основните области на приложение на анотираните корпуси е отстраняването на многозначността. Разбира се, анотацията е необходима и при еднозначните структури за съотнасянето им с коректните лингвистични стойности, но при многозначните структури анотацията показва не само коректните лингвистични стойности, но и избора им от множеството възможни. Семантично анотираният корпус на българския език съдържа следните видове лингвистична анотация като компонент или необходимо условие за семантична анотация: морфологична анотация (основна форма), морфосинтактична анотация (част на речта), синтактична анотация (граница на съставни думи и главна част), и самата семантична анотация (синонимно множество от БулНет).

1.3. Подходи при семантична анотация на значения

При досегашната практика са известни два основни начина за създаване на семантично анотирани корпуси с анотация на лексикални значения – чрез избор от множеството значения, които се предлагат в даден езиков източник – тълковен речник, тезаурус, лексикално-семантична мрежа (wordnet) (Фелбаум и др. 2001), или чрез превод на семантично анотиран корпус от един език на друг.

При създаването на БулСемКор основно е следвана методологията, приета при изработването на Английския семантично анотиран корпус СемКор (SemCor), създаден в Принстънския университет (Ландерс и др. 1998). С термина *семантичен конкорданс* се означава асоциацията между корпус и лексикон, така че всяка дума в текста да е свързана с подходящо значение от лексикона (Милър и др. 1993:303-308; Милър 1995: 92-94). Според тази дефиниция *семантичен конкорданс* е или корпус с текстове, в които семантичната многозначност е отстранена, като думите са анотирани семантично, или лексикон, в който към всяка тълковна дефиниция има примерни изречения.

СемКор включва част от американския Браун корпус и един допълнителен текст – новела (Ландерс и др. 1998:199-200), като се състои от 352 текста и съдържа около 700 000 думи. Думите, принадлежащи към

четирите части на речта – съществително, прилагателно, глагол и наречие, са ръчно анотирани със значения от Принстънския wordnet с помощта за специално конструирана за целта помощна програма (първоначалната цел на създаването на анотирания корпус е да се подкрепи или отхвърли представянето на значенията в wordnet). В 186 текста (359 732 думи) думите от отворените класове (192 639 думи) са тагирани за част на речта, лематизирани и семантично анотирани. В останалите 166 текста (316 814 думи) са анотирани само глаголите със съответната лема и значение (41 497 глаголни срещания). Използваната процедура е следната: ако думата има само едно значение – значението се проверява и ако е необходимо, се коригира; ако думата има повече от едно значение – задачата е да се избере правилното значение от възможните (същата процедура се използва при анотирането и на Българския семантично анотиран корпус).

Подобна методология се прилага в различни проекти за семантична анотация. Например в корпус със семантична анотация, базирана на wordnet (Нг и Лее 1996:44), най-често срещаните многозначни съществителни и глаголи в английски (121 съществителни и 70 глагола) са анотирани с коректното си значение в 192 800 изречения от Браун корпус и списанието *Уол стрит*.

В корпуси от около 100 000 думи за испански и каталански и от около 50 000 думи за баскски се предлагат три нива на анотация – синтактична, семантична и прагматична (Агире и др. 2006; Наваро и др. 2003). На семантично равнище анотацията се извършва спрямо синонимните множества от wordnet за съответните езици, като се анотират само съществителните, глаголите, прилагателните и наречията (причината за по-малкия корпус на баскски е по-малкият wordnet – около 35 000 единици, спрямо около 100 000 за испански и 65 000 за каталански). Особеното е, че се приема асоциация с повече от едно синонимно множество, ако подобна интерпретация се допуска от контекста.

За италиански (Монтемани и др. 2000) в корпус от около 300 000 думи (наред със синтактична анотация) се анотира значението на съществителни, глаголи и прилагателни и съответните типове съставни думи спрямо италиански wordnet. Особеното е, че се включва и допълнителна лексикално-семантична информация, например при случаите на преносна употреба, както и информация за процеса на анотация, например бележки за проблематичните случаи.

За шведски в корпус от около 1 000 000 думи, организирани по модела на Браун корпус, са анотирани над 250 000 думи със значения от Лексикалната база от данни Готенбург (Йаеборг и др. 2002:1494). Първоначалната цел – да се провери хипотезата дали е възможно значенията на думите да се опишат по такъв начин, че да се предвиждат еднозначно

при реалната си употреба, прераства в създаването на достатъчно голям семантично аотиран корпус за нуждите на компютърната обработка. Особеното е, че източникът за семантична аотация не е wordnet.

За китайски думите от корпус (Бу и др. 2006) се реферират със значения, които се дефинират специално за целите на аотацията, като всяко значение е свързано и със съответното синонимно множество от лексикално-семантична мрежа, подобна на wordnet. Към момента на написването на статията 813 съществителни и 132 глагола са включени в специално конструирания речник и 60 895 техни срещания в текстове от *Пипълс дейли* са семантично аотирани. Особеното е, че семантичната аотация се извършва едновременно със създаването на речника – ресурс за аотация.

Научната инициатива SENSEVAL (Evaluation Exercises for the Semantic Analysis of Text – Тестове за оценка на семантичен анализ на текст), която се провежда периодично, разполага с няколко (повечето за английски) семантично аотирани корпуса (спрямо различни версии на wordnet или тълковни речници – по-рядко) за проверка и оценка на системи за автоматично отстраняване на многозначност (Килгариф и Розенцвайг 2000:18). Особеното е, че в някои от тях са аотирани само определени думи.

По втория възможен начин – чрез превод – е създаден паралелният семантичен корпус МултиСемКор (MultiSemCor) (Пианта и Бентиволи 2003; Бентиволи и Пианта 2005). Хипотезата, която е в основата на този подход, е, че при превод от един език на друг семантичната информация до голяма степен се запазва. Следователно, семантичните аотации могат да се прехвърлят от изходния текст към преводния му неанотиран еквивалент, като се използва автоматично съотнасяне по думи. МултиСемКор се състои от 116 текста от СемКор и техните еквиваленти на италиански. Текстовете са преведени от английски на италиански от професионални преводачи, след което са автоматично съотнесени на равнище дума и английската семантична аотация е пренесена към италианските им преводни еквиваленти (срещу 119 802 семантично аотирани думи на английски има 92 820 прехвърлени значения на италиански – няма пълно съответствие не само по лингвистични причини, но и поради точността на автоматичното съотнасяне по дума).

Подобен подход е използван и при създаването на семантично аотиран корпус за румънски (Лупу и др. 2005). 34 английски текста от СемКор, състоящи се от 65 926 думи, са преведени на румънски. Получените паралелни текстове са съотнесени на равнище дума и английската семантична аотация е пренесена към румънските преводни еквиваленти. Особеното е, че се прилага метод за частично разпознаване

на съставни думи или фразеологични изрази като една лексикална единица и обратно – сегментация на лексикална единица, ако е приложимо.

Основната трудност (а оттам и недостатък) при трансферния подход е автоматичното съотнасяне по думи между паралелни текстове на различни езици – известно е, че няма еднозначна кореспонденция между единични и съставни думи (фразеологични изрази), между морфологични и синтактични конструкции, както и между степените на лексикализация, а съвсем отделно е субективното влияние на преводача при избора на лексикални, граматични и стилистични (прагматични) средства.

В обобщение известните семантично анотирани корпуси за различни езици се характеризират със следните белези:

- Изходните корпуси са малки по обем, в много случаи това е Браун корпус или корпуси, съставени по негов модел, в други случаи това са периодични публицистични издания.
- Семантичната анотация се извършва от човек, в повечето случаи от експерти, което предопределя както очакваното добро качество, така и малкия обем на анотираните корпуси и продължителността на процеса (разбира се в някои случаи се прилагат различни техники за оптимизиране на анотацията).
- Доминиращият ресурс за семантична анотация е wordnet, но има и примери, когато анотацията се извършва спрямо друг речник или тезаурус, който съдържа значения. Wordnet е предпочитан в сравнение с останалите възможни източници за семантична анотация поради структурата, големината и достъпността си.
- Една от основните особености при известните примери за семантична анотация е, че се анотират само глаголите, съществителните, прилагателните и наречията, а в някои случаи се избират само дадени (честотно определени) думи от тези класове – това ограничение се налага от структурата на wordnet, където са включени само изброените части на речта.
- В повечето случаи семантичната анотация се комбинира с друг тип анотация – обикновено морфологична (лематизация) и морфо-синтактична (част на речта), но в някои случаи и със синтактична или дискурсна анотация.
- Обикновено при семантичната анотацията се избира само едно подходящо значение, но в някои подходи се допуска и повече от едно, ако значенията са дефинирани твърде гранулирано и са подходящи за интерпретация в контекста, както и допълнителни бележки за степен на увереност и колебания на анотаторите.

Българският семантично аотиран корпус предлага експертна семантична анотация на думите, които се срещат в извадка от Българския Браун корпус. В нея всяка дума се асоциира с едно синонимно множество от БулНет, представящо най-достоверно значението, реализирано в съответния контекст. Семантичната анотация се комбинира с морфологична, морфо-синтактична и частична синтактична анотация.

1.4. Съставяне на изходния корпус

Българският Браун корпус (Коева и др. 2006) е създаден в съответствие с методологията на известния Браун корпус, която представлява добре балансиран модел за синхронното състояние на даден език. По тази причина Браун корпус е по-подходящ за семантична анотация, отколкото произволна колекция от текстове – например поредица от броевете на даден вестник. Българският Браун корпус съдържа 1 001 286 думи и включва 500 текста с дължина от около 2 000 думи, разпределени в 15 категории от два типа – художествени и информативни. Корпусът обхваща текстове от интернет от периода 1990-2005 г., предимно след 2000-та година. При съставянето на Българския Браун корпус са пренебрегнати някои от принципите на Браун корпус (всички текстовете да са оригинални, да са съвременни и др.), поради невъзможност да се илюстрират всички категории с електронни публикации.

Изходният корпус за семантична анотация запазва оригиналната структура на Българския Браун корпус, като съдържа извадка с приблизително 100 думи от всеки един от 500-те текста (извадките се разширяват наляво и дясно до край на изречение, по тази причина броят на думите в тях не е точно 100).

Извадките в първата част от изходния корпус са избрани според критерий за добра балансираност на гъстотата на срещанията на думите от отворените класове. За да се осигури оптимален брой пълнозначни думи, е съставен честотен речник на основните форми от отворените класове, базиран на два корпуса с отстранена граматична многозначност: 71 876 думи от българския превод на *1984* на Оруел (Ерявец 2004) и 328 964 думи от три тематични области – икономика, право и политика (общо 400 840 думи). Търсят се тези извадки от Българския Браун корпус, които съдържат най-голям брой думи от конструирания честотен речник. За да се постигне по-добър баланс в пропорцията между съществителни и глаголи спрямо прилагателни и наречия, на думите в честотния речник се приписват тегла, както следва: 0.4 на съществителните, 0.3 на глаголите, 0.2 на прилагателните и 0.1 на наречията. За да се осигури максимално разнообразие от думи в извадките, се приписват допълнителни тегла

на думите в честотния речник, така че най-рядко срещаните думи в Българския Браун корпус и БулНет получават тегло с най-висока стойност. Процедурата по селекция включва няколко експеримента с последователно изключване от честотния речник на думите с най-ниска честота и най-висока стойност на относително тегло. Извадките от Българския Браун корпус, които имат най-добро лексикално покритие, разбирано като най-голям брой различни лемми в комбинация с най-голям брой кореспондиращи значения в БулНет, са избрани за включване в първата част на изходния корпус за семантична анотация.

При селекцията на втората част от изходния корпус за семантична анотация също се запазва оригиналната структура на Българския Браун корпус, като се селектират части от приблизително по 100 думи от всеки един от 500-те текста. Критериите, които са заложили при избора на втората част от изходния корпус, са: максимален брой от думите в извадките да не се срещат в БулНет, както и максимален брой от думите в извадките, които се срещат в БулНет, да участват в повече от едно синонимно множество. Целта е да се създаде изходен корпус, който, от една страна – да съдържа думи, чиито значения не са били анотирани в първата част, а от друга страна – да съдържа максимален брой многозначни думи, така че да се илюстрира контекстният избор на различни значения.

Изходният корпус се състои от 811 извадки, в които се съдържат 101791 токъна (символи между спейсове), които от своя страна представляват 95119 лексикални единици (от които 89341 думи и 5778 съставни думи). Изходният корпус е подложен на предварителна обработка – нормализация (например отстраняване на някои печатни грешки) и лематизация (автоматично съотнасяне с основна форма), като връзката между първоначалното състояние и променените варианти е възстановима и може да послужи за създаването на правила за автоматична нормализация.

Макар че голямо количество от известните семантично анотирани корпуси използват Браун корпус или корпуси, създадени по неговата методология, доколкото ни е известно, не е прилагана евристика при определянето на това, кои части от корпуса да бъдат анотирани – напротив (представители на) отворените класове думи се анотират последователно в текстовете.

1.5. Значение на думата

(Автоматичното) отстраняване на многозначност представлява асоцииране на дадена лексикална единица с подходящо значение сред множеството на възможните значения, които са свързани с нея. Това

предполага ограничението, че значението на словосъчетание или изречение се базира само на едно от значенията на дадена многозначна дума, употребена в тях, и че няма интерпретация на значението в зависимост от екстралингвистични фактори, както и изискването за източник, в който всички възможни значения на дадена дума да са описани пълно и непротиворечиво. Разбира се, нито ограничението е валидно (за определяне на правилното значение са важни както контекстът, така и екстралингвистичните фактори), нито изискването е удовлетворено (няма пълен и непротиворечив езиков ресурс с описание на значенията за даден език), така че задачата се свежда до това да се направи най-адекватният избор сред съществуващите възможности (изборът на езиков ресурс, който съдържа множество от дефинирани значения, формулирани по определен начин, в някаква степен предопределя резултата).

При дефинирането на значението на думата могат да се обособят няколко основни подхода: на традиционните тълковни речници, на теориите, базирани на примитиви, и на релационните теории. Традиционните тълковни речници представят значението посредством дефиниции, които трябва да интерпретират уникално тълкуваното понятие (Фонтенел 2008). Теориите, базирани на примитиви, приемат, че значението може да се дефинира с помощта на фиксирано множество от примитивни елементи (Катц 1972), като различията основно се свеждат до множествата от примитиви, до които могат да се декомпозират значенията. Според релационните теории значението се представя посредством експлицитно изразени релации между думите (и съответните понятия) (Фодор 1975), категоризирани посредством множества от свойства.

1.5.1. Традиционни подходи при дефиниране на значението

В тълковните речници се приема, че значението на думите може да се дефинира посредством (подредено) множество от значения за дадена дума, кито нямат сечение. Дефинициите се състоят от генерализация (класа, към който принадлежи лексикалната единица) и диференциация (какви свойства различават тази дума от останалите членове на класа). Различните речници включват различни множества от значения – дадено значение в един речник може да е разделено на няколко значения, да е обединено с друго значение или изобщо да не фигурира в друг речник. Множеството от значения зависи от предназначението на речника, от избрания словник (може да се основава на честотата на срещанията или на „важността”) и от възприетото равнище на гранулираност (например дали абстрактната и конкретната употреба на дадена дума (Ханкс 2000:207) – *‘банка като финансова институция’* и *‘банка като сграда на финансова институция’* –

трябва да се интерпретират като отделни значения), като в много случаи решението за по-общо или по-детайлно представяне на конкретните значения е субективно мотивирано (Килгариф 1997:100).

Използването на традиционните речници за отстраняване на многозначност се затруднява и от интерпретацията на съставните думи (основно при съществителните) и алтернациите (основно при глаголите). Представянето на съставните думи (например *кръвна банка, зърнена банка, банка за стволони клетки*) може да бъде включено към значението на главната дума, в случая *'банка като финансова институция'* (Ханкс 2000:207), да бъде съотнесено с по-обща дефиниция, която се отнася за група съставни единици, например *'всяка една от институциите за складиране и съхраняване на различен тип обекти'*, или съставните думи може да се разглеждат като отделни лексикални единици със собствени значения. Макар че най-последователният подход към съставните думи е да се интерпретират като отделни лексикални единици, практика в традиционните речници е да се игнорират или да се изброяват към речниковата статия на главната дума, без да се тълкуват.

Глаголните алтернации – промяна на реализацията на глаголните аргументи по отношение на една базова структура (Коева 2006а; 2008) – биват диатези (редуцира се семантичната релация на елемента, който кореспондира със субекта в неутрално изречение, съпроводено с промяна на лексикалното значение на глагола, както и на синтактичните категории и функции при реализацията на (част от) останалите елементи); семантични алтернации (редуване на семантични релации на елементи, което води до (възможна) промяна на синтактичните им категории и функции) и синтактични алтернации (промяна на синтактични категории и функции). Диатезите са включени частично и непоследователно в тълковните речници (подобно на системните деривации) (Коева 2008). Възможно обяснение за това е, че диатезите са закономерни редувания, които могат да се предвидят, ако изходната структура удовлетворява определени изисквания. Системността обаче не променя факта, че диатезите представят различни значения и трябва да бъдат описани по подходящ начин.

Наблюдават се две основни тенденции в лексикографията – да се диференцират повече значения, за да се обхванат всички случаи на употреба, или да се търси едно обединяващо по-общо значение за всички употреби. В първия случай е възможно да няма ясна разлика между множеството диференцирани значения, които могат в някаква степен да се припокриват, а от друга страна, обобщените значения обикновено не са достатъчно конкретни. Фундаменталният въпрос дали съществува значение на думата (Ханкс 2000:205) може да се преформулира по

следния начин – дали значението на думата е това, което се формулира в тълковните речници, или тълковната дефиниция може да се разглежда като част от значението на думата.

Самите дефиниции, дори да спазват добре познатите модели от общо (хипероним) към частно, да се придържат към формални модели за описание и свързване на компонентите си, не представляват достатъчно непротиворечиво описание, подходящо за автоматичен анализ. Следователно разделянето на значения, така както е прието да се представят в тълковните речници, не е подходящо при компютърна обработка на езика (Килгариф 1997:108). Речниците обикновено са създадени да реферират към определени значения и тяхното предназначение не е семантична анотация – в много от случаите значенията от речниците не съответстват на значенията на думите при тяхната реална употреба (Фелбаум и др. 2001).

Корпусната лингвистика се стреми да намали субективността на традиционните подходи при дефиниране на значенията по отношение на гранулираността, като изследва езикови данни за множество срещания на дадена лексикална единица в големи по обем корпуси (Синклер 1987), а значението на думите в речниците се описва след анализ и синтез на информация от контекста (Аткинс 1993). В етапа на анализ трябва да се намерят колкото може повече релевантни лингвистични факти за значенията на дадена дума (Аткинс 1993:33) чрез изследване на употребата на думата в корпус, в съществуващи архиви и речници (които също може да се разглеждат широко като конкорданс). Данните се разделят така, че лексикалните единици в дадена група примери да имат общо значение и да се различават от значенията в останалите групи, като за всяка група се определя какво я обединява. В етапа на синтез направените заключения се обобщават в ограниченията на езика на тълковните дефиниции.

Подобен, но ориентиран за целите на компютърната лингвистика, е подходът за отстраняване на семантична многозначност, при който основна единица не е значението на думата, а срещанията на думата в контекст. В най-простия случай примерите за употреба в корпус попадат в една или повече различни групи и всяка група, ако е достатъчно голяма и достатъчно различна от останалите групи, оформя ново значение (Килгариф 1997:108) – значенията трябва да се разбират като абстракции по отношение на групи от примери за употребата на дадена дума.

Според някои автори думата има различно значение, само когато всяко значение има различен превод на друг език (Ресник и Яровски 2000), така че проблемът за идентификация на значението се свежда до това да се намери преводен еквивалент – такъв подход е приложим само при автоматично определяне на значението, базирано на паралелни корпуси,

и в него се мултиплицират недостатъците от автоматичното съотнасяне по изречения и лексикални единици.

1.5.2. Теории за значението, базирани на примитиви

Най-общо семантичните примитиви се разбират като единици, които се използват, за да се обяснят понятията, но от своя страна не подлежат на тълкуване с помощта на други единици. Стремещът е да се дефинират думите посредством ограничен език за репрезентация. Положителните страни са системното представяне на дефинициите по непротиворечив начин и избягване на цикличността, а ограниченията на подхода могат да се обобщат като липса на единно мнение по отношение на примитивите (семантичния език) и невъзможност да се представят някои семантични разлики (Джакендоф 1990; Вежбицка 1996).

Подобен е подходът, при който думите се дефинират посредством вероятностни семантични компоненти – отделни, комбиниращи се, възможни за употреба единици, които формират едно или друго значение.

Приема се, че извън контекста на събитието, в което се проявява значението, думите се характеризират по-скоро с потенциали за значения, отколкото със самите значения (Ханкс 2000:211). Потенциалът на значението на дадена дума се разглежда като множество от прототипни и вероятностни семантични компоненти, като при употреба в контекста се активират само някои от тях. Например значението на думата *банка* може да се представи по следния начин (Ханкс 2000:209):

- Е ИНСТИТУЦИЯ • Е ГОЛЯМА СГРАДА • ЗА СЪХРАНЕНИЕ • ЗА ПАЗЕНЕ • НА ФИНАНСИ/ПАРИ • ОСЪЩЕСТВЯВА ТРАНЗАКЦИИ • СЪСТОИ СЕ ОТ ПЕРСОНАЛ ОТ ХОРА

При всяка употреба на думата се реализира поне един от тези компоненти, по-често комбинация от тях – вероятностният компонентен подход дава възможност да се отчете не само многозначността, но и неопределеността. Компонентите могат да се разглеждат като свързани в мрежа, която осигурява потенциал за създаване на нови значения (Ханкс 2000:214). От такава позиция на отстраняването на многозначност може да се гледа като на конкуренция между различни компоненти в множеството от възможни – различните употреби активират различни комбинации на компоненти.

По отношение на дефинирането на значението на думата, вероятно това е най-адекватният подход, който, комбиниран с организирането на думите в лексикално-семантични мрежи от типа на *wordnet*, би предоставял продуктивен модел за компютърна обработка, включително при (автоматичното) отстраняване на многозначност. Към момента дори за

английски няма практическа реализация на лексикон, в който потенциалът от значенията на думите се дефинира като прототипни вероятностни семантични компоненти. Освен това проблем ще представлява не само формулирането на прототипните компоненти, но и на процедурите за съотнасяне на дадена форма с множеството от компоненти, които са валидни за нея в даден контекст.

В някаква степен концептуално към прототипния вероятностен модел може да се причисли и така нареченият генеративен лексикон, в който семантично свързаните значения не се изброяват, а се генерират посредством правила, които описват зависимостите при образуването на значения. Контекстът на дадено изказване се разглежда като съвкупност от множество генеративни фактори, които оказват влияние върху начина, по който може да бъде разбрано изказването (Пустейовски 1995: 410-413). Вместо предефинирано множество от значения се предлага комплекс от репрезентативни равнища, които дефинират семантичните граници, подходящи за употребата на дадено значение (Пустейовски 1995: 419-436), както следва:

Лексикално-типизираща структура – определя начините, по които дадена дума се свързва с останалите думи (понятия) в структурирана система, и осигурява експлицитна връзка към общото знание за света (би могло да се разглежда като аналог на wordnet).

Аргументна структура – свързва синтактичната реализация на думите с броя и вида аргументи, които са идентифицирани.

Събитийна структура – идентифицира типа събитие за даден глагол или фраза: състояние, процес или преход.

Qualia структура – дефинира съществените свойства на обектите, събитията и релациите, които се асоциират с лексикалната единица.

Четирите равнища се съотнасят посредством множество от генеративни правила (субселекция, промяна на типа и съ-композиция), за да се осигури композиционната интерпретация на дадена дума в контекста. Или с други думи, генеративният лексикон се разглежда като генеративна система, където семантичната интерпретация на думата се изразява с логически операции, дефинирани посредством добре формулирани семантични правила. Съществуват и правила за композиция, които позволяват включването на тази информация в семантичната интерпретация на по-големи изрази. Подходът към контекстно обусловено представяне на значението също не може да се прилага към момента при (автоматичното) отстраняване на семантична многозначност по простата причина, че няма практическа реализация за конкретен език.

1.5.3. Релационни теории за значението

Най-популярното приложение на схващането, че значението се представя посредством експлицитно изразени релации между думите (и съответните понятия), е wordnet. Принстънският wordnet (Милър 1990; Фелбаум 1998) представлява лексикално-семантична мрежа, чиито възли са синонимни множества (наречени синсети), които съдържат думи или съставни думи (наречени литерали), а дъгите изразяват семантични, морфо-семантични словообразователни и екстралингвистични релации между обектите, разположени във възлите. Значението на литералите, както и значението на синонимните множества се разглежда като реализация на езиково независимите концепти. При семантичната анотация под значение на дадена дума се разбира значение на синонимно множество. Значението на лексикалните възли в wordnet се изразява посредством релациите към другите възли в мрежата, от една страна, и чрез свойствата на самия възел (имплицитно чрез синонимната релация между единиците в синонимното множество и експлицитно чрез тълковното значение и примери за употреба), от друга страна. В Принстънския wordnet са включени два типа релации: между литерали (наречени лексикални) и между синонимни множества (наречени семантични). Семантичните релации се отнасят за всички литерали в двете синонимни множества, които свързват. Лексикалните релации свързват отделни думи в рамките (по-рядко) на едно или две синонимни множества. Противопоставянето между литерална релация и релация между синонимни множества е по-подходящо, тъй като литералните релации също могат да бъдат семантични (Коева 2007а). Релациите между синонимни множества са семантични, морфо-семантични и екстралингвистични, а релациите между литерали – семантични и деривационни. Семантичните, морфо-семантични, словообразователните и екстралингвистичните релации в wordnet отразяват реални отношения (между множества от обекти или абстрактни същности) от действителността.

Синонимните множества в Принстънския wordnet включват литерали, принадлежащи към четири части на речта: съществително, прилагателно, глагол и наречие, така наречените отворени класове са включени частично (числителните са класифицирани като прилагателни или съществителни – по-рядко, някои местоимения също са включени като съществителни, прилагателни или наречия).

Wordnet е един от най-пълните лексикални ресурси (за сравнение литералите в Българския wordnet са много повече от словника на стандартен правописен речник), а синонимните множества от различни езици са свързани посредством междуезикова релация на еквивалентност,

чрез която се образува многоезиковата лексикално-семантична мрежа wordnet.

Wordnet комбинира качествата на съществуващите езикови ресурси: съдържа дефиниции и примери, както обикновените речници, но наред с това организира синонимните множества в концептуална мрежа посредством семантичните релации, които съществуват между тях. Не случайно повечето семантично анотирани корпуси и повечето системи за автоматично отстраняване на семантична многозначност използват wordnet като семантичен ресурс – при определянето на правилната семантична анотация се има предвид принадлежността на дадена лексикална единица към определен възел от лексикално-семантичната мрежа, тоест не само дефинираното тълковно значение, но и всички семантични, морфо-семантични, словообразователни и деривационни релации, в които участва, примерите за употреба, бележките към литерала или синонимното множество, с други думи – комплексното място на лексикалната единица по отношение на останалите лексикални единици (и понятия) от езика.

В структурата на wordnet думите и съставните думи са включени равнопоставено, тоест преодолян е един от недостатъците на традиционните речници съставните думи да не се описват последователно. Същото може да се твърди в някаква степен и за алтернациите – wordnet е структура, в която алтернациите имат своето място (Кол и др. 1998), макар че към момента са представени ограничено и в някои случаи противоречиво – само алтернациите, които образуват нова лексикална единица, могат да се включват в wordnet като отделно синонимно множество.

Структурата на wordnet позволява добавянето на нови релации от различен тип, както между синонимни множества, така и между литерали. В момента се означат експлицитно значенията на т.нар. морфо-семантични връзки (Фелбаум и др. 2007), като се изхожда от идеята, че словообразователните правила могат да бъдат класифицирани в краен относително малък брой семантични категории от типа на семантичните роли (например агент: *откривател* – *откривам*; инструмент: *стържа* – *стъргало*; предназначение или функция: *чета* – *читанка*; резултат или причина: *греша* – *грешка* и др.). Наблюденията показват, че не всички съществуващи деривационни връзки са изразени в Българския wordnet – нито тези, отнасящи се до релациите *произведен*, *произхождащ* и особено *причастие*, които са налице и в Принстънския wordnet, нито повечето от останалите (езиково специфични) деривационни релации. Наличието на продуктивни словообразователни механизми, които имат относително регулярни свойства, е добра отправна точка за обогатяване на wordnet, където се включват понятия, за които има лексикализация

поне в един език, като се спазва принципът за запазване на структурата на Принстънския wordnet, изпълняващ ролята на междуезиков индекс.

Основното преимущество на wordnet е именно релационната му структура. Благодарение на нея, високата степен на гранулираност на wordnet може не само да се преодолее, но и да се използва от различните подходи за автоматично отстраняване на семантична многозначност (Иде и Веронис 1999): посредством приписване на тегла в зависимост от типа релация – хиперонимия, синонимия и т.н., и отчитане на дължината на пътя и броя релации от един тип към даден възел (търси се най-краткият път между възлите за измерване на семантична близост – хипотезата е, че за дадено множество от лексикални единици, които са в съседство в текста, се избират значения, които са с максимална семантична близост); посредством измерване на специфичността на понятията, организирани в хиперонимна релация (колкото по-специфично е понятието, толкова повече неговите хипоними са семантично свързани); посредством формулиране на семантични и вероятностни правила с цел обединяване на близки по значение синонимни множества за намаляване на многозначността (Михалкеа и Молдован 2001) и др.

В заключение, от една страна, съществуващите традиционни речници не предлагат адекватно описание на значението на думите от гледна точка на (автоматичното) отстраняване на многозначност. От друга страна, въпреки новите теории – генеративен лексикон, прототипни вероятностни компоненти – няма пълен и непротиворечив модел за описание на лексикалното значение, а още по-малко езиков ресурс, съставен по такъв модел, който да съдържа достатъчно количество лексикални единици, чието лексикално значение е описано адекватно и непротиворечиво. Основанията Българският wordnet да бъде избран като източник на значения при анотацията на Българския семантично анотиран корпус са комплексността му (значението не се изразява само от тълковната дефиниция и примерите, а и посредством всички релации на синонимното множество), пълнотата му и релационната му структура, която позволява различни типове обобщения и групиране по различни признаци на значенията.

1.6. Българският wordnet

Семантичната анотация в Българския семантично анотиран корпус осигурява връзка към синонимното множество от БулНет, включващо съответния литерал (дума или съставна дума), следователно към дефиницията (включваща примерите и бележките към множеството и литерала), семантичните, морфо-семантичните и екстралингвистичните релации на синонимното множество и семантичните и деривационните

релации на литерала. С други думи, семантичната анотация не е само анотация към значението, дефинирано в wordnet, а към цялата структура, в която участва съответният литерал.

1.6.1. Експлицитно изразяване на значението в БулНет

За разлика от Принстънския wordnet дефинициите за тълковното значение на синонимните множества са въведени в БулНет пълно, последователно и в максимална степен непротиворечиво. В Принстънския wordnet дефинициите не са или не винаги са тълковни дефиниции, а служат по-скоро като ориентир за значението, изразявано комплексно от синонимното множество. Дефинициите в БулНет се придържат към класическата структура от генерализация към диференциация, която рефлектира и във формалната им структура – различна за различните части на речта. Дефинициите не са взаимствани – съставени са така, че да съответстват на английския преводен еквивалент, да са съобразени със съществуващите тълковни дефиниции на български, но едновременно с това да са уникални към момента на съставянето си. Основните принципи, които трябва да се спазват при дефинирането на значенията на думите, могат да се формулират по следния начин (Ландау 1989; наред с много други): в дефиницията да не се използва дефинираната дума, нейни синоними или деривати; дефиницията да съдържа елементи, назоваващи по-общи понятия, например непосредствения хипероним; да се дефинира описваната дума, а не нейният контекст на употреба или друга странична информация; дефиницията да може да замени описваната дума (и нейните синоними) в даден контекст – това изискване е приложимо преди всичко за термините; да се избягват думи, които не са от общоупоредимата лексика; да се използва прост синтаксис и кратки дефиниции; да не се използват многозначни думи, чиято многозначност е неотстранима в контекста на дефиницията. Към синонимното множество може да се добави информация, която да показва употребата – един или повече примери (илюстрацията на всеки един от литералите с пример не е задължителна), една или повече бележки за стилистични, морфологични или синтактични признаци на синонимното множество или на литерала (някои от бележките, например за вид на глагола, са задължителни).

В wordnet няма връзка между значенията на многозначните думи – тоест не може да се направи разлика между омонимия и полисемия, освен ако не се въведе допълнителна класификация. В момента в българската лексикално-семантична мрежа номерата на значенията са произволни, те не съответстват на подреждането на многозначните думи в тълковен речник или на честотата на употреба на дадено значение

в семантично аотиран корпус. Семантичната анотация предоставя информация както за честотата на употреба на дадено значение, така и за честотата на употреба на даден литерал със съответното значение – тази информация се използва при подреждането на избираните значения по честота в системата за анотация и може да се припише към литералите и синонимните множества в БулНет. Твърдението за омонимите, че няма две или повече значения на дадена дума – има две значения на две думи, които се пишат и или произнасят еднакво (Ханкс 2000:205-6), може да се разшири и за полисемите – макар че има връзка между значенията, след като реферират към различни концепти, са различни думи. Следователно приложим е подходът значението на лексикалната единица да се разглежда като релация графична форма – синонимно множество. Таблица 1 представя броя на еднозначните и многозначните думи в БулНет към момента по части на речта.

синонимни множества	N	V	A	Adv	Pro	P	C	Part	Int
еднозначни литерали	25602	6615	5137	1217	44	85	65	24	3
двузначни литерали	2582	1669	1055	193	6	30	4	12	1
тризначни литерали	904	799	330	81	3	3	5	1	0
Nзначни литерали	760	1191	350	36	11	24	11	2	0

Таблица 1. Еднозначни и многозначни думи в БулНет

Общият брой еднозначни литерали е 38309, двузначни – 5564, тризначни – 2155 и Nзначни (повече от три значения) – 2442.

1.6.2. Релационно изразяване на значението в БулНет

Множеството от релации, които са присъщи на дадено синонимно множество, свидетелства за лексикалното значение и играе важна роля при вземането на решения при семантичната анотация. Следователно дефинирането на релациите и техните свойства има важно значение както за непротиворечивостта на структурата, така и за възможностите за нейното приложение.

Две думи W_1 и W_2 , денотиращи съответно множества от обекти E_1 и E_2 , могат да бъдат в една от следните четири релации (Крус 1986:87):

- идентичност: $E_1 = E_2$;
- включване: E_2 се съдържа в E_1 ;
- припокриване: E_1 и E_2 имат непразно сечение, но не се включват една в друга;
- изключване: E_1 и E_2 нямат общ елемент.

Следователно омографите са в изключваща се релация, полисемите – в релация на включване или припокриване, синонимите – в релация на идентичност, *антонимите, глаголните групи, също така и подобен на* – в релация на припокриване, *хиперонимите, меронимите и субсъбитията* – в релация на включване. В зависимост от свойствата си релациите образуват или не йерархични структури.

1.6.2.1. Йерархични релации

Йерархичните релации (релациите на включване) са три основни типа – таксономични (класификационни), мерономни и пропорционални серии (Крус 1986:112-135).

1.6.2.1.1. Таксономичните релации са релации, които асоциират обект от определен тип с друг обект от по-общ тип (Крус 1986:136-145):

Х е субтип на Y, ако един от следните изрази е истинен: Х е вид на Y, Х е тип на Y (за съществителните), Х е разновидност (начин) на Y (за глаголите).

Хиперонимията (хипонимията) в БулНет е обратна, транзитивна, нереклексивна и многократна семантична релация между синонимни множества (съществителни имена или глаголи), която изразява включване на класове (Коева 2007а). Транзитивността, нереклексивността и обратимостта (за разлика от синонимите) на хиперонимията позволяват йерархичната структура на мрежата, като хипонимите на по-долни нива наследяват (част от) семантичните свойства на хиперонимите директно или индиректно, което позволява да се правят семантични заключения за тях.

В таксономичните йерархии, за да се поддържа структурата на йерархията, даден възел може да няма кореспонденция в езика, например *крак* е хипоним на нелексикализираното понятие *външна част на тялото*. В Принстънския wordnet няма нелексикализирани възли, но в многоезиковата лексикално-семантична мрежа wordnet има, защото има междуезикова асиметрия при лексикализацията, което доближава структурата на wordnet до таксономия, но не напълно, тъй като не всички нелексикализиранни понятия намират място в нея, а само тези, които се лексикализират поне в един език.

В БулНет се допускат повече от един непосредствен хипероним при съществителните – причината за това е, че няколко таксономични релации са обединени в хиперонимната релация: род – вид и тип – представител на типа.

Непосредствените хипоними могат да влизат в различни отношения помежду си: има хипонимни класове, които са дизюнктивни и следователно комплементарни (*кон, камила, куче* по отношение на *животно*), и класове, които могат да се припокриват и да имат сечение (*домашно животно, домашен любимец, животно за разплод* по отношение на *животно*) (Восен 1999). В БулНет отношенията между непосредствените хипоними не са експлицитно изразени.

Идеалният случай би бил, ако значенията се класифицират към изграден непротиворечив таксономичен модел, трудностите произтичат от факта, че паралелно се създават и таксономичният модел и лексикално-семантичната мрежа, при това допълнителна трудност е условието, че БулНет се изгражда като огледало на класификация на значения за друг език.

1.6.2.1.2. Мерономията описва релацията част – цяло, т.е. диференциацията на частите по отношение на цялото (Крус 1986:157-179):

A е част от B, ако някой от следните изрази е истинен:

- *A съдържа B;*
- *B е част от A.*

Меронимията (холонимията) в БулНет е обратна (може да бъде и транзитивна, и многократна, ако отношенията са такива) семантична релация между синонимни множества – съществителни, изразяваща отношение между цяло и неговите части (Коева 2007а).

Могат да се определят няколко вида меронимия, три от които са дефинирани в wordnet: *част от*, *член на* и *съставка от*. Условие за дефинирането на релацията *част от* е съществуването на множество компоненти от един и същи тип. Релацията свързва части към тяхното цяло, само ако е налице топологическо включване чрез физически контакт: *книга е част от библиотека, библиотека е част от сграда, *книга е част от сграда*. При релацията *член на* съществува ограничението, че хипонимът трябва да е дума, означаваща конкретно съществително, а меронимът – събирателно съществително или съществително, обозначаващо множество от обекти. При релацията *съставка от* съществителното, означаващо цялото, трябва да изразява вещество, материя или субстанция, която съществува преди частите си. Обикновено съставките не се изразяват с отделна лексикална единица, а описателно, чрез фраза, която реферира към цялото: *торта – парче торта*.

Друга мерономна релация е *има субсъбитие (е субсъбитие на)* – обратна (може да бъде и транзитивна, и многократна) семантична релация между синонимни множества, изразяваща отношение на включване при глаголите (едното събитие започва, протича и приключва в рамките на другото) (Коева 2007а).

Мерономията и таксономията имат много общи свойства. Меронимията строи същите йерархични структури като хиперонимията, само че не толкова дълбоки. Мерономите по принцип не допускат транзитивност, но на лингвистично равнище транзитивността при меронимията е налице. Между непосредствените холоними също може да има дизюнктивни отношения, например *копчета* или *цип* спрямо меронима *панталон*, които не са експлицитно изразени в БулНет. Основната разлика е, че мерономията не класифицира елементите си по класове и йерархичната структура на мерономията не възниква заради йерархия на класове.

1.6.2.1.3. Пропорционалните серии – стойност, изразена като пропорция по отношение на всички стойности в дадена серия (например *кобила – жребец – овца – овен*), не са включени в БулНет (поне не експлицитно и последователно), но те представляват добър механизъм за представяне на деривационни релации в комбинация с друга семантична релация, например антонимия (*добър – добричък – лош – лошичък*), хиперонимия (*животно – животински – животинче – кон – конски – конче*) и т.н.

1.6.2.2. Неийерархични релации

Неийерархични релации са бинарни или троични (например тъй като има малко пълни синоними, синонимията се разглежда като троична релация – дума W_1 и дума W_2 са синоними в контекст C). В БулНет *синонимията* е рефлексивна, симетрична и транзитивна релация на еквивалентност между лексикални единици, които принадлежат към една и съща част на речта. В Принстънския wordnet при дефинирането на синонимите обикновено се прилага критерият за субституция:

Две думи са синоними в даден лингвистичен контекст C , ако субституцията на едната с другата не нарушава истинността на C (Милър 1990).

Следствията от подобен подход са, че е включена не само пълната синонимия, защото даден контекст не означава всеки контекст, каквото е изискването за взаимозаменяемост на пълните синоними. Има и много думи, които са подобни по значение и се приемат за синоними, но не

могат да се заместват една с друга в даден контекст по различни причини – синтактични, семантични, стилистични и др., например остаряла и съвременна дума (*токчетата на обувките ми са високи – *токчетата на опинците ми са високи*).

Подобно на синонимите антонимите са в силна зависимост от контекста и затова образуват троична релация.

Дума₁ и Дума₂ са антоними, ако Дума₁ в контекст С е отрицанието на Дума₂ в контекст С.

Антонимията в БулНет е симетрична и нетранзитивна релация на противоположност между два литерала от различни синонимни множества, принадлежащи към една и съща част на речта. В Принстънския wordnet антонимията се разглежда като релация между литерали, а не между значения. Антонимията, която се дефинира за всички членове на дадено синонимно множество, се обозначава като друга релация, наречена *близка антонимия*. В момента в БулНет *антонимията* е представена (не напълно коректно) като отношение между синонимни множества. Наблюдават се различни степени на антонимия, но разликите не са отразени в БулНет.

Другите бинарни нейерархични семантични релации в БулНет са (Коева 2007а):

Има атрибут (е атрибут на) е обратна и многократна семантична релация между синонимно множество – съществително име, и едно или повече синонимни множества – прилагателни имена, които изразяват негови стойности.

Каузира (е каузиран от) е обратна семантична релация между синонимни множества, която изразява отношението причинно следствие при глаголите.

Подобен на е симетрична, евклидова и многократна семантична релация между синонимни множества, която се основава на подобие на значенията между едно „централно” синонимно множество от прилагателни имена и останалите подобни по значение синонимни множества от прилагателни.

Глаголна група е симетрична, евклидова и многократна семантична релация между синонимни множества, която изразява подобие между едно „централно” и други глаголни множества.

Също така е симетрична, евклидова и многократна семантична релация между синонимни множества или от глаголи, или от прилагателни, които семантично се асоциират с „централно” синонимно множество. Релациите на подобие с „централно” синонимно множество не са последователно представени в wordnet.

Словообразователните релации между литерали от едно или различни синонимни множества също могат да се дефинират като нейерар-

хични бинарни релации. За момента словообразователните релации в БулНет са много малко в сравнение с реално съществуващите в езика.

Произхождащ е обратна (може да е многократна) словообразователна релация между съществителни и техни производни прилагателни, принадлежащи към различни синонимни множества.

Причастие е обратна (може да е многократна) словообразователна релация между глагол и причастие, образувано от него, принадлежащи към различни синонимни множества.

Производен е обратна (може да е многократна) словообразователна релация между глаголни множества, принадлежащи към различни синонимни множества, например.

Принадлежащ е обратна (може да е многократна) словообразователна релация между прилагателно и съществително, принадлежащи към различни синонимни множества. Наблюденията показват, че не всички съществуващи връзки, отнасящи се до релациите *производен*, *произхождащ* и особено *причастие*, са изразени в Българския wordnet, както и някои от най-продуктивните деривационни механизми в български все още не са представени в БулНет. Основната причина е езиково специфичният им характер и фактът, че се спазва принципът за запазване на структурата на Принстънския wordnet (Коева 2007а). Съществуват няколко подхода за включване на езиково специфичните начини за лексикализация при словообразуването (Кръстев и др. 2006; Коева 2007б):

- Да се смята, че дериватите се отнасят към различно понятие, за което се дефинира ново синонимно множество в wordnet (досега в Българския wordnet по този начин се третират някои съществителни от женски род от родовите двойки и някои относителни прилагателни).
- Дериватите да се включат в едно синонимно множество заедно с думата, от която са образувани (досега в Българския wordnet по този начин се подхожда към видовите двойки, някои умалителни и увеличителни съществителни, някои прилагателни).
- Дериватите да не се включват в wordnet освен ако значението им не се различава и по други признаци от значението на изходната дума, различни от т.нар. словообразователно значение (досега в Българския wordnet това са умалителните и увеличителните съществителни, които не се срещат в корпуса за анотация).
- Да се приложи към литералите описание не само на техните флективни форми, но и на словообразователните им форми.

Деривационните релации свидетелстват за наличието на морфосемантични връзки от различен характер между самите синонимни

множества (*асистентът асистира, участникът участва*). Нито деривационните връзки между литералите, нито морфо-семантичните релации между синонимните множества са достатъчно пълно и последователно представени в (Българския) wordnet.

В БулНет са включени три екстралингвистични релации: *Тематична област, Регион на употреба и Сфера на употреба* са обратни и многократни бинарни екстралингвистични релации.

В заключение, свойствата на релациите произтичат от универсалните свойства, които съществуват между същности, събития и свойства, дотолкова доколкото можем да ги наблюдаваме и описваме обективно. Правилното дефиниране на свойствата на релациите и пълното им и непротиворечиво представяне в wordnet е предпоставка за правилното изразяване на лексикалните значения посредством релационната структура, което от своя страна е предпоставка за коректна (автоматична) анотация.

1.7. Анотационна схема (общи принципи)

Няма унифицирани стандарти за анотационна практика, но има множество от препоръки и критерии, по отношение на които може да се съгласува изработването на дадена анотационна схема. За някои от равнищата на анотация съществуват предписания в поредица от документи, които описват добрите практики за това какво, до каква степен и как трябва да се аотира (ЕГСКОВОЕ 1996а; ЕГСКОВОЕ 1996б и др.). Препоръчват се следните общи принципи: да се използва формализмът **атрибут: стойност**; да не се спазва строга йерархия на атрибутите и стойностите; да се използват три равнища на анотация – задължителни, препоръчителни и незадължителни атрибути и стойности (например на морфо-синтактично равнище задължителен е атрибутът част на речта, атрибутът тип (нарицателно и собствено) е препоръчителен за съществителните, а атрибутът бройност (броимо и неброимо) е незадължителен). Коректното определяне на значението на думата зависи от по-ниските равнища на анотация, например ограничаването на възможните значения за думата *син* зависят от дефинирането на лемата, съответно като съществително или прилагателно, някои думи също така се срещат с определени граматични характеристики само като част от съставна дума или с различно значение (например *английска сол, минерални соли, солна киселина, солена вода*).

Следните общи препоръки за добра практика при анотация на текста са спазени при създаването на Българския семантично аотиран корпус (Лиич 1997; МакЕнери и Уилсън 2001):

- Оригиналният текст може да се възстановява лесно без анотацията, която е въведена.
- Анотациите могат да се извличат лесно от анотирания текст.
- Всеки анотиран текст се съпровожда с подходяща документация, включваща анотационната схема.

Следните общи препоръки също така са взети под внимание при изработването на анотационна схема за семантична анотация на Българския семантично анотиран корпус (Уилсън и Томас 1997:55-57):

- Семантичната анотация трябва да има отношение към значението в лингвистичен или когнитивен смисъл (Българският семантично анотиран корпус се анотира със значения от Булнет – структура, в която думите са организирани в семантични множества, свързани помежду си с различни семантични релации, така че да представят понятийната структура на езика).
- Анотацията трябва да обхваща голяма част от думите в корпуса, а не само някои (в Българския семантично анотиран корпус са анотирани всички думи).
- Схемата трябва да е гъвкава, така че да позволява да се описват различни детайли – периоди от развитието на езика, регистри и т.н. (анотацията към семантично множество от БулНет включва и всички бележки към литералите или синонимните множества, отразяващи особеностите или ограниченията при тяхната употреба).
- Трябва да се оперира с определено равнище на гранулираност на значенията, като въпросът кое равнище на гранулираност е най-подходящо остава отворен (транзитивността на хиперонимията позволява преходи към по-високо и по-нискоразнище, съответно към по-обобщено или по-гранулирано представяне на значенията).
- Семантичната система за анотация трябва да има йерархична структура, базирана върху нарастваща генерализация на отношенията между значенията (основната релация в БулНет е таксономична – хиперонимната релация).
- Анотацията трябва да съответства на стандартите, ако има такива (до момента са известни предварителни препоръки за семантична анотация (ЕГСКОЕ 1999)).

Предварителните препоръки за семантична анотация (ЕГСКОЕ 1999) включват някои общи критерии, които трябва да се имат предвид, когато се изработва анотационната схема. Те са:

- Адекватно покритие: трябва да бъдат включени всички лингвистични свойства, които са цел на анотацията, респек-

тивно – без елементи, които не са необходими. Целта на семантичната анотация в БулСемКор е да се опише уникалното значение, което се проявява при употреба на дадена дума, като се използва най-подходящият езиков ресурс от наличните в момента – БулНет. Анотацията не е към значението на думата от синонимното множество, а към цялото синонимно множество, като се наследяват свойствата му. Не се предвижда пълна морфосинтактична и синтактична анотация, защото не е релевантна за целите на семантичната анотация. Всяка дума в корпуса получава семантична анотация. Не се допуска асоциация на дума с повече от едно синонимно множество. Единичните и съставните думи се интерпретират по един и същи начин както в семантично аотириания корпус, така и в Булнет.

- **Непротиворечивост:** анотационната схема трябва да бъде организирана около непротиворечиви принципи, които да определят кои типове обекти са тагове, кои типове обекти са атрибути и кои типове обекти са стойностите на таговете. В Българския семантично аотириан корпус таговете са думите. Съответно атрибутите са, както следва: словоформа (стойност е употребената форма в корпуса), лема (стойност е приписаната лема към дума или съставна дума), значение (стойност е идентификационният номер от Булнет).
- **Възстановимост:** анотационната схема трябва да позволява възстановимост на първоначалния текст от аотирианата му версия. Условието е изпълнено в Българския семантично аотириан корпус – възстановим е не само текстът без анотацията, но и текстът преди нормализацията.
- **Проверимост:** трябва да е възможна проверка на анотацията, разбирана като процес, при който автоматично се проверява дали маркираността в документа следва стандартите, които са приети. Проверките за непротиворечивост са заложиени като функционалност на програмата за анотация, която се използва при създаването на Българския семантично аотириан корпус.
- **Възможност за извличане на информация от анотацията:** анотационната схема трябва да позволява различни равнища на анализ. Идентификационният номер показва мястото на синонимното множество в структурата на Булнет, следователно реферира не само към съдържанието на съответното синонимно множество, но и към всички равнища на йерархичните релации в структурата.

- Възможност за обработка: структурата на представяне на данните в анотационната схема трябва да отговаря на изискванията за компютърна обработка на текста. Програмата за анотация работи с текстове в определен формат.
- Разширяемост: структурата на анотационната схема трябва да допуска възможност за разширяване. Анотацията в Българския семантично аотиран корпус може да бъде разширена с информация на морфологично, морфо-синтактично и/или синтактично равнище.
- Компактност: това изискване е свързано с ограничаването на броя на символите, които се въвеждат в текста при анотацията. С развитието на технологиите това изискване става в голяма степен нерелевантно, но така или иначе е спазено, защото цялата информация, която се асоциира с думата като нейна семантична анотация, се свежда до идентификационния номер.
- Четаемост: аотирият текст трябва да е разбираем (което означава или да не е кодиран, или да може лесно да се трансформира като текст). Това изискване също е спазено, тъй като идентификационният номер сочи към уникално синонимно множество.

1.8. Проверка на анотацията

Могат да се различат два метода за анотация (Килгариф 1998): линеен (текстов) метод, при който думите се аотират последователно, без да се разглеждат едновременно лексикалните единици с еднакво значение или множеството от значения, с които се асоциира дадена лексикална единица. Вторият метод е лексикален (пресичащ се), при който се аотират всички срещания на значението на дадена лексикална единица в корпуса, след това се преминава към друга лексикална единица и т.н. – всички значения на дадена дума се анализират само един път. При линейния подход множеството от значения на дадена дума се проверява при всяка нова анотация. Предимствата на втория метод са, че е във фокус множеството от значенията на дадена многозначна дума и може да се обърне внимание на специфични разлики, зависимости и т.н. По този начин семантичната структура на множеството от значения, асоциирани с дадена дума, се анализира само един път, което би могло да доведе до по-голяма непротиворечивост. Недостатък е, че няма напълно аотиран фрагмент, докато не се завърши анотацията на целия корпус. От друга страна, това, че част от думите не са аотирани, може в някои случаи

да затрудни вземането на решение за коректния избор на значение. При анотацията на Българския семантично анотиран корпус двата основни метода се прилагат едновременно, така че да се използват предимствата им и да се неутрализират недостатъците.

Основно поради гранулираността на wordnet, понякога повече от едно значение може да е подходящо за анотация на дадена дума. Един от основните проблеми при всеки тип анотация е субективността за това кое е подходящото значение. Субективността при вземането на решение е в тясна връзка със съгласието между анотаторите за валидността на дадена анотация. Не случайно в СемКор анотаторите отбелязват и степента на увереност, с която избират дадено значение сред множеството възможни (Фелбаум и др. 1998), което се използва за изследване на съпадението на избора от експерти и неексперти.

При работата по Българския семантично анотиран корпус проверката се извършва посредством втори, понякога трети анотатор, който проверява решението за въведената анотация.

1.9. Основни резултати

Процесът на работа по създаването на Българския семантично анотиран корпус може да бъде описан така: формулиране на анотационна схема, създаване на програма, подпомагаща анотацията, компилиране на изходен корпус и нормализация на думите, срещащи се в него, изработване на начална конвенция за анотацията, семантична анотация на значенията в корпуса. По време на анотацията на значенията се извършва паралелно тестване на програмата за анотация и допълване ѝ с нови функционалности, конкретизиране на анотационната конвенция, проверка на вече съществуващите значения в БулНет и допълване с нови. Получените резултати съответстват на основните етапи на работа:

- Анотационна схема (описана в тази част)
- Програма за анотация (описана във втора част)
- Анотационна конвенция (описана от трета до единадесета част включително)
- Компилиране на изходен корпус (описано в тази част)
- Семантична анотация на значенията в корпуса (количествени данни се предлагат в тази част, а корпусът е достъпен за търсене на <http://dcl.bas.bg/semcor>)
- Допълване и проверка на Булнет (описано от трета до единадесета част включително)

Целият изходен корпус е аотиран за част на речта, лема и значение спрямо кореспондиращото синонимно множество от Булнет. Количествени данни за аотацията са представени в Таблицы 2 и 3.

N	V	A	Adv	Pro	P	C	Part	Int
27010	14161	11092	6115	6497	13165	6434	2540	9

Таблица 2. Брой аотирани думи по части на речта (без съставните думи)

брой значения	N	V	A	Adv	Pro	P	C	Part	Int
повече от 4	9026	9044	4729	2576	2062	12621	5691	1826	4
3	8725	1254	1414	890	116	75	436	3	0
2	4671	1225	1831	1016	586	172	248	598	0
1	3772	1546	3047	1545	3715	259	46	50	5

Таблица 3. Брой срещания на многозначни думи в корпуса по част на речта (без съставните думи)

БулНет се разширява по два начина. Първият начин е свързан с дефинирането на аотационната схема и възприетия принцип всяка дума от текста да получи семантична аотация – това налага добавянето в структурата на БулНет на синонимни множества за още пет части на речта (тъй като става въпрос за затворените класове, това означава добавяне и класифициране на всички лексикални единици от затворените класове). Дефинирането на значенията на думите от затворените класове се базира на анализ на синтактични доказателства и семантични признаци, които се наблюдават в корпуса, както и на съществуващите лексикографски и граматични описания. Съществуващите класификации на затворените класове, поради високата степен на полисемия при повечето от тях, в много случаи не са достатъчно прецизни и не са базирани на ясни критерии. За някои от думите от затворените класове има синонимни множества в английски и те се вземат под внимание. Тук спадат и аотационните конвенции, които налагат обогатяването на структурата на Булнет. БулНет се разширява с културно-специфични концепти, за които се намира позиция в структурата, както и с езиково-специфични концепти – относителни прилагателни, демюнитиви и т.н.

Вторият тип разширение на Булнет е свързан с необходимостта да се добавят нови значения, ако липсват, или да се променят вече съществуващи. В процеса на аотация еднозначните и многозначните думи

се проверяват и ако има подходящо синонимно множество, то се избира, ако има синонимно множество, което би било подходящо при съответна модификация и тази модификация отговаря на цялостната структура на wordnet, промяната се нанася, ако няма подходящо синонимно множество – въвежда се ново.

1.10. Предназначение на Семантично аотирания корпус

Основното предназначение на Българския семантично аотиран корпус (и причината да бъде създаден) е да служи като тренировъчен корпус и като стандарт при тестове за проверка при автоматично отстраняване на семантична многозначност (Фелбаум и др. 2001), което се прилага в много от областите на компютърна обработка на естествения език. Разбира се, процесът на създаване на корпуса е свързан с проверка и евентуално допълване или корекция на формулираните значения в Българския wordnet (разбирани комплексно не само като тълковна дефиниция, а като цялостното изразяване на значението на дадено синонимно множество посредством отношенията му с останалите). Не на последно място Българският семантично аотиран корпус предлага информация за (лингвистични) изследвания, тъй като значението на дадена лексикална единица е експлицитно свързано с нейната употреба.

Цитирана литература

- Агире и др. 2006: Agirre E., Aldezabal I., Etxeberria J., Izagirre E., Mendizabal K., Quintian. M., eta Pociello E. A methodology for the joint development of the Basque wordnet and Semcor. In: *Proceedings of the 5th International Conference on Language Resources and Evaluations (LREC)*, Genoa (Italia).
- Аткинс 1993: Sue Atkins Theoretical Lexicography and its relation to Dictionary-making. In: *Dictionaries: the Journal of the Dictionary Society of North America*, (guest editor) W. Frawley, DSNA, Cleveland Ohio. 4-43.
- Бентиволи и Пианта 2005: Luisa Bentivogli and Emanuele Pianta Exploiting Parallel Texts in the Creation of Multilingual Semantically Annotated Resources: The MultiSemCor Corpus. In: *Natural Language Engineering, Special Issue on Parallel Texts*, Volume 11, Issue 03, September 2005, 247-261.
- Вежбицка 1996: Wierzbicka, A. *Semantics: Primes and Universals*. Oxford University Press, Oxford, UK.

- Восен 1999: Vossen P. (Ed.) *Special Issue on EuroWordnet, Computers and the Humanities*, vol.32, nos.2-3, 221-251.
- Бу и др. 2006: Yunfang Wu, Peng Jin, Yangsen Zhang, Shiwen Yu: A Chinese Corpus with Word Sense Annotation. *ICCPOL 2006*: 414-421 .
- Джакендоф 1990: Jackendoff, R. *Semantic Structures*. MIT Press, Cambridge, USA.
- ЕГСКОЕ 1996а: *Expert Advisory Group for Language Engineering Standards Preliminary recommendations on corpus typology. EAG-TCWG-CTYP/P*. Pisa: Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale.
- ЕГСКОЕ 1996б: *EAGLES: Recommendations for the Morphosyntactic Annotation of Corpora. EAGLES Document EAG--TCWG--MAC/R*. Pisa: Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale.
- ЕГСКОЕ 1999: *EAGLES LE3-4244: Preliminary Recommendations on Semantic Encoding, Final Report*. <http://www.ilc.pi.cnr.it/EAGLES/EAGLESLE.PDF>
- Ерявец 2004: Erjavec T. MULTEXT-East Version 3: Multilingual Morphosyntactic Specifications. *International Conference on Language Resources and Evaluation, LREC'04*. Paris.
- Иде и Веронис 1998: Ide, N., Veronis, J., Word sense disambiguation: the state of the art. *Computational Linguistics* 24 (1), 1–40.
- Йаеборг и др. 2002: Jaerborg, J., Kokkinakis, D. and Toporowska Gronostaj, M. Lexical and textual resources for sense recognition and description. In: *Proceedings LREC 2002*. Las Palmas, 1492-1497.
- Катц 1972: Katz Jerrold J. *Semantic Theory*. New York: Harper and Row.
- Килгариф 1997: Adam Kilgarriff. I don't believe in word senses. *Computers and the Humanities* 31(2):91-113.
- Килгариф 1997: Adam Kilgarriff. I don't believe in word senses. *Computers and the Humanities* 31(2):91-113.
- Килгариф 1998: Kilgarriff A. Gold Standard Datasets for Evaluating Word Sense Disambiguation Programs. In: *Computer Speech and Language. Special Use on Evaluation* 12(4), 453-472.
- Килгариф и Розенцвайг 2000: Kilgarriff Adam and Joseph Rosenzweig. Framework and Results for English SENSEVAL. *Computers and the Humanities* 34(1-2) (Special Issue on SENSEVAL) 15-48.

- Коева 2006а: Коева, Св. Синтактични редувания в български – диатези и алтернатиции. В: Български език, кн. 4, 19-32.
- Коева 2006б: Коева, Sv. Inflection Morphology of Bulgarian Multiword Expressions. In: *Computer Applications in Slavic Studies*, Boyan Penev Publishing Center, Sofia, 201-216.
- Коева и др. 2006: Коева, Sv., Sv. Leseva, M. Todorova, Bulgarian Sense Tagged Corpus. In: *Proceedings of the 5th SALTMIL Confernece on Minority Languages: Strategies for Developing Machine Translation for Minority Languages*, Genoa, 79-87.
- Коева 2007а: Коева, Св. БулНет (лексикално-семантична мрежа на българския език) – част от световната лексикално-семантична мрежа, В: сп. Български език, брой 1, 34-50.
- Коева 2007б: Коева, Св. Морфо-семантични релации в БулНет (лексикално-семантична мрежа на българския език), сп. Български език, кн. 4, 36-59.
- Коева 2008: Коева, Sv. Semantic Nature of Diatheses, In: *Studia kognitywe = Etudes cognitives*, Warszawa, Vol. 8, 71-95.
- Кол и др. 1998: Kohl K. T., Douglas A. Jones, Robert C. Berwick, and Naoyuki Nomura Representing Verb Alternations in WordNet In: C. Fellbaum (Ed.), *WordNet: An Electronic Lexical Database*. Cambridge: MIT Press, 153-178.
- Крус 1986: Cruse D. A. *Lexical Semantics*. Cambridge: Cambridge University Press.
- Кръстев и др. 2006: Krstev C., S. Koeva, D. Vitas. Towards the Global Wordnet, in: *First International Conference of Digital Humanities Organizations (ADHO) Digital Humanities 2006*, Paris-Sorbonne, 5 –9 July 2006, 114 –117.
- Ландау 1989: Landau, S. I. *Dictionaries, the art and craft of lexicography*. Cambridge: Cambridge University Press.
- Ландерс и др. 1998: Landes Shari, Claudia Leacock and Randee Teng. Building Semantic Concordances. In: *Word-Net: An Electronic Lexical Database and Some of its Applications*. Ed. Christiane Fellbaum, Cambridge, Mass.: MIT Press, 199-216.
- Лиич 1977: Leech, G. Introducing corpus annotation. In Garside R., Leech, G., McEnery, A. M. (eds.) *Corpus Annotation: Linguistic Information from Computer Text Corpora*. London: Longman.
- Лупу и др. 2005: Lupu, M., Trandabăţ, D., Husarciuc M. A Romanian SemCor Aligned to the English and Italian MultiSemCor, *1st ROMANCE*

FrameNet Workshop, International Workshop held at EUROLAN 2005 Summer School, July, 26 – 28, 2005, University Babes-Bolyai, Cluj-Napoca, Romania, 20-27.

- МакЕнери и Уилсън 2001: McEnery A. M., Wilson A. *Corpus Linguistics: An Introduction*. Edinburgh: Edinburgh University.
- Милър 1990: Miller G. *Five Papers on Wordnet. Special Issue in International Journal of Lexicography*, vol.3, no.4.
- Милър и др. 1993 Miller G. A., Leacock C., Randee T., and Bunker R. A Semantic Concordance. *Proceedings of the 3rd DARPA Workshop on Human Language Technology*. Plainsboro. New Jersey, 303-308.
- Милър 1995: George A. Miller. Building Semantic Concordances: Dissambiguation vs. Annotation. In: AAAI Technical Report SS-95-01, 92-94.
- Михалкеа и Молдован 2001: Mihalcea R. and D.I. Moldovan. Automatic generation of a coarse grained Wordnet. In: *Proceedings of the SIGLEX workshop on „Wordnet and Other Lexical Resources: Applications, Extensions and Customizations”*, NAACL. Pittsburgh, USA.
- Монтемани и др. 2000: Montemagni S., Barsotti F., Battista M., Calzolari N., Corazzari O., Zampolli A., Fanciulli F., Massetani M., Raffaelli R., Basili R., Pazienza M.T., Saracino D., Zanzotto F., Mana N., Pianesi F., Delmonte R. The Italian Syntactic-Semantic Treebank: Architecture, Annotation, Tools and Evaluation. *Proceedings of the COLING Workshop on „Linguistically Interpreted Corpora (LINC- 2000)”*, Luxembourg, 6 August 2000, 18-27.
- Наваро и др.2003: Borja Navarro, Montserrat Civit, Antonia Martí, Raquel Marcos, Belen Fernandez. Syntactic, Semantic and Pragmatic Annotation in Cast3LB. *Corpus Linguistics 2003 Workshop on Shallow Processing of Large Corpora*, Lancaster, UK.
- Нг и Лее 1996: Hwee Tou Ng and Hian Beng Lee. Integrating multiple knowledge sources to disambiguate word sense: An exemplar-based approach. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics (ACL)*, 40-47.
- Пианта и Бентиволи 2003: Emanuele Pianta and Luisa Bentivogli „Translation as Annotation” In: *Proceedings of the AIIA 2003 Workshop „Topics and Perspectives of Natural Language Processing in Italy”*, Pisa, Italy, September 2003, 40-48.
- Пустейовски 1995: Pustejovsky, J. *The Generative Lexicon*, MIT Press, Cambridge, MA.

- Ресник и Яровски 2000: Resnik P. and D. Yarowsky. Distinguishing Systems and Distinguishing Senses: New Evaluation Methods for Word Sense Disambiguation, *Natural Language Engineering* 5(2), 113-133.
- Синклер 1987: Sinclair, John M., editor. *Looking Up: An Account of the COBUILD Project in Lexical Computing*. London: Collins.
- Уилсън и Томас 1997. Wilson, A., Thomas J. Semantic Annotation. In R. Garside, G. Leech & A. M. McEnery, (eds.) *Corpus Annotation: Linguistic Information from Computer Text Corpora*. London: Longman.
- Фелбаум 1998: Fellbaum, C. (ed.). *Wordnet: An Electronic Lexical Database*. Cambridge, Mass.: MIT Press.
- Фелбаум и др. 1998. Fellbaum, C., Grabowski, J. and Landes, S. Performance and confidence in a semantic annotation task. In Fellbaum, C. (ed.), *Wordnet: An Electronic Lexical Database*. Cambridge (Mass.): The MIT Press, 217-237.
- Фелбаум и др. 2001: Fellbaum, C., Palmer, M., Dang H., Delfs L. & Wolff, S. Manual and Automatic Semantic Annotation with Wordnet. In NAACL-2001 Workshop on Wordnet and Other Lexical Resources. Pittsburgh, Philadelphia. SIGLEX.
- Фелбаум и др. 2007: Christiane Fellbaum, Anne Osherson, Peter E. Clark. Putting Semantics into WordNet's „Morphosemantic” Links. In: *Proceedings from 3rd Language and Technology Conference: HLT as a Challenge for Computer Science and Linguistics*, Poznan.
- Фонтенел: 2008: Thierry Fontenelle. (ed.). *Practical Lexicography: A Reader* (Oxford Linguistics). Oxford: Oxford University Press.
- Фодор: 1975: Fodor, Jerry. *The Language of Thought*. Cambridge, MA: Harvard University Press.
- Ханк 2000: Patrick Hanks: Do word meanings exist? In: *Computers and the Humanities*, 34(1-2):205-215.

2. Система за аотиране Chooser

Борислав Ризов

2.1. Формат на корпуса

БулСемКор се състои от множество от файлове, представени в плитък xml формат. Текстът е кодиран под формата на поредица от xml тагове с наименование *word*, както е илюстрирано в (1).

(1) `<word l=„компас” p=„-1234468596” w=„компас,”/>`

Идентификацията на думата (словоформата), заедно с информация за основната ѝ форма (лема), значението ѝ, и автора на аотацията, се съхранява в отделни атрибути на тага *word*. Съставните думи са кодирани в атрибута *parent*, означен с ‘*p*’, като всички графични думи в съставната дума получават една и съща стойност за атрибута *p*.

Възможността за разширение на xml файловете ще бъде използвана, ако се наложи допълнително аотиране на същия корпус с информация от друго равнище, например синтактична аотация. За целите на потребителския интерфейс се съхранява информация която е последно аотираната дума.

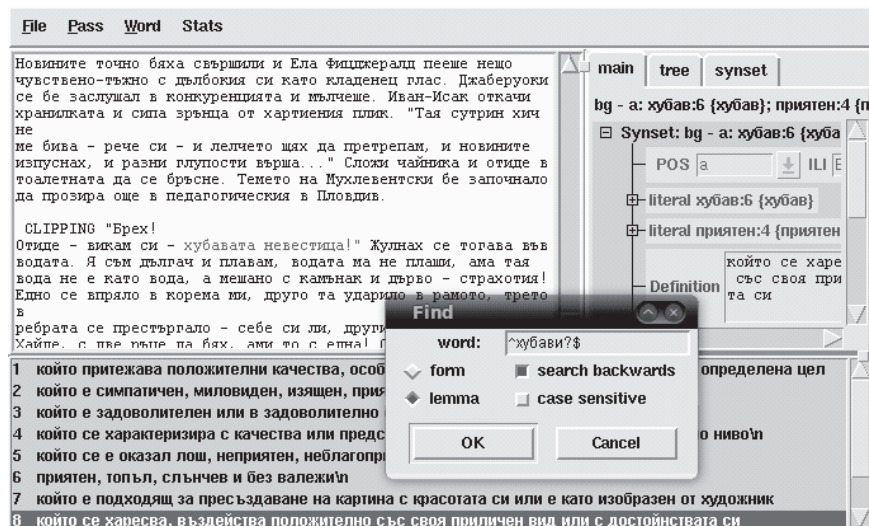
2.2. Системата за аотиране Chooser

Аотацията на БулСемКор се осъществява с помощта на версия на многофункционалната система за аотиране на корпуси Chooser. Chooser интегрира различни стратегии за навигация по текста, който се аотира: обхождане на (i) всички единици; (ii) на неанотираните единици; (iii) на всички срещания на конкретна (единична или съставна) дума. Основните функции на системата включват: (i) бърз и лесен избор на значение за дума или съставна дума; (ii) информация за множеството от кандидати, от които аотаторът избира; (iii) свързване на графични думи в съставни лексикални единици. За да бъде разпозната думата и свързана със съответните синонимни множества от БулНет, преди това лемата трябва да е правилно идентифицирана (автоматично или ръчно).

Системата Chooser използва и редица операции за редактиране на корпуса: (i) редакция на словоформа; (ii) редакция на основна форма (лема); (iii) изтриване и добавяне на думи в самия текст, ако това е необходимо (например за корекция на правописни и печатни грешки); (iv) редакция на неправилно форматирани текст. Предвидена е и опция за търсене по словоформа и по основна форма (лема) в корпуса с помощта на регулярни изрази (с Python синтаксис¹), например *това* – намира

¹ <http://docs.python.org/library/re.html>

всички единици съдържащи поредицата ‘това’ – товари, натовари и др.; **^хубави?§** – намира само ‘хубав’ и ‘хубави’, както е илюстрирано във Фигура 1, но не и ‘разхубави’.



Фигура 1. Търсене в системата за анотация Chooser

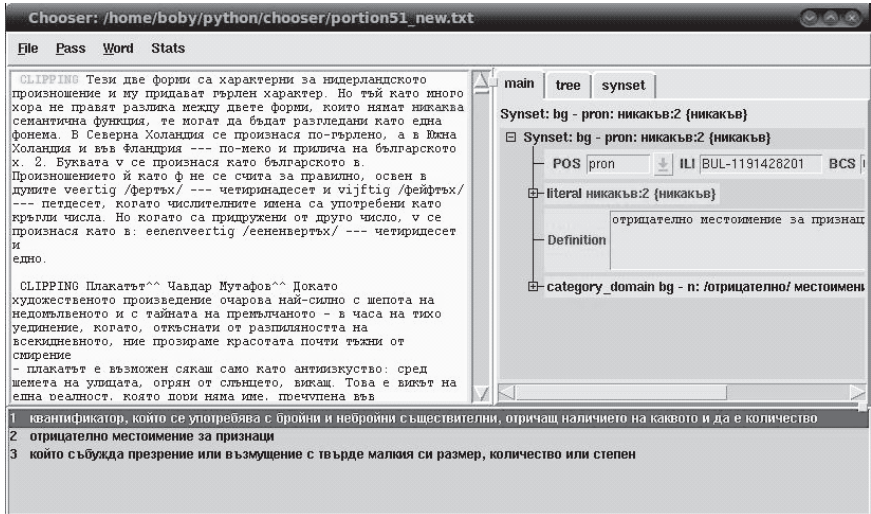
Системата достъпва множеството от значения за дума през API на многофункционалната система за обработка на wordnet Hydra² (Ризов 2008). Chooser използва визуализациите на синонимните множества и другите обекти от БулНет в Hydra. Унифицираният изглед на данните в Chooser и Hydra улеснява анотаторите, като позволява изграждането на лексикално-семантичната мрежа БулНет да се извършва успоредно с анотирането на БулСемКор. Добавените синонимни множества, литерали, текстови данни и всички промени на информацията в БулНет се актуализират в Chooser и Hydra в момента на въвеждането им за всички потребители на двете системи.

2.3. Потребителски интерфейс

При зарежданена файл за анотиране или редакция Chooser предоставя прозорец с три панела (Фигура 2): (i) текстово поле, в което се появява

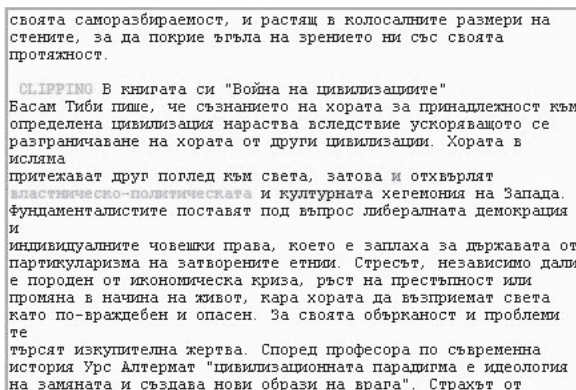
² Системата Hydra осигурява паралелен достъп до БулНет и до Принстънския wordnet (или до повече езици). Синонимните множества от лексикално-семантичните мрежи се зареждат в паралелен изглед, за да могат да бъдат обхождани, сравнявани, редактирани. Системата е с централизирано съхранение на данните и осигурява конкурентен и абстрактен достъп с помощта на модален логически език.

текстът, който трябва да се аотира (бялото поле вляво); (ii) списъчно поле с единиците (синонимните множества), от които анотаторът избира подходящ кандидат за анотация за конкретната единица от текста (полето отдолу с маркирана в червено дефиниция на лексикалната единица); (iii) информационно поле с панелите на Hydra, изобразяващи синонимните множества (полето вдясно с дървовидната структура на местоимението „никакъв“).



Фигура 2. Потребителски интерфейс на системата Chooser

2.3.1. Текстовото поле, показано във Фигура 3, изобразява текста за аотиране.



Фигура 3. Текстово поле на системата Chooser

Текущата единица, която трябва да се аотира или чиято анотация трябва да се проследи и редактира, е маркирана с различен цвят (червено във Фигура 3). Цветове различават характеристиките на единиците в текста, както следва:

- необходими единици (черно)
- аотирани единици (синьо)
- неанотирани единици
 - несвързани до момента със значение (лилаво)
 - еднозначни (зелено)
 - многозначни (лилаво)

2.3.2. Поле за избор

Поле за избор, илюстрирано във Фигура 4, е стандартен списъчен контрол, който показва частите на речта и дефинициите на синонимните множества, между които потребителят трябва да избере при аотиране на съответната единица в текста.

1	внимание, изразяващо се в концентрация на зрението върху определен обект
2	действие, свързано с насочването на очите към определен обект на възприемане
3	мисловна позиция, която предопределя определен тип отношение към света

Фигура 4. Поле за избор на системата Chooser

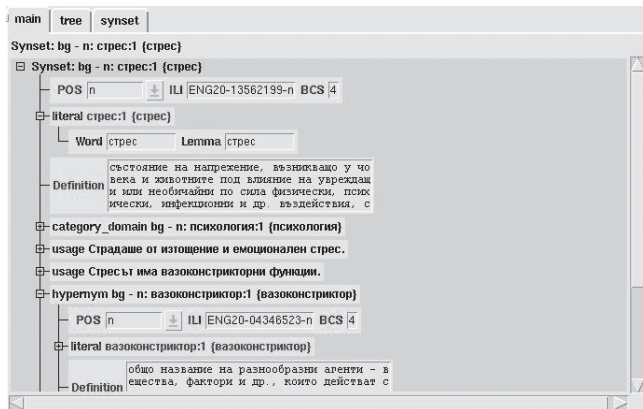
Синонимните множества са подредени в низходящ ред според това колко често са били избирани до момента. Така най-често избраните синонимни множества се намират на първите редове в полето за избор (във Фигура 4 това е редът с дефиниция ‘внимание, изразяващо се в концентрация на зрението върху определен обект’, пред която фигурира числото 1). Разглеждането и изборът на синонимните множества се извършва с помощта или на клавиатурата, или на мишката (във Фигура 4 избраното или текущо разглежданото синонимно множество е третото, маркирано с червено). Всяко синонимно множество е отбелязано с буква или цифра и може да се селектира чрез избиране с клавиатурата.

2.3.3. Информационно поле

Информационното поле съдържа три изгледа на БулНет с център синонимното множество, за което се показва необходимата информация. Използват се изгледите на системата Hydra, а както беше отбелязано, унифицирането между двете системи е важно улеснение за потребителите и значително ускорява аотирането.

Първият и най-подробен изглед е рекурсивен. Той представя информацията в дървовидна структура, чийто възли са синонимните множества и информацията в тях, както може да се види за синонимното

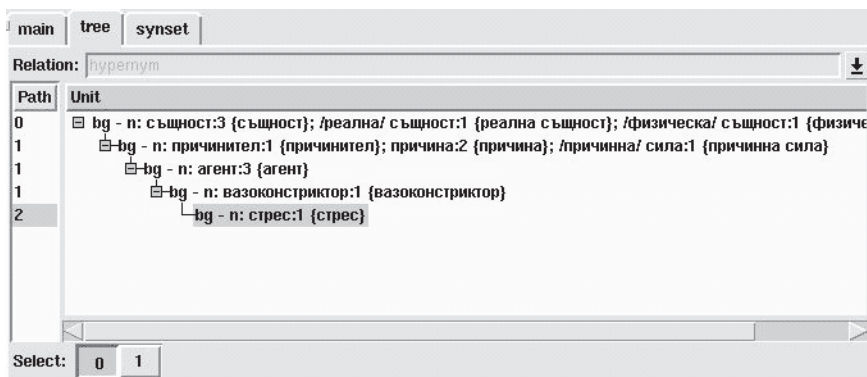
множество на {стрес:1} – Фигура 5а. Свързаните с релации синонимни множества, например хиперонимната релация към {вазоконстриктор:1}, заедно с литералите, изграждащи множествата, и текстовата информация се изобразяват като поддървета със същата структура като първото дърво.



Фигура 5а. Информационно поле на системата Chooser – основен изглед

Вторият изглед също е дървовиден – Фигура 5б. Тук по избрана от потребителя релация (например хиперонимната) се изгражда дървото от текущото синонимно множество надолу, както и път нагоре в графа на релацията.

Пътят по подразбиране е ‘най-левият’, т.е. избира се винаги първият от възможните предци в релацията. Потребителят вижда броя на предците за всеки възел и може да избере по кой от тях да се проследи пътят нагоре. По този начин е възможно да се обходи всеки възможен път в тази посока до върха. Релациите, за които може да се използва тази визуализация, са нецикличните релации в БулНет.



Фигура 5б. Информационно поле на системата Chooser – дървовиден изглед

Последният изглед представлява текстово поле, което представя информацията в дадено синонимно множество, както и на неговите директни съседи по релации, както е представено във Фигура 5в.

main	tree	synset
Pos: n ID: ENG20-13562199-n BCS: 4		
хурегпум bg - n: вазоконстриктор:1 (вазоконстриктор)		
хурегпум bg - n: /нервно/ напрежение:1 (нервно напрежение)		
category_domain bg - n: психология:1 (психология)		
literal стрес:1 {стрес}		
usage Страдаше от изтощение и емоционален стрес.		
usage Стресът има вазоконстрикторни функции.		
Definition: състояние на напрежение, възникващо у човека и животните под влияние на увреждащи и ли необичайни по сила физически, психически, инфекционни и др. въздействия, с увреждащ ефект върху телесните функции; действа като вазоконстриктор, води до повишаване на кръвната захар, телесната температура и др.		

Фигура 5в. Информационно поле на системата Chooser – синонимно множество

Тази информация включва: дефиниция на значението; част на речта; стойността по условната класификация *основни концепти*; идентификационния номер; литералите в синонимното множество; примерни употреби; бележки (към синонимното множество или към литерала); всички директни синонимни множества, съседи по релации в БулНет (представени според частта на речта и литералите от синонимното множество).

2.4. Архитектура на системата

При разработването на анотационната система Chooser са следвани най-добрите практики и техники в обектно ориентираното програмиране (Гама и др. 1995). Архитектурата на системата се основава на парадигмата Model – View – Controller, използва няколко популярни шаблона (design patterns) като Observer, Iterator, Composite, Strategy и други и следва основните принципи за абстракция, капсулация, слаба обвързаност на класове и елементи в системата и полиморфизъм. Достъпът до wordnet (БулНет, както и до лексикално-семантичните мрежи за други езици) се извършва чрез библиотеката на системата за wordnet Hydra. Представянето на wordnet остава скрито (абстрактно), но достъпът е централизиран, т.е. всички потребители на системата имат едновременен достъп до една и съща информация (в конкретния случай данните от БулНет и Принстънския wordnet). Това се улеснява и от абстрактния модел на wordnet – като релационна структура, както и от използването на модалния език за заявки към wordnet (Ризов 2006).

2.4.1. Класове

Представя се организацията на класовете и техните функции и отношения, като се пропускат тези класове от Hydra API, които са добре известни (Гама и др. 1995).

Едносъставните и съставните думи в текста, който се аниотира, са представени в класовете Word и Words. Words наследява Word и подменя част от методите на имплементация, като в същото време съдържа списък от обекти от типа Word (Composite). Така едносъставните и съставните думи имат общ интерфейс, като едносъставните единици могат лесно да се свържат в съставни. Word предоставя информация за словоформата, основната форма, избраното значение, данните за думите (те се използват например при оцветяването на текста), както и възможните значения за дадена лексикална единица – дума или съставна дума). Класът Words предоставя и функции за добавяне и премахване на графични думи в рамките на съставните.

Текстовете за аниотиране са организирани в обекти от класа Model като колекции от думи. Model има функции за зареждане от файл и запазване във файл, преминаване към следваща, предишна или произволна (определена с индекс) дума от текста, смяна на итератор.

Преминаването от една дума към следващата става с помощта на итератор. Обектите от този клас имат за задача да имплементират метод, който задава стратегията за това кога дадена дума да се пропусне и кога да стане текуща за модела. Итераторите се прилагат при извикване на методите за преминаване напред или назад в текста – при избор на значение или при навигация със стрелките наляво и надясно. При преминаването се прилага стратегия за оцветяване на текста. Тази стратегия е изведена в класа Colored (капсулация), който по даден Word обект дава определен цвят в зависимост от неговите характеристики.

Комуникацията между визуалните обекти (View) се извършва по шаблона Observer. Обектите, които подлежат на опресняване при въвеждането на определени промени, се абонира за тях и имплементират метода refresh. Обектите, в които се пораждат промени, наследяват Context класа, организиращ абонирането и изпращането на съобщения към абонатите. Много от визуалните обекти се появяват и в двете роли: на Context и на Observer. Така например информационният панел (ViewHolder), който визуализира синонимните множества от wordnet, съдържа три изгледа. Той се абонира за съобщенията на полето за избор (играе ролята на наблюдател). Когато получи съобщение за смяна на текущото синонимно множество с друго, информационният панел следва да се опресни. За целта изпраща съобщение към всичките си абонати (като контекст), които в конкретния случай са трите изгледа за БулНет.

2.2.5. Имплементация

Системата е имплементирана на Python. За GUI се използват стандартните Tkinter и Tix библиотеки. По този начин Chooser се изгражда като системно независимо приложение, което може да работи успешно под почти всяка платформа и към момента е използвана под Linux и Windows.

Цитирана литература

- Гама и др. 1995: Gamma, E., R. Helm, R. Johnson, and J. Vlissides. *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley.
- Ризов 2006: Ризов, Борислав. *Релационни структури за WordNet. Модален подход*. Магистърска теза, СУ „Св. Климент Охридски”.
- Ризов 2008: Rizov, Borislav. *Hydra: A Modal Logic Tool for Wordnet Development, Validation and Exploration*. In: Proceedings of LREC 2008.

3. Общи лингвистични конвенции при анотацията

Светлозара Лесева

Ще бъдат представени общите лингвистични конвенции, следвани при анотацията на БулСемКор. Възприетите правила засягат различни аспекти на работата по анотирането на корпуса и условно могат да се разделят на два етапа – нормализация и същинска анотация, които протичат отчасти паралелно. Същинската анотация от своя страна включва два процеса, които също се извършват успоредно: (i) избор на значение; (ii) попълване на нови значения или литерали в БулНет.

3.1. Нормализация

Нормализацията представлява привеждане на графичните думи в корпуса във вид, подходящ за анотация, и включва различни типове корекции на формите, срещащи се в корпуса, избиране на основна форма и други.

3.1.1. Нормализация на правописа

Нормализацията на правописа може да се извършва както в словоформата, която се среща в корпуса, така и в лемата. Освен коригиране на типографските грешки тук се включват и редица системни замени като:

- Заместване на думи с остарял правопис и подобни с техни съвременни правописни варианти, например. *чувствувам* с *чувствам*.
- Заместване на русизми и други дублети, които са маргинални за речника: *проблема* с *проблем*, *метода* с *метод*, *обряд* с *обред* и т.н., освен ако не са идиоматизирани или фиксирани в съставни думи само в тази форма.
- Заместване на архаизми, остарели думи, казионализми, авторови неологизми и други маргинални за речника думи с техни уместни еквиваленти, например *завъркулка* със *завъртулка*.
- Заместване на думи на друг език, когато е възможно, с уместни български еквиваленти, а в случаите, когато не се накрънява целостта и смисълът на текста – и отстраняването им. Например в (1) поредицата *Kunst der Sekunden und minutenaufnahme* се отстранява, още повече че представлява пояснение към

предходния текст; *Papa Hamlet* се замества с *Татко Хамлет*, както е преведено заглавието на сборника на Арно Холц и Йоханес Шлаф на български, а *Hanstein* се замества с *Ханщайн*.

(1) В произведението *Papa Hamlet* е приложен фотографичният начин на рисуване на последователните моменти „*Kunst der Sekunden und minutenaufnahme*” – методът на моменталните фотографически снимки, както щастливо го е нарекъл *Hanstein*.

- Не се заместват с българските им съответствия наименования на организации като *BBC*, *Deutsche Welle* и др., тъй като оригиналните им названия и/или абривиатури са се наложили като равностойни на, а понякога и предпочитани пред, транскрибираните. В тези случаи чуждоезичните варианти се въвеждат като литерали заедно с българските им съответствия в даденото синонимно множество.
- Допълнителни операции – сливане, разделяне, полуслято изписване на думи според кодифицираните правила в съвременния български книжовен език, когато в корпуса има разминаване с нормата. Частен случай представлява поставянето на разстояние между числителни и специални символи. Например знакът за процент (%) се разделя от предходното числително ($40\% > 40\%$), за да се разпознаят елементите на поредицата като отделни токъни със съответните лемите – 40 и %.

3.1.2. Нормализация на лемите

Нормализацията на лемите включва: (i) избиране на основна форма, тогава, когато на словоформата в корпуса поради омонимия е приписана повече от една лема (извършва се в преданотационния етап); или (ii) попълване на правилната лема, в случай че тя не фигурира (паралелно с анотацията).

Тъй като лемите на съставните думи и фразеологизмите не винаги са композиция от лемите на съставлящите ги словоформи, на този тип словосъчетания се приписва *абстрактна лема* (Савъри 2008), в която конституентите се поставят в съответната форма. Например, в съставното съществително *водосточна тръба* опората *тръба* изисква съгласуване по род на прилагателното *водосточен*, поради което в лемата то се поставя във формата за женски род *водосточна*.

По-долу са разгледани някои по-особени случаи.

3.1.2.1. Основната форма има различни съставлящи или строеж от формата, появяваща се в корпуса. Наблюдава се при съставни единици:

- Някои съставни предлози като *от страна на, в полза на*, следвани от име, длъжност и т.н. на лице, се появяват и във форма с притежателно местоимение (*от негова страна, в негова полза*) вместо с аналитичната форма *от страна на, в полза на + (име на) лице*. Това налага лемата да се приведе към основната форма, възприета в БулНет.
- Сходна конвенция за нормализация е възприета при определянето на основната форма на фразеологизмите, класифицирани като полуидиоми (Тодорова 2010), например *под неговия нос* се лематизира като *под носа на*.
- Местоименните форми в лемата, се поставят в най-неутралната използвана форма: *ми* вместо *аз* в *струва ми се*; *си* вместо *свой* в *затварям си очите*.

3.1.2.2. Замяна на една дума с друга в лемата:

- Заместване на *нямам* с *имам* и т.н. в лемата при отрицателни глаголни форми на лексикализирани глаголно-именни конструкции, например *нямам против > имам против*.
- Привеждане на деривативи към недеривативните им произвеждащи, когато не отговарят на възприетите критерии за въвеждане в wordnet.

3.2. Лингвистични конвенции

3.2.1. Анотационна единица

От гледна точка на анотацията, езиковите единици, които изразяват един концепт, представляват една анотационна единица. Анотационните единици може да са: прости думи (една графична дума); сложни думи (графична дума, чиято основа е образувана от два корена на самостоятелни думи) (ГСБКЕ 1982:28); съставни думи (поредици от две или повече графични думи, изразяващи едно понятие, като не е задължително съставлящите да са в съседство в линейния контекст (Коева 2005); фразеологизми; поредици, съдържащи специални символи; числени означения; и др. Под „дума” се разбира езикова единица, състояща се от една графична дума, или простите и сложните думи според традиционната

терминология. Съставните думи и фразеологизмите по-нататък ще бъдат наричани съставни единици.

Към думите, както и към разпознатите и нормализирани от анотатора съставни единици, са приписани всички значения от Българския wordnet, в които техните лемни участват като литерали. Анотаторът избира най-подходящото измежду тях или добавя ново синонимно множество. В случаите, когато възникват колебания, доколко дадена поредица от графични думи представлява едно понятие, се прилагат критерии за лексикализация (по-подробно в част 14).

3.2.2. Конвенции при въвеждането на синонимни множества в Българския wordnet

В организацията на лексикона в използваната лексикално-семантичната мрежа – Българския wordnet – са възприети редица конвенции, произтичащи от междуезиковата асиметрия най-вече по отношение на принадлежността към част на речта. При разработването на БулНет се следва принципът за запазване на структурата на Принстънския wordnet (Туфиш и др. 2004), който служи като посредник между лексикално-семантичните мрежи, разработени за различни езици. Чрез този принцип се осигурява еквивалентност между съответстващите си синонимни множества за всяка двойка езици.

Синонимните множества в Българския wordnet са разпределени в девет части на речта, четири от които в общи линии съответстват на следните пълнозначни класове думи – съществителни, прилагателни, наречия и глаголи, които съществуват и в Принстънския wordnet, а останалите пет части на речта (без числителните имена) – местоимения, предлози, съюзи, частици и междуметия, са допълнени в БулНет в процеса на анотацията.

3.2.2.1. Едно от основните разминавания между организацията по части на речта в лингвистичната литература и структурата на Принстънския wordnet се наблюдава при числителните, които в Принстънския wordnet са класифицирани по следния начин: числителните бройни и числителните редни с атрибутивна и предикативна функция, като в пример (2), са въведени в синонимни множества с част на речта „прилагателно име”, а числителните в самостоятелна употреба като число с определена стойност се отнасят към класа на съществителните, в едно синонимно множество със съответните съществителни имена, образувани от числителни, например *петица*, *десетица* в (3).

- (2) а. *На двора играеха пет деца.*
б. *Тя е във втори клас.*

в. *Ние сме петима.*

(3) а. *5+5 е равно на 10.*

б. *Числото 9 е едноцифрено.*

3.2.2.2. Друга особеност в структурата на лексикално-семантичната мрежа се наблюдава по отношение на местоименията. В Принстънския wordnet са въведени само част от тях – някои неопределителни, отрицателни и обобщителни местоимения като прилагателни, други като съществителни и като наречия – като са организирани в синонимни множества с част на речта съответно „прилагателно име”, „съществително име”, „наречие”. В еквивалентното множество в БулНет се въвежда бележка към синонимното множество, показваща, че литералите в него се отнасят към частта на речта „местоимение”. По подобен начин чрез бележка към литерала може да се посочи, че само определен литерал от синонимно множество принадлежи към местоименията. Обособяването по части на речта отразява вътрешната организация на категорията според това, дали членовете ѝ субституират: (i) модификатор в именната група; (ii) именна група или (iii) модификатор в глаголната група или в групата на прилагателното (вж. също класификацията у Ницолова 1986). Тази вътрешна подялба е взета предвид при дефинирането на значенията на синонимните множества и формулирането на тълковните дефиниции на местоименията, нефигуриращи в Принстънския wordnet. Те се определят с част на речта „местоимение”, а дефиницията им експлицитно посочва категориално-синтактичната им спецификация.

3.2.2.3. Причастията, които в лексикографската традиция се причисляват към прилагателните въз основа на функциите си – атрибутивна (модификатори в именната група) или предикативна, в Принстънския wordnet са въведени в синонимни множества с част на речта „прилагателно име”, например {*седнал:1, приседнал:1, седящ:1*} – *седящата Мадона*. С бележка към литерала се обозначава, че по произход са нелични глаголни форми.

3.2.2.4. Някои наречия и адвербиални предложни фрази са включени в Принстънския wordnet в синонимни множества като наречия и като прилагателни. Причисляването им към една от двете части на речта се определя на базата на функцията им – модификация в именна група (прилагателни) или модификация в глаголна група (наречия). С оглед на запазването на еквивалентността между БулНет и Принстънския wordnet и коректната анотация, съобразяваща се с преводните еквиваленти и възможностите за заместване със синоними, в тези случаи българските наречия се включват в съответните множества от двете части на речта. Например *презглава* се въвежда като литерал в множествата {*стремителен:2, неудържим:2, главоломен:1, презглава:2*} и

{*стремително:2, неудържимо:1, главоломно:1, презглава:1*}, като в литерала на първото се поставя бележка към литерала, че презглава е наречие. Различното аотиране в примерите (4а) и (4б) отчита възможността за замяна на наречието презглава с всеки от българските синоними от даденото множество, както и с коректните английски съответствия при превод *hasty* (прилагателно) и *precipitately* (наречие).

- (4) а. *бягство презглава*
{*стремителен:2, неудържим:2, главоломен:1, презглава:2*}
б. *Хвърли се да бяга презглава*
{*стремително:2, неудържимо:1, главоломно:1, презглава:1*}.

Аналогично, при количествените наречия *много, малко, повече, по-малко* и др. във функцията им на квантификатори (прилагателни в Принстънския wordnet) принадлежността към класа на наречията също се означава с бележка към синонимното множество или към литерала според случая.

3.2.2.5. Сходен е подходът при частици и съюзи в български, които в английския език (и съответно, ако са включени в Принстънския wordnet) се категоризират като наречия.

3.2.2.6. Словесни формули като *здравей, довиждане, сбогом, благодаря*, които в българската лексикографска традиция са отнасяни към междуметията, са въведени като съществителни по аналогия с класифицирането им в Принстънския wordnet.

3.2.2.7. Тъй като в БулНет фразеологизмите се третират наравно с лексикалните единици, те също получават стойност за част на речта въз основа на категориалните си признаци и синтактичната си функция. В съответствие с това, например фразеологизирани изречения като *ща не ща, искам не искам; пък каквото ще да става* и др., са отнесени към наречията.

3.2.2.8. При създаването на синонимни множества в БулНет, които нямат съответствие в Принстънския wordnet, стойността за част на речта на новосъздадените множества се определя според категориалната им принадлежност в българския език. Значенията се формулират, като се вземат под внимание функцията, синонимите, когато има такива, и коректните преводни еквиваленти.

3.3. Лингвистични конвенции при аотацията

В съответствие с принципите, залегнали при изграждането на БулСемКор, се аотират всички несъставни и съставни думи и фразео-

логизми, като се следват приети общи конвенции. Особеностите на анотацията при различните части на речта са разгледани в съответните раздели.

3.3.1. Общи принципи

Всяка идентифицирана от анотатора лексикална единица в корпуса трябва да има поне едно кореспондиращо значение от БулНет. Когато между наличните кандидати няма подходящ, се проверява дали синонимно множество, изразяващо даденото значение, съществува в Принстънския wordnet. Ако има такова, то се въвежда в БулНет, а когато няма, в Българския wordnet се създава ново синонимно множество без съответствие в Принстънския. Изборът на синонимно множество при семантичната анотация следва да отразява лексикалното значение на единицата в корпуса максимално точно, като се вземат предвид няколко критерия за еквивалентност: (i) заменимост с останалите синоними от синонимното множество, когато има такива; (ii) тълковната дефиниция; (iii) мястото на синонимното множество в структурата на лексикално-семантичната мрежа. Помощна функция имат примерите за употреба, а в някои случаи и формата (при дефективност на парадигмата и др.).

3.3.1.1. (Не)заменимост с други литерали от синонимното множество

Заменимостта на дума от корпуса с останалите литерали на дадено синонимно множество е един от критериите за еквивалентност на значението. Така например литералът *злато* фигурира в четири синонимни множества:

{*златен медал:1, злато:1*}

Дефиниция: *'най-високото отличие в спорта и други състезателни дейности под формата на медал, изработен от злато'*

{*злато:2*}

Дефиниция: *'монети, изсечени от злато'*

{*злато:3*}

Дефиниция: *'богатство, особено във вид на златни монети, накити и други предмети'*

{*злато:4, Au:1*}

Дефиниция: *'мек, ковък благороден метал (с атомен номер 79) със светложълт цвят, висока електро- и топлопроводимост, с*

променлива валентност (първа или трета), химически слабо активен (реагира с халогени и се разтваря в царска вода)'

Литералите *златен медал:1* и *Au:1* позволяват недвусмисленото идентифициране на съответните значения и без справка с тълковната дефиниция. Така в (5) въз основа на невъзможността другият член на тези синонимни множества да замести литерала злато в конкретния контекст, тези две значения са елиминирани от списъка с кандидатите за анотация. Изборът между *злато:2* и *злато:3* е възможен след проверка на дефинициите.

(5) *...твоите идеи са вятър, а моето злато {злато:3} е наслаждение...*

Синонимите в дадено множество не винаги осигуряват необходим и достатъчен критерий за разграничаване на едно значение от друго, главно когато се наблюдава пълно графично съвпадение между членовете на различни множества.

3.3.1.2. Тълковна дефиниция

Поради посочените по-горе причини тълковната дефиниция има основна роля при определянето дали конкретна лексикална единица от корпуса принадлежи към дадено значение.

Съставянето на дефиниции към синонимните множества в БулНет следва определени принципи. Например, в съответствие с принципа по-конкретните значения да се формулират чрез надредни (Милър 1990; Фелбаум 1990), {*злато:4*} се дефинира чрез хиперонима си {*благороден метал:1*}, а {*злато:3*} чрез {*богатство:1*, *имане:1*}. По този начин дефинициите отразяват принадлежността на синонимните множества към по-общи класове, които в конкретния случай са ясно разграничени и улесняват избора на анотатора. От друга страна, безусловното придържане към установените принципи би могло да доведе до неестествени и/или подвеждащи формулировки, както би било в случая със {*злато:2*}, чийто хипероним е {*ценен метал:1*}.

Дефинициите може да съдържат семантични ограничения. Те от своя страна може да бъдат имплицитни – чрез онтологичната категория, към която принадлежи синонимното множество, или експлицитни – чрез условия в самата дефиниция, например:

{*чудесен:2*, *прекрасен:4*}

Дефиниция: *'за време (метеорологични условия) – който е много приятен, или за ден и под. – който се характеризира с приятни или подходящи за определена цел атмосферни условия'*

Важна особеност на wordnet е, че поради организацията на лексикалната информация в синонимни множества, използването на синоними в структурата на тълковната дефиниция или като пояснение към нея се избягва.

3.3.1.3. Позицията на синонимното множество в структурата на лексикално-семантичната мрежа

Синонимните множества, с които даден кандидат за анотация влиза в определени релации, както и самият тип на релациите, в редица случаи улеснява избора. Отчетлив пример за това представлява принадлежността на синонимни множества със съвпадащи членове към различни онтологични класове, например естествени обекти и артефакти {крило:4} (орган за летене) и {крило:3} (крило на самолет), артефакти от различен тип, например части от конструкцията на сгради и на превозни средства {прозорец:3} и {прозорец:2}; институции и сградите, в които се помещават – {училище:1} и {училище:2}; организации, населени места, помещения и т.н. и хората в тях – {село:1} и {село:2}; съдове и тяхната вместимост {бутылка:2, шише:2} и {бутылка:1, шише:1} и мн. др. На свой ред различната позиция на синонимните възли в структурата се отразява в дефиницията.

3.3.1.4. Примери за употребата

При избора между различни значения в процеса на аотирането примерите за употребата на членовете на дадено синонимно множество имат помощна функция, като илюстрират и поясняват допълнително конкретното значение. Нерядко анотаторът може да прецени кой от предложените кандидати за анотация е най-подходящ, съдейки само по примерите. В процеса на работа към наличните примери в БулНет (преводни от Принстънския wordnet, конструирани и извадени от корпуси) се добавят и засвидетелствани в БулСемКор.

3.3.2. Анотация на съставни единици

По-долу ще бъдат разгледани особеностите на анотацията на съставните единици от синтагматична гледна точка, т.е. при реализацията им в текст. Най-общо спецификите произтичат от различните степени на композираност между отделните компоненти.

Под съкращаване ще се разбира всяка поява на съставна дума в нейна непълна, редуцирана форма. Най-общо причините за редукцията са две – текстово и лексикално обусловени (Колет 1997 и др.). Към първия тип се

отнасят контекстовите замени за избягване на повторението в рамките на текста, елипсата при координация, към втория – лексикализирана употреба на главна или подчинена част вместо съставна дума.

3.3.2.1. Съкращаване, обусловено от контекста

При наличие на контекстово обусловена елипса на съставни единици се възприема възстановяването в лемата на пълната (нередуцираната) форма. Мотивацията за това е следната: (i) в същия контекст е възможна замяна на съставната единица с несъставен синоним (ба, бб); (ii) преводът на съставната лексема в други езици може да е несъставна дума.

- (6) а. *Посадиха по няколко ябълкови {ябълка:2, ябълково дърво:1} и крушови дървета {круша:1, крушово дърво:1}.*
б. *Посадиха по няколко ябълки {ябълка:2, ябълково дърво:1} и круши {круша:1, крушово дърво:1}.*

Преводните еквиваленти също съдържат както съставни, така и несъставни единици – {круша:1, крушово дърво:1} – {pear tree:1, pear:2}.

Този подход не се прилага при координация с елипса на свободни словосъчетания, които се анотират индивидуално катко в (7):

- (7) *Те се изработват от бял {бял:1} и червен {червен:1} конец {конец:1}.*

От гледна точка на синтагматичната реализация на съставните единици различаваме следните основни случаи, установени в БулСемКор:

3.3.2.1.1. Елипса на модификатор при съставни съществителни

- (8) *...циркулацията на кръвта, лимфата и останалите течности {телесна течност:1, биологична течност:1}...*

Както се вижда от (8) течност означава течните вещества, произведени от тялото. Подходящият кандидат за анотация е {телесна течност:1, биологична течност:1}. Тъй като в случая няма самостоятелно развито значение на редуцираната форма, а изпускане на елемент, който се подразбира от контекста, основната форма *телесна течност* се възстановява в лемата и *течност* се анотира по посочения начин.

3.3.2.1.2. Елипса на опората при съставни съществителни

В случаите на елипса на опората при координация на съставни единици се възстановява съставната лема, още повече че нередуцираните форми също са приемливи в контекста (9):

- (9) *...следните дялове на мозъка – челен {челен дял:1}, тилен {тилен дял:1}, теменен {теменен дял:1}, слепоочен {слепоочен дял:1}...*

3.3.2.1.3. Елипса на други неопорни конституенти

Освен елипса на модификатор се наблюдават и изпускане на други неопорни конституенти като в примерите в (10) и (11):

- (10) *...възможност да стане друга личност, да се освободи {еманципирам се:1, освобождавам се:4, освободя се:4}, или еманципира {еманципирам се:1, освобождавам се:4, освободя се:4} от дискурса на контрола и господството...*
- (11) *Може да постигнете изненадващо добри резултати и да запазите талията си на русалка, без да {без да:1} се ограничавате и да {без да:1} оставате гладни.*

Допълнителен аргумент в подкрепа на този подход към съставните лексеми е фактът, че в (10) и (11) елипсата е езикова норма. Несъкратените форми са неправилни (*се еманципира*, освен ако не се повтори и съюзът *да*) или маркирани (*без да оставате гладни*).

3.3.2.1.4. Формално съвпадение между редуцирани форми и надредни понятия

Тук се разглежда замяната на еноцентричен хипоним (съставна дума, чиято опора е нейният хипероним) с надредното му понятие. Приема се, че по контекстово обусловени причини в случая се реализира не хипонимът в съкратена форма, а хиперонимът, тъй като е известно, че заместването на дадена лексема със синоними, хипероними, мероними, холоними, както и с единици, намиращи се в други релации с нея, е един от механизмите при изграждането на кохезията и кохерентността на дискурса (Халидей и Хейзън 1976).

- (12) *Около 1931 г. Ернст Руска (Ernst Ruska) построява първия електронен микроскоп {електронен микроскоп:1}. През 1986 г., заедно с други двама физици, създали микроскопа {микроскоп:1}, той получава Нобелова награда за физика.*

Поради изтъкнатите причини във второто изречение *микроскоп* се анигира по посочения начин.

3.3.2.2. Езиково обусловено съкращаване. Несъставни варианти

Несъставните варианти са форми, при които конституент на съставна единица самостоятелно развива значение, синонимно с нередуцираната форма. По-долу ще посочим два основни случая.

3.3.2.2.1. Несъставни варианти с експлицирана опора

Главната дума може самостоятелно да изразява същото значение като съставната. В такъв случай и двата варианта се въвеждат като литерали в синонимното множество. По-долу в (13а) са дадени три от синонимните множества, в които участва литералът *операция*. Поради лексикалната еквивалентност на съставния и несъставния вариант при срещане на съкратената форма в корпуса попълването на пълния вариант в лемата е редундантно (13):

- (13) а. {*операция:1*, *хирургическа операция:1*, *хирургична операция:1*}
{*операция:4*, *военна операция:1*}
{*операция:7*, *математическа операция:1*, *действие:13*,
математическо действие:1}.
- (13) б. *Успехът на операцията* {*операция:1*, *хирургическа операция:1*, *хирургична операция:1*} е ранното откриване на аномалията.

3.3.2.2.2. Несъставни варианти с редуцирана опора

В тези случаи модифициращ конституент се субстантивира, например {*мобилен телефон:1*, *клетъчен телефон:1*, *GSM:1*, *Джи Ес Ем:1*, *джиесем:1*, *мобилен:1*}. В бележката към литерала към субстантивираната форма се поставя означение за субстантивация на прилагателно.

Когато са налице повече от един вариант на съставна единица, следва всички те да се въведат като литерали в БулНет, и думата в корпуса да се аотира с формата, в която се появява в конкретния случай. Синонимията между съставни и несъставни лексеми от този тип открива насоки за изследване на механизмите за развиване на самостоятелни значения – специализация, метонимия и т.н. Наред с това тя представлява и предизвикателство пред лингвистични задачи като установяването на границите на лексемите и автоматичното разпознаване на съставни думи.

3.3.4. Образно-метафорични значения и преносни употреби

Образно-метафоричните значения имат постоянен езиков характер и влизат в системни отношения с други думи в езика. Възприето е да се въвеждат в wordnet като литерали в синонимни множества (Милър 1995). За разлика от тях преносните употреби имат индивидуален характер в конкретен контекст. При преносна употреба в корпуса се избира подходящо пряко значение, като пред формата на съответната дума се поставя специален символ – @, като в (14):

- (14) *Вагоните весело подскочиха, за да @помогнат на изтерзаните пътници да се придвижат до изходите.*

В случая между значенията на помагам като най-подходящо е избрано {помагам:1, помогна:1, оказвам помощ:1, окажа помощ:1} – ‘подкрепям, улеснявам, оказвам съдействие или съм от полза на някого в дейност, работа, ситуация, особено когато не може да се справи сам’.

3.3.5. Заключение

Един от основните резултати в процеса на работа по семантичния корпус е обогатяването на БулНет с нови синонимни множества, допълването и коригирането на съществуващи синонимни множества и създаването на синонимни множества на местоименията, предлозите, съюзите, частиците и междуметията. Извършени са редица оптимизации в организацията на лексикално-семантичната мрежа, например въвеждане на онтологична категория като релация при наречията, местоименията и служебните думи, обогатяване на синонимните множества с допълнителна лингвистична информация, със стилистични и граматични бележки и т.н.

В хода на анотацията са засегнати редица теоретични и практически въпроси, свързани с класифицирането на думите според част на речта и синтактична функция, по-специално категоризацията на фразеологизмите, идентифицирането на границите на съставните думи, разпознаването им в контекст и коректното им анотиране. Решенията, вземани в процеса на работа, са мотивирани от засвидетелстваните в корпуса езикови факти и явления, като са съобразени със схващанията в съвременната лингвистика.

Цитирана литература

- ГСБКЕ 1982: *Грамматика на съвременния български книжовен език. Том 2. Морфология.* Издателство на БАН.
- Коева 2005: Коева, Sv. Inflection Morphology of Bulgarian Multiword Expressions. In: *Computer Applications in Slavic Studies – Proceedings of Azbuki@net, International Conference and Workshop*, Sofia, 201-216.
- Колет 1997: Collet, T. La réduction des unités terminologiques complexes de type syntagmatique. *Meta: Journal des Traducteurs/Meta: Translators' Journal* 42 (1), 193-206.

- Милър 1990: Miller, G. A. Nouns in WordNet: A Lexical Inheritance System. *International Journal of Lexicography* 3 (4), 245-264.
- Милър 1995: Miller, G. A. Building Semantic Concordances: Disambiguation vs. Annotation. In: *AAAI Technical Report SS-95-01*, 92-94.
- Ницолова 1986: Ницолова, Р. *Българските местоимения* София
- Савъри 2008: Savary, A. Computational Inflection of Multi-Word Units. A Contrastive Study of Lexical Approaches. *Linguistic Issues in Language Technology* 1(2), 1–53.
- Тодорова 2010: Тодорова, М. Организация на лингвистичната информация в компютърен морфологичен речник на българските глаголни фразеологизми. В: *Доклади от Петата конференция по лексикология и лексикография 2009*, София, 331-344.
- Халидей и Хейзън 1976: Halliday, M. A. K., and R. Hasan. *Cohesion in English*. London, Longman.
- Туфиш и др. 2004: Tufis, D., D. Cristea and S. Stamou. BalkaNet: Aims, Methods, Results and Perspectives. In: *Romanian Journal on Information Science and Technology* 7 (1-2), 1-32.
- Фелбаум 1990: Fellbaum, C. English Verbs as a Semantic Net. In *International Journal of Lexicography* 3 (4), 278-301.

II. Част

Българският семантично аотиран корпус: ЛИНГВИСТИЧНИ КОНВЕНЦИИ

В тази част са представени някои от по-важните лингвистични конвенции (в рамките на общите), които са приети при решаването на конкретни случаи. Представянето е разделено между интерпретацията на думи и съставни думи, а критериите за аотиране на несъставните думи са групирани по части на речта.

Втора част също не обхваща описанието на всички лингвистични решения, които са взети при създаването на Българския семантично аотиран корпус. Стремехът е бил, от една страна, да се представят по-общите конвенции, а от друга страна, тези, които са свързани с решаването на някои спорни или недостатъчно изяснени въпроси в българската лингвистика.

Както общите, така и частните конвенции, се базират на начален модел, който е прецизиран в процеса на работа от колектива на проекта. В съставянето на Българския семантично аотиран корпус са участвали Светла Коева, Борислав Георгиев (автор на програмата за аотация), Никола Обрешков (компиляция на изходния корпус) и аотаторите (с различна интензивност): Екатерина Търпоманова, Катя Алахверджијева, Мария Тодорова, Николай Раднев, Светлозара Лесева, Христина Кукова и Цветана Димитрова.

Българският семантично аотиран корпус е достъпен на <http://dcl.bas.bg/semcor>

4. Лингвистични конвенции при анотация на съществителните

Екатерина Търпоманова

Повечето класификации на съществителните имена приемат разделението на съществителните на съществителни собствени и съществителни нарицателни (ГСБКЕ 1993; Буров 2004; Ницолова 2008). Наричателните реферират към класове от предмети или отделен представител от класа, а собствените – към уникални същности. Според Ницолова нарицателните се делят на дуални – предмет, съставен от две части, и недуални. Недуалните от своя страна се категоризират като броими, които имат форма за единствено и множествено число и могат да се съчетават с бройни числителни, и неброими – обикновено имат форма само за единствено или множествено число и не се съчетават с бройни числителни. Броимите са два вида – за лица и нелица, а неброимите – за маса (веществени или абстрактни) и събирателни (Ницолова 2008:48-52). Куцаров разделя съществителните на понятия за материални обекти, понятия за абстрактни явления, понятия за качества и признаци и понятия за действия и състояния. Авторът представя лексико-граматичните разрези при съществителните въз основа на противопоставянията – лични и нелични, нарицателни и собствени, броими и неброими, съществителни с веществено и с отвлечено значение, основни и оценъчни (умалителни и увеличителни) съществителни (Куцаров 2007:59-65).

При описанието на съществителните имена в настоящата работа отделно са разгледани съществителните нарицателни и съществителните собствени имена, тъй като това основно разграничение е релевантно за подхода при аотиране. Представена е и опозицията между конкретни и абстрактни названия, която също предполага различен подход на анотация. При нарицателните са представени само подвидовете, които могат да представляват предизвикателства при процеса на анотация.

4.1. Нормализация на текста с оглед на съществителните

В някои откъси от корпуса се срещат остарели, диалектни или разговорни съществителни, както и съставени от автора на текста думи или чужди думи, които не са възприети в българския език. Подходът към тях зависи от преценката на аотиращия, като стремежът е, от една страна, да се запази автентичността на текстовете, а от друга – да не се нарушава структурата на БулНет и да не се създават синонимни множества с маргинално значение. Най-приемливият вариант е замяна с дума, която

е възприета в книжовната норма и съществува в БулНет, но при условие че се запазва смисълът на текста.

- (1) *Сложността и дълбочината на „50/50” идва не от бекграунда, а от @инфилтрацията на образите и знаците, различните светове и култури заедно с предразсъдъците им.*

В пример (1) думата *бекграунд* е заменена с по-разпространената *фон* и е свързана със синонимното множество {фон:1} ‘декорът, оформящ задния план на сцената’ с паралел в английски {backdrop:1, background:1, backcloth:1}, докато инфилтрация се интерпретира като преносна употреба и е съотнесена с {прецеждане:1, просмукване:1, инфилтриране:1, инфилтрация:1} ‘бавно преминаване, навлизане на течност през филтрираща среда или твърда преграда’.

За съпоставка съществителното *пангар* в (2) се добавя в БулНет, въпреки че е остаряла дума, тъй като назовава конкретна реалия, която не може да се замени с друга дума. Множеството с литерал {пангар:1} ‘помещение в църква, в което се продават свеци’ се представя като хипоним на {стая:1, помещение:1}.

- (2) *В черква хаджи Генчо ходи всеки ден; в неделя и по-големите празници той седи в големия трон при пангара и продава свеци...*

Изречението *На ъпсурта* в (3) се отстранява от текста – несъществуващата дума *ъпсурт* не се добавя в БулНет, а премахването ѝ не накърнява целостта на текста.

- (3) *Не се замисляш дали това е черна комедия или бяла трагедия, защото отдавна не осъзнаваш темповете на абсурда. На ъпсурта. И вървиш ...*

4.2. Съществителни нарицателни

Анотацията на съществителни нарицателни имена в Българския семантично аотиран корпус следва възприетите правила. За основна форма на съществителните се приема нечленуваната форма в единствено число, с някои изключения. Следващата стъпка е избор и свързване със съответното синонимно множество от БулНет, което представя значението на съществителното, като се вземат предвид синонимите, тълковната дефиниция, положението на множеството в структурата на wordnet – хиперонимните и хипонимните релации, както и останалите релации, съответствията в Принстънския wordnet, примерите за употреба, бележките към литералите и синонимните множества.

4.2.1. Съществителни за конкретни понятия

Определящи за анотацията на конкретните понятия обикновено са синонимите и дефиницията, а помощна роля играе дефиницията на значението към съответното множество в английски, като по-рядко се налага да се вземат под внимание мястото на възела в мрежата и релациите с други множества.

За съществителното *витрини* в (4) има два кандидата за аотиране със съответните дефиниции:

- (4) *Обхождаш с очи фасади и натруфени витрини.*

{витрина:1}

Дефиниция: ‘голям прозорец – остъклена част от фасадата на магазин, където обикновено се излагат стоки, които се продават в магазина’

{витрина:2}

Дефиниция: ‘изцяло или отчасти остъклен шкаф, обикновено в музей или изложбена зала, в който се съхраняват и излагат ценни и редки предмети’

От контекста в изречението се разбира, че става въпрос за витрина на магазин, следователно се избира {витрина:1}.

Омонимията и полисемията при конкретните съществителни рядко създават проблеми при анотацията, тъй като контекстът в откъса обикновено е достатъчен, за да се определи значението на думата – (5) и (6).

- (5) *По масата се въргаляха свеци, чаши, препълнени пепелници и, разбира се, публиката.*

- (6) *В тази минута бе изпръскан с вода, защото огромен бизон излезе от потока на брега, след това бе повдигнат във въздуха и хвърлен назад сред подвижната маса.*

За съществителното *маса* изборът е между тринадесет варианта:

{множество:7, куп:7, маса:12}

Дефиниция: ‘голям неопределен брой обекти от даден вид’

{маса:9, маси:1, народ:3}

Дефиниция: ‘обикновените хора, схващани като цяло’

{маса:10}

Дефиниция: *‘свойство на телата, което се изразява в силата, с която гравитационното поле действа върху даден обект’*

{*маса:1, низина:1*}

Дефиниция: *‘основната част, широките слоеве от обществена група или обикновените членове на дадена организация’*

{*куп:1, купчина:1, маса:2*}

Дефиниция: *‘голям брой, количество, размер, обем, степен и т.н. от нещо’*

{*маса:3*}

Дефиниция: *‘обем от вещество, материал с неопределена форма’*

{*трапеза:1, маса:4, храна:2, ядене:3*}

Дефиниция: *‘храна, ястия, сервирани при дадено хранене’*

{*маса:5*}

Дефиниция: *‘мебел, част от обзавеждането, върху която се сервира храната’*

{*маса:7*}

Дефиниция: *‘мебел с плоска хоризонтална повърхност, поддържана от крака, която се използва за хранене, писане и др.’*

{*трапеза:2, маса 8, маса за хранене:1*}

Дефиниция: *‘маса със сервираните върху нея ястия’*

{*маса:11, струпване:4*}

Дефиниция: *‘хаотична съвкупност от множество подобни неща’*

{*дузини:1, десетки:1, камара:1, куп:5, купища:1, грамада:1, планина:2, маса:6, тонове:1, сума:3, сюртия:1*}

Дефиниция: *‘голям брой или количество от нещо’*

Тринайсетте варианта могат да бъдат обособени в три групи значения – мебел, физическа величина и голямо количество. В пример (5) думата *маса* е конкретно название и изборът се свежда до множествата, обозначаващи мебел. Избира се синонимното множество {*маса:5*}. В пример (6) става въпрос за голямо количество бизони, избира се множеството {*маса:11, струпване:4*}.

4.2.2. Съществителни за абстрактни понятия

Анотацията на абстрактни понятия изисква прегледа на всички критерии за избор на значение. Често синонимите и дефиницията им се оказват недостатъчни и е необходимо подробно изследване на лексикално-семантичната мрежа, както и на съответствията в Принстънския wordnet. Примерите за употреба също са ориентир за анотацията. Многозначността е съществен проблем, тъй като много често подзначенията са твърде разслоени, а тълковните дефиниции са много близки и не допринасят за изясняването им. Това се дължи, от една страна, на мрежовидната структура на wordnet – синонимни множества с (поне един) еднакъв литерал и сходна дефиниция попадат в различна йерархия по отношение на хиперонимната релация. В тези случаи определящ за анотацията на съществителното е хиперонимът на избраното множество, както е в (7), (8), (9), (10). От друга страна, многозначността в БулНет се поражда при превод на синонимни множества от английски – думи, които принадлежат на различни множества в английски, или думи в хиперонимно-хипонимни отношения се превеждат на български с една и съща дума. В този случай от решаващо значение е съпоставката на избраните варианти в БулНет с английските им съответствия, например в (11).

(7) *...интелектуално ниво над средното и художествен усет около гениалното...*

За съществителното *усет* съществуват 4 възможности за анотация:

{*усещане:1, усет:1, сетивна способност:1*}

Дефиниция: *‘вродена способност на организма да възприема и обработва стимули от външния свят’*

Хипероним: [*хипероним: умствена способност:1*]

{*чувство:2, усет:2*}

Дефиниция: *‘интуитивно познаване или способност за разбиране в дадена област или вроден талант за нещо’*

Хипероним: [*хипероним: интуиция:1*]

{*чувство:4, усет:3*}

Дефиниция: *‘инстинктивно разбиране, способност или умение за нещо’*

Хипероним: [*хипероним: осъзнаване:1, съзнание:2*]

{*усет:4; нос:5; нюх:1*}

Дефиниция: *‘вродено умение за правилна или подходяща преценка’*

Хипероним: [*хипероним: умение:1, вещина:2*]

Четири синонимни множества, в които думата *усет* участва като литерал, са в различни разклонения на дървовидната хиперонимохипонимна структура на БулНет и според хиперонима на множествата *усет* се определя като проява на сетивна способност, на интуиция, на съзнание и на умение. В конкретния случай е избрано множеството {чувство:2, усет:2}.

За абстрактното съществително *среща* в (8), (9) и (10) вариантите за анотация са шест:

(8) *Европейският съвет определи на **срещата** си в Копенхаген критериите, които трябва да удовлетворява дадена страна...*

(9) *Хората, които лежаха в тревата, се изправиха и погледите на всички се насочиха към новодошлите. Отначало не изглеждаха много доволни от тази **среща**...*

(10) *Отговорите ще научите при следващата ни **среща**.*

{среща:5, събрание:1}

Дефиниция: 'събиране на множество хора с цел обсъждане на някакъв въпрос, вземане на решение и др.'

{мач:1, среща:1, схватка:3}

Дефиниция: 'двубой между отбори или състезатели в даден спорт или бойно изкуство'

{среща:2, уговорка:4, ангажимент:3}

Дефиниция: 'предварително насрочено събиране, най-често между двама потенциални партньори или приятели'

{среща:3, случайна среща:1, неочаквана среща:1}

Дефиниция: 'случайно събиране или идване на едно място с други хора'

{среща:4, попадане:1, натъкване:1}

Дефиниция: 'случайно и неочаквано намиране, откриване на или влизане в досег с някого или нещо'

{срещане:2, среща:6}

Дефиниция: 'събиране на хора с някаква цел'

Подходящи кандидати за свързване със съществителното *среща* в пример (8) са синонимните множества {среща:5, събрание:1} и {срещане:2, среща:6}. Избрано е множеството {среща:5, събрание:1}, като освен тълковната дефиниция е взето предвид и разположението на думата в дървото: хипероним {събиране:4, събор:1, сбор:3, множество:2,

събрание:4} и хипоними {конференция:1}; {съвет:1, съвещание:1, заседание:1}; {форум:1}; {среща на високо равнище:1} и {конгрес:2}. В пример (9) става въпрос за случайна среща, затова изборът е {среща:3, случайна среща:1, неочаквана среща:1}. От контекста в (10) се разбира, че срещата е планирана, но не официална и не приятелска, следователно се избира множеството {срещане:2, среща:6} с хипероним {събрание:3, събиране:14} и с хипоними {контактуване:2, общуване:2}; {сесия:1}; {посещение:2, посещаване:1, гостуване:2, визита:2} и съответствие в английски {meeting:2, coming together:2}.

Съществителното *пътуване* от (11) участва като литерал в три синонимни множества, които са свързани с хиперонимна релация.

(11) *Ще имате шансове да реализирате **пътуване** по работа още към средата на разглеждания месец.*

{пътуване:1}

Дефиниция: 'придвижване (с или без транспортно средство) между две точки в пространството, от едно място на друго'

Хипероним: [хипероним: движение:1, придвижване:4]

Еквивалент в английски: {travel:2, traveling:2, travelling:1}

{пътуване:2, пътешествие:2}

Дефиниция: 'придвижване от едно място на друго, обикновено с определена цел'

Хипероним: [хипероним: пътуване:1]

Еквивалент в английски: {journey:1, journeying:1}

{пътуване:3}

Дефиниция: 'отделен случай на пътуване донякъде (и обратно) с определена цел'

Хипероним: [хипероним: пътуване:2, пътешествие:2]

Еквивалент в английски: {trip:3}

В случая {пътуване:1} е хипероним на {пътуване:2}, което от своя страна е хипероним на {пътуване:3}, като значението се стеснява от {пътуване:1} към {пътуване:3}. Думата *пътуване* от (11) е свързана с най-общото понятие – синонимното множество {пътуване:1}.

4.2.3. Съществителни за лица

Проблемите при аотирането на съществителни имена за лица засягат предимно названията за професии или други отношения между

хората, които са свързани с разделянето по пол, което влияе върху граматическия род на думата. Анотацията на такива названия налага, от една страна, съобразяване със структурата на Принстънския wordnet и съответно вземане под внимание на спецификата на английския език, а от друга – обогатяването на БулНет със съществителни от женски род, които липсват в английски.

Случаите, в които в английски са възникнали названия от мъжки и женски род за полово разграничение на едно и също понятие, обикновено не създават проблеми при анотацията.

(12) ... *учителят по български му доверил за своя предишна любов към ученичка.*

В Българския wordnet съществуват отделни синонимни множества {ученик:4} с еквивалент {schoolboy:1} и {ученичка:1} с еквивалент {schoolgirl:1}, с общ хипероним {ученик:3} – {schoolchild:1, school-age child:1, pupil:2}. Думата ученичка от пример (12) естествено се свързва с множеството {ученичка:1}.

По-специфичен е подходът към думи от мъжки род в български, които означават общо понятие без оглед на пола и представител от мъжки пол, какъвто е случаят с {ученик:3} и {ученик:4}. Английските съответствия са отделни думи, обикновено в хиперонимно-хипонимни отношения. Тези случаи се разрешават чрез изследване на конкретния контекст, за да се установи кой от двата варианта се има предвид – (13).

(13) *Съпругът ѝ пристигна в тия затънтени земи като независим представител на няколко страни...*

За съществителното *съпруг* съществуват два кандидата за анотация, като първото множество е хипероним на второто:

{съпруг:1, брачен партньор:1, партньор:4}

Дефиниция: 'партньор в живота, с когото човек е встъпил в брак'

Еквивалент в английски: {spouse:1, partner:3, married person:1, mate:7, better half:1}

{съпруг:2, мъж:2, законен съпруг:1, женен мъж:1}

Дефиниция: 'мъж, встъпил в брак, партньорът от мъжки пол в брачната двойка'

Еквивалент в английски: {husband:1, hubby:1, married man:1}

Смисълът на изречението показва, че става въпрос за мъжа в брачната двойка, следователно думата *съпруг* от пример (13) се аотира като {съпруг:2, мъж:2, законен съпруг:1, женен мъж:1}.

Случаите, в които на български е възникнала дума от женски род, която е свързана с представител от женски пол на професия, занимание и под., а в английски не съществува такава дума, се разрешават, като думата от женски род се добавя в синонимното множество заедно с думата от мъжки род, като в полето *бележка към синонимно множество* се отбелязва, че множеството обединява лексикални единици от мъжки род и женски род, например {рецепционист:1, рецепционистка:1}. Думи от женски род, които се срещат в разговорната реч, но не са приети в езиковата норма, се привеждат в мъжки род като част от нормализацията на текста, например *министърка, адвокатка*. Към БулНет са добавени само думи от женски род, които са били срещнати в корпуса. Добавянето е по преценка на аотиращия, когато думата от женски род е възприета в книжовния език и може да бъде включена в синонимното множество, т.е. отговаря на тълковната дефиниция – (14).

(14) *Секретарката на Тес седеше на подреденото емайлирано бюро и работеше с компютър.*

Съществителното от женски род *секретарка* се включва като литерал в множеството {секретар:1, секретарка:1} ‘помощник, чиито задължения са да води кореспонденцията и документацията, да планира и да организира бизнес срещите на човека или организацията, за които работи’, но не и към множеството {секретар:2} ‘официално лице, оглавяващо някои от висшите органи на държавната власт със статут на министерство, в дадени страни, напр. САЩ и Великобритания’.

4.2.4. Съществителни в множествено число

Освен съществителните *pluralia tantum* (напр. *очила*), съществуват съществителни, които имат форма за единствено число, но формата им за множествено число реферира към различно понятие. В такъв случай се избира основна форма в множествено число. В корпуса са засвидетелствани два основни случая: субстантивирани прилагателни, които означават партийна принадлежност или принадлежност към спортен отбор (Динева 2005), и съществителни или субстантивирани числителни, означаващи голямо количество. Първият случай може да се илюстрира със синонимните множества {Съюз на демократичните сили:1, СДС:1, сини:1} и {Левски:1, сини:2}, а вторият – с множеството {дузини:1, десетки:1, камара:1, куп:5, купища:1, грамада:1, планина:2, маса:6, тонове:1, сума:3, сюрия:1}. Думата *левити* от корпуса се привежда в основна форма *леви* и се свързва с множеството {леви:1, левица:1} – (15).

- (15) Докато мнозина от **левите** днес говорят с езика на свободата, тяхната реална политика ограничава индивидуалния избор...

4.2.5. Умалителни съществителни

Характерно за българския език е, че много съществителни образуват умалителни форми чрез добавяне на суфикс. Към добавянето на умалителните форми в БулНет се подхожда с внимание, за да не се стига до немотивирано разрастване на лексикално-семантичната мрежа. В семантично отношение умалителният суфикс може да означава физическа умаленост на денотата (*маса – масичка*) или да изразява субективно отношение към денотата (*кафе – кафенце*) (Зидарова). В редица случаи деминутивите имат собствена лексикална семантика, различна от тази на произвеждащата дума: *луна – ‘небесно тяло, спътник на Земята’, и луничка – ‘кафяво петънце по кожата’* (Зидарова). Умалителните съществителни се добавят в БулНет по преценка на аотирацията, при условие че отговарят на следните критерии: съществува паралел в английски; деминутивът представлява отделно понятие от съществителното, от което е произведен; участва в устойчиви изрази, в които не може да се трансформира в думата, от която е произведен. Деминутивът може да се добави към синонимното множество с думата, от която произхожда, ако аотирацията прецени, че честотата му на употреба в езика е висока. В тези случаи основната форма на съществителното в корпуса остава умалителна. Ако деминутивът не отговаря на критериите, се избира за основна форма съществителното, от което е образуван, и се свързва със съответното синонимно множество в БулНет.

В БулНет под хиперонима {*малко:3*} са добавени някои умалителни съществителни, които обозначават малките на животни и имат паралели в Принстънския wordnet. В английски обаче това са суплетивни форми по отношение на названията за възрастно животно, докато преводите в български са както суплетивни – *агне, агънце/овца, теле, теленце/крава*, така и деминутивни форми, образувани чрез прибавяне на умалителен суфикс – *конче/кон, прасенце/прасе*. В пример (16) думата *кончето* може да се аотира по два начина според преценката на аотирацията.

- (16) ...*после се отбил към Борунчешма да напои кончето.*

Ако от контекста се разбира, че става въпрос за малко животно, основната форма остава умалителна и думата от корпуса се свързва със синонимното множество {*конче:3*} ‘*малко на животните от вида Кон (Equus caballus)*’. Ако текстът подсказва, че думата е гальовно обръщение към животното, се избира основна форма *кон* и думата се свързва

с множеството {кон:4} ‘представител на едноименния род (*Equus*) нецифтокопитни тревопасни животни от едноименното семейство, опитомени от човека и използвани за езда, спорт, лов, впрегатна работа и под.’. Случаят е разрешен, като думата от текста е свързана с множеството {кон:4}, тъй като от контекста в корпуса става ясно, че деминутивът е употребен като гальовно название.

От деминутивите в пример (17) в умалителна основна форма остава само *брадичка*, която се свързва с множеството {брада:1, брадичка:1} ‘издадената напред крайна част на долната челюст при човека’, различно от множеството {брада:3} ‘окосмение в долната част на лицето (бузите и брадичката) при мъжете, което обикновено се бръсне или оформя чрез подстригване’. За носленце и гушка се избира основна форма съответно нос и гуша.

(17) ...това хубаво **носленце**, тия мънички алени уста, тая мъничка **брадичка**, под която се крие бяла **гушка** гургуличина.

Умалителната форма *масичка* е добавена в БулНет като хипоним на синонимното множество {маса:7}, тъй като участва в устойчиви изрази като *масичка за кафе*, *холна масичка*, в които замяната с *маса* не е приемлива – (18).

(18) Съдейки по разгърнатото списание „Вог”, оставено на малката **масичка** за кафе, Бен предположи, че в момента пациентката на Тес е жена.

4.2.6. Преносна употреба на съществителните

Съществителните имена, употребени преносно в текстовете от корпуса, се анализират според общите правила за анотация на преносна употреба. Ако думата има установено преносно (образно-метафорично) значение, разпространена е в езика, участва в устойчиви изрази или има паралел в Принстънския *wogdnet*, тя се добавя към БулНет, например:

{ръка:5}

Дефиниция: ‘оказване на помощ за извършването на нещо’

Примерна употреба: *И той ни е помагал по-рано, трябва да му подадем ръка.*

{празни думи:1, брътвеж:1, вятър:2, празни приказки:2}

Дефиниция: ‘безсъдържателни, обикновено претенциозни твърдения’

Ако не отговаря на тези критерии, думата се свързва с множеството, което отговаря на основното ѝ значение, като се маркира със знака @ – (19).

- (19) *Хиляди училища, @разсадници на доброто, пълни с отбрани педагози, тъй майсторски преобръщат лошите деца в добри, че не можеш да ги познаеш.*

Съществителното *разсадници*, което е употребено метафорично, е свързано с множеството {разсадник:1} 'леха, в която се отглеждат млади растения, преди да се разсадят'.

4.2.7. Обогаляване на БулНет със съществителни нарицателни

Освен с културно специфични и езиково специфични думи, в процеса на анотация БулНет е разширен със съществителни нарицателни, които съществуват в английски език, но не са добавени към Принстънския wordnet. Различават се два случая. В първия става въпрос за съществителни, които обикновено са производни от прилагателни или глаголи, и в Принстънския wordnet съществува само думата, от която произхождат, но не и самото производно съществително. Във втория случай в Принстънския wordnet няма подходящо значение, съответстващо на значението в корпуса. И в двата случая съществителното се добавя в БулНет на подходящото място в дървото.

- (20) *...напомня за изповед, кодирана в поетична образност.*

В Принстънския wordnet съществуват две синонимни множества, съдържащи литерал *figurative* (*преносен*), но производното му съществително *figurativeness* (*образност*) не е добавено, затова в БулНет се създава ново множество:

{образност:1}

Дефиниция: *'представяне чрез образни, поетични, метафорични средства'*

Хипероним: [хипероним: представяне:7, изобразяване:2]

Бележка към литерала: *преводен еквивалент figurativeness*

Синонимното множество {образност:1} се добавя под хипероним {представяне:7, изобразяване:2}, който е най-подходящ в структурата на БулНет.

Синонимните множества {интрига:1, машинация:1} 'заплетени тайни действия, често нечестни, за постигане на някаква цел' и {интрига:3} 'тайна любовна афера' не могат да бъдат асоциирани със значението на думата *интрига*, срещнато в корпуса (21), затова се създава ново множество {интрига:2} 'интересни, ключови моменти от действието, които оформят сюжетната линия на художествено произведение', което се поставя под хипероним {действие:14}.

(21) ...отвъд дисциплината на обективните ситуации и масовата привлекателност на фабулната **интрига**...

Справката с английски речници показва, че думата *intrigue* (*интрига*) се среща в това значение и в английски език и това се отразява като бележка към литерала: *преводен еквивалент intrigue*.

4.3. Съществителни собствени

Ако съществителното собствено не е включено в БулНет, първата стъпка е да се потърси еквивалент в Принстънския wordnet и да се създаде съответното синонимно множество в български, например (22).

(22) *Отишъл Буш при Господ и го попитал: „Господи, кога ще свърши войната в Афганистан, кога ще спре арабският тероризъм, кога ще бъде овладяна престъпността...”*

В Принстънския wordnet съществува множеството {*Bush:3, George Bush:2, George W. Bush:1, George Walker Bush:1, President Bush:2, President George W. Bush:1, DUBYUH:1, DUBYA:1*}, следователно в българския се създава множеството {*Буш:1, Джордж Буш:1, Джордж Уокър Буш:1*}, за основна форма на името от корпуса се избира *Буш* и се свързва със съответното синонимно множество в БулНет. От контекста е ясно, че става въпрос за Буш-син, 43-ия президент на САЩ, а не за Буш-баща, 42-ия американски президент.

Ако съществителното собствено име не е включено в Принстънския wordnet, има два подхода. Първият е името да се добави в БулНет без паралел в английския като хипоним на вида обект, който назовава – например множеството {*Иван Вазов:1, Ив. Вазов:1, Вазов:1, Иван Минчов Вазов:1, патриарх на българската литература:1*} се добавя като хипоним на множеството {*писател:1, писателка:1, автор:2, авторка:1*}. Ако анотиращият прецени, че името не е достатъчно важно или значително, за да се добави към БулНет, се възприема вторият подход – за основна форма се избира онтологичната категория на името, след което се свързва със съответното множество в БулНет – например за *Грозда* се избира основна форма *малко име* и се анотира със съответното множество {*малко име:1, собствено име:1*}, а за *Паскалев* се избира основна форма *фамилия* и се свързва с множеството {*фамилия:1, фамилно име:1*} и т.н.

Благодарение на работата с корпуса, Българският wordnet е обогатен с множество съществителни собствени имена:

- географски названия – имена на градове под хипероним {*град:2, метрополис:1, голям град:1*} – *Плевен, Ямбол* и др.; малки градове с хипероним {*град:5, градче:1, неголям град:1*} –

Банско, Козлодуй, Перуцица и др.; реки под хипероним {река:1} – *Струма, Марица* и др.; области под хипероним {географска област:1} – *Добруджа* и др.;

- имена на личности – писатели под хипероним {писател:1, писателка:1, автор:2, авторка:1} – *Иван Вазов, Захари Стоянов* и др.; поети под хипероним {поет:1, стихоплетец:1, стихотворец:1} – *Христо Ботев, Никола Вапцаров, Пенчо Славейков* и др.; политици и общественици под съответните хипероними – *Петър Стоянов, Георги Първанов, Симеон Сакскобургготски, Бойко Борисов* и др.;
- имена на литературни герои под хипероним {герой:3, персонаж:2, образ:5, фикционален герой:1, измислен герой:1} – *Андрешко, бай Ганьо, Крали Марко* и др.;
- имена на институции – добавени са някои институции на Европейския съюз, които не са отразени в структурата на Принстънския wordnet, например множеството {Европейска комисия:1, ЕК:1, Еврокомисия:1} с хипероним {комисия:1, комитет:1}, свързано и с релация *Част от* с множеството {Европейски съюз:1, Европа:1, ЕС:1, ЕУ:1, Европейска общност:1}; някои важни български институции като *СУ „Св. Климент Охридски”, БНТ, БНР*, като се свързват със съответните хипероними – *университет, телевизия, радио* и т.н.

По-интересни са случаите, в които съществителното собствено име е добавено като синонимно множество в БулНет, но конкретната форма, с която се среща в корпуса, не се добавя към множеството – (23).

(23) *Попитали един човек: „Ако бяхте в една стая със Саддам Хюсеин, Милошевич и Бил Гейтс и имате пистолет с 2 патрона, какво бихте направили?” Отговорът бил: „Двата патрона в **Бил** и после с дръжката по главата, по главата...”*

В лексикално-семантичната мрежа съществува синонимно множество {*Бил Гейтс:1, Гейтс:1, Уилям Хенри Гейтс:1*}, но не е приемливо да се добави и литерал *Бил* (за разлика от *Саддам* – множеството е {*Саддам Хюсеин:1, Саддам:1, Саддам бин Хюсеин ал Такрити:1*}, тъй като личността е популярна и с малкото си име). В случая *Бил Гейтс* се свързва със съответното синонимно множество, а *Бил* се аотира като *малко име*.

Интересен е и проблемът със съществителни собствени имена, които са значещи думи – обикновено това са прякори и някои видове названия (на компании, кораби, квартали и др.). В тези случаи се взема под внимание дали името може да бъде преведено – (24) и (25).

(24) ...и когато гласът на **Ястребово око** пресекна...

Името *Ястребово око* се аотира според значението на лексикалните единици – *Ястребово* се свързва с множеството {ястребов:2} ‘който се отнася или е присъщ на ястреб’, а *око* – с множеството {око:2} ‘чифтен орган на зрението’.

(25) ...на Столичния център за работа с деца в ж. к. „**Младост 1**”.

Названието *Младост* се аотира като {име:3, наименование:1}.

Особена категория са названията на обществени реалии, които се променят във времето – имена на политически партии, политически или други публични личности, министерства, държавни органи, институции, обществени организации. Поради невъзможност да бъдат проследявани и коригирани при промяна, е взето решение добавянето им в БулНет да става според актуалната към момента на добавяне ситуация. В процеса на работа, при повторно срещане на такъв тип название, което е претърпяло някаква промяна, тя се отразява в БулНет. Например към синонимното множество {Национално движение Симеон Втори:1, царско движение:1, НДСВ:1} е добавен литерал *Национално движение за стабилност и възход:1*, а в тълковната дефиниция е уточнено, че името на партията е сменено на конгресите през 2007 и 2008 г.

4.4. Съкращения

Съкратените форми на съществителни имена и именни групи се добавят в БулНет като литерали към синонимното множество, което съдържа съответната разгърната форма, ако имат широко разпространение в езика. Добавени са често употребявани съкращения на титли и обръщения – *д-р, г-н, г-жа*; съкращения на мерни единици – *см, сек., кг*; абривиатури на институции и организации – *БНР, БНТ, ФБР*; имена на известни личности със съкратена форма на малкото име – *Хр. Ботев, Ив. Вазов* и др. Допустимо е добавянето на съкратени форми на латиница, ако се употребяват в български – *kg, EU, FBI, bTV*. Изписването на съкратените форми следва правилата, възприети в българския правопис. При аотирането на съкращения се запазва основната форма и думата се свързва със съответното синонимно множество в БулНет. Ако съкращението не се е установило в езика, то не се добавя към БулНет, а се аотира като {съкратена форма:1, съкращение:1, абривиатура:1} – (26) и (27).

(26) *Пред **bTV** в неделя той каза, че...*

(27) ...създаваненапо-добриусловияидостатъчноресурсизаизграждане на постоянно действаща Система за контрол, поддържане и непрекъснато повишаване на качеството (СКПНПК).

В пример (26) основната форма остава *bTV* и думата се свързва с множеството {*bTV:1, Би Ти Ви:1*}. В (27) абревиатурата СКПНПК не е разпространена, избира се основна форма *абрeвиатура* и думата се свързва със синонимното множество {*съкратена форма:1, съкращение:1, абрeвиатура:1*}.

4.5. Заключение

При аотирането на съществителните имена се спазват общите анотационни конвенции, като се вземат предвид някои по-особени случаи, произтичащи от междуезиковата асиметрия. В тези случаи аотиращите преценяват внимателно дали думата трябва да се добави към БулНет, избора на основна форма и коректното аотиране. Работата с корпуса допринася и за разширяването на БулНет с нови синонимни множества без паралел в английски.

Цитирана литература

Буров 2004: Буров Ст. *Познанието в езика на българите. Граматично изследване на концептуалната категоризация на предметността*. Велико Търново.

ГСБКЕ 1993: *Граматика на съвременния български език, том II, Морфология*. София.

Динева 2005: Динева М. *Възникване на хибридни граматични форми от взаимовръзката на прилагателни имена с други части на речта в три славянски езика*.

http://www.belb.net/doktoranti/dineva_monika/referat7.htm

Зидарова: Зидарова В. *Деминутиви и лексикална номинация*. <http://www.belb.net/cat.php?personal=zidarova>

Куцаров 2007: Куцаров Ив. *Теоретична граматика на българския език. Морфология*. Пловдив.

Ницолова 2008: Ницолова Р. *Българска граматика. Морфология*. София.

5. Лингвистични конвенции при анотация на глаголите

Светлозара Лесева

В настоящото изложение ще бъдат разгледани както глаголни, така и неглаголни предикати, които условно причисляваме към тази част на речта. Анотацията при глаголите се подчинява на общите принципи, изложени в част 3, а именно: (i) (значителна степен на) семантична еквивалентност с други членове на даденото синонимно множество; (ii) съответствие с тълковната дефиниция; (iii) позицията в структурата на лексикално-семантичната мрежа. Особеност на предикатите е, че поради характера им на свойства на същности или на релации между същности „един от най-важните критерии за разграничаване на значенията на даден глагол е чрез определяне на броя и лексикално-синтактичната реализация на неговите аргументи,“ (Коева и др. 2008: 60). В хода на анотацията отнасянето на дадена глаголна форма или съчетание на глагол и друга част на речта в корпуса към определено глаголно значение се извършва чрез анализиране на елементите от аргументната структура в контекста и интерпретиране на отношенията между тях и предиката.

Както при останалите части на речта, така и при глаголите, ключова роля играе правилното идентифициране на графичните думи, които лексикализират дадено значение. Поради характера на формообразуването, словообразуването и синтаксиса им една от особеностите на предикатите е значителната омонимия между различни форми на една и съща лексема и между форми на различни лексеми. Следователно, разпознаването на границите на лексикалните единици на синтагматично ниво изисква анализ не само на лексикалните елементи, които са част от съставните глаголи, напр. *имам предвид*, *имам против*, *вземам страна*, но и такъв, чрез който да се определи дали дадена словоформа от близкото обкръжение е част от граматичната форма, съставка на лексемата или синтактичен конституент. Установяването на групата на глагола се прилага паралелно с и неразделно от морфо-синтактичния анализ. На практика, в процеса на работа анотаторът извършва процедури, чрез които търси съответствие между непосредствения контекст на глаголната форма и елементите на ситуацията, описвана от предиката. Изложението представя някои основни случаи и решенията, които се вземат в процеса на анотация.

5.1. Идентифициране на основна форма – несъставни и съставни предикати

Най-напред ще бъдат разгледани няколко типични случая при определянето на основната форма на глаголните лексеми и явленията, които се проявяват и си взаимодействат на това равнище.

5.1.1. *Се/си*-глаголи

Се-формите могат да бъдат лексеми: (i) постоянно рефлексивни *се*-глаголи, които представляват диатези (отделни значения), кореспондиращи с изходни глаголи без *се*, например инхоативни *се*-глаголи, медиални *се*-диатези и др.; (ii) постоянно реципрочни *се*-глаголи; (iii) постоянно рефлексивни *се*-глаголи; или глаголни форми: (i) *се*-пасив; (ii) същински рефлексиви; (iii) същински реципрочни. Комбинациите от глаголна форма и *си* позволяват следните интерпретации: (i) постоянно рефлексивни глаголи; (ii) същински рефлексиви; (iii) постоянно реципрочни глаголи; (iv) същински реципрочни глаголи; (v) етичен датив; (vi) кратка форма на притежателно местоимение, намираща се в съседство с глаголна форма. Изчерпателна класификация на диатезите и алтернациите в българския език е предложена у Коева (2006), а на *се/си*-глаголите у Коева (2004).

Поради това, че се срещат по-рядко като част от глаголната лексема или глаголната форма (най-често са притежателно местоимение), *си*-формите няма да бъдат разглеждани отделно, а в случаите, в които е релевантно, ще бъдат споменати успоредно със *се*-формите. *Си*-глаголите, въведени в БулНет, са много по-малко в сравнение със *се*-глаголите. Добавените *си*-глаголи включват: (i) постоянно рефлексивни като *въобразявам си*, *отивам си* (връщам се въкъщи), *отивам си* (умирам), *признавам си* (правя самопризнания), *позволявам си* (смея, дръзвам), *позволявам си* (имам достатъчно средства за нещо), *отказвам си* (лишавам се), *говоря си*; (ii) постоянно реципрочни: *пиша си* (кореспондирам), *взаимодействам си*, *противореча си*.

В 1.1.1. са представени по-честите случаи на лексикално-граматична омонимия при *се*-формите.

5.1.1.1. Основни типове омонимия

При омонимия между синтактични (алтернации) и лексикално-семантични редувания (диатези) (терминологията е на Коева 2006) първата стъпка е да се установи основната форма. Диатезите се въвеждат и аотират като самостоятелни значения, а алтернациите и свободните съчетания (възвратно-притежателна клитика за притежание или за етичен датив¹ в съседство с глаголна форма) – като комбинация от глагол и местоимение (рефлексивно, реципрочно, притежателно) или от глагол и частица (напр. за личен или безличен пасив). По-долу са илюстрирани по-често срещаните случаи на омонимия, засвидетелствани в корпуса:

¹ В БулНет двете значения на кратките форми на възвратно-притежателното местоимение са разделени в отделни синонимни множества.

5.1.1.1.1. Постоянно рефлексивен глагол и *се*-пасив

- (1) а. *Постепенно хранителният режим се разширява.* (*се*-пасив)
б. *Прави го част от едно цяло, в което той се механизира, започва да мисли по начин, различен от собственото му желание, става чужд на себе си...* (постоянно рефлексивен)

Когато е възможна интерпретация като постоянно рефлексивен глагол и като *се*-пасив, се проверява дали експлицирането на агентивна *от*-фраза е граматично (Коева 2004). В случай че това е възможно, формата е за пасив *Постепенно хранителният режим се разширява от лекуващия лекар*, където *се разширява* е *се*-пасив от глагола *разширявам*, който принадлежи към следното множество:

{*разширявам:3, разширя:3*}

Дефиниция: *‘увеличавам пространствените граници, обхвата, обсега, в които нещо съществува, простира се, проявява се, прилага се и т.н.’*

Примерна употреба: *Действието на закона бе разширено, така че да обхваща всички граждани.*

Формата *се* се аотира като {*се:5*} с дефиниция: *‘граматикализирана частица за образуване на възвратно-страдателни форми от лични глаголи’*.

Отрицателният резултат от прилагането на теста с *от*-фраза не елиминира *се*-пасива (Коева 2004). В такъв случай се преценява дали ситуацията, описвана от формата, представлява пряк резултат от нечий действия или от действието на определена причина, аналогично на (1а), или възниква и съществува самостоятелно без специфициран причинител или вършител (постоянно рефлексивен) (1б), където *се механизира* е формата на глагола *механизирам се*:

{*механизирам се:1*}

Дефиниция: *‘ставам механичен, машинален, монотонен, лишен от съзнателност или творчество’*

Примерна употреба: *Душата ти се механизира като един часовник, който нито напред, нито назад върви.*

Между двойките постоянно рефлексивни и преходни глаголи (*разширявам се – разширявам, механизирам се – механизирам*) съществува отчетлива смислова разлика – първите са инхоативни глаголи, които изразяват претърпяване на промяна в състоянието, спонтанно или без определен причинител, със събитийна структура [у *BECOME <STATE>*], а вторите – каузативни глаголи за предизвикване на промяна в състоянието

със събитийна структура [*x CAUSE* у [*BECOME <STATE>*]] (Рапапорт Ховав и Левин 1998). Съответствието между двата типа глаголи се изразява в това, че вътрешният аргумент на каузативния глагол, който се проецира в прякообектна позиция, е идентичен със субекта на инхоативния. В семантичната структура на каузативния е налице каузиращо подсъбитие.

5.1.1.1.2. Постоянно рефлексивен глагол и същински рефлексив

- (2) а. *Да, скука, скука, плесен, от която трябва да се освободим срещу всички жертви.* (постоянно рефлексивен)
б. *„Лекаро, изцери се сам.“* (същински рефлексив)

В (2а) е засвидетелствана формата на инхоативния глагол *освободя се* в значението:

{*отървавам се:1, отърва се:1, освобождавам се:1, освободя се:1*}

Дефиниция: *‘отказвам се от, отстранявам, отхвърлям нещо нежелано или ненужно’.*

Примерна употреба: *Освободи се най-после от тези стари обувки.*

Инвариантната интерпретация на значението е *‘ставам свободен’*, аналогично на инхоативните глаголи, посочени по-горе. За разлика от (2а) в (2б) *изцери се* не е форма на инхоативния *изцери се* (ставам отново здрав), а е същински рефлексив от каузативния глагол *изцери*, тъй като се има предвид агентивно действие от страна на субекта, насочено към субекта:

{*лекувам:2, излекувам:1, целя:3, изцелявам:1, изцеля:1, цери:1, изцериavam:1, изцери:1*}

Дефиниция: *‘довеждам до или подпомагам възстановяването на някого от заболяване, неразположение, нараняване’*

Примерна употреба: *Така шарлатаните изцериавт всички болезки.*

Събитийните структури на каузативния и инхоативния глагол съответстват на посочените по-горе. В (2б) *се* се анимира като рефлексивно местоимение {*се:1, себе си:1*}.

5.1.1.1.3. Постоянно реципрочен и същински реципрочен

- (3) а. *Там те започнаха да си подхвърлят новата придобивка.* (същински реципрочен)
б. *Преди края на седмицата социалистите ще се консултират и с останалите коалиционни партньори.* (постоянно реципрочен)

Постоянно реципрочните глаголи се разграничават от същинските реципрочни чрез тест със заместване в единствено число, тъй като

вторият тип форми са дефективни в това отношение (Коева 2004). Проверката чрез трансформация, например **Той започна да си подхвърля новата придобивка със съседа си*, показва, че в (3а) е налице форма на преходния глагол *подхвърлям*. Заместването на глагола в (3б) *групата на социалистите ще се консултира с партньорите си*, потвърждава че *се консултират* е форма на постоянно реципрочния глагол *консултирам се* от синонимното множество:

{*съветвам се:2, посъветвам се:2, консултирам се:3, допитвам се:1, допитам се:1*}

Дефиниция: *‘обсъждам, планирам или решавам въпрос, проблем съвместно с някого’*

В Принстънския wordnet са представени някои глаголи с реципрочна семантика като *бия се, състезавам се, съревновавам се, съизмервам се, срещам се, виждам се* (срещам се), *взаимодействам си, пиша си, противореча си*. За други глаголи като *шушукам си, говоря си* с примерна употреба: *Те си шушукаха и изглежда решаваха съдбата на пленника*, е възприето компромисно решение да се аотират като глагол + реципрочно местоимение (вж. т. 5.1.1.2.6.).

5.1.1.1.4. Постоянно рефлексивен и същински реципрочен

От една страна възможността за замяна на множественото число с единствено в (4) показва, че формата не е същински реципрочна, а от друга, тъй като в единствено число семантиката на изречението не е реципрочна (описваната ситуация е, че охранителите ще пийнат от чуждото вино, а не че всеки ще почерпи другия), глаголът не е и постоянно реципрочен.

(4) а. *Охранителите на конвоя не пропускат възможността да се почерпят за сметка на плахите винари.* (постоянно рефлексивен)

Следователно, в случая е употребен постоянно рефлексивният глагол *почерпя се* със следното значение:

{*черпя се:1, почерпвам се:1, почерпя се:1, гостя се:1, угощавам се:1, угостя се:1*}

Дефиниция: *‘участвам в гощавка, ям и пия, предложени или платени от някого ястия и напитки, най-вече по някакъв повод’*

Примерна употреба: *Гостите на събитието се почерпиха пребогато.*

който се различава от транзитивния:

{*черпя:3, почерпвам:2, почерпя:2, гостя:2, угощавам:2, угостя:2*}

Дефиниция: *'предлагам или купувам за своя сметка и предлагам напитки и ястия на някого, често по определен повод'*.

Примерна употреба: *Както винаги тя почерпи гостите с отбрани деликатеси.*

Тъй като постоянно рефлексивният глагол няма съответствие в Принстънския wordnet, следва да бъде въведен в синонимно множество, уникално за българската лексикално-семантична мрежа.

5.1.1.2. Класове се-глаголи

Както се вижда от анализа на примерите (1-4), омонимията се разрешава въз основа на морфологични, синтактични и семантични критерии. По-долу ще бъдат посочени основни класове глаголи, участващи в каузативно-инхоативната и други каузативни диатези (Левин 1993: 27-32), които имат отношение към омонимията на *се*-формите. Там, където е необходимо, ще бъдат представени и конкретните решения по отношение на въвеждането на глаголите в БулНет.

5.1.1.2.1. Каузативни (транзитивни) глаголи за предизвикване на промяна в състоянието със събитийна структура [*x CAUSE y [BECOME <STATE>]*], евентуално позволяващи същински рефлексив, вж. *изцера се* в пример (2а) по-горе, и инхоативни постоянно рефлексивни глаголи, описващи претърпяване на промяна в състоянието, със събитийна структура [*y BECOME <STATE>]*.

Тази група включва разнообразни класове предикати, участващи в каузативно-инхоативната и други каузативни диатези. Предложена е хипотезата (Левин и Рапапорт Ховав 1995: 102-110), че предикатите, които участват в това редуване, концептуализират промяна в състоянието с външен причинител (с тематична роля Външна причина), като при непреходните корелати той е неспецифициран. Това е възможно, когато глаголите допускат различни по семантичната си природа причинни сили, и затова стриктно агентивните глаголи не участват в диатезата.

В тази група се категоризират и глаголите за преминаване в психическо или емоционално състояние или за предизвикване на преминаване в такова състояние, като *развълнувам се – развълнувам, възбуждам се – възбуждам, затрогвам се – затрогвам, изненадвам се – изненадвам²* и др., много от които нямат еквиваленти в английския език (вж. 5.1.1.2.2.).

² Тъй като наблюденията са общовалидни както за свършените, така и за несвършените по вид глаголи от дадена глаголна двойка (когато такава съществува), за краткост в текста ще бъдат посочвани само имперфективните корелати. Изключение се прави в част 5.4., там, където се демонстрира начинът на въвеждане на глаголите в БулНет.

5.1.1.2.2. Стативни постоянно рефлексивни глаголи, описващи състояния, и каузативни нерелексивни глаголи за предизвикване на състояние.

Събитийната структура на рефлексивните и каузативните глаголи от този тип се различава от тези в 5.1.1.2.1. както следва: [*y* <BE-IN-STATE>] и [*x* CAUSE *y* <BE-IN-STATE>].

Към този клас принадлежат предикатите за психически и емоционални състояния и каузативните глаголи за предизвикване на такива състояния като *плаша се*, *стракувам се* – *плаша*, *вълнувам се* – *вълнувам*, *забавлявам се* – *забавлявам* и т.н.

В редица случаи българските инхоативни и стативни глаголи за психически и емоционални състояния нямат лексикализирани английски еквиваленти (превеждат се чрез комбинации от копула *съм*, *ставам* и под. + прилагателно) и съответно не са въведени в Принстънския wordnet. При засвидетелстването в корпуса на такива глаголи, те се въвеждат в БулНет (вж. т. 5.4.1).

5.1.1.2.3. Постоянно рефлексивни глаголи за самостоятелно движение и каузативни глаголи за предизвикване на такова движение (частен случай на каузативно-инхоативната диатеза), напр. *търкалям се* – *търкалям*, *нося се* (*по вълните*) – *нося*, *топам се* – *топам* и т.н.

Следвайки анализа на логическите структури на Ван Валин (2007: 35-36), като запазваме означенията в духа на Левин и Рапапорт Ховав, събитийната структура би имала вида съответно *y*(ACT) и [*x* CAUSE *y* <ACT>].

Тук следва да се отбележат префигираните предикати за достигане до крайна точка в резултат от самостоятелно движение със събитийна структура: *y* [BECOME <BE-IN-PLACE *z*>] и в резултат от предизвикано движение: [*x* CAUSE *y* [BECOME <BE-IN-PLACE *z*>]] – образувани от глаголи за начин на движение, например *дотъркалям се* – *дотъркалям*. Тъй като обикновено тези глаголи в английски се образуват с помощта на предложна фраза от глаголи за начин на движение, те не фигурират в Принстънския wordnet като отделни значения и следва да се въведат в БулНет.

Друга подгрупа са префигираните глаголи за насочено движение без достигане до крайна точка като *оттласквам се* – *оттласквам*, *изтласквам се* – *изтласквам* и т.н. Глаголите за насочено движение не се лексикализират системно в английски и също се въвеждат като уникални множества за БулНет (вж. т. 5.4.6.3.).

5.1.1.2.4. Постоянно рефлексивни инхоативни глаголи за заемане на пространствено разположение и каузативни глаголи за поставяне в

пространствено разположение като: *подпирам се – подпирам, навеждам се – навеждам, изправям се – изправям, настанявам се – настанявам, намествам се – намествам*. Събитийната им структура е съответно у [BECOME <BE-IN-POSITION>] и [x CAUSE у [BECOME <BE-IN-POSITION>]]. В някои случаи, като например *настанявам се* и *намествам се* рефлексивните глаголи в Принстънския wordnet се смятат за рефлексивна употреба на съответния преходен глагол, поради което в БулНет са въведени в едно синонимно множество с уточнение в бележка към синонимното множество, че това са различни глаголни значения (вж. т. 5.4.4.).

5.1.1.2.5. Каузативни преходни глаголи и постоянно рефлексивни за (неволю) нараняване – *удрям – удрям се, блъскам – блъскам се, порязвам – порязвам се, наранявам – наранявам се, ухапвам – ухапвам се* и т.н.

Изходните предикати може да принадлежат към клас, различен от класа на диатезите, например *удрям, блъскам* са глаголи за контакт чрез удар. Постоянно рефлексивните се образуват в резултат от диатезата Unintentional interpretation with reflexive object (Левин 1993:101-102) и изразяват (неволю) нараняване на субекта от самия себе си. Събитийната структура кореспондира с тази в 5.1.1.2.1. Тук също важи ограничението, че стриктно агентивни глаголи не участват в диатезата, напр. **рит(в)ам се*. Тъй като рефлексивните глаголи не са отразени в Принстънския wordnet, в тези случаи е необходимо създаването на нови синонимни множества в БулНет (вж. т. 5.4.2.).

5.1.1.2.6. Други конвенции

- Някои глаголи, които приемат като обект неотделима част от субекта – *сбърчвам се (аз) – сбърчвам (вежди); цупя се (аз) – цупя (устни); втрещвам се (аз) – втрещвам (поглед)* се разглеждат като варианти, които се включват в едно синонимно множество;
- рефлексивните употреби на глаголи, отнасящи се до грижа за цялото тяло или за част от тялото като *мия, забулвам, обличам, реша, напудрям, начервявам* са същински рефлексиви – субектът извършва действие върху тялото си или част от тялото си – и срещанията им се аотират със съответния преходен глагол и рефлексивно местоимение, ако *се* е употребено;
- глаголи с реципрочна семантика, които имат съответстващи неречипрочни, в които субектът е идентичен на субекта в реципрочния (*говоря си* (от *говоря с някого*), *шушукам си* (от *говоря тихо с някого*), *целувам се* (от *целувам някого*), компромисно се аотират като форма на неречипрочния предикат + реципрочното местоимение, тъй като при съответните

английски нерещипрочни глаголи реципрочността е имплицитна и се изразява чрез множествено число (*те се целуват*) (+ реципрочна конструкция), а не чрез отделен предикат;

- възвратно-притежателното местоимение за етичен датив се анулира отделно от глагола със съответното значение {*си:4*} – ‘кратка форма на притежателното местоимение, участваща в образуването на конструкции с така наречения етичен датив, изразяващи субективно отношение на говорещия към ситуацията’.

5.1.2. Акузатива тантум и/или датива тантум

Тук ще бъдат разгледани два типа глаголи: третолични глаголи със задължителна клитика, на които съответстват глаголи с пълна парадигма (5.1.2.1.), и третолични глаголи със или без задължителна клитика със съответстващи им безлични предикати с или без клитика (5.1.2.2.)

5.1.2.1. Глаголи със задължителна клитика с кореспондиращи глаголи с пълна парадигма

В синтагматичен план основното затруднение при глаголите акузатива и датива тантум (за подробно описание вж. Коева 2004) е, че в някои случаи имат съответстващи лични глаголи с пълна парадигма, като разликата между двата реда предикати е различната перспектива, от която се концептуализира ситуацията. Например при глаголите *допада ми*, *нрави ми се*, *харесва ми*, участващи в синонимното множество:

{*харесвам:3*, *харесам:3*, *обичам:8*, *допада ми:1*, *допадне ми:1*,
нрави ми се:1, *харесва ми:1*, *хареса ми:1*}

Дефиниция: ‘*намирам нещо за приятно, хубаво, интересно, по вкуса ми*’
Примерна употреба: *Допадат ѝ руските класически романи.*

и *допадам*, *нравя се*, *харесвам се* от множеството:

{*привличам:4*, *харесвам:2*, *харесам:2*, *харесвам се:1*, *харесам се:1*,
нравя се:1, *допадам:1*, *допадна:1*}

Дефиниция: ‘*предизвиквам одобрение, симпатия у някого или доставям удоволствие, наслада на някого*’

Примерна употреба: *Идеята за ваканция на тропически остров много ми допада.*

се забелязва следното: и при двата реда предикати аргументът с тематична роля Стимул се реализира като субект, а Експериментаторът – като дативен

обект. Илюстрираните глаголи от първия ред обаче са третолични със задължителна дателна клитика (5в), докато вторите са с пълна парадигма и незадължителна дателна клитика (5а), (5б).

- (5) а. *Идеята допада/се нрави/се харесва/харесва на всички.*
б. *Зная, че аз много му допадам/се нравя/харесвам/се харесвам.*
в. *Не ми допада/ми се нрави/ми харесва тази идея.*

В трето лице с клитично изразен Експериентор формите на третоличните и личните съвпадат вж. трансформациите на (5а) и (5б) в (6а) и (6б), съответно:

- (6) а. *Идеята им допада/се нрави/се харесва.*
б. *Зная, че тя много му допада/се нрави/харесва.*

От примерите личи, че наред с лицето на глаголната форма словоредът също има значение за определянето на глагола, но въпросът се нуждае от по-задълбочено проучване. В някои случаи обаче словоредът също не е показателен. Така например в (7) трябва да се реши дали *харесват ти* е форма на *харесвам* (на някого) с пълна парадигма, или на *харесва ми* (третоличен със задължителна клитика). В отсъствието на морфологични и синтактични маркери въз основа на контекста е избрано {*харесвам:3, харесам:3, обичам:8, допада ми:1, допадне ми:1, нрави ми се:1, харесва ми:1, хареса ми:1*}:

- (7) *Човекът от Запада би познал в тях тайфата нехранимайковци, за които е създаден законът на линча. – Харесват {харесвам:3, харесам:3, обичам:8, допада ми:1, допадне ми:1, нрави ми се:1, харесва ми:1, хареса ми:1} ли ти? – попита Дебелия тихо. – Толкова, колкото и на тебе, тоест никак.*

Аналогична е ситуацията с предикати за необходимост, като *трябвам, нужен съм, необходим съм, потребен съм* и *трябва ми, нужен ми е, необходим ми е, потребен ми е*.

5.1.2.2. Глаголи, изразяващи физически усещания, *боли ме, сърби ме, смъди ми* (за свойствата и ограниченията при тези класове вж. Коева 2004), участват в диатеза, при която третоличен глагол, чийто субект е част от тялото – източник на болката, появяващ се обикновено със задължителна винителна или дателна клитика, проектираща Експериентора, (съществува и вариант без клитика – *как болят тези зъби, очите ми смъдят*), се трансформира в безличен или в безличен акузатива или датива тантум. От примерите в (8) се вижда, че се обособяват два класа значения – третолични с клитика или без клитика (8а), когато глаголят изразява значението, че определена част на тялото причинява

или е източник на болка, и безлични с клитика и без клитика, изразяващи изпитване (от Експериентора) на физическа болка (8б).

(8) а. *Смъдят ми очите/Много смъдят тези очи.* (очите причиняват болка)

б. *Смъди ми на очите./Смъди.* (аз усещам болка/усеща се болка)

Двете значения са отразени в БулНет, както следва:

{боли ме:1, боли:2}

Дефиниция: *‘за орган или част от тялото – представлявам източник или мястото, в което някой чувства болка’*

Примерна употреба: *Болят ме краката.*

{боли ме: 2, боли:1}

Дефиниция: *‘изпитвам неприятно, болезнено усещане, чувствам физическа болка’*

Примерна употреба: *Болеше ли те след катастрофата?*

Пример (9) от корпуса илюстрира и двата варианта:

(9) *Децата обичат да удрят с ръчички водата. Нея не я боли {боли ме:2, боли:1}. После децата плачат – болят ги {боли ме:1, боли:2} ръцете.*

Причината за близостта на двата типа значения е слабата „полярност” на аргументите Стимул и Експериентор. Според Даути (1991) двата аргумента притежават еднакъв брой протоагенсни свойства, което обяснява наличието на два класа психически предикати (Крофт 1991: 214) – глаголи, които проектират Експериентора като субект (Experiencer-subject verbs), и глаголи, които проектират Еспериентора като обект (Experiencer object verbs). Особеностите, наблюдавани при българските предикати, в известна степен може да се разглеждат като аналогични. В 5.1.2.1. задължителността на клитиката и ограничената парадигма на третоличните глаголи граматично маркират концептуализирането на ситуацията от гледна точка на Експериентора, който се кодира чрез клитиката. По този начин третоличните имат аналогична семантика с глаголите със субектно изразяван Експериентор. Пълната парадигма на личните глаголи заедно с незадължителността на клитиката маркират концептуализация, фокусирана върху Стимула, изразен субектно, и съответстват на английските глаголи с Експериентор обект. В 5.1.2.2. глаголите, съответстващи на предикати с Експериентор в позицията на субекта, са безлични, т.е. изключват изразяването на Стимула.

5.1.3. Пасив

При аотирането на форми за пасив обкръжението на спрегаемата глаголна форма помага за разграничаването на причастния от безличния причастен пасив и на *се*-пасива от безличния *се*-пасив. При безличните форми на причастния пасив се наблюдават следните морфологични и синтактични особености: (i) миналото страдателно причастие е само в среден род единствено число; (ii) невъзможно е да се експлицира подлогът с агентивна *от*-фраза (Коева 2004); (iii) липсва формален подлог.

Аналогично в (10) средният род на миналото свършено деятелно причастие има същата различителна функция:

(10) *Из кафенетата се {се:6} заговорило за политика...*

Се се аотира със значението {се:6} – ‘граматикализирана частица за образуване на безлични възвратно-страдателни форми от лични глаголи’

При *се*-пасива трябва да се провери дали има изразен подлог – *философският камък* в (11), чрез съгласуване по число между глаголната форма и предполагаемия субект *се създават философските камъни*. Когато е налице такова съгласуване, *се*-формата се аотира със синонимното множество {се:5} – ‘граматикализирана частица за образуване на възвратно-страдателни форми от лични глаголи’, а при установена невъзможност за съгласуване – със {се:6} :

(11) *...онези чудни материали от които се {се:5} създава философският камък.*

В 5.1.4. ще бъде обърнато внимание на пасива с оглед на предикатите с опора причастие.

5.1.4. Предикати с опора, различна от личен глагол

Друг тип случаи, при които анализът на граматичната информация, носена от формантите на глаголните форми или лексеми, служи за правилното им разпознаване, е при разграничаването на лексикализираните стативни предикати от свободни съставни именни сказуеми и/или пасивни конструкции (12):

(12) *Резерватът е разположен в лонгозна гора по долното течение на река Батова.*

В този случай възможните подходи са два – първият е съчетанието да бъде аотирано като съставно именно сказуемо, композирано от

глагола *съм* + причастието *разположен* в предикативна функция, или като се вземе предвид еквивалентността на глаголното значение в (12) и (13), изразът да се анулира като стативния предикат, който принадлежи към синонимното множество {*лежа:1, разполагам се:1, разположен съм:1*} с дефиниция '*съм, намирам се върху определено пространство, имам определено местоположение или заемам определено място, позиция, положение*':

(13) *Град Н се разполага на брега на големия язовир.*

В БулСемКор и в БулНет се приема, че наличието на синонимни глаголи в български и на преводни еквиваленти, изразяващи същото значение в други езици, говори за некомпозиционния характер на комбинациите от копула и форма от друга част на речта или нелична глаголна форма в разглежданите тук случаи. Това дава основания те да се смятат за самостоятелни предикати.

Възможно е да възникнат колебания при анотацията, в случаите когато на стативния предикат кореспондира каузативен глагол, чийто причастен пасив е омонимен със статива. В илюстрирания пример се наблюдава съвпадение на формата на стативния предикат с пасива на глагола за поставяне в пространствено положение (Левин 1993) *разполагам*. При съпоставката на (14а) и (14б) се вижда, че между формите в двете изречения има ясно изразена смислова разлика – стативният предикат изразява пространствено свойство, а формата за пасив – резултат от промяна на състоянието, предизвиквана от съответния преходен глагол.

- (14) а. *Лагерът бил разположен от Помпей тъкмо там. > Тъкмо там разположил лагера си Помпей.*
б. *Резерватът е разположен в лонгозна гора... > *Резерватът е разположен в лонгозна гора от някого...*

Предикатната декомпозиция на каузативните глаголи от този тип показва, че в ядрото на резултативното им подсъбитие има стативен предикат за пространствено разположение. Това е допълнителен довод *разположен съм* да се смята за самостоен предикат, подобно на *седя, стоя, вися* и т.н.

Аналогично, въз основа на тяхната некомпозиционност, като самостоятелни се разглеждат адиективни (*нужен ми е, потребен ми е*), причастни (*роден съм*), номинални (*страх ме е*), адвербиални предикати (*добре ми е, наясно съм*), предикати от копула и предложна фраза (*на себе си съм, в сила съм – важа*).

5.2. Аргументна структура. Диатези

Както бе посочено, аргументната структура на предикатите също играе роля при разграничаване на значенията. Стремежът при дефинирането на глаголните синонимни множества в БулНет е те да съдържат общо описание на концептуализираната ситуация и по възможност задължителните участници в нея, както и основни техни семантични характеристики (имплицитно – следващи от мястото на множеството в общата структура на лексикално-семантичната мрежа; експлицитно – като пояснение към дефиницията, когато е необходимо). Водещи при вземането на решения за анотацията са общите принципи, а аргументната структура и семантичните свойства имат помощна функция.

5.2.1. Аргументна структура – брой на аргументите, селективни ограничения

Броят на аргументите и тяхната синтактична реализация могат да бъдат показателни за разграничението на едно значение от друго или поне за елиминирането на част от кандидатите. Поради големия брой значения няма да бъдат разглеждани всички налични синонимни множества, съдържащи даден литерал, а само онези от тях, които са релевантни за разглежданото явление. В (15) са илюстрирани две близки значения на *давам/дам*, които се различават по броя на аргументите си.

(15) а. *Дадох ѝ пари.*

б. *Давам ти три за двата диска, които изгубих.*

Глаголът в (15а) принадлежи към синонимното множество {*давам:12, дам:12*} с дефиниция '*правя така, че някой да получи у себе си или да започне да притежава нещо в конкретен или абстрактен смисъл*', и има три аргумента, а формата в (15б) – към множеството {*давам:42, дам:42*} – '*предлагам нещо на някого като компенсация за или в замяна на нещо друго*' – и е четириаргументен глагол. Наличието на четвъртия аргумент елиминира {*давам:12, дам:12*}.

В редица случаи броят на аргументите съвпада и значението се определя с помощта на имплицитни или експлицитни семантични специфики на участниците в ситуацията и на самата ситуация.

Експлицитно селективните ограничения се посочват като пояснения в дефинициите. Например в дефиницията на синонимното множество {*стрелям:2, гръмвам:4, гръмна:4, изгърмвам:2, изгърмя:2, давам изстрел:2, дам изстрел:2, произвеждам изстрел:2, произвежда изстрел:1*} се формулира ограничение за субекта '*за огнестрелно*

оръжие – изхвърлям заряд по цел'. По този начин глаголите по-лесно се разграничават от каузиращите ги предикати от синонимното множество {стрелям:1, гръмвам:2, гръмна:2, давам изстрел:1, дам изстрел:1, произвеждам изстрел:1, произвежда изстрел:1} с дефиниция: 'изхвърлям снаряд от огнестрелно оръжие'.

В други случаи освен експлицитно изразени селективни ограничения се анализират и други специфики на глаголните аргументи и на ситуацията изобщо, като е възможно да се вземат предвид и останалите релации, в които влизат разглежданите синонимни множества. За глаголната двойка *убивам, убия* в БулНет фигурират четири значения, отнасящи се до отнемането на живот:

{убивам:6, убия:6, умъртвявам:1, умъртвя:1}

Дефиниция: '*причинявам смъртта на човек или друго живо същество (често предумишлено)*'

Примерна употреба: *Престъпникът убил няколко души при опита да обере банката.*

{убивам:8, убия:8}

Дефиниция: '*причинявам смъртта на човек или друго живо същество неволно, например при инцидент или в резултат на случайни действия*'

Примерна употреба: *Родителите ѝ били убити в автомобилна катастрофа.*

{убивам:3, убия:3, погубвам:2, погубя:2}

Дефиниция: '*за болест, бедствие и др. – довеждам до или причинявам смъртта на някого*'

Примерна употреба: *За последните две десетилетия СПИН е убил милиони.*

{убивам:10, убия:9}

Дефиниция: '*представявам заплаха или имам гибелно въздействие върху живота на някого*'

Примерна употреба: *Цигарите убиват.*

Двойките лимерали {убивам:6, убия:6}, {убивам:8, убия:8}, {убивам:3, убия:3} описват ситуация, в която човек и/или друг агент или причинител предизвиква смъртта на човек или друго живо същество. Трите значения се различават въз основа на наличието на възможна интенция и пряко действие, което отличава {убивам:6, убия:6} от

другите две; наличието на пряка каузация, отсъствието на интенция и наличието на неодоушевен вършител – {убивам:8, убия:8}; отсъствието на интенция и пряко действие (причиняваща последица) – {убивам:3, убия:3}. Може да се заключи, че първото значение е агентивно и (потенциално) волево, второто – каузално с неодоушевени неагентивни субекти или с неконкретизиран субект в пасив, третото – каузално в по-широк смисъл, чиито субекти в общия случай са природни сили или причинители на или следствия от човешки действия. Допълнителна яснота относно разграничението между тези значения внасят релациите с други синонимни множества. Множеството {убивам:6, убия:6} включва още синонимите *умъртвявам* и *умъртвя*, а част от неговите хипоними са *избивам*, *екзекутирам*, *ликвидирам*, *обезглавявам*, *линчувам*, *набивам на кол* и т.н.; {убивам:8, убия:8} има един хипоним в Принстънския wordnet, невъведен в БулНет *electrocute* (*убивам с електрически ток*), с примерна употреба: *Тя бе убита от ток, след като изпусна сешоара във ваната*; {убивам:3, убия:3} има един хипоним – {покосявам:1, покося:1}.

Четвъртото разглеждано значение {убивам:10, убия:9} спада към друга подструктура в организацията на wordnet, тъй като се отнася до свойството на определени причинители (например фактори, действия) да довеждат до смъртта, да са гибелни за живота на някого или нещо. Ситуацията, описвана от глаголите {убивам:10, убия:9}, може да се разглежда като производна на инвариантната ситуация, засвидетелствана в предните три значения. Субектите, които това глаголно значение селектира, са вредни фактори, действия, въздействия и т.н., които са в състояние да оказват отрицателен или пагубен резултат. То се различава от останалите и по ограничената си парадигма по време, тъй като се употребява без актуална темпорална референция, срв.:

- (16) а. *Алкохолът убива* {убивам:10, убия:9}.
 б. *Алкохолът я уби* {убивам:3, убия:3, погубвам:2, погубя:2}.

Изборът на анотатора между четирите значения в конкретния контекст се мотивира от това, спецификите на кое от тях съответстват на ситуацията, описвана от глагола в корпуса, като се съди по обкръжението му. В (17а) е избрано {убивам:6, убия:6}, а в (17б) {убивам:3, убия:3}:

- (17) а. *Няма значение дали става дума за Едип, който убива* {убивам:6, убия:6} *баща си, или за Павлик Морозов, който действа срещу дядо си.*
 б. *...може да ги ударят терористи, може да ги убият* {убивам:3, убия:3} *земетресения.*

5.2.2. Диатези

Диатезите представляват промяна, която засяга семантичните релации между аргументите и предиката (вж. Коева 2006 и за изчерпателна класификация на българските диатези според аргументната структура на изходните и производните предикати, семантичните и синтактичните условия и типа на промяната). Системно наблюдаваните съответствия допринасят за разграничаването на отделни глаголни значения, които сами по себе си може да са трудно разпознаваеми. Особеностите на диатезите и класовете глаголи, които участват в тях, от една страна, улеснява вземането на решение при анотацията, а от друга – води до последователност при въвеждането на значенията в БулНет и при определянето на семантичните спецификите на аргументите, когато това е необходимо в дефиницията.

5.2.2.1. Редувания, свързани с проекцията на вътрешните аргументи

Поради разпространеността ѝ и относителната хомогенност на глаголите, участващи в нея, най-напред ще бъде разгледана диатезата, известна като *Spray/load*, която засяга проекцията на вътрешните аргументи при два класа триаргументни глаголи – за поставяне (*verbs of putting*) и за запълване (*verbs of filling*) (Левин 1993: 50), макар да не може да се твърди, че се ограничава до тях. В първото значение глаголите изразяват предизвикване на промяна в местоположението на един или повече обекти или на вещество или материал, като в резултат от тази промяна те започват да се намират на дадено място, а във второто – запълване на съд, повърхност, място с различни обекти или с материал или вещество. Комплементът, който променя местоположението си, е известен под различни наименования в различните теории: *Locatum* (Левин 1993 и в много други изследвания); *Theme* (Рупенхофер и др. 2006). Тук се приема наименованието *Тема*, а мястото или пространството, което е крайна точка на движението на обекта/обектите ще бъде означавано с *Местоположение* – *Location* (Левин 1993), *Goal* (Рупенхофер и др. 2006). При глаголите за поставяне аргументът с роля *Тема* се проектира в прякообектна позиция, а аргументът с роля *Местоположение* – като комплемент на предлозите *в*, *върху*, *на* и др. При глаголите за запълване – аргументът с роля *Местоположение* е в прякообектна позиция, а *Темата* е комплемент на предлога *с*. В *wordnet* с малки изключения тези два класа са последователно отразени.

Близостта на двата семантични класа, засвидетелствана и в голямото припокриване между членовете на синонимните множества, би могла да бъде подвеждаща при анотацията. Следователно анотаторът трябва да вземе предвид семантиката и синтактичното обкръжение на глаголната

форма при срещането ѝ в корпуса. Измежду значенията на *натоваря* следните две се отнасят до поставяне на или запълване с товар:

{*товаря:2, натоварвам:5, натоваря:5*}

Дефиниция: '*поставям товар на определено място или в превозно средство*'

Примерна употреба: *Натовариха партидата на камионите.*

{*товаря:1, натоварвам:1, натоваря:1*}

Дефиниция: '*запълвам дадено място или пространство с товар*'

Примерна употреба: *Натовариха автомобила с багаж.*

Въз основа на синтактичното обкръжение в контекста в (18) е избрано {*товаря:2, натоварвам:5, натоваря:5*}:

(18) *...трябвало да ги натоварят {товаря:2, натоварвам:5, натоваря:5} на ремаркета...*

5.2.2.2. Диатези, свързани с редуциране на външен аргумент

Общото за тези диатези (съответстващи на Тип 3 у Коева 2006) е, че включват редуване на агентивен с неагентивен субект, който е идентичен с вътрешен аргумент на агентивния глагол, съпроводено с редуциране на агентивния аргумент. Тъй като ситуацияте, описвани от двата варианта, са производни една на друга и двете аргументни структури имат общи елементи, то и значенията са близки. За по-ясното им разграничаване чрез дефиницията обикновено е необходимо поясняване на семантичните свойства на различаващия се аргумент, проектиран в субектна позиция.

5.2.2.2.1. Locatum subject

Част от глаголите, които участват в тази диатеза са предикатите за запълване. Семантичната промяна се изразява в това, че аргументът с тематична роля Тема се проектира в субектна позиция, аргументът с тематична роля Местоположение – в прякообектна, а аргументът, заемащ субектната позиция на предиката за запълване – Агенс или Причина, не фигурира в аргументната структура. По-долу са представени две синонимни множества, демонстриращи диатезата:

{*украсявам:2, украся:2, декорирам:1*}

Дефиниция: '*правя нещо да стане по-красиво, по-привлекателно чрез добавяне на орнаменти, украшения, цвят и др.*'

Примерна употреба: *Украсете стаята за тържеството.*

{*крася:1, украсявам:3, украся:3*}

Дефиниция: ‘за декорация, украса, орнаменти и т.н. – придавам похубав, изящен вид на някого или нещо’

Примерна употреба: *Цветята красят масата.*

В (19) въз основа на синтактичното обкръжение (и на съпоставката с другото значение, участващо в диатезата) може да се заключи, че конкретното срещане на *украсява* е реализация на второто значение.

(19) *Черна като катран коса украсява {крася:1, украсявам:3, украся:3} Гроздиното бяло лице...*

5.2.2.2.2. Abstract cause alternation

Агентивните глаголи, които участват в тази диатеза, включват предикати, описващи ситуация, при която някой постига дадена абстрактна цел (убеждава, доказва, потвърждава, отрича, обяснява и под.) с помощта на някакъв абстрактен продукт на мисловна дейност, а съответстващите глаголи, при които абстрактната причина заема субектната позиция, изразяват действието или приноса на този продукт за осъществяване на целта. Диатезата се наблюдава в следните две синонимни множества:

{*потвърждавам:5, потвърдя:5, подкрепям:11, подкрепя:10*}

Дефиниция: ‘подсилвам определено твърдение с доказателства, аргументи’

Примерна употреба: *Свидетелите потвърдиха разказа на жертвата.*

{*потвърждавам:4, потвърдя:4, затвърждавам:1, затвърдя:2, подкрепям:9, подкрепя:9*}

Дефиниция: ‘подсилвам, придавам фактологична тежест на мнение, твърдение и т.н. (за или с помощта на свидетелства, доказателства и под.)’

Примерна употреба: *Историята му потвърди съмненията ми.*

Глаголните форми *потвърди* и *потвърждава* в (20а) и (20б) са анотирани съответно с {*потвърждавам:5, потвърдя:5, подкрепям:11, подкрепя:10*} и с {*потвърждавам:4, потвърдя:4, затвърждавам:1, затвърдя:2, подкрепям:9, подкрепя:9*}:

- (20) а. *Новината потвърди {потвърждавам:5, потвърдя:5, подкрепям:11, подкрепя:10} пред „Сега,, доц. Румен Златев.*
б. *Избелялата, на места прокъсана фланелка с небрежно навити ръкави... само потвърждава {потвърждавам:4, потвърдя:4, затвърждавам:1, затвърдя:2, подкрепям:9, подкрепя:9} първоначалното впечатление.*

5.2.2.2.3. Container subject alternation

В диатезата участват глаголи, изразяващи релация на включване между цяло, изразено в субектна позиция, и неговите части – в прякообектна. Ситуацията може да е предизвикана от агентивен субект, тогава цялото се проектира като комплемент на предлога *в*. В нея участват глаголи като *включвам*, *интегрирам*, *инкорпорирам* и под.

{*включвам*:6, *включва*:5}

Дефиниция: *'прибавям нещо като съставна част, компонент на нещо друго или отнасям нещо към група, множество, категория'*

Примерна употреба: *Включихме всички присъстващи в списъка.*

{*включвам*:5, *включва*:4}

Дефиниция: *'съдържа нещо като съставна част или компонент'*

Примерна употреба: *Изложбата включва черно-бели и цветни фотографии.*

Аналогична е диатезата Location subject, при която релацията е не между част и цяло, а между пространство или контейнер и неговата вместимост. В български в тази диатеза участват глаголните двойки *побирам*, *побера* и *събирам*, *събера*.

5.2.2.3. Други диатези

При някои диатези агентивна ситуация се реконцептуализира като статичен признак, подобно на примера с {*убивам*:10, *убия*:9} по-горе. Общата характеристика на тези редувания се изразява в следното – глагол, описващ ситуация, при която някой или нещо предизвиква промяна в състоянието на друг човек, животно, растение, или извършва някакво действие, е преосмислен като глагол, изразяващ качество, свойство или тенденция на аргумента, който се проектира като субект, да предизвика съответната промяна в състоянието на някого или нещо. В общия случай се променят аспектиалните характеристики на глаголите, като употребата в актуални времена е силно ограничена. Повечето от посочените подолу значения не се срещат в корпуса, тъй като са относително редки, но са налични в БулНет и следователно се появяват сред кандидатите за анотация.

5.2.2.3.1. Предикати, изразяващи разрязване, разделяне чрез инструмент, и такива, лексикализиращи съответното свойство на инструмента (Левин 1993: 39):

{*режа*:3, *разрязвам*:2, *разрежа*:2, *отрязвам*:7, *отрежа*:7}

Дефиниция: *'разделям нещо на части с остър инструмент или по подобен начин'*

Примерна употреба: *Разрязохме хляба с нож.*

{режа:4}

Дефиниция: *‘служва като инструмент за рязане’*

Примерна употреба: *Ножът реже добре.*

Това значение не бива да се изравнява с диатезата, при която инструментът се проектира като субект, защото при нея глаголите се изразяват свободно в актуални времена.

5.2.2.3.2. Диатеза с глаголи за финансови транзакции (sum of money subject alternation). Производното значение описва ситуация, при която дадена сума се разглежда като достатъчна за извършването на транзакцията (Левин 1993: 83). Аргументът, реализиращ сумата, за чиято равностойност се извършва търговската транзакция, се проектира в субектна позиция, Агенсът на изходния глагол – като дативен обект, а аргументът, изразяващ обекта, който променя собственика си, запазва обектната си позиция.

{купувам:2, купя:2, закупувам:1, закупя:1}

Дефиниция: *‘придобивам стока или ползвам услуга срещу определено заплащане’*

Примерна употреба: *Семейството си купи нова кола за 5000 долара.*

{купувам:3, купя:3}

Дефиниция: *‘представявам паричната равностойност на стока или услуга или правя нещо финансово достъпно за някого’*

Примерна употреба: *Парите купуват всичко.*

Аналогичен характер има и отношението между *продавам/продам* в основното му значение, като в този случай се наблюдава промяна в аргументната структура – триаргументният глагол се трансформира в двуаргументен (реципиентът е редуциран), а на аргумента в субектна позиция са наложени различни селективни ограничения – от лице или организация в абстрактна същност.

{продавам:5, продам:5}

Дефиниция: *‘предлагам стоки или услуги срещу определено заплащане’*

Примерна употреба: *Той продаде къщата си.*

{продавам:2, продам:2}

Дефиниция: *‘за нещо абстрактно – имам основен принос за продажбата на даден продукт’*

Примерна употреба: *Общозвестен факт е, че рекламата продава стоката.*

5.3. Предлози в състава на глаголната лема

Предлозите се класифицират като непредикативни и предикативни, като първите се управляват от глаголите и свързват аргументи с предикат, а вторите – адюнкти с предикат (Ницолова 2005: 151, Ницолова 2008: 457-459). Непредикативните предлози се определят от лексикалната семантика на предикатите, като изборът им може да бъде строго специфичен или да обобщава по-големи класове от аргументи с определена семантична роля. В някои случаи се говори за формални или граматикализирани предлози, при които степента на абстракция е толкова висока, че предлогът се превръща в маркер за граматична позиция, като свързва аргументи с различни семантични роли с предиката. Например предлогът *на* се е превърнал в маркер за индиректен аргумент и свързва непреки аргументи с различни роли, като реципиент (*давам ябълка на детето*), адресат (*пиша на Мария*), стимул (*радвам се на природата*), експериментор (*на детето му е добре*), екатив (*детето прилича на майка си*) и др., аргумент-адюнкти (*отивам на работа*) (примерите са от Ницолова 2005: 144-146). Предлозите свързват и аргументи на имплицитни предикати (Ницолова 2005: 144-146) (*трудът на родителите му, мнението на сестра му, книгата на Набоков, шофьор на такси*) (примерите са на автора) и т.н.

В някои случаи семантиката на даден предикат се дължи на комбинацията му с определен предлог, като от една страна значението на предиката не може да се интерпретира без предлога, а от друга, самият предлог трудно може да се дефинира от гледна точка на дадено предикативно или непредикативно значение, срв. например *радвам се на* 'изживявам благоприятен развой или етап' (21а) и *радвам се* (наслаждавам се) в (21б) и *разбирам от* 'извличам полза или наслада от нещо' и *разбирам* в (22а) и (22б):

- (21) а. *Индустрията се радва на истински бум.*
б. *Радвам се на топлото време.*
- (22) а. *Какво разбира ти от живота?*
б. *Какво разбира от пиесата?*

Когато налице е идиоматизирано съчетание от глагол и предлог, предлогът се включва като част от лемата.

5.4. Въвеждане на езиковоспецифични значения в БулНет

В процеса на анотацията са засвидетелствани класове глаголи, които нямат съответствие в Принстънския wordnet.

5.4.1. Стативни и инхоативни глаголи за психически и емоционални състояния (т. 5.1.1.2.1. и 5.1.1.2.2.) се въвеждат като хипоними на {изпитвам:1, изпитам:1, чувствам:1, почувствам:1} – {feel:11, experience:6}, аналогично на лексикализираните глаголи от тези класове.

5.4.2. Постоянно рефлексивни за (неволно) нараняване (5.1.1.2.5) се въвеждат като хипоними на {наранявам:1, нараня:1} – {injure:1, wound:6}.

5.4.3. Стативни и инхоативни глаголи за промяна на състоянието, образувани от прилагателни:

- стативни за признак – чернея (се), белея (се), се въвеждат като хипоними на {съм:9} – {be:4}; съответните инхоативни и каузативни се лексикализират в английските глаголи и са засвидетелствани в Принстънския wordnet.
- инхоативни – оглушавам – оглушея; онемявам – онемея – каузативните варианти (където е релевантно) са лексикализирани в английски.

В общия случай инхоативните глаголи в 5.4.3. са предикати за претърпяване на промяна и се включват като хипоними на {променям се:1, променя се:1} – {change:11}.

5.4.4. Други случаи – инхоативни глаголи за промяна в състоянието като образовам се, еманципирам се, освобождавам се – освободя се; други постоянно рефлексивни глаголи и т.н.

В много от тези случаи глаголите се смятат за рефлексивна интерпретация на преходен глагол (често се образуват с рефлексивно местоимение), например *изтръгвам се, издърпвам се*. Някои от тях спадат към така наречените задължително рефлексивни у Левин (*Obligatorily reflexive object*) (Левин 1993:107) – *браня се, излагам се – изложа се, унижа(вам) се, опозоря(вам) се*. В зависимост от структурата на синонимните множества в Принстънския wordnet, съдържащи съответните преходни глаголи, рефлексивите се въвеждат компромисно или като техни литерали, като в бележка към синонимното множество се посочва, че значенията са различни, или като самостоятелни синонимни множества.

5.4.5. Глаголи, изразяващи фаза от действие, процес и т.н.

Начинателни (*запявам – запея, засмивам се – засмея се*), финални (*достроя(вам), допявам – допея*) и др. глаголи, лексикализиращи фаза

от процес, действие и т.н., се въвеждат като хипоними на синонимното множество, в което се намира произвеждащият глагол, и на синонимното множество, съответстващо на фазата.

5.4.6. Глаголи (префигирани), които лексикализират „скала” (Филип 2008), по която се измерва прогресът на действието или процеса, описван от тях. Въвеждат се в съответствие с възприетия подход, различен за различните класове, изброени по-долу (Лесева 2008).

5.4.6.1. Глаголи, които лексикализират резултат от действието или процеса, описван от непрефигирания глагол. Въвеждат се като литерали в същото синонимно множество, например: *чета – прочитам – прочета*.

5.4.6.2. Глаголи, които принадлежат към същия семантичен клас като произвеждащите глаголи (например за ограничена продължителност или за степен) се въвеждат като хипоними на произвеждащите глаголи – (*помълчавам – помълча; пооздравявам – пооздравея; повъзстановя(вам) се, посъвзема(м) се*).

5.4.6.3. Глаголи, които в резултат от префиксацията преминават в друг семантичен клас, се въвеждат като хипоними на подходящо множество в съответния семантичен клас.

Глаголите в 5.4.6.2. и в 5.4.6.3. лексикализират разнообразни отношения – адвербиални: локативни (*облиствам – облистя; облъхам – облъхна; полазвам – полазя*), за насочено движение – отдалечаване (*оттласквам (се) – оттласна (се), отблъсквам (се) – отблъсна (се), отлитам/отлетя(вам)*), доближаване (*долитам/долетия(вам)*), движение навътре (*втривам – втрия*), движение навън (*излитам/излетия(вам), избяг(в)ам*), обхващане на период от време (*пролежда(вам)*) и др. В тази категория спадат каузативните и инхоативните глаголи за достигане до крайна точка в резултат от насочено движение, както и глаголите за насочено движение (без задължително достигане до крайна точка) (5.1.1.2.3.)

5.5. Заключение

Статията не изчерпва въпросите, свързани с анотацията на глаголите, а се стреми към обобщение на основни проблеми и явления, с които анотаторите са се сблъскали в процеса на работа. Изложени са основните стратегии, които подпомагат вземането на лингвистични решения, като се търси лингвистична мотивация за тях и се очертават редица важни специфики на българските глаголи.

Цитирана литература

- Ван Валин 2007: Van Valin R. The Role and Reference Grammar Analysis of Three-Place Predicates. – *Suvremena Lingvistika* 63, pp. 31-63.
- Даути 1991: Dowty D. Thematic Protoroles and Argument Selection. – *Language*, Vol. 67, № 3, pp. 547-619.
- Коева 2004: Коева Св. Синтактично и семантично описание на диатезите в българския език. В: *Българско езикознание*, т. 4, Академично издателство „М. Дринов”, С., с. 182-231.
- Коева 2006: Коева Св. Синтактични редувания в български – диатези и алтернации. – *Български език*, XIII, № 4, с. 19-31.
- Коева и др. 2008: Коева Св., Р. Влахова, Р. Декова, П. Несторова, А. Атанасов. Представяне на езиковата информация в Българския ФреймНет – лингвистична мотивировка. – В: Св. Коева (съст.) *Българският Фреймнет. Семантико-синтактичен речник на българския език*, С., с. 58-78.
- Крофт 1991: Croft W. *Syntactic Categories and Grammatical Relations*. University of Chicago Press.
- Левин 1993: Levin B. *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago Press.
- Левин и Рапапорт Ховав 1995: Levin B. and M. Rappaport Hovav. *Unaccusativity: At the Syntax-Lexical Semantics Interface*. Linguistic Inquiry Monograph 26, MIT Press, Cambridge, MA.
- Лесева 2008: Leseva S. Towards Encoding Event Structure in Wordnet. – In: M. Tadic, M. Dimitrova-Vulchanova and S. Koeva (eds.) *Proceedings of the Fifth International Conference on Formal Approaches to South Slavic and Balkan Languages*. Croatian Language Technologies Society – Faculty of Humanities and Social Sciences, Zagreb, pp. 74-78.
- Ницолова 2005: Ницолова Р. За значението на предложните съчетания в българското изречение. – В: Св. Коева (съст.) *Аргументна структура. Проблеми на простото и сложното изречение*. СемаРШ, С., с. 141-152.
- Ницолова 2008: Ницолова Р. *Българска граматика. Морфология*. Университетско издателство „Св. Кл. Охридски”, С.
- Рапапорт Ховав и Левин 1998: Rappaport Hovav M. and B. Levin. Building Verb Meanings. – In: M. Butt and W. Geuder (eds.) *The Projection of Arguments: Lexical and Compositional Factors*. CSLI Publications, Stanford, CA, pp. 97-134.

- Рупенхофер и др 2006: Rupenhofer J., M. Ellsworth, M. R. L. Petruck, C. R. Johnson, J. Scheffczyk. *Framenet II. Extended Theory and Practice*. Berkeley, CA: International Computer Science Institute. <http://framenet.icsi.berkeley.edu/>.
- Филип 2008: Filip H. Events and Maximalization: the Case of Telicity and Perfectivity. – In: Rothstein, S. (ed.) *Theoretical and Crosslinguistic Approaches to the Semantics of Aspect*. John Benjamins Publishing Company: Amsterdam, pp. 217-256.

6. Лингвистични конвенции при анотация на прилагателните

Христина Кукова

6.1. Анотационни решения при прилагателните

Семантичната анотация на прилагателните имена в Българския семантично анотиран корпус се извършва съобразно възприетите общи правила. Двете основни стъпки включват (автоматично) привеждане на словоформите до основната им форма и определяне на лексикалното значение, с което е употребена думата. За разлика от останалите отворени класове прилагателните имена не участват в хиперо-хипонимни отношения. Изборът на значение зависи от: мястото на синонимното множество в мрежовидната структура на БулНет; релациите, с които е свързано с други множества; тълковната дефиниция; съответстващия му преводен еквивалент в Принстънския wordnet; и не на последно място, посочените типични или специфични примери за употреба. Например прилагателното *стар* от (1):

- (1) *След като завърших гимназия, изкарах курсове за готвач и започнах работа в малки ресторанти, където се учех от опита на **старите** майстори.*

се асоциира в БулНет с пет синонимни множества:

{стар:1}

Дефиниция: *'който съществува от дълго време, не е нов или нововъзникнал'*

{стар:2, възрастен:2, дърт:1}

Дефиниция: *'който е живял дълго време и по-голямата част от живота му е преминала'*

{стар:3, застоял:2, престоаял:1}

Дефиниция: *'който с времето е загубил качествата си или е станал негоден за употреба'*

{предишен:4, преден:4, стар:4}

Дефиниция: *'който предшества непосредствено някого или нещо друг(о) в хронологичен план или по ред или който принадлежи към отминало време'*

{стар:5}

Дефиниция: *'който е добре познат и близък'*

Примерна употреба: *стар приятел*

В този пример контекстът и тълковните дефиниции показват с кое синонимно множество се съотнася конкретната употреба, а именно със {стар:2}. Тук следва да се отбележи, че замяната между литералите на едно синонимно множество не винаги е възможна в съответния контекст, като в случая причините са предимно стилистични.

При голяма част от синонимните множества на прилагателните се наблюдава по-висока гранулираност на лексикалните значения, което усложнява аотирането. В някои случаи е налице не само сходство в дефинициите на синонимните множества, но и пълно съвпадение на литералите, които ги изграждат. Ето защо е необходимо да се вземат под внимание и други свойства на синонимните множества. За прилагателните най-често използваните ориентири са семантичните релации на подобие *подобен на* и *също така*, както и деривационната релация *произведен*. Релациите на подобие предлагат добра стратегия за точен избор на значение, тъй като групират синонимните множества по сходни семантични свойства. Това значително редуцира броя на подходящите кандидати и улеснява аотационния процес. Например прилагателното *тежък* от (2) фигурира в петнайсет синонимни множества в БулНет, някои от които изразяват близки значения.

(2) *Решиха да оставят тази **тежка** задача на Златния зъб, техния прекупвач.*

Анализът на подходящите за аотация синонимни множества показва, че някои от значенията, приложими в конкретния контекст, могат да бъдат отхвърлени посредством обобщение в следните класове:

- (i) значения, свързани с тегло в тесния смисъл на думата;
- (ii) значения, характеризиращи описваното с висока степен или интензивност;
- (iii) значения, засягащи неблагоприятно физическо състояние или бременност.

Заслужават по-внимателен преглед дефинициите, свързани с полагане на усилия, като по този начин потенциалните синонимни множества се свеждат до две:

{*тежък*:5, *усилен*:2}

Дефиниция: '*който изисква или е свързан с голямо физическо усилие*'

Примерна употреба: *тежък физически труд*

{*труден:3, тежък:10*}

Дефиниция: '*за справянето с който се изисква голямо физическо или умствено усилие или който е свързан с преодоляването на проблеми, неприятности, несгоди*'

Примерна употреба: *тежка задача*

Дефинициите на двете синонимни множества, членовете, които ги представляват, и примерите за употреба ограничават избора за анотация до {*труден3, тежък 10*}.

6.2. Самостоятелна употреба на прилагателни имена

Освен атрибутивна и предикативна функция прилагателните имена относително често демонстрират и самостоятелна употреба. Подобни употреби могат да се дължат на елипса на съществителното, което се пояснява от прилагателното, или на субстантивация на прилагателното (Ницолова 2008:118 и сл.). От синтактична гледна точка елипсата може да бъде дефинирана като синтактична операция, при която се изпуска елемент от изреченската структура, тъй като се подразбира. От семантична гледна точка елипсата може да се интерпретира като процес на изпускане на възстановим елемент, като подчиненият експлицитно изразен конституент запазва общата семантика на разпадналата се фраза – това явление се означава и като субстантивация (вж. по-подробно и у Якубович 2004: 37).

6.2.1. Елипса

При елипса съществителното може да бъде изпуснато само когато е възстановимо от обкръжаващия го контекст, най-общо предходно или следходно споменаване на съществителното.

(3) *Има бели (рози) и червени рози. Дайте ми една бяла (роза).*

При елипса на главна част съществително име са възприети два подхода при анотацията на прилагателното модификатор – или се избира синонимно множество, съответстващо на значението на прилагателното, или се възстановява липсващата опора.

Прилагателни, които са модификатори в именни словосъчетания с елипса на главната част, които имат подобна синтактична структура с друго именно словосъчетание в изречението без елипса, се анотират като прилагателни.

- (4) *Когато първата съгласна упражнява влияние върху втората¹, имаме прогресивна асимилация.*

Литералът *втори* (вж. и 8.1. по-долу) се среща в четири синонимни реда:

{*втори:1; 2-ри:1; II:1; 2:5; ii:3*}

Дефиниция: *‘който следва първия и прехожда третия по позиция, време и т.н.’*

{*втори:2*}

Дефиниция: *‘нещо или някой, който следва началния или челния в поредица или йерархия’*

{*повторен:1, втори:3*}

Дефиниция: *‘който се извършва, става след друг’*

{*последен:II, втори:4*}

Дефиниция: *‘който се отнася до по-късно споменатия от два обекта или групи обекти’*

Избира се последният вариант, тъй като най-точно отразява значението на думата съобразно контекста, в който е използвана.

Съществителните имена могат да бъдат елиптирани и когато са част от съставни думи (най-често при координация). За разлика от случаите с елиптични свободни съчетания от прилагателни и съществителни, при съставните думи се възстановява изпуснатата дума, за да се припише точното значение.

- (5) *Такива процеси са превръщането на енергията на химичните реакции в **електрична (енергия)** и обратно, пряко участие на електричеството в химични процеси – производството на различни вещества по електрохимичен път.*

(5) е пример за съставна лексикална единица с елипса на опората, която следва да бъде възстановена при аотирането, като *електрична* се разширява в лемата до *електрична енергия*.

¹ В съгласие с общите конвенции, възприети за БулНет, редните числителни се въвеждат като част на речта прилагателно име.

6.2.2. Субстантивирани прилагателни

Субстантивирани прилагателни реферират към съществителни имена. При субстантивиранието на прилагателни имена е в ход процес на номинация, при която от двусъставни атрибутивни словосъчетания остава само атрибутивната част, която функционира като субстатив и поема номинативната функция на цялото словосъчетание (Зидарова; за ролята на определеността, изразена чрез членуване, притежателни форми и форми в множествено число, при субстантивация, вж. у Ницолова 2008: 66, 70, 118).

Субстантивирани прилагателни функционират в текста като съществителни и следва да бъдат разграничавани от елиптичните словосъчетания, тъй като в лексикалното си значение включват и думата, с която биха се свързали. Субстантивирани прилагателни често имат фиксирана семантична интерпретация, например *възрастните* (членувано, мн.ч.) най-често се отнася до *възрастните хора*, а не до *възрастните кучета*.

Прилагателните, които са развили субстантивно значение, се въвеждат в лексикално-семантичната мрежа като съществителни имена, като според значението им синонимното множество може да включва само субстантивирани прилагателни или съществителни и синонимните им субстантивирани прилагателни. Следователно, омонимни словоформи биха могли да бъдат анотирани съгласно употребата си в конкретния контекст – като прилагателни, като в (6), или като съществителни като в (7).

- (6) *И Тарилйомът и Селямсъзът са **учени** и почтени, но нямат съгласието, а без съгласието, като няма него, всичко пропада.*

В този пример *учен* е прилагателно в предикативна употреба, свързано съчинително с друго прилагателно *почтен*, затова се анотира като прилагателно име със съответното значение.

{*учен:2, обучен:2, школуван:1*}

Дефиниция: *‘който е получил знания и е развил умения в процеса на специализирано обучение в конкретна област’*

- (7) *Присъствието на доц. Аргирова, заместник-министър на здравеопазването и експерт на Световната здравна организация по СПИН, един от шестимата европейски **учени**, член на ръководството на Световната програма по СПИН на Организацията на обединените нации, е неприемливо.*

В (7) за анотация на думата *учен*, която е пояснена от модификатор прилагателно име и за чиято референтност сигнализируют числителните, отнасящи го към един индивидуален референт от множеството (за определеността и неопределеността в подобни фрази, вж. Грозева 1979, Станков 1998, Стаменов 1985), се избира синонимното множество {*учен: I*}, което е определено с част на речта съществително име.

{*учен: I*}

Дефиниция: *‘човек с висока квалификация в някоя област на познанието, обикновено притежаващ съответния ценз’*

Субстантивирани прилагателни са въведени в БулНет като литерали в синонимни множества с част на речта съществително име, като в бележка към литерала се отбелязва, че прилагателните са субстантивирани. Най-често срещаните субстантивирани прилагателни са от среден род единствено число с абстрактно значение (*модерното, непознатото*, вж. Ницолова 2008: 118). Названията на цветове представляват отделна семантична подгрупа, в която също са представени субстантивирани прилагателни.

6.3. Качествени и относителни прилагателни

Една от важните класификации на прилагателните имена е подялбата им на качествени и относителни (СБЕ 1998: 325-326; вж. и у Ницолова 2008). Няма да бъдат разглеждани подробно всички разлики между тях, а само тези, които засягат анотацията и оказват влияние върху избора на значение.

6.3.1. Качествени прилагателни

Качествените прилагателни обикновено назовават „първични, естествени и постоянни характерни белези на явленията, които не зависят от отношенията им с други явления,, докато относителните се мислят като съдържащи „връзки или отношения с други явления,, (вж. у Куцаров 2007).

Качествените прилагателни най-често изразяват размери, цветове, вкусове, физически свойства и интелектуални, волеви и нравствени качества и могат да бъдат мислени като проявяващи се в по-голяма или по-малка степен. Това е причината по-голяма част от тази група прилагателни, за разлика от относителните, да имат форми за сравнителна и превъзходна

степен (СБЕ 1998:325; Ницолова 2008:115). В БулНет някои прилагателни имена са въведени във формите си за сравнителна и превъзходна степен като литерали в отделни синонимни множества, обикновено при наличието на еквиваленти в Принстънския wordnet при случаи на по-необичайно формообразуване и значение на степенуваните форми, което е различно от това на нестепенуваните. Подобни синонимни множества са свързани със синонимното множество {сравнителна степен:1} или {превъзходна степен:1} с помощта на релацията *сфера на употреба*. В (8) *най-бърз* се разглежда като форма на *бърз* от съответното синонимно множество.

(8) *Напомняме ви как стояха нещата през седмицата: вторник Михаел Шумахер бе най-бърз с 1:18,444, следван от Баричело.*

{бърз:2}

Дефиниция: *‘който се извършва с голяма скорост или е способен да извършва движения с голяма скорост, за кратко време’*

От качествените прилагателни се образуват нови (оценъчни) прилагателни, с които се изразяват различни степени на означеното качество или емоционално отношение към определяния обект. В зависимост от типа и семантиката на словообразователния афикс оценъчните прилагателни са умалителни или увеличителни, които се интерпретират по следния начин:

i) Добавят се към синонимното множество на прилагателното, на което са дериват (9).

(9) *Гласът е грубоват.*

{груб:4, грубоват:1, стържец:1, дразнец:2, остър:8}

Дефиниция: *‘за глас, звук – който има неприятно дрезгаво звучене’*

ii) Когато в структурата на Принстънския wordnet са налице прилагателни с оценъчно значение, в БулНет се дефинира еквивалентно синонимно множество.

6.3.2. Относителни прилагателни

Тази подгрупа прилагателни обхваща най-често производни от други думи, предимно собствени или нарицателни съществителни имена. Когато в Принстънския wordnet съществува синонимно множество с такова прилагателно, в БулНет се въвежда негов еквивалент.

{*wooden:2*}

Дефиниция: '*made or consisting of (entirely or in part) or employing wood*'

{*дървен:1*}

Дефиниция: '*който е изработен или се състои изцяло или отчасти от дърво*'

По-често значението на българските относителни прилагателни в английски се предава чрез атрибутивно употребени съществителни имена. По тази причина в голяма част от случаите българските относителни прилагателни нямат съответствия в Принстънския wordnet и по тази причина в БулНет се въвежда ново синонимно множество, което се свързва със съответното съществително с помощта на деривационната релация *произведен*.

Езиково специфичните относителни прилагателни могат да бъдат образувани от собствени и нарицателни съществителни. Например прилагателни имена – производни от лични и географски наименования: {*ямболски:1*} с дефиниция '*който се отнася до град Ямбол или до жителите му*' е произведен от *Ямбол* {*Ямбол:1*} с дефиниция '*голям град, разположен в югоизточната част на България*'.

6.4. Сложни прилагателни

Относително често срещани в текстовете от определен стил са сложните прилагателни, образувани от две основи. Когато са обединени в сложна дума, прилагателните се въвеждат в ново синонимно множество, свързано с двете основи, от които е образувано, посредством релацията *произведен*. Например *културно-етнически* (10) от {*културно-етнически:1*} се свързва чрез релацията *произведен* както с {*културен:3*}, така и с {*културен:1, етничен:1, етнически:3*}.

(10) *Хората, изпаднали в несигурност, търсят нови социални връзки и все по-силно се ориентират към привидно природообосновани категории като нация, раса, религия, а социални проблеми като безработица, жилищна криза, дефицит в образователната система и свързаните с това социални неравенства свеждат до културно-етнически неравноценности.*

6.5. Атрибутивно употребявани съществителни от чужд произход

Някои съществителни от чужд произход са развили атрибутивна употреба. Част от тях се въвеждат като нови синонимни множества, други се добавят във вече съществуващи. Думата *бизнес* например е добавена в множеството {*делови:1, работен:2, бизнес:1*}. Интересен е случаят с *офис*, където се наблюдава и образуване на прилагателно с помощта на българска наставка: {*офисен:1, офис:3, служебен:2*}.

Една от специфичните особености на атрибутивно употребяваните съществителни е, че те не изразяват формално изменението по род и число и съвпадат със съответните съществителни. В тези случаи анотацията в Българския семантично анотиран корпус се основава на синтактичната функция на думата в изречението – (11) и (12).

- (11) *Той се е покачил с 1,5 пункта, което се дължи почти изцяло на повишения оптимизъм в строителството, свързан с очакваната бизнес ситуация през следващите шест месеца.*
- (12) *Специалистите, с които се консултира, преди да се захване с този бизнес, излязоха прави.*

В пример (11) думата „*бизнес*„ е употребена с атрибутивна функция, т.е. определя следходното съществително *ситуация*, което изисква анотирането ѝ като {*делови:1, работен:2, бизнес:1*} с дефиниция: ‘*който е свързан с изпълнението на дадена работа или с професионални задължения*’. В пример (12) същата дума е използвана с характеристиките на съществително (самата тя бива определена от показателното местоимение в атрибутивна функция *този*), затова се анотира като {*брани:2, бизнес:3*} с дефиниция ‘*професия или сфера на дейност*’.

6.6. Причастия

„Причастията са нелични глаголни форми, чрез които на лица и събития могат да се приписват признаци от глаголни действия„ (ГСБКЕ 1982), наричани още „глаголно-именни форми„ (Куцаров 2007:114). Някои причастията се въвеждат в БулНет с част на речта прилагателно, когато са възникнали нови и различни значения. Това налага разнородна интерпретация на причастията при конкретните употреби, като при анотацията е възможно да се асоциират с основна форма глагол като в (12), където причастие *гледан* е част от аналитична глаголна форма (страдателна), или прилагателно като в (13), където *разглеждан* може да се замести с *въпросен* и е с еквивалент прилагателно име в английски.

- (12) *Работейки по него, преди да го подпиша, се убедих, че възникват въпроси, които наистина трябва да бъдат гледани в единство, за да не правим след това още буквално на следващия ден поправки, за да не правим корекции, които биха спрели нужното за тази част от обществото.*

{възприемам:4, възприема:4, гледам:8, имам:8}

Дефиниция: ‘смятам нещо или някого за някакъв според вътрешната си, субективна оценка’

- (13) *Ще имате шансове да реализирате пътуване по работа още към средата на **разглеждания** месец.*

{въпросен:1, разглеждан:1, разискван:1, обсъждан:2, дискутиран:1}

Дефиниция: ‘за който/когото се говори, който е предмет на обсъждане, дискусия и т.н.’

Девербалните прилагателни, образувани с наставките -им, -ем (-аем, -уем), които са били форми на сегашни страдателни причастия, се въвеждат в БулНет и се анотират в корпуса като прилагателни (повече за девербалните прилагателни вж. у Иванова 2007, Радева 1991).

6.7. Заключение

При анотацията на прилагателните наред с общите конвенции – например анотацията на сложните думи, се открояват някои специфични само за тази част на речта особености – например интерпретацията на езиково специфичните относителни прилагателни, получени в резултат на продуктивна деривация.

Цитирана литература

ГСБКЕ 1983: *Граматика на съвременния български книжовен език. Том II.* Издателство на БАН. София.

Грозева 1979: Грозева, М. *Употреба на словоформата един за предаване на неопределеност в българския език, съпоставена с употребата на неопределителния член ein в немския език.* — Съпоставително езикознание, 1979, № 4, 13—21.

- Зидарова: Зидарова, В. *Прояви на лексикална икономия в съвременния български език*. На: http://www.belb.net/personal/zidarova/Lex_economy.htm
- Иванова 2007: Иванова, Ц. *Образуване на прилагателни имена от глаголи в съвременния български книжовен език*. – Български език, 2007, кн. 2, 63-74.
- Куцаров 2007: Куцаров, Ив. *Теоретична граматика на българския език*. Университетско издателство „Паисий Хилендарски”, Пловдив.
- Ницолова 2008: Ницолова, Р. *Българска граматика. Морфология*. Университетско издателство „Св. Климент Охридски”, София.
- Радева 1991: Радева В. *Словообразуването в българския книжовен език*. Университетско издателство „Св. Климент Охридски”, София.
- Стаменов 1985: Стаменов, Хр. *За употребата на един като показател за неопределеност в български език (в сравнение с английски)*. – Език и литература, 1985, № 3, 33—44.
- Станков 1998: Станков, В. *Особености на генеричната употреба на детерминаторите за определеност/неопределеност в българския език*. – Във: Александров, Ал., Р. Русинова Помагало по българска морфология. Имена. Шумен, „Епископ Константин Преславски”, 170-182.
- Якубович 2004: Yakubovich, I. *Head-noun ellipsis in Hittite and elsewhere (A study in recoverability conditions)*. The Panels. 37-49.

7. Лингвистични конвенции при анотация на наречията

Христина Кукова

Според ГСБКЕ наречието е „самостояйна неизменяема дума, която означава признак на действие, на друг признак или на предмет” (с.387). Куцаров стеснява определението до „признак на друг признак” (Куцаров 2007, с. 127).

Основната граматична характеристика на наречията е флективната им неизменяемост, т.е. представителите на тази група не получават различни граматични форми (като се изключат формите за степенуване при някои наречия).

Наречията, които се срещат в корпуса и до този момент не са въведени в лексикално-семантичната мрежа, се добавят като нови синонимни множества. Те, от своя страна, се свързват с множество, означаващо най-общо време, място, начин, степен, количество, причина, честота, състояние и др., чрез релацията *тематична област*. Друга възможност представлява свързването им със синонимни множества според функцията, която изпълняват, или сферата на употреба – адюнкт, интензификатор, разговорна дума, диалект, идиом, формалност и др. посредством релацията *сфера на употреба*.

7.1. Видове наречия в БулНет

Таблица 1. илюстрира всички групи наречия в БулНет с посочени примери и актуални числени данни за множествата от всяка група.

Групи наречия	Пример за синонимно множество	Брой
{място:4}	{там:2} 'означава конкретното място, на което или в което се извършва посоченото в изказването'	68
{начин:2, стил:2, маниер:1}	{бързо:1, скоростно:1, живо:1, светкавично:1} 'характеризира посоченото в изказването като извършващо се с бързи движения'	355
{време:9}	{тогава:1, по това време:1} 'отнася посоченото в изказването към точно определен минал момент'	115

{емфаза:1}	{обаче:2} 'означава, че посоченото в изказването контрастира, противопоставя се на предишно твърдение'	6
{честота:1, честотност:1, фреквентност:1}	{рядко:2, нарядко:1, изрядко:1} 'означава, че посоченото в изказването се реализира през неопределено дълги интервали от време'	17
{степен:3, ниво:3, равнище:6}	{съвсем:1, напълно:1} 'характеризира посоченото в изказването като реализиращо се в най-голяма възможна степен или до възможния си предел'	77
{модалност:1}	{вероятно:3, навярно:1, сигурно:3, по всяка вероятност:1} 'изразява категорично положителна оценка относно възможността посоченото в изказването да се реализира или да се е реализирало'	18
{количество:1}	{малко:7} 'означава, че посоченото в изказването се реализира слабо или рядко'	21
{основание:4, причина:5, мотив:3}	{незнайно защо:1, кой знае защо:1, бог знае защо:1} 'представя причината за извършването на дадено действие като неизвестна за говорещия'	4
{ориентация:5, посока:4}	{хоризонтално:1} 'характеризира посоченото в изказването като разположено успоредно на земната повърхност'	2
{сравнение:2}	{нито:3, нито пък:1} 'изтъква, че по подобие на нещо друго, посочено по-напред в изказването, и това (въвеждано от съюза) не се осъществява, не протича, не е налице и под.'	1

{оценяване:1, оценка:1, преценяване:1, преценка:1}	{лично:3} 'означава, че говорещият прави бележка, коментар, оценка и под. по отношение на посоченото в изказването от свое име или от собствената си гледна точка'	4
{условие:4, предусловие:2}	{само ако:1, само когато:1} 'означава, че посоченото в изказването се реализира или е вярно при наличието на определени условия'	1
{фокус:2}	{особено:3} 'характеризира посоченото в изказването като осъществяващо се в подчертано по- голяма степен, обхват и т.н. в сравнение с обичайното или нормалното'	1
{заключение:3, извод:2, логически извод:1}	{оттам:3, следователно:2, оттук:2} 'въвежда логическо заключение в рамките на простото или сложното изречение, при което второто от посочените твърдения е резултат или следствие от първото'	1
{следствие:2, последствие:2, резултат:3, последица:2, ефект:3}	{тогава:3, в такъв случай:1, значи:1} 'представя действието като следствие от нещо'	1
{отношение:2, вързка:8}	{по този въпрос:1, по този пункт:1, в това отношение:2, там:5} 'адвербиален израз, който замества въпрос, тема, аспект, по отношение на които посоченото в изказването се преценява'	3
{логическо отношение:1, логическа релация:1}	{обратно:8, наопаки:2} 'означава, че логическата последователност или редът на посочените участници или елементи на ситуацията са разменени, преобърнати'	1

Таблица 1. Групи наречия

Особеност на наречията е развиването на функции, характерни за други части на речта, например на предлози или съюзи (*сякаш, прочее, следователно* и др.; вж. част 11.3.). В БулНет тази разлика в граматичната характеристика е отразена чрез въвеждането им като отделни синонимни множества и чрез свързването им със съответната категория с релацията *тематична област*. В бележка към синонимното множество е отбелязана съответната част на речта.

Релация <i>тематична област</i> с:	Пример за синонимно множество	Брой
{ <i>съчинителен съюз:1</i> }	{ <i>освен това:1, при това:2, а пък и:1, пък и:1, и без това:2, а и:1, плюс това:2</i> } ' <i>въвежда присъединителни изречения, с които се допълва посоченото в изказването</i> '	5
{ <i>подчинителен съюз:1</i> }	{ <i>така че:1, така че да:2, за да:2, да:8</i> } ' <i>подчинителен съюз за въвеждане на подчинително обстоятелствено изречение за цел</i> '	5
{ <i>въпросително/местоимение:1</i> }	{ <i>докога:1</i> } ' <i>въпросително местоименно наречие, с което се пита за периода или крайната точка във времето, до която ще продължи да съществува, да се реализира, проявява или протича посоченото в изказването</i> '	7
{ <i>неопределително местоимение:1</i> }	{ <i>отнякъде:1</i> } ' <i>неопределително местоименно наречие за означаване на неопределено или неизвестно място, източник, посока, откъдето идва или произхожда посоченото в изказването</i> '	3

{частица:2}	{дори:1, дори и:1, и:3, даже:1} 'подчертава съдържанието, смисъла на посоченото в изказването, към което се отнася, като насочва към сравнение с нещо друго, изразено или подразбиращо се, особено когато то е неочаквано или изненадващо'	3
{предлог:1}	{на прага на:1} 'предложен израз, който означава непосредствено доближаване до някой или нещо или приближаване до някаква цел, резултат, изход, без обаче да са постигнати'	3

Таблица 2. Групи наречия, свързани с екстралингвистичната релация тематична област

Таблица 3 показва групите наречия според функцията, която изпълняват в изречението, или сферата на употреба, свързани помощта на релацията *сфера на употреба*. Възможно е едно синонимно множество да фигурира в повече от една класификация:

Релация <i>сфера на употреба</i> с:	Пример	брой
{адюнкт:1}	{така да се каже:1, тъй да се каже:1} 'адюнкт, който служи за изразяване на възможна, вярна или валидна гледна точка или оценка на посоченото в изказването'	48
{разговорна дума:1, жаргонна дума:1, разговорен израз:1, жаргонен израз:1, колоквиализъм:1}	{ужасно:1, адски:1, страшно:2} 'интензификатор на количество или качество'	13
{интензификатор:1}	{направо:1, просто:1} 'модификатор за подчертаване, наблягане на силата и пълната реализация на посоченото в изказването'	24

{идиом:1, идиоматичен израз:1, фразеологизъм:1, фразеологично словосъчетание:1}	{като по часовник:1} 'характеризира посоченото в изказването като отличаващо се с изключителна редовност и прецизност'	1
{комбинативна форма:1}	{лошо:2, зле:1} 'характеризира начина на извършване на посоченото в изказването като незадоволителен, недостатъчно добър, неподходящ или неуместен'	2
{диалект:2, диалектна реч:1, наречие:1, говор:2}	{напред:2} 'характеризира посоченото в изказването като осъществяващо се, намиращо се, движещо се или ориентирано по посоката, в която се намира предната част или лицето' Пристъпете напред.	1
{формалност:1, официалност:1}	{там:2} 'означава конкретното място, на което или в което се извършва посоченото в изказването'	1

Таблица 3. Групи наречия, свързани с екстралингвистичната релация сфера на употреба

Сравнителните и превъзходните степени на наречията, които ги образуват, както и на прилагателните, се свеждат до основната форма на думата. В сравнителна или превъзходна степен се запазват само тези, които изразяват различно значение в сравнение с положителната степен. В тези случаи формите за сравнителна или превъзходна степен се въвеждат отделно в лексикално-семантичната мрежа – (1).

(1) **Най-после** разбра.

{накрая:2, най-накрая:1, в края на краищата:2, най-после:1, най-сетне:1, в крайна сметка:3}

Дефиниция: 'характеризира посоченото в изказването като краен резултат от процес или поредица от действия'

В останалите случаи думата се аотира съобразно вариантите, които се предлагат в основната форма.

Сложните наречия (Ницолова 2008:445) се интерпретират по същия начин като останалите сложни думи – свързват се със синонимните множества, към които принадлежат съставните им части.

7.2. Заключение

В заключение следва да се отбележи, че тази част от изложението засяга предимно наречията, функциониращи в примерите от корпуса като адюнкти към глаголни фрази. Някои наречия, въвеждащи подчинени изречения, и адвербиални устойчиви изрази са разгледани по-подробно в другите части.

Цитирана литература

ГСБКЕ 1983: *Граматика на съвременния български книжовен език. Том II*. Издателство на БАН. София.

Куцаров 2007: Куцаров, Ив. *Теоретична граматика на българския език*. Университетско издателство „Паисий Хилендарски”, Пловдив.

8. Лингвистични конвенции при анотация на числителните

Светлозара Лесева

8.1. Обща класификация на числителните в Принстънския wordnet и БулНет

8.1.1. Числителни за точно количество

Както бе посочено в част 3, в Принстънския wordnet числителните имат следната организация: (i) числителните бройни – примери (1а) и (1б); и числителните редни – примери (2а) и (2б), с атрибутивна и предикативна функция, изразяващи количествен признак на същностите, са въведени като синонимни множества с част на речта „прилагателно име“; (ii) числителните бройни в самостоятелна употреба като число с дадена стойност или като графичното означение на такова число, като в примери (1в) и (1г), както и производните им съществителни (*двойка*, *тройка* и т.н.), се въвеждат като литерали в синонимни множества с част на речта „съществително име“.

- (1) а. *Сервизът се състои от **шест** чаши.*
б. *Чашите са само **пет**.*
в. ***Девет** е едноцифрено число.*
г. ***Едно** е по-малко от **две**.*
- (2) а. ***Вторият** студент влезе на изпита.*
б. *Тя бе **пета** в класацията.*

Така в структурата на лексикално-семантичната мрежа числителните са разделени в три класа – числителни бройни, условно наречени прилагателни, които се свързват чрез релацията *подобен на* със синонимното множество {броен:1} с дефиниция ‘който означава определено количество единици от нещо’, числителни редни, също условно наречени прилагателни, които се свързват със същата релация с множеството {реден:1} с дефиниция ‘който означава реда на даден елемент в поредица от еднородни предмети, явления и пр.’, и числителни, условно отнесени към класа на съществителните, които се въвеждат като хипоними на синонимното множество {число:1} – като преки или косвени хипоними на {цяло число:1} или на {дроб:1, дробно число:1}.

Това разделение до голяма степен съответства на възприетите принципи на класификация в българистиката. Така например, Ницолова (2008:130-134) посочва, че бройните числителни „означават числа в рам-

ките на десетичната система” (съответстващи на числителните „съществителни” в wordnet), или „конкретизират количеството на предметите с помощта на числата” (съответстващи на числителните „прилагателни”). Употребяват се самостоятелно или в съчетание със съществителни имена атрибутивно или предикативно. Следователно двата типа употреби имат аналог в дистрибуцията на „съществителни” и „прилагателни”. Числителните редни означават „мястото на предмет или група предмети в реда на числата” (т.е. имат поясняваща функция към имена). Отбелязва се, че формообразуването им е подобно на това на прилагателните, както и фактът, че семантичните и морфологичните им особености дават основание на някои автори дори да ги отнасят като подвид на относителните прилагателни (както е направено в wordnet).

Графично числителните се въвеждат както с езиковите, така и с числените си означения (с арабски и/или с римски числа), например: {сто:1, 100}, {десети:1, 10-и:1, 10:1, X:4, x:5}

8.1.2. Числителни за относително количество

Съществуват два модела за образуване на числителните бройни, означаващи приблизително количество: (i) безсъюзно (графично отбелязвано с дефис) – 5-6, *седем-осем*; (ii) чрез наставка *-ина* – *десетина*, *двайсетина* и т.н. (Ницолова 2008:136-137; Куцаров 2007:80-81)

В Принстънския wordnet не са въведени числителни, изразяващи приблизително количество, тъй като тези значения се изразяват чрез комбинации с наречия или прилагателни *or so*, *odd*, *some*, *about*, *approximately*, аналози на които има и в български (*около* *двайсет*, *приблизително* *двайсет*, *трийсет* и *няколко*).

Числителните за приблизително количество с наставка *-ина*, засвидетелствани в корпуса за анотация, се въвеждат като езиково специфични синонимни множества, които се свързват в структурата на лексикално-семантичната мрежа аналогично на останалите числителни бройни. Образованите безсъюзно числителни за приблизително количество се отделят като отделни графични думи (токъни) и се анотират като съответните числителни бройни за точно количество.

8.2. Числителните в състава на думи от друга част на речта

Числителни влизат и в състава на някои: сложни прилагателни – напр. *петчленен*, *триетажен*, *едноседмичен*, *двугодишен*; наречия –

двойно, петкратно и т.н.; съществителни – витамин В12, 1-ви април, стотица, петдесетачка; фразеологизми – *правя се на две и половина*, няма **шест-пет**, по първи петли. Когато сложните прилагателни от БулСемКор не фигурират в Принстънския wordnet, те се въвеждат като уникални множества в БулНет.

Освен сложните в българоезичната литература се посочват още три типа прилагателни, образувани чрез суфиксация: (i) „за означаване на броя на съставките на предмети”, образувани със суфиксите *-оен/-оин- -орен/-орн-*; (ii) прилагателни, които „означават единично или мултиплицирано действие или резултат от действие”, образувани със суфикс *-кратен*; и (iii) прилагателни „за означаване на повече от един различни начини на действие” със суфикс *-оак* (Ницолова 2008: 139, подобна е формулировката и у Куцаров 2007: 83).

От числителни бройни се образуват и: (i) наречия за кратност с наставка *-ж* (за числителните *едно*, *две* и *три*) и наставка *-кратно*; (ii) наречия, които означават „количествени признаци на свързаните с дадено действие предмети” със суфикс *-ойно*, *-орно*.

От числителни редни чрез конверсия се образуват и наречия за изброяване *първо*, *второ*, *трето* (Ницолова 2008: 140-141), както и съществителни имена, които обозначават определена стойност на банкнота или монета – *десетачка*, *стотачка*. Посочените класове прилагателни, съществителни и наречия са системно представени в Принстънския wordnet и БулНет и анотацията им следва общите принципи при съответната част на речта.

По-долу ще бъде разгледана анотацията на числителните в състава на числени и темпорални изрази.

8.3. Анотация на изрази с конституенти числителни имена

В корпуса се срещат различни типове фрази за време, количество и т.н., които съдържат числителни имена. Ще бъдат посочени спецификите на анотацията в основните случаи, срещащи се в БулСемКор.

Ако числителните в състава на такива изрази са въведени в Принстънския wordnet, те се анотират със съответстващото им значение – като бройно или редно числително прилагателно или като съществително според функцията им. Поради големия брой на числителните последователното им въвеждане в wordnet е ограничено: (i) за бройните (версия 2.0 на Принстънския wordnet) – всички до 99, кратните на 10 до 200, кратните на 100 до 500, 1000, 10000, 100000, 1 милион, 1 милиард;

(ii) за редните – всички до 31-ви, кратните на десет от 40-и до 190-и, кратните на сто от 200-ен до 500-ен. Произволни числени стойности освен посочените не са въведени в тази версия на Принстънския wordnet. В тези случаи е възприето да не се създават нови множества в БулНет (срв. примери (3) и (4)), а числителните да се лематизират като „число” и да се аотират със синонимното множество {число:1} с част на речта „съществително име” независимо от функцията им и от онтологичната категория (време, количество и т.н.), към която принадлежат изразите, в които участват.

(3) *30* {трийсет:1, тридесет:1, 30:2} *етажа* {етаж:1}

(4) *кабинет* {кабинет:1} *356* {число:1}

Безсъюзно свързаните числителни в изрази за приблизителна стойност се сегментират и аотират поединично, както в (5):

(5) *в последните 35* {трийсет и пет:1, тридесет и пет:1, 35:1} – {-:1} *40* {четирийсет:1, четиридесет:1, 40:1} *години* {година:1, г.:1, год.:1}

Тъй като има идеографска функция, дефисът за изразяване на приблизителност е въведен в БулНет и аотиран по съответния начин.

8.3.1. Фрази, съдържащи числителни, които обозначават количествени стойности

Целите и дробните числа се аотират според общите критерии за числената част на свободните фрази, съдържащи числителни имена, посочени по-горе. Мерните единици и други думи, означаващи вида на измерваната същност, се аотират като съответното множество, както е посочено в (6) и (7):

(6) а. *30* {трийсет:2, тридесет:2, 30:2} *м* {метър:2, м:1, т:1}

б. ... *като през него преминават около 70 000* {число:1} *литра* {литър:1, l:1, л:1}

(7) *две трети* {две трети:1}

При наличието на специални означения – проценти, валута, температурна скала и т.н., специалните символи също се аотират със съответстващото им множество – (8). Тъй като често числената част и специалният символ се изписват без интервал, най-напред трябва да бъдат сегментирани.

- (8) а. 40 {четирийсет:1, четиридесет:1, 40:1} % {процент:3, знак за процент:1, %:1}
 б. На осветената от Слънцето страна температурата достига 430 {число:1} °C {градус по Целзий:1, ° C:1}.

8.3.2. Изрази за време

В корпуса се срещат различни фрази с числена част, обозначаващи времеви периоди, например *3-9 октомври*, *14 ноември 1989 г.*, *април 89-а*, *21 юни т. г.*, *в периода 1997 – 2001 г.*, дати и други темпорални изрази.

8.3.2.1. Дати

Според конвенцията в БулСемКор датата от месеца се аностира като числителното редно, което съдържа литерал с дадената числена стойност, а месецът – със съответното синонимно множество. Тъй като е невъзможно, а и ненужно всички възможни години да се въведат в wordnet, в случай че числителното не фигурира в лексикално-семантичната мрежа, то се лематизира като *година*. Например фразата *14 ноември 1989 г.* се аностира по следния начин в (9):

- (9) 14 {четиринайсети:1, четиринадесети:1, 14:4, 14-и:1, XIV:2:, xiv:3} ноември {ноември:1} 1989 {година:1, г.:1, год.:1} година {година:1, г.:1, год.:1}

8.3.2.2. Периоди от време

При периоди, означени с конкретна начална и крайна година (*в периода 1863-1923*) се постъпва както в (9), като числата се разделят, както в (5).

- (10) *в периода 1863* {година:1, г.:1, год.:1} – *1923* {година:1, г.:1, год.:1}
 (11) *Двайсет* {двайсет:1 двадесет:1 20:1} години {година:1, г.:1, год.:1} *чакам тоя ден*.

Десетилетия, например *20-те*, *30-те години* и т.н., са въведени в Принстънския wordnet и в БулНет и се аностират по съответния начин, като различните варианти за изписване – съкратени, несъкратени, с числени означения, са въведени като литерали в съответното множество – (12):

- (12) *...60-те години* {шейсетте:1, шейсетте години:1, шестдесетте:1, шестдесетте години: 1, шейсетте години на деветнайсети век:1, шестдесетте години на деветнайсети

век:1, 60-те:1, 60-те години:1, 60-те години на деветнайсети век:1, 60-те години на деветнадесети век:1, 60-те години на XIX век:1... } са време на една небулоза на съзнанието...

3.2.3. Периоди, означаващи възраст

Периодите от човешкия живот, равняващи се на десетилетия, са въведени в wordnet като хипоними на синонимното множество {*период от човешкия живот:1; етап от човешкия живот:1*}, чиито хипоними са също *детство, юношество, пубертет, младост...* – (13).

(13) *Беше висока, около 50-те {петдесетте:1, 50-те:1} и още красива.*

{*петдесетте:1, 50-те:1*}

Дефиниция: '*период от човешкия живот, обхващащ от петдесетата до шестдесетата година*'

8.3.3. Други случаи

Числителните участват в състава на разнообразни други словосъчетания, например при цитиране, в идентификационни номера, телефонни номера и т.н. При цитиранията е възприето името на автора (авторите) да се анулира според това, дали е въведено в wordnet, или не. Във втория случай се анулира със синонимното множество, обозначаващо онтологичния клас, към който принадлежи, напр. {*фамилия:1, фамилно име:1*}. Номерът на страницата се лематизира и анулира като {*страница:2, номер на страница:1*} (14):

(14) *Сиксу {Сиксу:1, Елен Сиксу:1} 1998 {година:1, г.:1, год.:1}: 389 {страница:2, номер на страница:1}*

Номера на договори, регистрационни номера и т.н. се лематизират и анулират като {*номер:4, идентификационен номер:1, №:1*} поради факта, че може да са както поредици от цифри, така и низове от числа и други символи (15), а телефонните номера като {*номер:3, телефонен номер:1, телефон:1, тел.:1*}, както в (16):

(15) *договор {договор:2, контракт:1} 332/17.12.1997г. {номер:4, идентификационен номер:1, №:1}...*

(16) *+35952/453729 {номер:3, телефонен номер:1, телефон:1, тел.:1}*

При по-специфични случаи се следва общата конвенция, като се извършват модификации, където се налага.

8.4. Заключение

В последните години на идентифицирането и класифицирането на числените и темпоралните изрази се отделя все по-голямо внимание във връзка с разпознаването на именуванни същности (Named Entity Recognition). Разглеждането на числителните тук има за цел да набележи основните случаи, които са срещнати в хода на анотацията на БулСемКор, заедно с решенията с оглед на описанието на този тип езикови изрази за практически цели.

Цитирана литература

- Куцаров 2007: Куцаров, Ив. *Теоретична граматика на българския език. Морфология*. Университетско издателство „Паисий Хилендарски”, Пловдив.
- Ницолова 2008: Ницолова, Р. *Българска граматика. Морфология*. Университетско издателство „Св. Кл. Охридски”, С.

9. Лингвистични конвенции при анотация на местоименията

Екатерина Търпоманова

9.1. Местоименията в БулНет

Местоименията като относително затворен клас думи, отличаващ се от пълнозначните думи, са добавени към Българския wordnet за целите на анотацията на корпуса. Възприета е концепцията за местоименията като единен клас думи, включващ същинските местоимения (субститути на съществителни имена и именни групи), местоименията прилагателни и местоименните наречия (Ницолова 1991:30), като същевременно е взета под внимание и синтактичната им функция, както и семантичните характеристики на отделни представители на класа. Въвеждането на семантични редове местоимения в Българския wordnet е съобразено с особеностите на местоименната система в английски език, тъй като едно от приложенията на Българския семантично анотиран корпус е провеждането на експерименти за автоматно отстраняване на семантична многозначност. Това налага създаването на отделни синонимни множества за една и съща местоименна форма в Българския wordnet с оглед на превода ѝ на английски, например {*мой:1, ми:1*} '*притежателно местоимение, което определя поясняваната от него дума*' с преводен еквивалент на английски *tu* и {*мой:2*} '*предикативна употреба на притежателното местоимение*' с преводен еквивалент *mine*.

Видовете местоимения (лични, показателни и т.н.) са добавени в Българския wordnet като хипоними на синонимното множество {*местоимение:1*}, а всяко отделно множество с част на речта „местоимение” се свързва чрез релацията *тематична област* със съответния вид, към който принадлежи, например {*аз:1*} е свързано с релацията *тематична област* с {*лично местоимение:1*}, {*мой:1, ми:1*} – с {*притежателно местоимение:1*} и т.н. В синонимното множество местоименията се добавят в основната си форма (първо лице, единствено число, мъжки род, номинатив). В БулНет видовете местоимения имат следния вид (в малки скоби е посочен броят на синонимните множества местоимения, свързани към съответния вид с релацията *тематична област*):

{*местоимение:1*}

{*хипоним: притежателно местоимение:1*} (2 синонимни множества)

[хипоним: възвратно притежателно местоимение:1] (1 синонимно множество)

[хипоним: въпросително местоимение:1] (15 синонимни множества)

[хипоним: обобщително местоимение:1] (8 синонимни множества)

[хипоним: относително местоимение:1] (15 синонимни множества)

[хипоним: показателно местоимение:1] (6 синонимни множества)

[хипоним: анафорично местоимение:1] (по аналогия със съществуващо синонимно множество в Принстънския wordnet, с цел да бъде запазена симетрията – 0 синонимни множества)

[хипоним: неопределително местоимение:1] (14 синонимни множества)

[хипоним: отрицателно местоимение:1] (8 синонимни множества)

[хипоним: лично местоимение:1] (1 синонимно множество)

[хипоним: възвратно лично местоимение:1] (1 синонимно множество)

[хипоним: реципрочно местоимение:1] (1 синонимно множество)

Систематизацията на синонимните множества, съдържащи местоимения, е затруднена от факта, че в Принстънския wordnet са добавяни местоименни думи като други части на речта – съществителни, прилагателни или наречия (вж. 3.2.2.2.). При оформянето на Българския wordnet първата стъпка, преди да бъде добавено синонимно множество с част на речта „местоимение”, е да бъде направена проверка, дали съответното множество не съществува в Принстънския wordnet. Ако бъде открит такъв паралел, в БулНет се създава еквивалентно множество, което има идентична позиция в мрежата с тази на английския еквивалент, като допълнително се свързва чрез релацията *тематична област* с множеството, съдържащо вида местоимение. При липса на съответствие в английски, множеството се добавя само в БулНет чрез релацията *тематична област*. Проблематични са множествата с местоименни наречия, които обикновено са включени в Принстънския wordnet, но съдържат няколко литерала, някои от които са местоименни наречия, а други са наречия с неместоименен произход, например {*винаги:1, всякога:1*}. В тези случаи не е направена връзка с множеството, съдържащо съответния вид местоимение, но в полето *бележка към синонимно множество* или *бележка към*

литерал се посочва неговата част на речта – местоимение. При въвеждането на местоимения в БулНет, които нямат съответствие в Принстънския wordnet, се добавя *бележка към синонимно множество* със съдържание преводния еквивалент на английски език.

9.2. Анотация на местоименията в БулСемКор

Анотацията на местоименията в корпуса следва определени критерии. В процеса на анотация се вземат предвид семантиката на местоимението в контекста на изречението или в контекста на целия откъс от текста, синтактичната му функция (атрибутивна, предикативна и т.н.) и евентуалният превод на английски език, независимо дали съществува еквивалентно множество в Принстънския wordnet.

Примерите (1), (2), (3) и (4) представят варианти на обобщителното местоимение *всеки*. При анализа на формите се определя синтактичната им функция – в (1), (2) и (3) *всеки* е в атрибутивна позиция (местоимение прилагателно), а в (4) – със субстантивна функция (местоимение съществително).

- (1) *Всеки* начинаещ студент изучава основните принципи на социалната икономика.

{*всеки*:2}

Дефиниция: 'който се отнася до, свързан е с, характеризира без изключение членовете на дадено множество'

Еквивалент в английски: {*every*:1}

- (2) Напрегна до краен предел *всеки* мускул по едрото си тяло.

{*всеки*:1, *всеки един*:1}

Дефиниция: 'който се отнася до, свързан е с, характеризира членовете на дадено множество, разглеждани поотделно'

Еквивалент в английски: {*each*:1}

- (3) Здравните картони на пациентите може да бъдат централно съхранявани, за да бъдат теглени от Мрежата от *всяка* точка на света.

{*който и да е*:2, *какъвто и да е*:2, *кой да е*:1, *какъв да е*:1, *всеки*:3, *всякакъв*:1}

Дефиниция: 'който означава или се отнася до случаен, произволен член от дадено множество'

Еквивалент в английски: {*any*:2, *whatever*:1, *whatsoever*:1}

- (4) ...*всеки*, вместо да говори с другия, включва компютъра и се понася някъде в Мрежата.

{*всеки*:4}

Дефиниция: 'обобщително местоимение съществително (заместващо именна фраза), което се означават лица'

без паралел в Принстънския wordnet

Бележка към синонимно множество: преводен еквивалент *everyone*

За коректното аотиране на (1), (2) и (3) се разглежда типът множество, към който се реферира, и начинът, по който местоимението се отнася към членовете на множеството. В пример (1) става въпрос за хомогенно множество, при което членовете се третират без изключение равнопоставено по отношение на признака, който им се приписва; в пример (2) членовете на множеството се разглеждат поотделно по отношение на признака; в пример (3) става въпрос за произволен член от множеството. Както се вижда от примерите, разгледаните значения на *всеки* имат отделна лексикализация в английски език и коректният превод от български на английски играе важна роля за правилното аотиране на българското местоимение.

9.3. Местоимения интензификатори

В българския език като интензификатори се употребяват показателните, въпросителните и относителните местоимения, както и неопределителното *един*. При интензификация местоименията не се употребяват със своето основно значение – с въпросителните не се пита за признак, с показателните не се посочва признак, а с относителните не се споменава повторно вече споменат признак, т. е. налице е полисемия с голямо семантична разлика между основното значение и значението интензификация (Ничолова 1978:182-183). Това дава основание интензифициращата функция на местоименията да бъде отразена в БулНет и да се взема предвид при аотирането им, като се прави разграничение между стилистично неутралната и стилистично маркираната употреба на една и съща местоименна форма.

В пример (5) въпросителното местоимение *какъв* е аотирано със стилистично неутралното си значение, докато в (6) е избрано маркираното значение на *какъв* като интензификатор.

- (5) *Какъв е всъщност романът „Балкански грешник“?*

{какъв:1}

Дефиниция: *‘въпросително местоимение прилагателно (поясняващо именна фраза или самостоятелно употребено), използвано в преки и косвени въпроси, с което се пита за качества, признаци, свойства на лица, предмети, ситуации, явления и т.н.’*

- (6) *Я проследи в **какъв** ужас изминаха твоите цветущи години.*

{какъв:2}

Дефиниция: *‘въпросително местоимение прилагателно, изразяващо висока степен на някакво качество’*

9.4. Неутрален среден род на местоименията

Местоименията в неутрален среден род формално съвпадат с местоименията в среден род, но всъщност са неутрални по отношение на категорията род – не се съгласуват по род и число с референта си (Пашов 1965:100). Формите в неутрален среден род са добавени като отделни синонимни множества в БулНет. В процеса на аотиране следва да се направи разлика между местоимение, съгласувано с референта си в среден род, и местоимение в неутрален среден род. При аотирането на местоимение, употребено в неутрален среден род, се избира основна форма в среден род (8), а не в мъжки род (7), както е в общия случай.

- (7) *Погледът на тия очи очарова с чудна благост, ей, **това** лице, **това** хубаво носленце...*

{онзи:3, оня:4, този:4, тоя:4}

Дефиниция: *‘показателно местоимение прилагателно (поясняващо именна фраза) за изтъкване на нещо характерно, добре познато или разпознаваемо от главното изречение, което се пояснява от подчинено определително изречение’*

- (8) ***Това** не са котки за хора, които очакват от тях да седят безучастно на канапето...*

{това:1}

Дефиниция: *‘неутрална форма на показателното местоимение, която не се съгласува по род и число с референта си, употребява се и с просентенциална функция – обобщава посоченото в изказването в цялост’*

В пример (7) *това* е местоимение прилагателно, съгласувано по род и число със съществителното в именната група, избрана е основна форма *този* и местоимението е свързано със съответното синонимно множество, което отразява семантиката му. В пример (8) е употребен неутралният род на местоимението и основната форма остава *това*.

9.5. Субстантивирани местоимения

При субстантивация на местоименията се следват общите правила – субстантивираното местоимение се аотира като съществително, като от контекста в изречението или в целия отрязък от текст се прави разграничение между собствено местоименна и субстантивирана употреба на дадено местоимение. По-особен случай е неопределителното местоимение *нещо*, което в субстантивираната си форма е развило множествозначения. БулНет съдържа едно синонимно множество слитерал *нещо* като местоимение ({*нещо:8*} със съответствие в Принстънския wordnet {*something:1*}), едно синонимно множество с литерал *нещо* като наречие ({*донякъде:1, до известна степен:1, нещо:13, малко нещо:1*} със съответствие в Принстънския wordnet {*rather:1, kind of:1, kinda 1, sort of:1*}) и десет синонимни множества с литерал *нещо* като съществително, от които девет съответстват на *thing* и едно – на *matter* в Принстънския wordnet. Първата стъпка при аотирането на *нещо* е да се определи частта на речта, в която е употребено – местоимение, съществително или наречие. За местоименната и адвербиалната употреба има съответно по един кандидат и при правилното им разпознаване в корпуса, аотирането не създава проблеми. Субстантивираната форма може да се членува и ако има модификатори (прилагателно, неопределително местоимение, числително), те са предпоставени, докато местоименната форма не може да бъде членувана, а при наличие на прилагателно то е в постпозиция (Ницолова 1986:177; Пашов 1965:105). За субстантивната употреба се избира най-подходящият кандидат измежду 10-те синонимни множества в БулНет, като за коректния избор основна роля имат тълковната дефиниция и хиперонимът на всяко множество, както и примерите за употреба.

В пример (9) *нещо* е употребено като местоимение, за което говори постпозицията на прилагателното. В примерите (10)-(13) *нещо* е съществително – може да се членува и прилагателното е предпоставено. В (10) и (11) семантиката на думата е определена като реално съществуващ обект, а в (13) – като признак. В (12) изречението е недостатъчно за правилното определяне на значението, но от по-широкия контекст се изяснява, че става въпрос за ситуация.

- (9) В категоричното членуване на думата „хартия” има **нещо** показателно.

{нещо:8}

Дефиниция: ‘материална или абстрактна същност от определен тип, но неконкретизирана индивидуално’

Тематична област: {неопределително местоимение:1}

- (10) ...там, където **нещо** не върви, са нужни промени, независимо дали **нещото** се казва Конституция...

{нещо:1, вещь:5, предмет:5}

Дефиниция: ‘отделна, самостоятелна реална същност’

Хипероним: [хипероним: същност:3, реална същност:1, физическа същност:1]

- (11) ...но когато след два часа напразни усилия да заспя отново се опитам да включа и проклетото **нещо** въобще не разпозна диска, разбрах, че проблемът е много по-сериозен.

{нещо:4}

Дефиниция: ‘реално съществуваща или абстрактна същност, която не е или не може да бъде точно определена или назована’

Хипероним: [хипероним: същност:3, реална същност:1, физическа същност:1]

Примерна употреба: Не мога да кажа какво бе това **нещо**.

- (12) И понеже не знаят, те не рядко изпадат във вечен дълг. Мисълта за подобно **нещо** би съсипала всеки здравомислец българин и повечето европейци. В Америка обаче състоянието на дълг се възприема за **нещо** нормално.

{нещо:6}

Дефиниция: ‘ситуация, положение, което по някакъв начин се отличава от останалите’

Хипероним: [хипероним: ситуация:3, положение:10, състояние на нещата:1]

- (13) Първото **нещо**, което привлича вниманието при Корниш Рекс, е нейната космена покривка.

{нещо:9}

Дефиниция: ‘качество или признак, разглеждани самостоятелно от притежателя’

Хипероним: [хипероним: признак:3, атрибут:1]

Примерна употреба: *Нещото, което харесвам у нея, е...*

9.6. Местоимения в устойчиви изрази

Когато участват в устойчиви изрази, местоименията се аотираат като част от израза, както е възприето според общите правила. В корпуса са аотирани като изрази *у нас, все още, имам нещо против, нещо повече, нещо като* и др.

9.7. Заключение

Трудността при аотирането на местоименията произтича от факта, че в Принстънския wordnet са включени някои местоимения като различни части на речта – съществителни, прилагателни или наречия. Затова първата стъпка при аотиране е да се прецени дали местоимението, срещнато в корпуса, отговаря на съществуващо синонимно множество в wordnet или трябва да бъде създадено ново множество, уникално за БулНет. Други решения, които трябва да вземе аотиращият са свързани с разграничаването на стилистично неутрална и стилистично маркирана употреба на местоименията, правилната лематизация при неутрален среден род и коректното аотиране при субстантивизиране. Важно е и правилното определяне на границите на думата, когато местоимението е част от устойчив израз.

Цитирана литература

Ницолова 1978: Ницолова, Р. *Интензифициращите местоимения в славянските езици*. – Славистични изследвания, № 4, с. 181-189.

Ницолова 1986: Ницолова, Р. *Българските местоимения*. София.

Пашов 1965: Пашов, П. *Наблюдения върху граматическия род в български и испански език (неутрален среден род на българските местоимения в сравнение с испанските neutro)*. – Годишник на Софийския университет, Филологически факултет, т. LIX, № 1, с. 89-130.

10. Лингвистични конвенции при анотация на затворените класове

Цветана Димитрова

В тази част се разглеждат принципите на анотация на предлозите, съюзите, частиците и междуметията в БулСемКор. Принстънският wordnet не включва затворените класове думи, за разлика от БулНет, където за затворените класове са създадени нови синонимни множества, принадлежащи към съответните части на речта. Преводният еквивалент на дадено значение в английски се предава в бележка към синонимното множество. Например за синонимното множество {и:8} с дефиниция 'съчинителен съюз за въвеждане на поредното или последното от две или повече последователни събития, действия и т.н.', е посочена следната бележка: преводен еквивалент: *and*.

В настоящето изложение, освен БулСемКор и лексикално-семантичната мрежа БулНет (Коева 2008), с помощта на която е анотиран корпусът, за сравнение е използван предимно Българският тълковен речник (БТР: Андрейчин 2002)¹, а при анализа на английските предлози са взети предвид значенията, дефинирани в рамките на The Preposition project (PP)².

Освен тълковната дефиниция основният фактор за определяне и

¹ Не е изведена съпоставка с по-пълния Речник на българския език (РБЕ), тъй като за тази цел следва да се направи критичен преглед на класификацията на значенията на интересуващите ни лексикални единици в РБЕ спрямо тези в БулНет, а целта на настоящето изложение е да се представи анотацията на БулСемКор, не толкова принципите на изграждане на лексикално-семантичната мрежа БулНет.

² В The Preposition Project значението на всеки предлог се снабдява с характеристики, които да послужат за целите на обработката на естествения език. В базата данни са включени 334 предлога, към тях са прикрепени общо 673 значения, като е отбелязана семантична роля или име на релация, както и синтактичните и семантичните характеристики на комплемента. Всяко значение е допълнено от: тълковната дефиниция и примерните употреби по Oxford Dictionary of English; позицията в семантичната йерархия на предлозите по A Comprehensive Grammar of the English Language; плюс други предлози, които са синоними на разглеждания или изразяват подобна семантична роля. Семантичните фреймове на ФреймНет са използвани за описание на комплемента, както и на значението на предлозите в други синтактични позиции.

анотиране на значението на затворените класове е контекстът в изречението или в по-голямата текстова единица, като например позицията в началото на изречението за съчинителните съюзи. Значението на затворените класове думи може да бъде уточнено и чрез проследяване на езиковите релации на синонимните множества *подобен на* и *антоним*, например предлогът {зад:4}, ‘предлог означаващ местоположение откъм или движение покрай гърба или задната част на някого или нещо’ фигурира като антоним на предлога {пред:3} с дефиниция ‘предлог, означаващ местоположение откъм или движение покрай лицевата страна или предната част на някого или нещо’, но не и на останалите значения на *пред*. Показателни са още някои екстралингвистични релации, като *сфера на употреба* (в едно от синонимните множества на наречieto *пък* фигурира *сфера на употреба: адюнкт*), *тематична област* (в която се отбелязва принадлежността на думата към съответната част на речта, например частица, предлог, или към нейна подкатегория – подчинителен съюз, съчинителен съюз и други), както и примерните употреби.

10.1. Предлози

Анотираните в БулСемКор предлози се състоят от една (едносъставни предлози) или повече от една думи (съставни предлози). В БулНет повечето предлози са свързани с релацията *тематична област*: {предлог:1}. Съставните предлози са съчетания от съществително, прилагателно, причастие или наречие с прост предлог или съюз. Анотират се като съставна единица. Предлози, които са част от по-големи словосъчетания (например адвербиални като *по никакъв начин*, *до голяма степен*), се анотират заедно с останалите части на словосъчетанието.

10.1.1. Едносъставни предлози

В тази категория са простите предлози, съставени от една морфема, като *на*, *в/във*, *с/със*, *за*, *по*, *при*, *над*, *зад*, *след*, *върху*, и сложните предлози, образувани от свързването на предлози, като *пomeжду*, *пoкрай*, *oтпpeд*, *зapaди*, прости предлози с имена, като *вместо*, *около*, и прости предлози с наречия, като *извън*, *въпреки*, *отвъд* (ГСБКЕ 1983а: 409-411).

В(1) са посочени примери за срещанията на предлога *по* в БулСемКор (в скоби са включени литералите от синонимното множество).

- (1) а. *превози по {но:24} вода*
- б. *... но се различават по {но:26} своята вътрешна молекулна структура.*

- в. Факултета **по** {но:14} педагогика
- г. Всяко дърво получава **по** {но:12} малко вода от специално направена водна инсталация.
- д. Трябва да търся наред. Отдолу нагоре, стая **по** {но:6} стая.

Примерите илюстрират употреби само на част от синонимните множества на *по*. Синонимните множества за *по* са общо 37. От тях 35 са предлози, а 2 са наречия: {но-малко:3, по:33} и {повече:5, по:32}.

В (1а) {но:24} е с дефиниция ‘предлог за въвеждане на среда, в която се извършва движение’; Бележка към синонимното множество: преводен еквивалент: *by*; Примерна употреба: *Войските ще бъдат транспортирани по вода.*

(1б) представя пример за употреба на {но:26} с тълковна дефиниция ‘предлог, с който се въвежда критерий, по отношение на който се определя, измерва, сравнява някой или нещо’; [подобен на: на:2]; Бележка към синонимното множество: преводен еквивалент: *in*; Примерна употреба: *Двата предмета са еднакви по големина.*

{но:14} в (1в) е ‘предлог, означаващ предназначение или качествена характеристика’; Бележка към синонимното множество: преводен еквивалент: *of*; *by means of a noun modifier*; Примерна употреба: *учител по биология.*

(1г) представя употреба на {но:12} ‘предлог, означаващ разпределение на определено количество или брой от нещо поравно между членовете на дадено множество’; Бележка към синонимното множество: преводен еквивалент: *each*; Примерна употреба: *Раздаде им по пет чаши.*

Примерът в (1д) илюстрира значението {но:6} с дефиниция ‘предлог, участващ като свързваща дума в устойчива конструкция с повтарящо се име, със значение последователност, поредност’; Примерна употреба: *Учениците влизаха един по един.*

Изреченията в (2) представят различните значения на предлога *пред*. След примерите са изредени синонимните множества в БулНет, които им съответстват, заедно с тълковните им дефиниции и някои релации, ако има такива.

- (2) а. семейните връзки остават на заден план **пред** {пред:1} обществените
- б. (...) каза тя, като коленичи **пред** {пред:3} електрическата печка (...)

- в. Близко до финализиране е процедурата **пред** {пред:6} Акредитационния съвет на Националната агенция за оценка и акредитация.
- г. Искам да знам **пред** {пред:9} кого точно министър-председателят е изразил изненада.

{пред:1, по отношение на:1, в съпоставка с:1, в сравнение с:1, в сравнение със:1}

Дефиниция: 'предлог за въвеждане на съпоставка, сравнение с нещо или някой с по-голямо значение, важност'

Бележка към синонимното множество: преводен еквивалент: *compared to, compared with, in comparison with*

Примерна употреба: *Пред останалите той е едно нищо.*

{пред:2}

Дефиниция: 'предлог за въвеждане на обект, по отношение на който се преценява перспектива, възможност за нещо в бъдеще или необходимост от избор или някакво действие'

Бележка към синонимното множество: преводен еквивалент: *before*

Примерна употреба: *Пред нас са открити два пътя.*

{пред:3}

Дефиниция: 'предлог, означаващ местоположение откъм или движение покрай лицевата страна или предната част на някого или нещо'

[антоним: зад:4]

Бележка към синонимното множество: преводен еквивалент: *before, in front of*

Примерна употреба: *Точно пред посолството стои един човек.*

{пред:4, преди:2}

Дефиниция: 'предлог, означаващ непосредствена пространствена близост или предхождане'

Бележка към синонимното множество: преводен еквивалент: *before*

Примерна употреба: *Перник е точно пред София.*

{пред:5}

Дефиниция: 'предлог, въвеждащ причина, обстоятелства и т.н., които са налице или с които някой се сблъсква'

Бележка към синонимното множество: преводен еквивалент: *at*

Примерна употреба: *Тя изтръпна пред мисълта за изпита.*

Примерна употреба: *Няма да се спре пред нищо.*

{*пред:6*}

Дефиниция: '*предлог за въвеждане на обект, към който е насочено отношение, действие, или на страна в отношения и т.н.*'

Примерна употреба: *сляпо преклонение пред авторитетите*

Примерна употреба: *Оправдават се пред нас за престъпленията, които извършиха.*

{*преди:3, пред:7*}

Дефиниция: '*предлог, означаващ непосредствено предхождане във времето*'

Бележка към синонимното множество: преводен еквивалент: *before*

Примерна употреба: *Пристигнаха десет минути преди началото на мача.*

Примерна употреба: *затишие пред буря*

{*пред:8*}

Дефиниция: '*предлог за въвеждане на вероятен изход от ситуация*'

Бележка към синонимното множество: преводен еквивалент: *on the verge of*

Примерна употреба: *В момента са пред развод.*

{*пред:9*}

Дефиниция: '*предлог, означаващ действието като извършващо се в присъствието на някого или нещо*'

Бележка към синонимното множество: преводен еквивалент: *before, in front of*

Примерна употреба: *Разказа всичко пред следователя.*

Примерна употреба: *Все още се смущаваше да говори пред много хора.*

В примерите само в (2б) *пред* е антоним на *зад*, както показва антонимната релация, въведена към синонимното множество {*пред:3*}. Друго по-сложно разграничение се прави между {*пред:6*} в (2в) и {*пред:9*} в (2г). При ситуацията, описана в (2г), действието (споделяне на мнение) се случва в присъствието и на субекта, и на обекта (който трябва да си състави мнение и който е въведен с {*пред:9*}). Описаната в

(2в) процедура е предназначена специално за посрещане изискванията на указаната институция/правоимащо юридическо (или физическо) лице, без да е изрично необходимо физическото присъствие на свързаните (с {пред:6}) засегнати обекти.

По-долу е дадена Таблица 1 на значенията на *пред* в БТР в сравнение с тези в БулНет и с английските еквиваленти според базата данни на The Preposition Project (PP).

БТР пред, <i>предл.</i>	БулНет пред, <i>p</i>	The Preposition Project in front of/before
1. Откъм основната, лицевата или началната страна, част на нещо. Прот. з а д; <i>Пред къщата. Пред мене. Това стана пред очите ми. Запетаята се поставя пред тази дума.</i>	{пред:3} 'предлог, означаващ местоположение откъм или движение покрай лицевата страна или предната част на някого или нещо' [антоним: зад]	<i>Before: 2a. in front of: Matilda stood before her, panting. / Spatial</i> <i>In front of: 1a. in a position just ahead of or at the front part of: the lawn in front of the house / Spatial</i> <i>In front of: 1b. In a position facing someone or something: she sat in front of the mirror / Spatial</i>
2. В непосредствена близост по време – по-рано от настъпването, започването на нещо; преди. <i>Пред Нова година. Затишие пред буря. Те са пред развод. Страната е пред катастрофа.</i>	{преди:3, пред:7} 'предлог, означаващ непосредствено предхождане във времето' {пред:8} 'предлог за въвеждане на вероятен изход от ситуация' {пред:2} 'предлог за въвеждане на обект, по отношение на който се преценява перспектива, възможност за нещо в бъдеще или необходимост от избор или някакво действие'	<i>Before: 1. During a period of time preceding (a particular event or time): the day before yesterday. / Temporal</i>

<p>3. В присъствието на някого или нещо (като причина, предпоставка или пречка за извършване или неизвършване на нещо, по отношение на който някой или нещо се проявява по определен начин). <i>Пред мене не смее да лъже. Той не се стира пред нищо. Равни сме пред закона. Срам ме е пред детето. Чист съм пред съвестта си.</i></p>	<p>{пред:5} 'предлог, въвеждащ причина, обстоятелства и т.н., които се актуализират или с които някой се сблъсква' {пред:9} 'предлог, означаващ действието като извършващо се в присъствието на някого или нещо'</p>	<p><i>in front of: 2. in the presence of: the teacher didn't want his authority challenged in front of the class. / Backdrop</i></p>
<p>4. По отношение на, в съпоставка, в сравнение със. <i>Какво е жабата пред вола? Предпочитам го пред другите.</i></p>	<p>{пред:1, по отношение на:1, в съпоставка с:1, в сравнение със:1} 'предлог за въвеждане на съпоставка, сравнение с нещо или някой с по-голямо значение, важност</p>	<p><i>Before 3. in preference to; rather than: a woman who placed duty before all else. / Tandem</i></p>
<p>5. По отношение на, към който е насочено дадено действие. <i>Обещавам пред народа. Давам отчет пред избирателите си. Отсрахме се пред госта.</i></p>	<p>{пред:6} 'предлог за въвеждане на обект, към който е насочено отношение, действие, или страна в отношения и т.н.'</p>	<p><i>Before: 2b. in front of and required to answer to (a court of law, tribunal or other authority): he could be taken before a magistrate for punishment. / Party</i></p>

Таблица 1. Сравнение на значенията на пред в БулНет, БТР и РР

В първия ред на таблицата е представен предлогът за посочване на място, местоположение на един обект спрямо друг обект (*Spatial*), като единствено в РР се прави разграничение на позициите „местоположение пред основната част на нещо” (*in front of: 1a*) и „местоположение откъм лицевата страна на някого или нещо” (*before: 2a*).

Във втория ред на таблицата са предлозите, означаващи най-общо разположение във времето (*Temporal*), където са създадени нови синонимни множества, принадлежащи към съответните части на речта. В РР се указва, че ситуацията може да се отнася както за период от време, така и за събитие (като нещо, което се случва във времето). При тези предлози в БулНет се прави по-детайлно разграничение на абстрактното разположение във времето. {*пред:7*} е предлог за непосредствено предхождане във времето, а {*пред:8*} – за предварително очакване на изход от ситуация/събитие. Значението на {*пред:2*} е по-специфично и насочва към ситуация в бъдеще време, за която е възможно да има алтернативни варианти за изход/разрешение.

В третия ред е предлогът за въвеждане на условия, при които нещо или някой присъства и с които се очаква среща лице в лице (*Backdrop* в РР). В БулНет се прави разграничение между предлога, изрично въвеждащ причина {*пред:5*}, и предлога, въвеждащ присъствие на някого или нещо, което не е задължително да е причина, предпоставка и прочие {*пред:9*}. Четвъртият ред включва предлога за съпоставка {*пред:1*}, свързващ два обекта/ситуации, които подлежат или са подлежащи на сравнение и единият се оказва предпочитан пред/преди другия (*Tandem* в РР).

Предлогът в петия ред въвежда цензор, инстанция (физическо или юридическо лице със съответните правомощия и авторитет), пред която нещо се случва и която, евентуално, би следвало да даде оценка на поведението/действието (*Party* в РР).

Примерът в (3) илюстрира употребата в БулСемКор на сложния предлог {*относно:1*}:

- (3) **Относно** {*относно:1*} болестите госпожа Еди съветва категорично да се изхвърлят всички симптоми и съмнения, че човекът може наистина да е болен.

{*относно:1*, по адрес на:1}

Дефиниция: ‘предлог, въвеждащ отношения на тематична свързаност’
[подобен на: за:4]

Бележка към синонимното множество: преводен еквивалент: *concerning, relating to*

Примерна употреба: *Има много научни спорове относно ползата и вредата от този процес.*

10.1.2. Съставни предлози

Съставните предлози се анотират както всички съставни думи. Предлогът *с/със* например участва в съставните предлози заедно със съществителни имена, например *в съгласие с/със, с оглед на*. Има случаи, когато предлогът е изпуснат, например в (4б):

- (4) а. *Подготвя се система за перманентно оценяване на качеството на дейността **съобразно с** {**съобразно с:2**} европейските практики и процедури.*
б. *Вие ще можете (...) да изберете вида помощ и грижи, които да оказвате, **съобразно** {**съобразно:1**} вашите възможности.*

{съответно с:3, съответно със:3, съответно на:3, в съгласие със:2, в съгласие с:2, в съответствие със:2, в съответствие с:2, съобразно със:2, съобразно с:2, по силата на:2, съобразно:1, съгласно със:3, съгласно с:3, съгласно:4, по:16, според:5}

Дефиниция: *'предлог за въвеждане на фактор, условие и т.н., които обуславят, определят или на основата на които нещо се извършва, предприема'*

Бележка към синонимното множество: преводен еквивалент: *according to*

10.1.3. Предлози в състава на устойчиви словосъчетания

В редица случаи предлозите се анотират като част от глаголната форма, както в (5), където предлогът *на* се анотира като част от пълнозначния глагол *приличам на*.

- (5) *(...) по-меко и **прилича на** {**приличам на:1**} българското х.*

{наподобявам:1, наподобя:1, напомням:1, напомня:1, приличам на:1, подобен съм:1}

Дефиниция: *'сходен съм с нещо или някого по външност, устройство, впечатление и пр.'*

Примерна употреба: *Тази статия наподобява моята собствена работа.*

Еквивалент в английски: *look like*

В тези случаи предлозите въвеждат аргументи, чиято поява и синтактични и семантични характеристики се обуславят много тясно от лексикалното значение на предиката. Предлозите, които се изискват от

един предикат, обикновено са ограничени само до един (Ницолова 2005: 143), но при включването на тези фразеологизми в БулНет се взимат предвид както интерпретацията на думата/израза със и без предлог, така и еквивалентът в английски, както в (5). В част 5.3. например се разглеждат случаите, в които значението на глагола не може да се интерпретира без предлога, например *радвам се на истински бум*.

Предлозите се маркират и като част от адвербиални устойчиви изрази, например {*по никакъв начин:1*} (анотиран с част на речта наречие) в (6).

- (6) *Днес, за съжаление, станах свидетел на една сцена, която не прави чест по **никакъв начин** {по **никакъв начин:1**} на Тридесет и седмото Народно събрание.*

{*никак:3, съвсем не:3, ни най-малко:1, в ни най-малка степен:1, по никакъв начин:1*}

Дефиниция: '*означава, че посоченото в изказването се проявява в много малка или незначителна степен*'

Предлозите участват и в съставните съюзи, например {*след като:1*} в (7).

- (7) *След като {след като:1} разделя полимерите на фракции с различна молекулна маса, Щаудингер обосновава своето твърдение за (...)*

{*след като:1, щом като:2, щом:3, като:8*}

Дефиниция: '*подчинителен съюз, който въвежда обстоятелствено изречение за определяне на времето на първото от две поредни действия*'

[*тематична област: подчинителен съюз:1*]

Бележка към синонимното множество: преводен еквивалент: *after, (on) + gerund*

Примерна употреба: *Като помисли малко, той заговори.*

Най-често срещаният едносъставен предлог в БулСемКор е {*на:1*}, '*предлог, изразяващ принадлежност, притежание, собственост*'; Примерна употреба: *Шапката на момчето литна, с 1474 срещания.*

Вторият по честота в корпуса предлог с 870 срещания е {*на:11*}, '*предлог, въвеждащ обект на действие, изразено чрез предикативно съществително (например отглаголно съществително), произлизащо от съответния преходен глагол*'; Примерна употреба: *анализиране на ситуацията.*

Трети с 482 срещания е {в:б, във:б}, ‘предлог, означаващ движение на някого или нещо отвън към вътрешността на някакъв предмет или среда, в резултат от което става част или е обграден от него/нея’; Примерна употреба: *В стаята е тихо.*

10.2. Съюзи

В БулНет категоризацията на съюзите като съчинителни и подчинителни фигурира както в тълковната дефиниция, така и като стойност на релация *тематична област: {съчинителен съюз:1}* и *тематична област: {подчинителен съюз:1}*. И подчинителните, и съчинителните съюзи се анотират в зависимост от това, какво главно или подчинено изречение въвеждат.

10.2.1. Прости съюзи

Съчинителните съюзи като *и* и *или* може да въвеждат както главно изречение в състава на сложното съчинено изречение, така и отделно самостоятелно изречение, което пояснява предходното самостоятелно изречение, като в (8), където въведеното от *или* изречение е уточнение на предходното твърдение:

- (8) ... *Бог, бидейки причина, която внезапно произвежда своето действие, не го предхожда непременно във времето. Или {или:1} „всеки път, когато се мисли нещо съществуващо, може да бъде мислен и принципът на неговото създаване.“ /Тома Аквински/*

{или:1}

Дефиниция: ‘*съчинителен съюз, който въвежда пояснение или перифраза на предходното твърдение*’

Бележка към синонимното множество: преводен еквивалент: *or*

Примерна употреба: *Направи си сам, или какво не трябва да правите вкъщи.*

Като {или:2} се анотира второто *или* в (9), докато първото *или* е определено като {или:3}.

- (9) *Право на организиране с цел да се осигури или {или:3} съдейства за упражняването на свободата на работниците и на работодателите да образуват местни, национални или {или:2} международни организации за защита на техните икономически и социални интереси (...)*

{или:2} е ‘съчинителен съюз за присъединяване на еднородни части на изречението при изреждане’; Бележка към синонимното множество: преводен еквивалент: *and*; Примерна употреба: *Плъзнали са по къщи, дворове, улици или мегдани.*

{или:3} е ‘съчинителен съюз за изразяване на избор между алтернативни възможности’; Бележка към синонимното множество: преводен еквивалент: *or*; Примерна употреба: *Можеш да вземеш сака или раницата.*

В корпуса много рядко се среща {или:4}, ‘съчинителен съюз, свързващ две числа, за изразяване на приблизително количество или брой’; Примерна употреба: *Взemi три или четири, но не повече.* Причината за малкото регистрации вероятно е фактът, че същото значение може да се изрази с помощта на дефис: *три-четири.*

{или:5}, регистриран в примера в (10), е ‘съчинителен разделителен съюз, с който се въвеждат две или повече взаимно изключващи се възможности’; Бележка към синонимното множество: преводен еквивалент: *either ... or*; Примерна употреба: *Или аз имам халюцинации, или той има брат близък.*

(10) ... защото от играта излизаш **или** {или:5} спечелил, **или** {или:5} загубил, среден изход няма.

Таблица 2 по-долу съпоставя значенията на съюза *или* в БТР и тълковните дефиниции на синонимните множества в БулНет. БулНет включва {или:4} за изразяване на приблизителен брой с колебание при определянето на количеството или качеството на определени обекти.

БТР или, сз.	БулНет или, сопј
1. Замяна. <i>Аз или ти. Днес или утре.</i>	{или:3}, ‘съчинителен съюз за изразяване на избор между алтернативни възможности’
2. Пояснение. <i>Предикат, или сказуемо.</i>	{или:1}, ‘съчинителен съюз, който въвежда пояснение или перифраза на предходното твърдение’
3. Присъединяване. <i>По къщи, дворове, улици или мегдани [т.е. и мегдани].</i>	{или:2}, ‘съчинителен съюз за присъединяване на еднородни части на изречението при изреждане’

4. Изреждане на възможности, всяка от които изключва другите. Или аз, или ти, или той.	{или:5}, ‘съчинителен разделителен съюз, с който се въвеждат две или повече взаимно изключващи се възможности’
	{или:4}, ‘съчинителен съюз, свързващ две числа, за изразяване на приблизително количество или брой’

Таблица 2. Значенията на съюза или в БулНет и БТР

При аотиране на съотносителните съюзи, като или, ... или, нито, ... нито, хем, ... хем, двата члена на съюзната дума се маркират заедно. В (11) е илюстрирана употребата на {и и:1}, ‘съотносителен съюз, свързващ еднакви части при изреждане и изтъкване на действия, събития или предмети’.

(11) Спадът на доверието в правителството постави ДПС в още по-неизгодна позиция от БСП, тъй като за разлика от левицата движението е **и** {и и:1} де факто, **и** {и и:1} де юре коалиционен партньор.

Думата и фигурира в единайсет синонимни множества като съюз и в две като наречие, едно от които е илюстрирано в (12):

(12) В България (също **и** {също и:1} Русия **и** {и:6} Румъния „червеният квазиетнос” е относително доста голям.)

Синонимното множество {също:3, още:6, също така:1, също и:1, също така и:1, още и:1, и:4} е въведено с част на речта „наречие” и:

Дефиниция: ‘означава, че посоченото в изказването представлява добавка или допълнение към нещо друго (подобно)’

[тематична област: количество:1]

[тематична област: степен:3, ниво:3, равнище:6]

Примерна употреба: Той има също и мерцедес.

В същия пример съюзът {и:6} е ‘съюз, свързващ еднородни части в простото изречение или равнопоставени прости изречения в рамките на сложното изречение’; Примерна употреба: Всички котки и кучета побягнаха.

{и:6} е и най-често засвидетелстваният в корпуса сред едносъставните съюзи, с 2556 регистрации, следван от доста по-рядко

срещания (495) {да:9}, 'съюз, въвеждащ подчинено допълнително изречение'; Примерна употреба: *Искам да се уча.*

10.2.2. Съставни съюзи

Съставните съюзи се анотират както останалите съставни думи, като се маркират едновременно всички техни елементи. Литералите *макар и* (13а), *въпреки че* (13б) и *независимо че* (13в) принадлежат към следното синонимно множество:

{*макар и:1, въпреки че:1, макар че:1, макар да:1, макар да:1, при все че:1, независимо че:1, ако и да:1, нищо че:1, независимо от това, че:1*}

Дефиниция: 'подчинителен съюз, който въвежда подчинено обстоятелствено изречение за отстъпка и означава, че действието в главното изречение се осъществява въпреки действието или обстоятелствата в подчиненото или изразява подобни отношения в рамките на простото изречение'

Бележка към синонимното множество: POS=conj

Примерна употреба: *Въпреки че знаеше отговора, той си замълча.*

- (13) а. *Приложението им, **макар и** {макар и:1} във военновременни условия, показва стремежа на царя, правителството и ...*
б. *И **въпреки че** {въпреки че:1} личността не се цени много, нека разгледаме живота на ацтеките в един по-личен аспект ...*
в. ***Независимо че** {независимо че:1} някои от тях подават молби за българско поданство, повечето не смятат за уместно и необходимо ...*

Съставен е и съюзът {*тъй като:1*}, чиято употреба е илюстрирана в (14).

- (14) *Спадът на доверието в правителството постави ДПС в още по-неизгодна позиция от БСП, **тъй като** {тъй като:1} за разлика от левицата движението е и де факто, и де юре коалиционен партньор.*

{*тъй като:1, че:10, защото:1, понеже:1*}

Дефиниция: 'подчинителен съюз, въвеждащ подчинено обстоятелствено изречение за причина'

[тематична област: подчинителен съюз:1]

Примерна употреба: *Отивай си, че става късно.*

Съставен предлог е и {без да:1}, илюстриран в (15).

- (15) *В него се предвижда създаването на съвет за координация, който да получи мандат от отделните партии за вземане на решения по определени политически въпроси, без да {без да:1} подлежат впоследствие на препотвърждаване (...)*

{без да:1}

Дефиниция: ‘подчинителен обстоятелствен съюз за начин, който показва, че действието в главното изречение се извършва при отсъствие на действието в подчиненото’

[тематична област: подчинителен съюз:1]

Бележка към синонимното множество: преводен еквивалент *without + -ing clause*

Примерна употреба: *Заминал, без да се обади.*

Предлози като *без да*, *преди да* и др. позволяват двата елемента на съставния предлог да не се намират в съседство, т.е. между тях може да се появи друг конституент на изречението, като например в *Помагам на хората, без те да разбират*. Съставките на съюза *без* и *да* се анотират заедно като съюза {без да:1}, а *те* – отделно.

10.2.3. Местоимения и наречия, въвеждащи подчинени изречения

Относителните и въпросителните местоимения и местоименни наречия, които въвеждат подчинени изречения, не се анотират като съюзи, каквато функция изпълняват в рамките на сложното съставно изречение, а според функцията си в простото изречение, съответно като местоимения или наречия.

Например {като:4} в (16) е ‘относително местоименно наречие за въвеждане на подчинено обстоятелствено сравнително коментиращо изречение, изтъкващо съдържанието на главното’; Бележка към синонимното множество: преводен еквивалент: *as*; Примерна употреба: *Както вече знаете, от утре сме в отпуск.*

- (16) *Дали {дали:1} това са „децата животни”, или {или:3} „децата растения”, както {както:4} по-грубо или {или:3} по-прилично ги наричат специалистите?*

В пример (16) {дали:1} е част от множеството {дали:1, да:3, нима:1, мигар:1}, ‘модална или въпросителна частица, която изразява

учудване, недоволство или недоумение'; [тематична област: частица:2]; [тематична област: модалност:1]; Примерна употреба: Дали вярваш, че това е възможно.

Комбинацията от *дали*, ... или в (16) не се аностира като съставен съюз, защото двата израза, свързани с *или* {или:3}, не се противопоставят един на друг, а са две алтернативни наименования на една и съща реалия. Местоименни наречия свързват и изреченията в (17):

- (17) а. *Но доколкото {доколкото:1} системата има много защитни механизми, не допускащи силна и ярка личност да навлезе в нея и още повече да заеме силна позиция, издигането на Горбачов ...*
б. *Числото е хармония дотолкова, доколкото {дотолкова доколкото:1} тя е обединение на противоположности (...)*

{доколкото:1, дотолкова доколкото:1, толкова:4, до такава степен:1}

Дефиниция: 'означава, че посоченото в изказването е степен или граница за реализиране на признак или действие'

[тематична област: степен:3, ниво:3, равнище:6]

Примерна употреба: *Доколкото знам, вие сте роднини.*

В БТР наречията, въвеждащи подчинени изречения, се категоризират като наречие-съюз. Тъй като наречията фигурират в Принстънския wordnet, за да се запази структурата на семантичната мрежа, в тези случаи в БулНет категоризацията по част на речта се запазва, като се въвежда уточняване на функцията като съюз в Бележка към синонимното множество: POS=conj. При наречията може да се отбележи и принадлежност към тематични области, различни от подчинителен съюз, като например [тематична област: време:9].

10.2.4. Омонимия между предлози и съюзи

Един от най-честите примери за омонимия между предлог и съюз е *като*. В БулНет литералът *като* фигурира в единайсет синонимни множества, от които шест са за съюза *като*. Наречието *като* е въведено в:

{като:21, като че ли:2, като че:2, сякаш:2, като да:1}

Дефиниция: 'изразява предположение или констатация относно това, как изглежда или какво е привидното впечатление от посоченото в изказването'

Примерна употреба: *Като че ли един откряхнат и прибулен прозорец трябва да бъде широко отворен, за да ни донесе светлина и въздух.*

В (18) са илюстрирани част от срещанията на предлога *като* в БулСемКор.

(18) а. ... *стремеж към противоречие като {като:1} преднамерена позиция на необходима другост ...*

б. *бял като {като:2} мляко*

в. ... *започват да се възприемат като {като:10} нелично дело ...*

{като:1} – (18а)

Дефиниция: ‘*предлог, въвеждащ временно или характерно качество, функция, роля, статут и т.н.*’

[подобен на: *в качеството си на:1*]

Бележка към синонимното множество: *преводен еквивалент as*

Примерна употреба: *Той беше нетъртим като човек.*

{като:2} – (18б)

Дефиниция: ‘*предлог, свързващ двете части на сравнение между лица, предмети, състояния, качества и т.н.*’

Бележка към синонимното множество: *преводен еквивалент: (as) ... as*

Примерна употреба: *бял като мляко.*

{като:10, за:16} – (18в)

Дефиниция: ‘*предлог, въвеждащ функция, служба, роля на някого или нещо за конкретна цел*’

Бележка към синонимното множество: *преводен еквивалент: as*

Примерна употреба: *Това ще ни служи като назидание.*

В (19) са посочени примери от корпуса за употреби, които реферират към прилика, уподобяване между два обекта.

(19) а. *Те предизвикват такива „мъжки категории” като {като:7} собствеността и правилността*

б. ... *с издут корем, с глава като котел; и изглежда като {като:12} всички наоколо*

в. ... *да не говорим за автори като {като:4} Лео Шпитцер или Нортърн Фрай.*

{*като*:7, *такъв като*:1, *такъв какъвто*:1} – (19а)

Дефиниция: ‘*предлог, въвеждащ принадлежност, отнесеност към посочван тип, вид*’

Бележка към синонимното множество: преводен еквивалент: *such as*

Примерна употреба: *Такъв план като предложения от теб изисква огромни ресурси от време и пари.*

{*като*:12} – (19б)

Дефиниция: ‘*предлог, означаващ наличие на прилика, сходство между участник в описваната ситуация и обекта, въвеждан от предлога*’

Бележка към синонимното множество: преводен еквивалент: *like*

Примерна употреба: *Тя среса косата си като Грета Гарбо.*

{*като*:4} – (19в)

Дефиниция: ‘*предлог, въвеждащ пример, илюстрация на нещо*’

Бележка към синонимното множество: преводен еквивалент: *such as, like*

Примерна употреба: ‘... и други представители на семейство *Котки*, като *лъвове, тигри* и т.н.,’

Съюзът *като* може да въвежда различни подчинени изречения. Част от аотираните употреби като съюз са илюстрирани в (20):

- (20) а. *Той седна, като {като:3} се хвана за перилата.*
б. *След неуспешна обсада на Виена турците се оттеглили, като {като:19} оставили значително количество кафе в чували.*
в. ... *циганите - роми - са юридически и финансово подчинени само на своя барон, като {като:9} запазват териториалния си статут във Византия.*
г. ...*да изберете човека, за когото да се грижите, (след) като {като:8, след като:1} се срещнете с него.*

Синонимното множество {*като*:3} в (20а) е ‘*подчинителен съюз, въвеждащ подчинено обстоятелствено несравнително изречение за начин*’; Бележка към синонимното множество: преводен еквивалент: *by*; Примерна употреба: *Той седна, като се хвана за перилата.*

{*като*:19} в (20б) е ‘*присъединителен съюз за въвеждане на изречение, което уточнява, допълва предходното*’; Бележка към синонимното множество: преводен еквивалент: *gerund*; Примерна употреба: *Президентът подаде оставка от политическия живот, като една година няма да има право да оглавява стари и нови партии.*

{като:9} в (20в) е част от синонимното множество {като:9, докато:1}, ‘подчинителен съюз за въвеждане на подчинени обстоятелствени изречения за време, с които се изразява едновременност на действието в главното и подчиненото изречение’; Бележка към синонимното множество: преводен еквивалент: *while*; Примерна употреба: *Като говорехме за нея, си спомних една подробност.*

В (20г) {като:8, след като:1, щом като:2, щом:3} е ‘подчинителен съюз, който въвежда обстоятелствено изречение, с което се съобщава за първото от две последователни действия’; Бележка към синонимното множество: преводен еквивалент: *after, on+gerund*; Примерна употреба: *Като помисли малко, той заговори.*

10.2.5. Омонимия между съюз и частица

Сред най-често срещаните примери за подобна омонимия е думата *да* (за по-подробно представяне на дебата и противоречията в научната литература за функцията на *да*, вж. Викторова 2006). В (21) са представени някои употреби на съюза *да*.

- (21) а. *Желяя неговата памет да {да:9} бъде обща на ухото ни в бъдещата ни работа.*
б. *Съмнителното щастие да {да:13} бъде превърнат в образ не споходи своеобразния Роман Якобсон или харизматичния Пол де Ман.*
в. *(...) лесно ли е човек да {да:19} стане пророк, светец и прочее в родното си село.*

{да:9} – в (21а)

Дефиниция: ‘съюз, въвеждащ подчинено допълнително изречение’

Примерна употреба: *Искам да се уча.*

{да:13} – (21б)

Дефиниция: ‘подчинителен съюз, който въвежда подчинено определително изречение’

Примерна употреба: *Бях му благодарен за идеята да ми помогне.*

{да:19} – (21в)

Дефиниция: ‘подчинителен съюз, който въвежда подчинено подложно изречение’

Примерна употреба: *Не е необходимо да идваш.*

Примерите в (22) представят употребите на частиците *да*.

- (22) а. *Бог да {да:4} пази България!*
б. (...) *дори може да {да:5} се твърди, че ученикът е благоволият да не накаже своя учител (...)*

{да:4, нека:1, нека да:1} – (22а)

Дефиниция: ‘*формообразуваща частица, с която се образуват сложни глаголни форми за повелително наклонение*’

[тематична област: частица:2]

Бележка към синонимното множество: преводен еквивалент *let me/you/him/her/it/us/you/them*

Примерна употреба: *Да разгледаме това следствие!*

{да:5} – (22б)

Дефиниция: ‘*частица, свързваща фазов или модален с пълнозначен глагол в съставно глаголно сказуемо*’

[тематична област: частица:2]

Примерна употреба: *Още утре започвам да чета.*

Думата *да* може да е част от сложни съюзи, като фигуриращите в примерите в (23):

- (23) а. ... *такава роля могат да играят и действително случили се събития, стига да {стига да:1} се остава в границите на възможното и вероятното.*
б. *Сцелда {сцелда:1} се осигури или съдейства за упражняването на свободата на работниците и на работодателите (...), договарящите страни се задължават*
в. ... *положи всичкото си старание, за да {за да:1} премахнеш този главен порок.*

{стига да:1} – (23а)

Дефиниция: ‘*съюз, въвеждащ подчинено обстоятелствено изречение за условие*’

Примерна употреба: *Стига да искам ще го направя.*

{така че да:1, за да:1, да:1, с цел да:1} – (23б) и (23в)

Дефиниция: ‘*подчинителен съюз за въвеждане на подчинено обстоятелствено изречение за цел*’

въведено като част на речта „наречие” (б)

Бележка към синонимното множество: *POS=conj*

Примерна употреба: *Той изучаваше змиите, за да опише поведението им.*

В БТР да е категоризирана само като частица, но не и като съюз.

10.3. Частици

Частиците са неизменяеми думи с разнороден произход, специфични семантични и граматични характеристики. В БулНет те са въведени като част на речта „частица” и са свързани с релацията *тематична област* със синонимното множество {*частица:2*}.

10.3.1. Словообразуващи частици

Към тази подгрупа принадлежат частици като *ще, не, ли, си, се* и други (ГБСКЕ 1983а: 477). Словообразуващата частица {*ще:2*} е въведена с дефиниция ‘*частица за изразяване на желание*’; Бележка към синонимното множество: преводен еквивалент *would*; Примерна употреба: *Какво ще желаете?*, като в (24).

(24) *Какво **ще** {ще:2} обичате, бензин или вода за цветята?*

В (25) {*ли:1*} е ‘*въпросителна частица в пряк или непряк въпрос*’; Примерна употреба: *Далеч ли е Ню Йорк?*

(25) *Имам **ли** {ли:1} тук някаква самоличност и **няма** {няма:1} **ли** {ли:1} да се изгубя в тъга и самота?*

В (26) са илюстрирани употреби на частицата *не*:

(26) а. *Главни действащи субекти във „Войната в Залива” (1991 г.) срещу Ирак бяха **не** {не:1} самите американски войници и иракски войници и офицери (...)*

б. *Макар следствието да се съдържа потенциално в причината, потенциално **не** {не:3} се актуализира веднага.*

{*не:1*} – (26а)

Дефиниция: ‘*частица, с която се отрича следхождащата дума или словосъчетание; произнася се с ударение*’

Примерна употреба: *Той е **не** добър, а отличен работник.*

{не:3} – (26б)

Дефиниция: ‘частица, обикновено предхождаща сказуемото, с която се отрича цялото изказване; произнася се без ударение’

Примерна употреба: *Той не говори английски.*

Бележка към синонимното множество: POS=*particle* (в Принстънския wordnet *not* е наречие, така че, за да се запази структурата на wordnet, в случая *не* също е въведено като наречие, но със съответното уточнение)

В (27) *се* е елемент от глаголната форма на {*препирам се:1, карам се:1, разправям се:2*}, ‘споря разпалено, шумно и непримируемо’, и {*изпарявам се:4, изчезна:9, изчезвам:9, изпаря се:4, чупя се:1, пръждосам се:1, пръждосвам се:1*}, ‘заминавам отнякъде незабавно, обикновено по принуда’.

(27) *Докато се разправяше {препирам се:1, карам се:1, разправям се:2} сам със себе си, колежката му се бе изпарила {изпарявам се:4, изчезна:9, изчезвам:9, изпаря се:4, чупя се:1, пръждосам се:1, пръждосвам се:1} неизвестно къде и как, без да остави никакви материални следи – дори мирисът на парфюма ѝ не се усещаше вече.*

10.3.2. Служебни частици

Към служебните частици спадат *ама, ето, пък* и други (ГБСКЕ 1983а: 477). Например *пък* в (28) е частицата {*пък:3*}, която ‘изтъква смисъла на определени думи и изрази’; Примерна употреба: *с бохчи или пък с китки цветя.*

(28) *А там, в кутийката, която по сателитно чудо или пък {пък:3} по милостта на кабелните оператори хваща музикалната телевизия ММ, вилнее другото (...)*

В БулНет фигурира и наречието {*пък:4*} в синонимното множество:

{*от друга страна:1, пък:4, но пък:1, а пък:1*}

Дефиниция: ‘характеризира посоченото в изказването като представящо гледна точка или аспект на въпрос, проблем и т.н., различни от вече представените или изтъкнатите’

[сфера на употреба: адюнкт:1]

[антоним: от една страна:1]

Примерна употреба: *Но пък може и да не дойде.*

10.3.3. Частици с формообразуваща функция

Към тази подгрупа се числят *няма, ще, да, нека, нека да* и други (*по-* и *най-* се маркират заедно с прилагателните или наречията, с които образуват сравнителни и превъзходни форми, или отделно като наречия, заедно със синонимите си *повече* и *по-малко*).

Частицата {*ще:1*} е въведена като ‘*частица за образуване на бъдеще време на глаголите*’ (29а), а {*няма:1*} като ‘*формообразуваща частица, с която се образуват отрицателни форми за бъдеще време*’ (29б).

- (29) а. *Аз също **ще** {ще:1} ги проследя, но за да защитя мислите, които ми хрумнаха.*
б. *Имам ли тук някаква самоличност и **няма** {няма:1} ли да се изгубя в тъга и самота?*

В корпуса са отбелязани и граматикализирани частици като *се* и *си*. Частицата {*се:6*} в (30а) е въведена с дефиницията ‘*граматикализирана частица за образуване на безлични възвратно-страдателни форми от лични глаголи*’; Примерна употреба: *Тук се говори бързо и ясно.*

{*си:4*} в (30б) е ‘*кратка форма на притежателното местоимение, участваща в конструкции с така наречения етичен датив, изразяващи субективно отношение на говорещия към ситуацията*’; Примерна употреба: *Хаввам си локум от време на време.*

- (30) а. (...) *някои бяха запечатани – от застрахователя „Ятаган”, ако се {се:6} съди по печата.*
б. *Това си {си:4} е чистата истина.*

Си може да бъде и местоимение, като в (31):

- (31) а. *Сега държа в ръцете си {си:1} друга картичка.*
б. *Докато се разправяше сам със себе си {себе си:1}, колежката му се бе изпарила неизвестно къде и как, без да остави никакви материални следи – дори мирисът на парфюма ѝ не се усещаше вече.*

{*си:1, свой:1*} – (31а)

Дефиниция: ‘*възвратно притежателно местоимение за означаване на принадлежност, обвързаност, отнасяне на някого или нещо със субекта на изречението*’

[*тематична област: възвратно притежателно местоимение:1*]

Примерна употреба: *Пие си кафето рано сутрин.*

{се:3, си:3, себе си:1} – в (31б)

Дефиниция: *‘лично местоимение, заместващо обект, съвпадащ със субекта’*

[тематична област: възвратно лично местоимение:1]

Примерна употреба: *Човек познава себе си най-добре.*

Други формообразуващи частици са {да:4, нека:1, нека да:1}, *‘формообразуваща частица, с която се образуват сложни глаголни форми за повелително наклонение’* в (32).

(32) а. **Нека** {нека:1} вземем за пример кръста – в аритметиката означава събиране (...)

б. **Бог да** {да:4} пази България!

Сред единиците, анотирани като частици в БулСемКор, най-често регистрираната, с 688 срещания, е {да:5}, *‘частица, свързваща фазов или модален с пълнозначен глагол в съставно глаголно сказуемо’*; Примерна употреба: *Още утре започвам да чета.*

Втора по честота е {се:5} с дефиниция *‘граматикализирана частица за образуване на възвратно-страдателни форми от лични глаголи’*; Примерна употреба: *Ремонтът се извършва бавно и некачествено*, с 511 срещания.

На трето място е {ще:1}, *‘частица за образуване на бъдеще време на глаголите’*; Примерна употреба: *Ще ходя на кино утре.*, с 484 срещания.

10.4. Междуметие

Междуметията в БулСемКор (само 9 думи са анотирани с тази част на речта) са свързани с релацията *тематична област: {непълнозначна дума:1, служебна дума:1}*. Например {ха:1} е *‘междуметие за наподобяване на смях’*, а преводният му еквивалент е отбелязан в Бележка към синонимното множество: преводен еквивалент: *ha*.

Междуметието {ало:1} е с тълковна дефиниция *‘междуметие, което служи за поздрав при обаждане по телефона или за привличане вниманието на някого’*.

10.5. Заключение

Анотацията на затворените класове се подчинява на общите анотационни принципи за корпуса. Тук трябва да се отбележи, че

БулСемКор е един от малко семантично анотирани корпуси, които включват анотация на тези класове думи. Интерпретацията им зависи от семантичните и синтактичните им характеристики и до известна степен е улеснена от факта, че допълването на затворените класове в БулНет се извършва предимно с оглед функциите и значенията им в българския език, а английските еквиваленти са добавени в превод на българските.

Цитирана литература

- БТР: Андрейчин, Л., Л. Георгиев, Ст. Илчев, Н. Николов, Ив. Леков, Ст. Стойков. Цв. Белчев. *Български тълковен речник*. Изд. „Наука и изкуство,,. София.
- Викторова 2006: Викторова, К. Да като компонент в сложни глаголни форми. - В: Ив. Куцаров, В. Райнов, М. Лакова, Вл. Мурдаров и Св. Коева (съст.) *Езиковедски изследвания в чест на 75-годишнината от рождението на ст.н.с. I ст. д.ф.н. Йордан Пенчев Пенчев*. Артграф, С., с. 172-180.
- ГСБКЕ 1983а: *Граматика на съвременния български книжовен език, т. II, Морфология*. Издателство на БАН, С.
- ГСБКЕ 1983б: *Граматика на съвременния български книжовен език, т. III, Синтаксис*. Издателство на БАН, С.
- Ницолова 2005: Ницолова, Р. За значението на предложните съчетания в българското изречение. - В: Св. Коева (съст.) *Аргументна структура. Проблеми на простото и сложното изречение*. СемаРШ, С., с. 141-152.
- Коева 2008: Koeva, S. Derivational and Morpho-Semantic Relations in Wordnet. - In: *Intelligent Information Systems*, 359-368.
- The Preposition Project*. 2002-2007. At: <http://www.clres.com/prepositions.html>
- Oxford Dictionary of English*. Revised Edition. C. Soanes and A. Stevenson (eds). Oxford University Press. 2005.
- РБЕ: *Речник на българския език*, 1977 – 2009.

11. Съставни единици в Българския семантично аотиран корпус

Мария Тодорова

11.1. Въведение

Една от основните цели при семантичното аотиране на корпуса е обхващането на всички типове лексикални единици, които се появяват в текста, включително и тези, които се състоят от две или повече графични думи, независимо от състава им и от спецификите на компонентите им с оглед на структурните и конституентните вариации, които позволяват. Един от основните проблеми, които възникват при аотирането на съставните единици е синхронизирането на данните в българската и английската лексикално-семантична мрежа. Идиомите, фразеологизмите, колокациите, както и липсата на лексикализация на определени синонимни множества са част от проблема за границата между композираното и некомпозираното значение на думите и преводните съответствия на думи във фрази и обратно. Избирането на определени комбинации от думи като съставни единици, а оттам и въвеждането им като отделни синонимни множества в лексикално-семантичната мрежа се определя главно от синтактичната им функция и употреба, както и от наличните лексикални данни в Принстънския wordnet и от структурата на БулНет. Докато при несъставните единици в корпуса основен проблем е многозначността и гранулираността на значението, то повечето съставни литерали имат не повече от две или три значения. Спецификите при аотирането на съставните думи произтичат от различните степени на композираност между техните компоненти, от особеностите на основната им форма и възможните вариации в структурата и формите при употребата в контекста. Основна трудност при тяхното аотиране е определянето им като съставни, намирането на преводното им съответствие и ако няма такова, определянето на мястото им в рамките на лексикално-семантичната мрежа като езиково специфични единици. Теоретично наличието на преводно съответствие на друг език е един от основните критерии за доказване на съставността на дадена единица (Мелчук 1995; Коева 2005).

11.2. Класификация на съставните единици с оглед на степента на тяхната композираност за целите на разрешаването на семантичната многозначност

Терминът *съставна единица*, обхваща голямо разнообразие от езикови конструкции, в които две или повече смислови цялости изразяват обща семантика (в общия случай едно понятие). От графична гледна точка съставните думи са последователност от две или повече графични единици, разделени от небуквени символи. Такива са както несвободните фрази (Мелчук 1995; Тодорова 2006; Стоянова 2009), които представляват „(не)прекъсната последователност от графични думи, разделени с интервали” (Коева 2005), така и думи или опори, съчетани морфемно и/ или с различни разделители. Степента на композираност е основният маркер, който определя решението при аотиране и добавяне в БулНет.

11.2.1. Сложни думи

Определянето на границата между съставни и сложни думи е в голяма степен относителна, като основната разлика между тях при аотирането е наличието или липсата на интервал в рамките на единицата. Морфемното комбиниране в рамките на думата, от рода на *ръкохватка* (*ръка* + *хватка*) в много изследвания се разглежда като аналогичен (на съставните думи) семантичен тип. Дори според някои автори като Чермак (2001) и Юлия Балтова (1971) това явление семантично напълно съответства на общото понятие за фразеологичност (идиоматичност). Такива примери съществуват в различните езици (вж. по-подробно в Чермак 2001) и могат да бъдат разглеждани като идиоматични комбинации на морфемите. В този смисъл терминът *лексикуализиран* е синонимен на *идиоматизиран* / *фразеологизиран* и означава, че на даден израз е приписано специфично значение.

Част от проблема за включването на композирани значения в структурата на БулНет, наред с останалите съставни единици, е въвеждането на времеви и числови изрази, съдържащи арабски или римски цифри, както и полуслетите комбинации от тях, изписвани с тире / дефис. В своето изследване Савъри (2008) подчертава ролята на разделителите като конституенти на съставните единици. В различните езици едни или други разделители могат да бъдат вътрешни членове на съставните единици, например някои препинателни знаци като тире и апостроф. Небуквените конституенти не са задължително разделители

на компонентите. Разделителите могат да служат за разрешаване на многозначността, но същевременно са и източник на ортографични варианти – те могат да участват и във формообразуването, например *SMS-и, 27-ми, най-голям* и т.н. По отношение на разделителите също са възприети определени принципи на аотиране и добавяне в БулНет. Спецификите произтичат от конвенцията за приписването на различни релации, ако разглежданата сложна дума не съществува в Пристънската лексикално-семантична мрежа.

- Сложни думи, съставени от два равнопоставени корена. В зависимост от частта на речта, към която принадлежат, композираността им се отразява при добавяне в лексикално-семантичната мрежа чрез релациите – *хиперонимия* или *производен*. В зависимост от значението и преводното си съответствие, съставната единица може да бъде хипоним на всеки от корените на сложната дума; с хипероним опорния корен или като синонимен литерал във вече съществуващ ред.

Например литералът *народонаселение:1* е въведен в синонимното множество {*гражданство:2, население:3, народ:3, народонаселение:1*}, а литералът *нервно-мускулен:1* – ‘*засягащ или характерен за мускулната и невралната тъкан*’ е свързан с релацията *производен* с {*мускул:1*} и {*нерв:1*};

- Сложни думи, съставени от опорен и модифициращ корен. Те се свързват с релацията *производен* с литералите, в които опорният корен съществува като самостоятелна, еднокоренна дума, с несвързано значение, или с релацията *хипоним на* с литерала, съдържащ самостоятелния опорен корен, при т.нар. еноцентрична хипонимия.

Например синонимното множество {*българо-руски:1*} е свързано с релация *производен* с множествата {*български:2*} и {*руски:1*}; {*вътрешнонароден:1*} – с релация *производен* с {*вътрешен:8; вътрешно групов:1*} и {*народен:2*}; {*социокултурен:1*} – с релация *производен* с {*култура:1*} и {*общество:1, социум:1*}; {*социално-политически:1, обществено-политически:1*} – с релация *производен* с {*политически:1*} и {*социален:2; обществен:3*}; {*неврохормон:1*} – с релация *хипоним* с {*хормон:1*}; {*фитотоксин:1*} е свързан с релация *хипоним* с {*токсин:1*}.

- Сложни думи, образувани с представките *полу- анти-*. В случай че нямат преводно съответствие, те се въвеждат в БулНет с релация *подобен на* с корена на думата.

Например множеството {полуприроден:1} е свързан с релация *подобен на* с {природен:1}.

Разбира се тези правила не са абсолютни и имат своите изключения, които налагат преди въвеждането на релациите при прилагателните (и наречията) да се провери дали не са производни от сложно съществително, както е при {невропсихиатричен:1}, свързан с релация *произведен* с {невропсихиатрия :1}.

11.2.2. Класове съставни единици

Ще бъдат разгледани основните класове съставни фрази, основавайки се отчасти на класификацията, представена в Тодорова (2006). Принадлежността на единиците към даден клас и степента на композираност, която той изразява, подпомага преценката при анотацията и може да улесни избора на подход при добавяне в лексикално-семантичната мрежа. Делението на подкласове се основава на типа референция, като се различават граматикализирани и лексикализирани съставни фрази. От тази гледна точка композираността се разглежда главно с лексикално-семантичната си функция. Тя се заключава в уникалното значение – което съчетанието от думи изразява в рамките на едно цяло – различно от значенията на компонентите на целостта, когато са употребени самостоятелно.

11.2.2.1. Граматикализирани словосъчетания

Граматикализираните словосъчетания – съставни предлози и съюзи – *за да, тъй като, благодарение на*, както и предложно именните словосъчетания – *с оглед на, в зависимост от* и др. не назовават понятие, а релация. При някои граматикализирани конструкции, като изрази, съдържащи модални и фазови глаголи, както и при аналитични глаголни форми и някои от глаголно-именните словосъчетания съставността се композира при самата анотация. Тук се има предвид, че чрез въвеждането на значения в БулНет за конституентите на определени типови комбинации, се избягват повторения и натрупване на съставни литерали в структурата на wordnet, като същевременно се разширява възможността за покритие при аотиране на съставни единици. Пример за това е изборът на значение на частицата *да* в примери (1). В (1а) изборът на модален глагол – *може*, съответното значение на частицата {*да*:5} – '*частица, свързваща фазов или модален с пълнозначен глагол в съставно глаголно сказуемо*' и значението на глагола *направя* композира сложната глаголна форма,

употребена в изречението. Изборът на значението {*да:2I*} с дефиниция ‘*формообразуваща частица, която участва като структурен елемент в сложните форми за бъдеще в миналото и бъдеще предварително в миналото и в някои отрицателни форми за бъдеще време*’ определя композирането на аналитичната глаголна форма на глагола *разкажа* в пример (1б).

- (1) а. *Чудесно, Чарли, това е най-доброто нещо, което може да се направи.*
б. *Но ако бяхме попитали Никое, той щеше да ни разкаже някоя нова история за тигъра, някое приключение.*

Някои от глаголно-именните конструкции, образувани с десемантизиран глагол, който проявява висока степен на съчетаемост с различни съществителни имена (например *правя, давам*), се композират в процеса на аотиране. Тези глаголи имат специфично значение в Българския wordnet, което се избира при съчетанието им с пълнозначна дума при т.нар. глаголно-именни перифрази. Това е илюстрирано в пример (2), в който за *направля* е избрано значението ‘*изпълнявам умствено или физическо действие, дейност, процедура*’.

- (2) *Бъчвите за добро вино са белязани с голям червен кръст. **Направи** избора си и мини на 394.*

11.2.2.2. Лексикализираны словосъчетания

Лексикализираните словосъчетания означават едно понятие. В текста те ще бъдат означавани и с термина *фразама*, заимстван от Мелчук (1995). Въз основа на фразеологичните маркери (виж Тодорова 2006) композираност и декомпозируемост, се различават три подкласа фраземи: неразложими (семантично цялостни), полуразложими и просто разложими, като посочените признаци се проявяват на различните езикови нива. Разграничаването на фраземите въз основа на признака композираност се представя в рамките на теорията „смисъл” <=> „текст”, в която езиковата единица се разглежда като съвкупност от следните елементи (i) означавано; (ii) означаващо; (iii) селективни особености на знака; (iv) ситуацията, в която се употребява; (v) концептуална репрезентация. Използва се операцията *лингвистично обединение* „O”, чрез която се съчетават всички типове лингвистични единици според стандартни езикови правила. Така, ако условно назовем компонентите на дадено словосъчетание А и В, то АОВ означава регулярна сума на

лингвистичните единици А и В. Лингвистичната операция *обединение* предполага и обхваща всички общи комбинаторни правила в езика, които не са в противоречие с обединяваните единици.

11.2.2.2.1. Съставни фраземи

Съставните фраземи представляват разложими (институционализирани по терминологията на Саг и др., (2002)) фрази, които са еднозначно определени в дадена сфера от обществения живот и следват определена конвенция. Те могат да бъдат декомпозирани на по-прости значения и позволяват висока синтактична вариативност. Въпреки че са семантично и синтактично композирани, тяхната комбинация е фиксирана към конкретен референт. Формално разгледана композицията ‘АОВ’ е регулярна сума от означаваните ‘А’ и ‘В’, но не може да бъде заместено от никое друго ‘У’, защото означавото от него ‘Х’ може да бъде назовано само от комбинацията от означителите А и В. По отношение на реализацията на фразеологичните маркери съставните фраземи отговарят на комбинацията от +буквалност, +композираност. Тъй като не позволяват вмъкване на компоненти и не се отличават с нехарактерно формообразуване, спецификите им може да произтичат главно от елиптична употреба в контекст. От своя страна биват:

11.2.2.2.1.1. Съставни термини

Образуват се чрез комбиниране на два или повече прости термина. Характерно за тях е, че са еднозначно използвани в определена сфера и контекст. Например: *бяла мечка*, *натриев хлорид* и др. Терминологичното им значение може да бъде комбинация от значенията на съставлящите ги прости термини. Анотирането им е относително безпроблемно и еднозначно, а добавянето им в БулНет се ограничава до сферата, в която се използват. В някои от случаите елиптичната употреба е равнозначна на контекстова замяна с хипероним, както е в пример (3а), в други обаче, както е в пример (3б), е необходимо попълване на липсващия компонент в лемата.

- (3) а. ...там засега има само една **кафява мечка**. *Мечката* ...
б. ...всички се втурнаха срещу **представителя** {народен **представител:1**; депутат:1} в парламента.

11.2.2.2.1.2. Съставни названия

Елементите на съставните названия (компаунди или конвенционални колокации по терминологията на Хауърт, 1998) запазват отделните си значения и имат композиран прочит. Значението на отделните елементи е явно, но тяхната комбинация е конвенционализирана и означава нов обект или същност. Тази особеност предопределя и възможността за елиптичната им употреба при координация – примери (4а) и (4б), както и замаяната им с опорната дума – хипероним – пример (4в), тъй като хиперонимът е изразен в термина.

- (4) а. *На връщане купи телешко и свинско месо.*
б. *Вече нямаше възможност да поддържа мобилен и стационарен телефон.*
в. *...сложи писмото в кутията (пощенска кутия) и тръгна.*

11.2.2.2.1.3. Съставни имена

Съставните имена са лексикални единици, които назовават единични, конкретни реалии. В процеса на аотиране, в зависимост от важността на съставните имена и с оглед на енциклопедичната информация в БулНет, са регламентирани два подхода. Добавяне на конкретната реалия в БулНет, в съответния синонимен ред ако има преводно съответствие на английски или в нов синонимен ред, съобразно значението и структурата на лексикално-семантичната мрежа (по-подробно вж. част 4). Съответно в корпуса са аотирани следните подтипове:

- Съставни имена, за които е възприета конвенция за обобщаване при аотиране – такива са *антропонимите* – поликомпонентни лични имена, прякори, псевдоними и т.н., аотирани със съответното синонимно множество или обобщено като {*малко име:1, собствено име:1*}, {*фамилия:1, фамилно име:1*}; и *топонимите*, аотирани със съответното синонимно множество или обобщено като {*топоним:1, географско название:1*} или негови хипоними;
- Съставни имена, които се въвеждат последователно в wordnet, без да бъдат обобщавани. Такива са *космонимите* (*Млечен път, Слънчева система*) и *хрононимите* – имена на исторически епохи и събития (*Втора Световна война, Ден на славянската писменост и култура*); *хремонимите* (собствени имена на предмети от материалната култура имена на картини, статуи,

търговски марки като – „Кока кола”; „Видима идеал стандарт”) и *ергонимите* – собствени имена на учреждения, организации, съюзи и др.

11.2.2.2.2. Идиоми

Идиомите представляват лексикално семантични единици с еднородна и неизменяема форма. Едновременно означават обект, действие или качество, имат метафорично значение и изразяват оценката на говорещия. Значението на целия израз е независимо от значенията на отделните компоненти и не е конструирано според общите правила на езика за семантични и морфологични характеристики или за синтактична конфигурация на елементите. Формално композицията може да се разгледа като ‘АОВ’, при която вместо регулярната сума ‘АОВ’ от означаваните от А и В, има означавано ‘С’, различно от тази сума, невключващо нито ‘А’, нито ‘В’. Например, ако означим компонентите на *ритвам камбаната* с АВ, където *ритвам* = А, а *камбаната* = В, значението на израза *умирам* ще бъде = С. Такива са: *от все сърце, от сутрин до вечер, втора цигулка*. Принципът на аотирането и въвеждането им в лексикално-семантичната мрежа се определя от тяхната семантика и възможността за замяната им с еднокомпонентен синоним. Критерият за неизменяемост на тази група стеснява значително обхвата и и я прави лесно различима при аотиране. При добавянето на този тип съставни единици в БулНет се взема предвид както общата семантика, която изразяват, и която може да бъде проверена чрез синонимна замяна с техния хипероним, така и допълнителното значение, изразяващо оценка и отношение. Например литералът *ритвам камбаната* е хипоним на синонимното множество {*умирам:1, умра:1, загивам:2, загина:2, почивам:5, почина:1*} и се включва в един възел, както със синонимните идиоми: *хвърлям топа:1, хвърля топа:1, гушвам букета:1, гушина букета:1, ритна камбаната:1, предам богу дух:1*, така и с еднокомпонентните: *пукясам:1, пукясвам:1, пукна:2, пуквам:2*. Тъй като дефинирането на множеството се определя и от разговорния характер на литералите, то е свързано чрез релацията *обичайна употреба* с {*разговорна дума:1, жаргонна дума:1*}.

11.2.2.2.3. Фразеологизми

Ербах (1992) нарича фразеологизмите полу-декомпозируеми или анализируеми фраземи. Те проявяват определена степен на вариация и / или носят семантика, която не е адекватна извън израза. Конституентите

на фразеологизмите са регулярни и значението им е преносно в определена степен, като формалната, морфологичната или синтактичната конструкция е частично фиксирана – позволени са вариации на словоредата, на лексикалните компоненти, но те не спазват общите, регулярни правила на езика. Изборът на означител зависи от ситуацията. От своя страна фразеологизмите биват:

11.2.2.2.3.1. Полуидиоми

Полуидиомите са фраземи, които проявяват определена степен на вариативност и имат частична фиксираност. Изразяват преносно и / или метафорично значение, като едновременно с това означават обект, действие или качество и изразяват отношение и емоция. Значението на целия израз е отчасти независимо от значенията на компонентите му, които от своя страна може да имат преносно значение или семантичното или лексикалното множество на възможните реализации да варира. Означаваното от изрече 'X' включва означаваните и на двете съставлящи го лексеми 'A' и 'B' и допълнително означавано 'C', което е различно от означаваните 'A', и 'B'. Например в *създавам семейство*, A= *създавам* и B = *семейство*, означаваното от обединението 'AOB' има стойност, както регулярното *омъжвам се, оженвам се, сключвам брак*, така и 'C'= *създавам семейни взаимоотношения и атмосфера*. 'C' може да бъде изразено от означаваните /D/ или /R/ – *дом, семейно огнище*, които принадлежат на означаваното 'B', но не и самостоятелно от 'A', 'B', 'R', 'D'. Този тип съставни единици създават най-много трудности при аотиране, тъй като позволяват структурни и компонентни вариации, дистантно разположение на компонентите в контекст, както и различни морфо-синтактични вариации. От друга страна при тях в хода на аотацията възниква и проблемът за определяне на границите на фраземата и за изключването на контекстовите елементи, вмъкнати в тяхната структура.

11.2.2.2.3.2. Прагматемите

Прагматемите са полудеконпозирани фраземи, в които компонентите са семантично композирани, значенията им поотделно са ясни, означаваният концепт се определя от екстралингвистична ситуация, която говорещият иска да вербализира. В този тип словосъчетания означаваното от комбинацията 'AOB' е регулярна сума от 'A' и 'B', но не може да бъде заменена частично или цялостно от никое еквивалентно означавано 'X', тъй като е определено от ситуацията, която описва. Множеството на прагматемите включва пословиците, поговорките,

крилатите фрази, езиковите формули и т.н. В прагматемата, ситуацията предоставя свободен избор на възможни значения. Понякога тя предопределя какво да се каже, или какъв конкретен израз да се даде на избраното значение. Докато при аотиране пословиците и поговорките се разглеждат главно като свободни изречения и отделните им компоненти се определят в зависимост от значенията, представени в БулНет, крилатите изрази и стилистично обусловените словесни формули като *добро утро, добър ден, добър вечер, добре дошъл* и др. се определят като съставни единици и се въвеждат в лексикално-семантичната мрежа според конвенцията в Принстънския wordnet, обединени като хипоними на {*поздрав:1; приветствие:1*}.

11.2.2.2.4. Колокации

Лексикално-функционалните фраземи, както още са известни колокациите, са регулярни комбинации от думи или фрази, при които обичайната композиционна семантична функция конструира значението на израза от значенията на съставлящите го думи. Включените думи имат склонност да се появяват в конкретната комбинация. Във фраземата означаваното е независимо от означаваното на някой от конституентите. Лексикалната единица, която пази своето означавано, е опорна, ключова дума на колокацията (означена с главни букви в примерите по-долу). Според класификацията на Мелчук (1995: 13) колокациите могат да се дефинират в четири подкласа според връзката между А, В и означаваното от тяхната комбинация ‘С’:

- Означаваното ‘С’ е различно от означаваното ‘В’.

(i) ‘С’ е празно, а лексемата В е полуспомагателна, използвана, за да осигури синтактичната конфигурация. Например: *Правя УСЛУГА, правя КОМПРОМИС, правя СРАВНЕНИЕ, давам ОТГОВОР, предприемам СЪПКА, вземам РЕШЕНИЕ.*

Този тип колокации, наричани още глаголно-именни названия или перифрази, съдържат т.нар функционални глаголи, които образуват девербални номинации. Глаголите в този тип конструкции като *правя, давам, вземам, имам (do, give, have, make, take)* се приемат като лексикално десемантизирани и имат специфично значение в лексикално-семантичната мрежа, което се използва при такъв тип комбинации (вж. 11.2.2.1). В много от случаите обаче композирането на някои от колокациите не е възможно поради спецификата или общия характер на дефиницията – взема се компромисно решение и целият израз се въвежда като литерал, синонимен на глаголите в подходящото синонимно множество или се добавя като

самостоятелно множество. Така например, {*вземам участие:1*} е синоним на {*участвам:2*}, {*полагам грижи:1*} – на {*обгрижвам:2*}, {*вземам решение:1*} – на {*решавам:7, достигам до решение:1*} и др. Колокацията {*вземам души:1*} е въведена като самостоятелно синонимно множество.

(ii) лексемата В изразява ‘С’ само в комбинация с А. Например: *черно КАФЕ, френски ПРОЗОРЕЦ, кисело МЛЯКО, пощенска КУТИЯ.*

При изразите с опорна и подчинена част се наблюдават различни вариации при аотиране. Тяхната композираност позволява елиптичната употреба на един от компонентите, като другият се подразбира от контекста. В такъв случай се наблюдава т. нар. елипса, която може да бъде предположена от наличието на опорния компонент, както е в (4в). Елипса на опората може да се наблюдава и при координирана употреба на колокации с обща опора в съседство (4б). В такъв случай липсващата опора се добавя в текста при лематизиране.

• Означаваното ‘С’ съвпада с означаваното ‘В’.

(i) ‘В’ не може да бъде изразен от никой друг възможен синоним. Например: *силно (*мощно, интензивно) КАФЕ, тежки (*трудни, непосилни) ПОСЛЕДИЦИ, пълно (*изчерпателно) ВЪЗСТАНОВЯВАНЕ.*

(ii) ‘В’ включва важна част от означаваното ‘А’. Например: *конско ЦВИЛЕНЕ, орлов НОС, гранясало МАСЛО.*

Този тип колокации не се аотират като съставни единици, значението им се композира от значенията на отделните им компоненти в wordnet.

11.3. Морфо-синтактичната идентификация на съставните единици

Идентифицирането на съставните фрази съответства на това при думите в лексикално-семантичната мрежа. Съставните единици са класифицирани в девет части на речта по модела на несъставните лексикални единици). Често отнасянето на съставните единици към една или друга част на речта е относително и се определя от преводното съответствие, от синтактико-семантичната им функция при употребата им в контекст (Ничева и др. 1974), а не от опората на израза. Това е особено изразено при идиомите и фразеологизмите и се дължи на спецификата на техните значения. Някои от тях изразяват оценка или отношение, един или повече от компонентите им са употребени в преносно значение.

Например фразеологизмът *кожа и кости* се състои от две съчинително свързани съществителни, но е включен в синонимно множество на прилагателно заедно с литералите *мършав:1, кльоцав:1, хилав:1,*

кокалест:2, дефинирани като ‘който е грозно слаб’ (Беше станала направо кожа и кости.)

Пагмематата *добре дошъл:1* (‘който е посрещнат с радост и е желан гост’) е определена като прилагателно, макар да е съставена от причастие и модифициращо наречие. Антонимното синонимно множество – {*нежелан:1, неприятен:4, неохотно приет:1*}, представя и друга гледна точка на съставността – невъзможността за точно и еднозначно предаване на значението на дадено синонимно множество, въведено в английската мрежа поради разминаване в начините за лексикализация на различните езици. Макар *неохотно приет* да има регулярно композирано значение, неговото въвеждане като литерал в множеството е необходимо по следните причини – точното преводно съответствие и запазването на антонимията. Този тип литерали изискват особено внимание при аотиране, отлично познаване на семантичните релации, представени в лексикално-семантичната мрежа, и обмисляне на преводното съответствие на дадена единица при превод.

Междуетиковите (не)съответствия по отношение на частта на речта при граматикализираните съставни фрази съответства на тези при несъставните единици. Аотирането и добавянето им в БулНет следват възприетите общи конвенции.

11.4. Критерии за аотиране на даден израз като съставен

11.4.1. Преводни съответствия

Наличието на преводен еквивалент и реализирането на изразяваното от съставната лема понятие в една дума в друг език е един от основните критерии, посочвани при проверка на съставността на дадена лексикална единица (Мелчук 1995; Коева 2005). Тук ще предложим разширен вариант – понятието се лексикализира в други езици (в случая отразено в Принстънския wordnet) като несъставна дума, съставна дума, фразеологизъм.

- Наличие на несъставно или съставно преводно съответствие в английската лексикално-семантична мрежа.

(i) При наличие на съставно преводно съответствие преводът на всеки от компонентите може да съвпада. Например: *вземам душ – take a shower; хващам бика за рогата – take the bull by the horns.*

(ii) В съставното преводно съответствие преводът само на някои от компонентите може да съвпада. Например: *вземам решение – take a decision.*

11.4.2. Семантични релации

- Наличие на синонимия с прости думи или съставни фрази, които лексикализират едно понятие, в същия език.
- Възможност за замяна с еднокомпонентен синоним и наличие на синонимно множество в лексикалната мрежа, дефинирано така, че напълно да отговаря на значението на разглежданата съставна единица.

В пример (5) фразеологизмът *идвам до гуша* е синонимен на едносъставните литерали {*писвам:1, писна:1, втръсвам:1, втръсна:1*} с дефиниция '*ставам много досаден*'.

- (5) ... когато за половин година животът стана песен, на човек му ***идва до гуша***.

11.4.3. Наличие на семантична композираност (Болдуин и др. 2003, Тодорова 2006; Стоянова 2009):

- непозволяваща декомпозиране и ограничение по отношение на възможните вътрешни модификации;
- позволяваща декомпозиране, като поне един от конституентите се реализира със значение, което не е допустимо извън рамките на лексикалната единица;
- позволяваща декомпозиране без композиционни алтернативи със сходна семантика.

Семантичната композираност се изразява в различни типове ограничения върху морфо-синтактичното поведение: ограничения върху заменимостта на всеки елемент; синтактична нерегулярност; значение, което изисква предварително знание; ограничения върху перифразиране с една дума или с определена конструкция; словоредни размествания и ограничена парадигматична продуктивност.

11.4.4. Наличие на ясно разграничима разлика в значението с омонимна граматична форма. Употребата в контекст и обкръжението на разглежданата фраза недвусмислено я различават от възможното й несъставно съответствие. В пример (ба) неодошевият аргумент *въпроси* недвусмислено определя фразата *бе разтънат на кръст* като съставна, носеща метафорично значение. На избраното значение съответства синонимното множество

{разпна:1, разпвам:1, разпна:1, разпвам на кръст:1, разпна на кръст:1}

Дефиниция: *‘критикувам някого остро, безмилостно за нещо казано или направено’*

Правилността на избора потвърждава и възможността за замяна на фразата с хиперонима на избраното множество – *Той бе критикуван заради глобалните промени, които предложи.*

В пример (6б) употребата в контекста на *от години* и *живот* илюстрира непрякото значение на *се е оставил на течението*. Изборът на синонимното множество {*оставя се на течението:1, оставям се на течението:1, нося се по течението:1*} със значение *‘живея безгрижно или безотговорно’* се потвърждава от възможността за замяна със синоним от същото множество – *нося се по течението*. Освен чрез замяната с хиперонима {*смайвам:2; смая:2; слисвам:2; слисам:2; удивявам:2; удивя:2; изумявам:2; изумя:2*} изборът на значението *‘смайвам някого толкова приятно, че му е трудно да си поеме дъх’* в пример (6в) се определя и от невъзможността за замяна на компонента *дъх* със синоними от друг синонимен ред {*дъх:4, дихание:2*}.

- (6) а. *Той бе разпнат на кръст от въпроси заради глобалните промени, които предложи.*
б. *Синът ѝ от години се е оставил на течението и този живот му харесва.*
в. *Направо ще ти спре дъха, като видиш картината.*

11.5. Лематизиране на съставните единици в Българския семантичен корпус

При определяне на основната форма на съставните единици в БулСемКор и при въвеждането им в БулНет се спазва принципът за минималност и неутралност на абстрактната лема: *Всеки компонент на фраземата се представя в най-ненатоварената с граматични признаци форма, в която при комбиниране с другите компоненти се запазва фразеологичността на израза. Фразеологичната конструкция се представя в най-слабо натоварената с контекст реализация, съдържаща само компонентите, задължителни за проявяване на фразеологичността* (Тодорова 2010).

Особеностите, произтичащи от съставния характер на съставните лексикални единици, поставят на преден план няколко проблема – дали

основната форма да е лингвистично коректна спрямо парадигмата на отделните компоненти, или като основна форма за дадена съставна единица да се приеме конкретна комбинация от микрословоформи. В много случаи компонентите на съставните фрази влизат в различни отношения помежду си и вътрешната им структура е свързана с различни типове съгласуване. От друга страна, компонентите на съставната фраза може да са ограничени в определена форма, която не съвпада с лемата на съответната самостоятелна лексикална единица. За целите на семантичното аотиране беше прието съставните фрази да се лематизират със своята абстрактна основна форма (Савъри 2008; Тодорова 2010). Тя представлява линейна последователност от компоненти, които не са леми сами по себе си. Компонентите могат да бъдат:

- лексикализирани (неизменяеми) компоненти, фиксирани в една граматична форма, т. е., конституенти без парадигматични изменения. Такива са всички елементи в състава на идиомите и съставните неизменяеми части на речта и някои компоненти във фразеологизмите, колокациите и в граматикализираните изрази (като частиците *се* при възвратните глаголи и т.н.);
- полулексикализирани (морфологично вариативни) компоненти, които проявяват парадигматични изменения. Такива са компонентите във фразеологизмите, колокациите и в граматикализираните изрази;
- нелексикализирани компоненти, чието специфично значение се основава на повторение и контраст или на специфични словоредни конструкции. Този тип отношения на композираност не се въвеждат в лексикално-семантичната мрежа, с изключение на някои граматикализираните конструкции, например съставни съюзи (*нито ...нито* и др.). В останалите случаи на предаване на специфична семантика чрез фразеологизирани конструкции (Ничева 1987) конструкциите се композират чрез аотиране на отделните компоненти с техните значения като отделни думи.

При аотирането и въвеждането в БулНет се следват критериите за определяне на изходната, основна форма на дадена фразема (Тодорова 2009).

Определянето на основната форма при аотиране на граматикализираните словосъчетания, съставните фраземи и идиомите е сравнително безпроблемно, тъй като техните компоненти са неизменяеми и основната им форма съвпада с формата, в която изразът е определен

в контекста. Възможностите за различни вариации при фразеологизмите и колокациите налага определянето на някои конвенции при свеждането им до основна форма. Принципът за въвеждането на съгласувана основна форма при лематизиране е свързан с т.нар. ендоцентрично (съгласуване между компонентите в рамките на фразеологизма) и екзоцентрично съгласуване (съгласуване, свързано с външната, морфо-синтактична съчетаемост с контекстови елементи, външни за конструкцията, при „отворените” фразеологизми (Ничева 1987:73; Тодорова 2009)). Възможността за трансформации и същевременно аномалните форми, в които може да е фиксиран даден фразеологизъм създават условия за появата на омонимни конструкции в контекста, които трябва да бъдат различавани при аотиране. Например при фразеологизмите, включващи притежателни конструкции, свеждането на употребената в контекста форма до основна следва няколко модела:

- При ендоцентрични фразеологизми свеждането до основна форма е свързано с преобразуване според приетите конвенции (Тодорова 2009). Например: *сърбам Ивановата попара* се лематизира като *сърбам попарата на; от нейната камбанария* се лематизира като *от камбанарията на*.
- При екзоцентрични употреби на фразеологизми лематизирането е свързано с преобразуване на компонентите до съответната неутрална граматична форма. Например: *... на детето му се брояха ребрата...* се лематизира като *ребрата му се броят*.

Специфично е и лематизирането на фразеологизми, съдържащи причастие:

- Фразеологизми – сравнения, формирани със съюза *като*, могат в контекста да бъдат сбъркани с пасивни трансформации в сравнителна конструкция. Те трябва да се различават от фразеологизми, които съдържат страдателно причастие в лемата си. Например фразеологизмите *като ударен с мокър парцал, като гръмнат, като на верига държан* и др. се свеждат до основна форма съответно *удрям с мокър парцал, гръмвам / застрелям някого, държа на верига*. Този тип конструкции се различават от конструкции, в които *като* конструкцията пояснява опорния глагол. Тяхната синонимна перифраза съдържа прилагателно или наречие: *спя като заклан ‘спя дълбоко’*.
- Фразеологизми, употребени в контекста в страдателен залог, трябва да бъдат разграничавани от фразеологизми, фиксирани

в страдателен залог (*спя като къпан, лъжа като съдран, писано е*). Например *песента ми е изпята* се лематизира като *изпя си песента*.

11.6. Проблеми при определянето на лексикалния състав на съставните единици в Българския семантично анотиран корпус. Структурни особености – дистантни компоненти, липсваща опора, вариации в структурата и формите

Съставните лексикални единици притежават ред специфични характеристики, произтичащи от структурата им. Графично, като линейна последователност от думи, те могат да бъдат: фиксирани в своята последователност, задължително непрекъснати, но податливи на разместване или да позволяват прекъсване от други словоформи. От своя страна компонентите им също могат да са напълно фиксирани или да подлежат на морфологични, синтактични и семантични вариации. Едни от основните проблеми при анотирането са съставни фрази, които не са фиксирани и подлежат на различни структурни видоизменения:

- Съставни единици, чиито конститuentи могат да се появят в различен ред и синтактични конструкции. Например фразеологизмът *давам възможност* в пример (7).
- (7) а. *На конференцията **щяла да се даде възможност** на младите хора.*
б. *Тази **възможност** по случайност му **беше дадена** от съдбата.*
- Съставни единици, които позволяват (i) вмъкване на външни контекстови елементи от различни лексикално-граматични категории и / или разместване на реда на конститuentите, (ii) наличие на повече от един конститuent, който може да варира. Например фразеологизмът *давам повод* в пример (8).
- (8) *След това изявление министърът **даде** още множество **поводи** медиите да се **вглеждат** изумено в лика му.*
- Понякога е трудно да се определи към коя част на речта принадлежи даден фразеологизъм и / или коя е лексикалната му опора с оглед на въвеждането му в БулНет (вж т. 11.3 и част 3).
 - Съставни единици, които позволяват (i) замяна и изпускане на компоненти; (ii) деривационни и перифразни трансформации; (iii) семантично мотивирани замени на компоненти.

(9) ... *мед ми закапа на сърцето* {*мед ми капе на сърцето:1*}, когато
чух как съседката плюе

11.7. Добавяне на нови думи и значения в Българския wordnet

В някои случаи в wordnet не е въведено синонимно множество, в което да фигурира дадена съставна единица, а в други английският еквивалент не включва значението, употребено в корпуса. Критериите за добавяне на нов литерал или синонимно множество в лексикално-семантичната мрежа изисква комбинаторен подход, при който се разглеждат потенциално подходящи синонимни множества в Принстънския wordnet, отговарящи на значението от корпуса, и преценка на релевантността на съставните названия в българския език. Невъзможността за композиране на аотираното словосъчетание от несъставни литерали, въведени в лексикално-семантичната мрежа, сигнализира, че е възможно дадена комбинация от думи да е съставна лексикална единица. Принципите за включване на съставни изрази в лексикално-семантичната мрежа съответстват на общите принципи – (i) синонимите в синонимното множество; (ii) релациите на избрания синонимен ред; (iii) дефиницията на лексикалното значение и др. Един от важните проблеми, които възникват във връзка с анотацията на съставните единици, е липсата на съответствие в Принстънския wordnet, което е особено често срещано при фразеологичните изрази (Фелбаум 1998). Кодифицирането на тези единици в БулНет е свързано с изграждането на критерии за избор на мястото им на включване в лексикално-семантичната мрежа, например чрез избор на хипероним. Изборът на надредно понятие от една страна може да се определи от наличието на общ семантичен признак в основата на значението на потенциалните хипоним и хипероним, а от друга – може да се избере словообразователно според опората на фразата.

11.7.1. Типове фраземи с оглед на включването им в лексикално-семантичната мрежа

- Фраземи, които имат точно преводно съответствие в Принстънския wordnet. Някои от тях са резултат от културни взаимодействия, например при калки като *take the bull by the horns* – *хващам бика за рога*. Тези примери имат смислово съответстващи конституенти, както и еднакво цялостно значение.

- Фраземи, които имат едно и също идиоматично значение и в двете семантични мрежи, но техните компоненти и етимология са различни.
- Случаи, в които преводният еквивалент на дадена фразама в единия или в другия език е самостоятелна дума. В такива случаи, когато идиомът е еквивалентен на синонимно множество, той трябва да бъде добавен като нов хипоним на синонимна речникова единица. Например в Българския wordnet фразеологизмът {*дим да ме няма:1*} е въведен като хипоним на {*изчезвам:6, изчезна:6, изгубвам се:4, изгубя се:4*}, съответстващо на английското *disappear*.
- Случаи, в които фраземите стават част от един и същ синонимен ред в структурата на речника. Например {*извеждам на преден план:1, изведа на преден план:1, изнасям на преден план:1, изнася на преден план:1, извеждам напред:1, изведа напред:1, изнасям напред:1, изнася напред:1*}.
- Случаи, в които фраземите стават част от синонимно множество заедно с други несъставни единици със същото значение. Например {*месарница:1, месарски магазин:1, магазин за месо:1*}, {*цветарница:1, цветарски магазин:1, магазин за цветя:1*}.

11.8. Заключение

Анотирането на съставни единици поставя редица въпроси за начините за представяне на тяхното значение и за подходите при формализирането на формите им. Явно е, че спецификите на съставността се изразяват и в уникалното значение на фразата, затова по отношение на синонимията на съставните единици можем да говорим за синонимна употреба, но не за взаимна заменяемост. Степента на композираност е основният критерий, който определя решението при анотиране и при добавяне в БулНет. Посочената класификация на съставните фрази, основана на признака композираност спомага при вземането на решение в спорни случаи.

Цитирана литература

- Балтова 1971: Балтова, Ю. *За границата между сложна дума и словосъчетание*. – БЕ, № 2 – 3.
- Болдуин Т. и др., 2003: Baldwin T, Bannard C, Tanaka T., Widdows D. *An Empirical Model of Multiword Expression Decomposability*. In: Proceedings of the ACL-2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment, Sapporo, Japan, pp 89-96.
- Ербах 1992: Erbach G. *Head-Driven Lexical Representation of Idioms in HPSG*. – In: Proceedings of the International Conference of Idioms, Tilburg.
- Коева 2005: Koeva S. *Inflection Morphology of Bulgarian Multiword Expressions*. – In: Computer Applications in Slavic Studies – Proceedings of Azbuki@net, International Conference and Workshop, Sofia, pp. 201-216.
- Мелчук 1995: Melchuk I. *Phrasemes in Language and Phraseology in Linguistics*. – In: Idioms: Structural and Psychological Perspectives, chapter 8. Lawrence Erlbaum Associates, pp. 167–232.
- Милър 1995: Miller G. A. *Building Semantic Concordances: Disambiguation vs. Annotation*. – In AAAI Technical Report SS-95-01., pp. 92-94.
- Ничева и др. 1974: Ничева К., С. Спасова-Михайлова И Кр. Чолакова. *Фразеологичен речник на българския език*. т. I. Издателство на БАН, София.
- Ничева 1987: Ничева К. *Българска фразеология*. София.
- Савъри 2008: Savary A. *Computational Inflection of Multi-Word Units. A contrastive study of lexical approaches*, CSLI Publications LiLT, vol. 1, issue 2.
- Саг и др. 2002: Sag I, Baldwin T., Bond F., Copestake A, and Flickinger D. 2002. *Multiword expressions: A pain in the neck for NLP*. In Proc. of the 3rd International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2002), pp 1–15.
- Стоянова 2009: Стоянова И. *Някои методи за автоматично разпознаване на съставни лексикални единици за български език*. – БЕ, кн. 3, стр. 48-69.

- Тодорова 2006: Todorova M. *On The classification of Bulgarian Non-Free Phrases*, Proceedings of the Fifth International Conference Formal Approaches to South Slavic and Balkan Languages, 18-20 October 2006, Sofia, Bulgaria, pp. 251-256.
- Тодорова 2010: Тодорова М. *Организация на лингвистичната информация в компютърен морфологичен речник на български глаголни фразеологизми*, Доклади от Петата конференция по лексикология и лексикография, 2009 г. стр. 331-344.
- Тодорова, М. 2009: Тодорова М. *Лексикализирана граматика на български глаголни фразеологизми*, В: Български език, 2009 г., кн. 3, стр 70 – 83.
- Фелбаум 1998: Felbaum C. *Towards a Representation of Idioms in WordNet*, In: Proceedings of the workshop on the Use of WordNet in Natural Language Processing Systems (Coling-ACL) 1998
- Фелбаум и др. 2004: Felbaum C., A. Pease. *Language to Logic Translation with PhraseBank*. In: Proceedings 2nd Global Conference (GWC'04). pp.187-192.
- Хауърт, 1998. Howarth P. 1998. *The Phraseology of Learners' Academic Writing*. In: Cowie A.P. (ed.) *Phraseology: Theory, Analysis, and Applications*. Oxford: OUP, pp. 161-186.
- Чермак Фр. 2001: Čermák Fr. *Substance of Idioms*. In: *Perennial Problems, Lack of Data or Theory?* In: *International Journal of Lexicography*, Vol 11, No 1, 1-20.
- Чолакова 1968: Чолакова Кр. *За лексико-граматичната еквивалентност на фразеологичната единица с думата*. Славистичен сборник. София, стр. 209.

Изречение: Момичето четеше на глас.

Дума: момичето

Избрано значение:

Синонимно множество: {момиче:1, младо момиче:1, девойка:1, госпожица:1, девойче:1}

Идентификационен номер: ENG20-09478614-n

Дефиниция: 'млада, обикновено неомъжена, жена'

Релации:

[хипоним: красавица:1, хубавица:1]

[хипоним: госпожица:2, г-ца:1]

[хипоним: девойка:2, девица:1, мома:1]

[хипоним: комсомолка:1]

[хипоним: гадже:2, мацка:1, кукла:3, мадама:1]

[хипероним: жена:2]

Примерна употреба: осемнадесетгодишна девойка

Останалите значения в БулНет:

Синонимно множество: {момиче:2, момиченце:1}

Идентификационен номер: ENG20-09437323-n

Дефиниция: 'дете от женски пол'

Примерна употреба: Бебето беше момиченце.

Дума: четеше

Избрано значение:

Синонимно множество: {чета:1, прочитам:1, прочета:1}

Бележка към литерала *чета*:1: имперф. т.

Бележка към литерала *прочитам*:1: несв. в.

Бележка към литерала *прочета*:1: св. в.

Идентификационен номер: ENG20-00604996-v

Дефиниция: 'възпроизвеждам, интерпретирам устно зрително възприета информация'

Релации:

[хипоним: зачитам:1, зачета:1]

[хипоним: прочитам бегло:2, прочета бегло:1, преглеждам бегло:1, прегледам бегло:1]

[хипоним: препрочитам:1, препрочета:1]

[хипоним: изчитам:1, изчета:1]

[хипоним: разчитам:1, разчета:1]

[хипоним: чета:5, прочитам:3, прочета:3]

[хипоним: сричам:1]

[хипероним: тълкувам:2, изтълкувам:2, интерпретирам:1]

[глаголна група: чета:2, прочитам:2, прочета:2]

[глаголна група: глася:1, пиша:7, казвам:7, кажа:7]

[субсъбитие: чета:4, уча:6]

[производен: четене:1]

[производен: читател:1]

Примерна употреба: *Прочети обявата.*

Примерна употреба: *Чел ли си Салман Рушди?*

Останалите значения в БулНет:

Синонимно множество: {чета:4, уча:6}

Бележка към литерала *чета:4*: имперф. т.

Бележка към литерала *уча:6*: имперф. т.

Идентификационен номер: ENG20-00587299-v

Дефиниция: *'усвоявам знания чрез четене от учебници, книги и т.н.'*

Примерна употреба: *В момента чете за изпита по геология.*

Синонимно множество: {чета:2, прочитам:2, прочита:2}

Бележка към литерала *чета:2*: имперф. т.

Бележка към литерала *прочитам:2*: несв. в.

Бележка към литерала *прочета:2*: св. в.

Идентификационен номер: ENG20-00606410-v

Дефиниция: *'оповестявам устно някаква информация'*

Примерна употреба: *Министърът прочете имената на победителите.*

Синонимно множество: {чета:5, прочитам:3, прочита:3}

Бележка към литерала *чета:5*: имперф. т.

Бележка към литерала *прочитам:3*: несв. в.

Бележка към литерала *прочета:3*: св. в.

Идентификационен номер: BUL-285899330-v

Дефиниция: *'по външни белези се досещам за чувствата, мислите, състоянието на някого'*

Примерна употреба: *В очите му прочете обзелия го ужас.*

Синонимно множество: {чета:6, разчитам:4, разчета:2}

Бележка към литерала *чета:6*: имперф. т.

Бележка към литерала *разчитам:4*: несв. в.

Бележка към литерала *разчета:2*: св. в.

Идентификационен номер: ENG20-00607499-v

Дефиниция: *'извличам информация от някакъв носител (магнитна лента, диск и т.н.)'*

Примерна употреба: *Речникът може да се чете от компютър.*

Синонимно множество: {чета:3, отряд:1, военен отряд:1, военен отред:1}

Идентификационен номер: ENG20-07717650-n

Дефиниция: *'малка сухопътна войскава част, сформирана за извършването на дадена бойна задача'*

Дума: на глас

Избрано значение:

Синонимно множество: {на глас:1, гласно:1, с глас:1}

Идентификационен номер: ENG20-00070546-b

Дефиниция: *’означава, че нещо се произнася не наум, а високо, така че да се чуе’*

Релации:

[тематична област: начин:2, стил:2, маниер:1]

Примерна употреба: *Моля прочетете пасажа на глас.*

Примерна употреба: *Тя се смееше с глас.*

Изречение: Хемингвей не беше предвидил тая работа.

Дума: Хемингвей

Избрано значение:

Синонимно множество: {*Хемингвей:1, Ърнест Хемингвей:1*}

Идентификационен номер: ENG20-10322287-n

Дефиниция: *’американски писател (1899-1961), носител на Нобелова награда за литература през 1954 г.’*

Релации:

[хипероним: писател:1, писателка:1, автор:1, авторка:1]

[производен: хемингвейски:1]

Дума: не

Избрано значение:

Синонимно множество: {*не:3*}

Идентификационен номер: ENG20-00024128-b

Дефиниция: *’частица, обикновено предхождаща сказуемото, с която се отрича цялото изказване; произнася се без ударение’*

Бележка към синонимното множество: POS=particle

Релации:

[тематична област: частица:2]

Примерна употреба: *Той не говори английски.*

Примерна употреба: *Тя не отиде на срещата.*

Останаха значения в БулНет:

Синонимно множество: {*не:1*}

Идентификационен номер: BUL-2119093311

Дефиниция: *’частица, с която се отрича следхождащата дума или словосъчетание; произнася се с ударение’*

Примерна употреба: *Той е не добър, а отличен работник.*

Синонимно множество: {*не:4*}

Идентификационен номер: BUL-1198182566

Дефиниция: *’частица, с която се дава отрицателен отговор’*

Примерна употреба: *Съгласен ли си? – Не.*

Синонимно множество: {*не:2*}

Идентификационен номер: ENG20-00024401-b

Дефиниция: *’частица за подсилване на отрицание, несъгласие, отказ и т.н.’*

Примерна употреба: *Не, вие грешите.*

Дума: беше

Избрано значение:

Синонимно множество: {съм:1}

Идентификационен номер: BUL-1743207879

Дефиниция: *‘спомагателен глагол, който е носител на граматичното значение
в рамките на аналитичните словоформи’*

Релации:

[тематична област: спомагателен глагол:1]

Примерна употреба: *Образът на автора е пресъздаден с лекота.*

Останалите значения в БулНет:

Синонимно множество: {съм:9}

Бележка към литерала съм:9: имперф. т.

Идентификационен номер: ENG20-02526983-v

Дефиниция: *‘имам определено качество, белег, отличителна черта’*

Примерна употреба: *Джон е богат.*

Синонимно множество: {съм:4}

Бележка към литерала съм:4: имперф. т.

Идентификационен номер: ENG20-02538502-v

Дефиниция: *‘идентичен съм с някого или нещо друго, изразено със сказуемното
определение’*

Примерна употреба: *Това е домът ми.*

Примерна употреба: *Президентът на компанията е Джон Смит.*

Синонимно множество: {съм:14}

Бележка към литерала съм:14: имперф. т.

Идентификационен номер: bg-0000130L

Дефиниция: *‘копулативен глагол, с който фраза или изречение, съдържащи
описание, оценка и под., се приписват на субекта’*

Примерна употреба: *Хубаво е подробностите да се обмислят предварително.*

Синонимно множество: {съм:6, намирам се:2}

Бележка към литерала съм:6: имперф. т.

Бележка към литерала съм:6: несл. в.

Идентификационен номер: ENG20-02576058-v

Дефиниция: *‘съществувам или в даден момент присъствам на определено
място в пространството’*

Примерна употреба: *Бараката за инструменти е в задния двор.*

Синонимно множество: {съм:13, представлявам:3, съставявам:1,
съставям:5}

Бележка към литерала съм:13: имперф. т.

Бележка към литерала представлявам:3: имперф. т.

Бележка към литерала съставявам:1: имперф. т.

Бележка към литерала съставям:5: несл. в.

Идентификационен номер: ENG20-02542637-v

Дефиниция: *‘еквивалентен или равнозначен съм на нещо или някого или имам
определен статут, качества, особености’*

Примерна употреба: *Тези пари са единственият ми доход.*

Примерна употреба: *Неколицината мъже представляват цялата му армия.*

Примерна употреба: *Писмото не представлява доказателство от правна гледна точка.*

Синонимно множество: {съм:15, намирам се:6}

Бележка към литерала съм:15: имперф. т.

Бележка към литерала намирам се:6: несв. в.

Идентификационен номер: bg-0000131L

Дефиниция: *'оказвам се, пребивавам (продължително) в определено положение или ситуация'*

Примерна употреба: *До неотдавна държавата беше в икономически и финансов колапс.*

Синонимно множество: {съм:3}

Бележка към литерала съм:3: имперф. т.

Идентификационен номер: ENG20-02372475-v

Дефиниция: *'работа на определено място на дадена длъжност, в дадена област или с определени функции'*

Примерна употреба: *Тя е ветеринар.*

Синонимно множество: {намирам се:3, съм:8, падам:20, падна:19, лягам:3, легна:3}

Бележка към литерала намирам се:3: несв. в.

Бележка към литерала съм:8: имперф. т.

Бележка към литерала падам:20: несв. в.

Бележка към литерала падна:19: св. в.

Бележка към литерала лягам:3: несв. в.

Бележка към литерала легна:3: св. в.

Идентификационен номер: ENG20-02651082-v

Дефиниция: *'намирам се в определено отношение спрямо други хора, неща или събития, обикновено включващо зависимост от тях или наличие на отговорност от тяхна страна'*

Примерна употреба: *Съдбата на Босна е в ръцете на Европа.*

Примерна употреба: *Отговорността за действията му пада върху нас.*

Синонимно множество: {съществувам:3, съм:12, има:5, битувам:1}

Бележка към литерала съществувам:3: имперф. т.

Бележка към литерала съм:12: имперф. т.

Бележка към литерала има:5: имперф. т.

Бележка към литерала има:5: impers.

Бележка към литерала битувам:1: имперф. т.

Идентификационен номер: ENG20-02526084-v

Дефиниция: *'представявам съставна част от света, имам реално битие'*

Примерна употреба: *Съществува ли Бог?*

Синонимно множество: {равнявам се:1, равен съм:1, съм:5}

Бележка към литерала равнявам се:1: несв. в.

Бележка към литерала равен съм:1: имперф. т.

Бележка към литерала *равен* съм:1: +р (с, на)

Бележка към литерала съм:5: имперф. т.

Идентификационен номер: ENG20-02584716-v

Дефиниция: *'имам еквивалентни характеристики или еднаква стойност с друг(и)'*

Примерна употреба: *Един долар се равнява на 1000 рубли.*

Синонимно множество: {съм:7}

Бележка към литерала съм:7: имперф. т.

Идентификационен номер: ENG20-02669122-v

Дефиниция: *'случвам се, ставам по определено време или на определено място'*

Примерна употреба: *Това беше посред ноц.*

Синонимно множество: {съм:16}

Бележка към литерала съм:16: имперф. т.

Идентификационен номер: bg-0000400L

Дефиниция: *'нещо съществува, налично е, протича във времето'*

Примерна употреба: *Бе мрак и студ.*

Синонимно множество: {съм:2}

Бележка към литерала съм:2: имперф. т.

Идентификационен номер: BUL-619109044

Дефиниция: *'присъствам на място или събитие по определено време'*

Примерна употреба: *На сватбата бяхме сто човека.*

Синонимно множество: {съм:10}

Бележка към литерала съм:2: имперф. т.

Идентификационен номер: ENG20-02650181-v

Дефиниция: *'изглеждам или се възприемам от някого като някакъв, най-вече чрез определени сетива'*

Примерна употреба: *Чаршафите са меки на пипане.*

Синонимно множество: {струвам:2, съм:11, коствам:2}

Бележка към литерала *струвам*:2: имперф. т.

Бележка към литерала съм:11: имперф. т.

Бележка към литерала *коствам*:2: имперф. т.

Идентификационен номер: ENG20-02621203-v

Дефиниция: *'имам определена парична стойност или се разменям срещу дадена сума'*

Примерна употреба: *Според наркополицая на уличния пазар грам кокаин е сто лева.*

дума: предвиждам

Избрано значение:

Синонимно множество: {предвиждам:3, предвидя:3, предусещам:2, предусетя:2, предугаждам:2, предугадя:2}

Бележка към литерала *предвиждам*:3: несв. в.

Бележка към литерала *предвидя*:3: св. в.

Бележка към литерала *преусещам*:2: несв. в.

Бележка към литерала *предусетя:2*: св. в.

Бележка към литерала *предугаждам:2*: несв. в.

Бележка към литерала *предугадя:2*: св. в.

Идентификационен номер: ENG20-00696899-v

Дефиниция: *'има предварително интуитивно или рационално знание, представа за бъдещото развитие на дейности, процеси, хода на събития и т.н.'*

Релации:

[хипероним: зная:1, знам:1]

Примерна употреба: *Скептиците отдавна предвиждаха, че това ще се случи.*

Останалите значения в БулНет:

Синонимно множество: {*определям:10, определя:10, уреждам:5, уредя:5, предвиждам:1, предвидя:1*}

Бележка към литерала *определям:10*: несв. в.

Бележка към литерала *определя:10*: св. в.

Бележка към литерала *уреждам:5*: несв. в.

Бележка към литерала *уредя:5*: св. в.

Бележка към литерала *предвиждам:1*: несв. в.

Бележка към литерала *предвидя:1*: св. в.

Идентификационен номер: ENG20-00985999-v

Дефиниция: *'изрично посочвам дадено положение като условие, изискване и под. в договор, споразумение или друг вид двустранни отношения'*

Примерна употреба: *Завециането предвижда тя да живее в къщата до смъртта си.*

Синонимно множество: {*предназначавам:2, предназнача:2, определям:11, определя:11, предвидя:4, предвиждам:4*}

Бележка към литерала *предназначавам:2*: несв. в.

Бележка към литерала *предназнача:2*: св. в.

Бележка към литерала *определям:11*: несв. в.

Бележка към литерала *определя:11*: св. в.

Бележка към литерала *предвиждам:4*: несв. в.

Бележка към литерала *предвидя:4*: св. в.

Идентификационен номер: ENG20-00686005-v

Дефиниция: *'планирам ангажирането или включването на нещо или някого в изпълнението на определена задача, цел, мисия, назначаването на някого на пост и т.н.'*

Примерна употреба: *Книгата е предвидена в помощ на учителите по математика.*

Синонимно множество: {*предвиждам:2, предвидя:2*}

Бележка към литерала *предвиждам:2*: несв. в.

Бележка към литерала *предвидя:2*: св. в.

Идентификационен номер: ENG20-00688745-v

Дефиниция: *'преценявам с увереност, сигурен съм, че нещо ще стане, ще се случи'*

Примерна употреба: *Предвиждаха да пристигнат достатъчно рано.*

дума: тая

Избрано значение:

Синонимно множество: {тоя:3, този:3}

Идентификационен номер: bg-0000703L

Дефиниция: *'показателно местоимение прилагателно (поясняващо именна фраза) за посочване на по-рано споменати или известни лица, предмети, ситуации, явления'*

Релации:

[антоним: онзи:6, оня:6]

[тематична област: показателно местоимение:1]

Бележка към синонимното множество: translation equivalent: *that*

Примерна употреба: *Това влудяващо бумтене и тропане продължаваше вече дни.*

Останалите значения в БулНет:

Синонимно множество: {тоя:5, този:5}

Идентификационен номер: bg-0000809L

Дефиниция: *'показателно местоимение прилагателно (поясняващо именна фраза) за посочване на лица, предмети, ситуации, явления, които се споменават, стават обект на разговор, обсъждане или посочване в настоящия контекст'*

Примерна употреба: *Чуй тази песен.*

Синонимно множество: {онзи:3, оня:4, този:4, тоя:4}

Идентификационен номер: bg-0000790L

Дефиниция: *'показателно местоимение прилагателно (поясняващо именна фраза) за изтъкване на нещо характерно, добре познато или разпознаваемо от главното изречение, което се пояснява от подчинено определително изречение'*

Примерна употреба: *Той я погледна с онази усмивка, която караше сърцето ѝ да се разтапя.*

Синонимно множество: {тоя:2, този:2}

Идентификационен номер: bg-0000288L

Дефиниция: *'показателно местоимение съществително (заместващо именна фраза) за посочване на известни или по-рано споменати лица, често пренебрежително'*

Примерна употреба: *Влиза, разбираш ли, тази, и иска 5 килограма сирене!*

Синонимно множество: {тоя:1, този:1}

Идентификационен номер: BUL-1094477356

Дефиниция: *'показателно местоимение прилагателно (поясняващо именна фраза) за посочване на пространствено близки лица, предмети, ситуации, явления'*

Примерна употреба: *Чие е това дете?*

дума: работа

Избрано значение:

Синонимно множество: {работа:12, нещо:12}

Идентификационен номер: ENG20-05344294-n

Дефиниция: *'неопределено или неконкретизирано състояние на нещата, занимание, интерес и т.н.'*

Релации:

[хипероним: работа:5, грижа:8, дело:7]

Примерна употреба: *Имам да свърша няколко неща.*

Примерна употреба: *Не е твоя работа.*

Примерна употреба: *Работите вървят добре.*

Останалите значения в БулНет:

Синонимно множество: {задача:4, работа:1}

Идентификационен номер: ENG20-05344294-n

Дефиниция: *'конкретна дейност, която някой е длъжен да извърши по необходимост или срещу заплащане'*

Примерна употреба: *Задачата да се оправи двигателят отне няколко часа.*

Синонимно множество: {работа:5, грижа:8, дело:7}

Идентификационен номер: ENG20-05343715-n

Дефиниция: *'дейност, занимание, което пряко засяга някого, в което той е пряк участник или е негова отговорност'*

Примерна употреба: *Сигурността на кораба е първа грижа на капитана.*

Синонимно множество: {изучаване:2, учене:2, занимание:6, работа:13}

Идентификационен номер: ENG20-05422945-n

Дефиниция: *'системно усвояване на знания в определена област или по определена дисциплина'*

Примерна употреба: *Овладяването на втори език означава непрестанно учене.*

Примерна употреба: *Подготовката за олимпиадата ще ви коства още много работа.*

Синонимно множество: {служба:5, работа:4}

Идентификационен номер: ENG20-00554938-n

Дефиниция: *'(професионално) занимание, свързано с точно определени задължения и заетост, което човек извършва срещу възнаграждение'*

Примерна употреба: *Вече втори месец търся работа.*

Синонимно множество: {работа:7}

Идентификационен номер: ENG20-00547741-n

Дефиниция: *'(начин на) изпълнение на дадена задача, задължение, дейност и т.н.'*

Примерна употреба: *Младата актриса свърши отлична работа в ролята на Офелия.*

Синонимно множество: {функциониране:1, работа:8, опериране:2}

Идентификационен номер: ENG20-12761712-n

Дефиниция: *'процесът или начинът, по който нещо (апарат, машина и под.) действа, функционира'*

Примерна употреба: *Моцността на двигателя предопределя работата му, щом колата излезе на пътя.*

Синонимно множество: {месторабота:1, работа:2, работно място:1}

Идентификационен номер: ENG20-04423637-n

Дефиниция: *'съвкупността от сгради, помещения, съоръжения, заедно с прилежащото земно пространство, където се провежда производствена, търговска или друга служебна дейност'*

Примерна употреба: *Звънни ми, щом пристигнеш на работа.*

Синонимно множество: {творба:2, творение:2, произведение:4, работа:3, дело:5, труд:3}

Идентификационен номер: ENG20-04421190-n

Дефиниция: *'завършен продукт, резултат от вложен физически или умствен труд или усилия'*

Примерна употреба: *Симфонията бе оценена като една от най-майсторските му работи.*

Синонимно множество: {работа:6}

Идентификационен номер: ENG20-00546605-n

Дефиниция: *'дейност, чиято цел е извършването или създаването на нещо'*

Примерна употреба: *Тя провери няколко детайла, които изискваха допълнителна работа.*

Синонимно множество: {занятие:3, работа:9, поминък:1}

Идентификационен номер: ENG20-00553013-n

Дефиниция: *'всеки от разнообразни видове трудова човешка дейност като източник на препитание'*

Примерна употреба: *С каква работа си изкарват хляба?*

Синонимно множество: {работа:11}

Идентификационен номер: ENG20-00555322-n

Дефиниция: *'задължение за извършване на нещо конкретно'*

Примерна употреба: *Поддържането на кореспонденцията ти не е моя работа.*

Синонимно множество: {работа:14}

Идентификационен номер: ENG20-05495750-n

Дефиниция: *'въпроси, интереси, отношения, отговорности, които се отнасят до определен човек или хора'*

Примерна употреба: *Ти по-добре си гледай твоите работи!*

В този миг отвън се чу някакво свирене.

Дума: в

Избрано значение:

Синонимно множество: {в:11, във:11}

Идентификационен номер: BUL-1710443708

Дефиниция: *'предлог, означаващ положение в определена точка във времето, отрязък или период от време'*

Релации:

[тематична област: предлог:1]

[тематична област: време:9]

[подобен на: на:9]

Примерна употреба: *В началото на века това е било много популярно.*

Примерна употреба: *В четири часа ще те чакам пред кафенето.*

Примерна употреба: *Този тип увреждания се откриват в по-късна възраст.*

Бележка към синонимното множество: translation equivalent at

Останалите значения в БулНет:

Синонимно множество: {в:4, във:4}

Идентификационен номер: BUL-1328967457

Дефиниция: *'предлог, означаващ област или сфера на дейност, действие или проява на нещо'*

Примерна употреба: *Тя много напредна в училището си.*

Синонимно множество: {в:19, във:19}

Идентификационен номер: bg-0000304L

Дефиниция: *'предлог за въвеждане на място, институция, където нещо съществува, намира се, осъществява се'*

Примерна употреба: *Концертът ще се състои в „Ройъл Албърт хол“.*

Синонимно множество: {в:14, във:14}

Идентификационен номер: bg-0000203L

Дефиниция: *'предлог за въвеждане на положение или принадлежност в рамките на абстрактно място, пространство или предмет'*

Примерна употреба: *В главата ми е пълен хаос.*

Синонимно множество: {при:9, в:32, във:32}

Идентификационен номер: bg-0000146L

Дефиниция: *'предлог за въвеждане на условията, ситуацията, в контекста на която се реализира посоченото в изказването'*

Примерна употреба: *Процесуални нарушения са установени и в случая на Иван Иванов.*

Синонимно множество: {в:13, във:13}

Идентификационен номер: bg-0000073L

Дефиниция: *'предлог, изразяващ включване в състава или строежа на нещо или участие в нещо'*

Примерна употреба: *Тази пиеса неизменно е в репертоара на трупата.*

Синонимно множество: {в:8, във:8}

Идентификационен номер: BUL-1548830153

Дефиниция: *'предлог за въвеждане на непрекия обект, към който е насочено действие или до който се отнася ситуацията'*

Примерна употреба: *Вярвам в истината.*

Синонимно множество: {през:3, в:22, във:22}

Идентификационен номер: BUL-1748092992

Дефиниция: *‘предлог за въвеждане на (относително кратък или разглеждан в неговата цялост) период от време в границите на който нещо се случва, протича, върши, съществува’*

Примерна употреба: *В зимата на същата година избухва въстание.*

Синонимно множество: {в:10, във:10}

Идентификационен номер: BUL-1845490103

Дефиниция: *‘предлог за въвеждане на начина на извършване или съществуване на нещо или формата, в която нещо или някой се намира, проявява и т.н.’*

Примерна употреба: *разстоянието, измервано в километри*

Синонимно множество: {в:16, във:16}

Идентификационен номер: bg-0000241L

Дефиниция: *‘предлог, означаващ движение на някого или нещо отвън към вътрешността на някакъв предмет или среда, в резултат от което става част или е обграден от него/нея’*

Примерна употреба: *Включи ютията в контакта.*

Синонимно множество: {в:3, във:3}

Идентификационен номер: BUL-2141654683

Дефиниция: *‘предлог, въвеждащ преминаване в състояние, положение или ситуация като резултат от някакво действие или процес’*

Примерна употреба: *Вулканът превърна града в димящи руини.*

Синонимно множество: {в:15, във:15}

Идентификационен номер: bg-0000204L

Дефиниция: *‘предлог за въвеждане на цел на движение (в конкретен или абстрактен смисъл)’*

Примерна употреба: *Преди да дойде в Ню Йорк, не бе виждал небостъргачи.*

Синонимно множество: {в:24, във:24}

Идентификационен номер: bg-0000450L

Дефиниция: *‘предлог за въвеждане на нещо, по отношение или в перспективата на което се разглежда даден обект, ситуация и т.н.’*

Примерна употреба: *В подобен план фокусира вниманието си и авторът.*

Синонимно множество: {в:30, във:30, за:40, при:15}

Идентификационен номер: bg-0000722L

Дефиниция: *‘предлог за посочване на действие, дейност, процес, към които нещо или някой има принос или в които изпълнява роля, функция’*

Примерна употреба: *Тези мазнини имат важна роля и за регулирането на обмяната на водата в организма.*

Синонимно множество: {в:9, във:9}

Идентификационен номер: BUL-1488481672

Дефиниция: *‘предлог за въвеждане на събитие, процес, период, в рамките на които се случват, протичат, наблюдават се дадени явления, състояния и под.’*

Примерна употреба: *В историята на клуба няма друг период на загуби.*

Синонимно множество: {в:5, във:5}

Идентификационен номер: BUL-1226964707

Дефиниция: *'предлог, въвеждащ положение във вътрешността или границите на дадено пространство или предмет като цел на насочвано отвън действие'*

Примерна употреба: *Отиваше в стаята на директора.*

Синонимно множество: {в:27, във:27}

Идентификационен номер: bg-0000654L

Дефиниция: *'предлог за въвеждане на ориентация или насоченост'*

Примерна употреба: *Хайдегер насочва изследванията си в посоката, застъпена от теологията.*

Синонимно множество: {при:13, в:31, във:31}

Идентификационен номер: bg-0000584L

Дефиниция: *'предлог за въвеждане на дейност, действие, ситуация, по време на чието осъществяване или в резултат от което се наблюдава дадена ситуация, положение и т.н.'*

Примерна употреба: *При изписването на иконите най-важното е предаването на духовното чувство.*

Синонимно множество: {в:1, във:1}

Идентификационен номер: BUL-952856811

Дефиниция: *'предлог за въвеждане на облекло, покритие'*

Примерна употреба: *Той ме посрещна във фрак с папийонка.*

Синонимно множество: {в:18, във:18, у:4}

Идентификационен номер: bg-0000277L

Дефиниция: *'предлог, въвеждащ обект (човек, предмет и т.н.), за който е типично или налично дадено качество, характеристика'*

Примерна употреба: *В нея изпъкват две основни качества: честността и чувството ѝ за справедливост.*

Синонимно множество: {в:17, във:17}

Идентификационен номер: bg-0000277L

Дефиниция: *'предлог за въвеждане на място или точка на съприкосновение при насочено движение или контакт, съпроводен с проникване в повърхността'*

Примерна употреба: *Пръстите ти успяват да се вкопчат в една пукнатина и ти прилепваш тяло към каменната стена.*

Синонимно множество: {в:23, във:23}

Идентификационен номер: bg-0000432L

Дефиниция: *'предлог за въвеждане на определен ден от седмицата или част от деня'*

Примерна употреба: *В такава нощ звездите светят ниско над града.*

Синонимно множество: {в:21, във:21}

Идентификационен номер: bg-0000425L

Дефиниция: *‘предлог за въвеждане на занимание или ангажираност с определена дейност’*

Примерна употреба: *Прекарват времето си в грижа за земните блага.*

Синонимно множество: {в:28, във:28}

Идентификационен номер: bg-0000673L

Дефиниция: *‘предлог за посочване на някого или нещо, за което е присъщо, характерно определено качество, черта, факт’*

Примерна употреба: *Тук се заключава основният нравствен конфликт в поезията му.*

Синонимно множество: {в:29, във:29}

Идентификационен номер: bg-0000707L

Дефиниция: *‘предлог за посочване на определена линия, курс на действие или поведение, които се следват’*

Примерна употреба: *В този ред на мисли бих искал да спомена...*

Синонимно множество: {в:2, във:2, о:1}

Идентификационен номер: BUL-233974173

Дефиниция: *‘предлог, означаващ място на повърхностен контакт, допир’*

Примерна употреба: *Пушката изтрака в камънака.*

Синонимно множество: {в:6, във:6}

Идентификационен номер: BUL-1247228446

Дефиниция: *‘предлог, означаващ пространствено положение в границите или вътрешността на някакво място или предмет’*

Примерна употреба: *В стаята е тихо.*

Синонимно множество: {в:12, във:12}

Идентификационен номер: BUL-1938629725

Дефиниция: *‘предлог за въвеждане на темпо на движение’*

Примерна употреба: *Конете препускаха в тръс.*

Синонимно множество: {в:20, във:20, у:3}

Идентификационен номер: bg-0000278L

Дефиниция: *‘предлог, изразяващ асоциация, свързване на образ, представа, качество и под. с даден обект (предложното допълнение)’*

Примерна употреба: *У нея имаше нещо много странно, но така и не разбрах какво.*

Синонимно множество: {в:7, във:7}

Идентификационен номер: BUL-607198122

Дефиниция: *‘предлог, въвеждащ състояние, положение или ситуация, в която някой или нещо се намира или от която е засегнат(о)’*

Примерна употреба: *Пациентът е в принудителна кома.*

дума: този

Избрано значение:

Синонимно множество: {тоя:5, този:5}

Идентификационен номер: bg-0000809L

Дефиниция: *‘показателно местоимение прилагателно (поясняващо именна фраза) за посочване на лица, предмети, ситуации, явления, които се споменават, стават обект на разговор, обсъждане или посочване в настоящия контекст’*

Релации:

[тематична област: показателно местоимение:1]

Примерна употреба: *Чуй тази песен.*

Бележка към синонимното множество: translation equivalent: *this*

Останалите значения в БулНет:

Синонимно множество: {*тоя:3, този:3*}

Идентификационен номер: bg-0000703L

Дефиниция: *‘показателно местоимение прилагателно (поясняващо именна фраза) за посочване на по-рано споменати или известни лица, предмети, ситуации, явления’*

Примерна употреба: *Това влудяващо бумтене и тропане продължаваше вече дни.*

Синонимно множество: {*онзи:3, оня:4, този:4, тоя:4*}

Идентификационен номер: bg-0000790L

Дефиниция: *‘показателно местоимение прилагателно (поясняващо именна фраза) за изтъкване на нещо характерно, добре познато или разпознаваемо от главното изречение, което се появява от подчинено определително изречение’*

Примерна употреба: *Той я погледна с онази усмивка, която караше сърцето ѝ да се разтапя.*

Синонимно множество: {*тоя:2, този:2*}

Идентификационен номер: bg-0000288L

Дефиниция: *‘показателно местоимение съществително (заместващо именна фраза) за посочване на известни или по-рано споменати лица, често пренебрежително’*

Примерна употреба: *Влиза, разбираш ли, тази, и иска 5 килограма сирене!*

Синонимно множество: {*тоя:1, този:1*}

Идентификационен номер: BUL-1094477356

Дефиниция: *‘показателно местоимение прилагателно (поясняващо именна фраза) за посочване на пространствено близки лица, предмети, ситуации, явления’*

Примерна употреба: *Примерна употреба: Чие е това дете?*

дума: миг

Избрано значение:

Синонимно множество: {*момент:5, миг:2, секунда:3, минута:4, час:5*}

Идентификационен номер: ENG20-14383479-n

Дефиниция: *‘конкретен момент, в който нещо се случва или наблюдава’*

Релации:

[хипоним: време:3]

[хипероним: момент:3]

Примерна употреба: *В мига, когато пристигна, купонът започна.*

Примерна употреба: *Дойде часът да се приеме суровата истина.*

Примерна употреба: *В тази секунда влезе приставът.*

Останалите значения в БулНет:

Синонимно множество: {момент:1, минута:2, миг:1, секунда:2}

Идентификационен номер: ENG20-14385646-n

Дефиниция: 'много кратък, субективно определян, период от време (не повече от няколко секунди)'

Примерна употреба: *Ще отнеме само секунда.*

дума: отвън

Избрано значение:

Синонимно множество: {навън:1, отвън:1, вън:1}

Идентификационен номер: ENG20-00112527-b

Дефиниция: 'означава, че посоченото в изказването се реализира в пространство, което е извън стените или пределите на дадено помещение или затворено място'

Релации:

[антоним: вътре:2, на закрито:1]

[тематична област: място:4]

Примерна употреба: *През лятото играем навън.*

Останалите значения в БулНет:

Синонимно множество: {отвън:2}

Идентификационен номер: ENG20-00112787-b

Дефиниция: 'означава, че посоченото в изказването е извън пределите, границите или повърхността на нещо'

Примерна употреба: *Отвън кутията беше черна.*

дума: се чу

Избрано значение:

Синонимно множество: {издавам звук:2, издам звук:2, издавам шум:1, издам шум:1, звуча:5, прозвучавам:4, прозвуча:4, раздавам се:1, раздам се:1, чувам се:2, чуя се:2, дочувам се:1, дочуя се:1}

Бележка към литерала *издавам звук:2*: неск. в.

Бележка към литерала *издам звук:2*: св. в.

Бележка към литерала *издавам шум:1*: неск. в.

Бележка към литерала *издам звук:1*: св. в.

Бележка към литерала *звуча:1*: имперф. т.

Бележка към литерала *прозвучавам:4*: неск. в.

Бележка към литерала *прозвуча:4*: св. в.

Бележка към литерала *раздавам се:1*: неск. в.

Бележка към литерала *раздам се:1*: св. в.

Бележка към литерала *чувам се:2*: неск. в.

Бележка към литерала *чуя се:2*: св. в.

Бележка към литерала *дочувам се:1*: неск. в.

Бележка към литерала *дочуя се:1*: св. в.

Идентификационен номер: ENG20-02113837-v

Дефиниция: 'ставам възприемаем чрез звука или шума, който произвеждам'

Релации:

[производен: звук:3, шум:4]

[производен: звук:4, звучене:2]

[производен: звук:5]

[произхождащ: произвеждам звук:1, произвежда звук:1, изваждам звук:1]

[хипоним: свиря:3, изсвирвам:1, изсвиря:1]

[хипоним: дрънкам:1, дрънча:1, звъня:1, звънтя:1]

[хипоним: отеквам:2, отекна:2, резонирам:2]

[хипоним: звънтя:2, звъня:2, позвъня:1, позвънявам:1]

[хипоним: бибиткам:1, свиря:4]

[хипоним: свиря:5, изсвирвам:2, изсвиря:2, свистя:1, изсвистявам:1, изсвистя:1]

[хипоним: шумя:1, про шумявам:1, про шумя:1]

[хипоним: почуквам:1, почукам:1, потропвам:1, потропам:1, барабаня:1]

[хипоним: ехтя:1, еча:1, проехтявам:1, проехтя:1, кънтя:1, прокънтявам:1, прокънтя:1, отеквам:3, отекна:3]

[хипоним: бръмча:1, фуча:3, свистя:2, жужа:1]

[хипоним: тракам:1, потраквам:1, потракам:1]

[хипоним: прекъсвам:1, прекъсна:1, давя се:2, задавам се:2, задавя се:2, кашлям:1, закашлям се:1]

[хипоним: потраквам:2, потракам:2, потропвам:2, потропам:2]

[хипоним: бия:4, тупам:2, думкам:1, барабаня:2]

[хипоним: гукам:2, загуквам:1, загукам:1]

[хипоним: изсвирвам:3, изсвиря:3]

[хипоним: писукам:1, изписувам:1, изписукам:1]

[хипоним: тиктакам:1]

[хипоним: ромоля:1, ромоня:1, бълбукам:1, бълбя:1]

[хипоним: шумоля:1, про шумолявам:1, про шумоля:1, шушина:2]

[хипоним: разцепя:2, разцепвам:2]

[хипоним: идвам:15, дойда:8, ида:5]

[хипероним: възприемам се:1, осезаем съм:1]

Примерна употреба: Тя издаде звук на задоволство.

Примерна употреба: Оръжието издаде звук „бум“.

Бележка към синонимното множество: често значението на „издавам звук/ издавам шум“ се конкретизира от съгласувано или несъгласувано определение Бележка към синонимното множество: „звуча“, „прозвучавам“, „раздавам се“, „чувам се“, „дочувам се“ и съответстващите им глаголи от св. в. изискват като субект самия звук или шум, а източникът му се реализира като обстоятелствено пояснение: „От пушката се раздаде „бум“.“

Останалите значения в БулНет:

Синонимно множество: {звуча:4, чувам се:1, чуя се:1}

Бележка към литерала звуча:4: имперф. т.

Бележка към литерала чувам се:2: несв. в.

Бележка към литерала чуя се:1: св. в.

Идентификационен номер: ENG20-02073211-v

Дефиниция: *издавам определен звук или звуци или предизвиква някакво слухово впечатление.*

Примерна употреба: Чува се слабо, но чудовището го възприема.

дума: някакво

Избрано значение:

Синонимно множество: {някакъв:2, някой:3, един:12}

Идентификационен номер: ENG20-01056112-а

Дефиниция: *‘неопределително местоимение прилагателно (поясняващо именна фраза или самостойно употребено), с което се създават неизвестни, непознати лица, предмети, ситуации, явления или такива, които не притежават съществени отличителни признаци’*

Релации:

[подобен на: неустановен:3, неуточнен:1]

[тематична област: неопределително местоимение:1]

Примерна употреба: *Някакъв човек те търси по телефона.*

Примерна употреба: *Въпреки всичко повечето момичета се надяват един ден принцът да се появи на бял кон.*

Примерна употреба: *Някоя вечер трябва да им идем на гости.*

Останалите значения в БулНет:

Синонимно множество: {някакъв:3, известен:6, определен:1, даден:3}

Идентификационен номер: ENG20-00671963-а

Дефиниция: *‘чието значение, съдържание, брой, обхват и т.н. са установени в някакви граници, познати са или могат да се предположат, без да са строго ограничени’*

Примерна употреба: *Заделям известна сума всяка седмица.*

Синонимно множество: {няколко:2, някакъв:1, някой:2, малко:6}

Идентификационен номер: ENG20-02185452-а

Дефиниция: *‘квантификатор – неопределително местоимение за количество, използвано с бройни и небройни съществителни за обозначаване на неопределено или неизвестно, но неголямо количество, брой’*

Примерна употреба: *Има няколко приятели.*

дума: свирене

Избрано значение:

Синонимно множество: {свирене:1, свистене:1, свирка:4}

Идентификационен номер: ENG20-06946820-п

Дефиниция: *‘звук, издаван от нещо движещо се бързо или от пара, излизаща през тесен отвор’*

Релации:

[производен: свиря:5, изсвирвам:2, изсвиря:2, свистя:1, изсвистявам:1, изсвистя:1]

[хипероним: звук:3, шум:4]

Изречение: Сякаш някой локомотив беше дерайлирал в съседната кръчма.

Дума: сякаш

Chooser: /home/cvetana/chooser/p51.txt

File Pass Word Stats End

<Clipping="II_P_Az_maniaka.txt"> Момчето четеше на глас. Хемингуей не беше преценил тая работа. В този миг откъм се чу някакво свирене. Сякаш някой локомотив беше дерайлирал в съседната кръчма. Руцунка се изправи и отиде до прозореца. Роксиската беше полетяла по облите й форми и тя небрежно я поправи. Отсрещна, точно под прозореца на третия етаж, се мъдреше един надпис, изписан с бяла боя в неопределен шват. Девойката се вторачи като кясоллед орел и някъде дълбоко в съзнанието й нещо избухна и се разкъса на милиони парчета. Ей там, на някакви си 30 метра, върху подалната от подините излизка на блока, с огромни цветни букви пишеше: Обичам те. - Ало! Добър ден.

</Clipping>

<Clipping="II_P_Baven_tanc.txt"> Там имаше някто само за ледло, хладилник и - нервата гордост и радост - старо пиано "Стайнуей", на което се упражняваше денонощно през времето, когато не беше в колежа. Една вечер към края на март, след особено тежък ден, Миди пристигна в квартирата на Себастиан. Щон й отвори вратата, отвърне я лъхна вриризната на застоял цигарен дим. - Здравей, миди. Визвай и слагай чайника. Миди

main tree synset

{сякаш}; като че ли:1 {като че ли}; като че:1 {като че}; все едно:1 {все едно}; вс

Synset: bg - conj: сякаш:1 {сякаш}; като че ли:1 {като че ли}; като че:1 {

POS [conj] ILLI [BUL-309394915] BCS [6]

literal сякаш:1 {сякаш}

literal като че ли:1 {като че ли}

literal като че:1 {като че}

literal все едно:1 {все едно}

literal все едно че:1 {все едно че}

Definition: подчинителен съюз за въвеждане на подчини еми обстоятелствени изречения за начин и сравнение, изразяващи предположение, не сигурност или приливо оприличаване

1 изразява предположение или констатация относно това, как изглежда или какво е привидното впечатление от посоченото в изказването/н/н/н

2 подчинителен съюз за въвеждане на подчинени обстоятелствени изречения за начин и сравнение, изразяващи предположение, несигурност или привид

дума: някой

Chooser: /home/cvetana/chooser/p51.txt

File Pass Word Stats End

<Clipping="II_P_Az_maniaka.txt"> Момчето четеше на глас. Хемингуей не беше преценил тая работа. В този миг откъм се чу някакво свирене. Сякаш някой локомотив беше дерайлирал в съседната кръчма. Руцунка се изправи и отиде до прозореца. Роксиската беше полетяла по облите й форми и тя небрежно я поправи. Отсрещна, точно под прозореца на третия етаж, се мъдреше един надпис, изписан с бяла боя в неопределен шват. Девойката се вторачи като кясоллед орел и някъде дълбоко в съзнанието й нещо избухна и се разкъса на милиони парчета. Ей там, на някакви си 30 метра, върху подалната от подините излизка на блока, с огромни цветни букви пишеше: Обичам те. - Ало! Добър ден.

</Clipping>

<Clipping="II_P_Baven_tanc.txt"> Там имаше някто само за ледло, хладилник и - нервата гордост и радост - старо пиано "Стайнуей", на което се упражняваше денонощно през времето, когато не беше в колежа. Една вечер към края на март, след особено тежък ден, Миди пристигна в квартирата на Себастиан. Щон й отвори вратата, отвърне я лъхна вриризната на застоял цигарен дим. - Здравей, миди. Визвай и слагай чайника. Миди

main tree synset

Synset: bg - a: някакъв:2 {някакъв}; някой:3 {някой}; един:12 {един}

Synset: bg - a: някакъв:2 {някакъв}; някой:3 {някой}; един:12 {един}

POS [a] ILLI [ENG20-01056112-a] BCS [4]

literal някакъв:2 {някакъв}

literal някой:3 {някой}

literal един:12 {един}

Definition: неопределително местоимение прилагателно (поясняващо именна фраза или самостоятелно употребено), с което се определят неизвестни, непознати лица, предмети, ситуации

similar: to bg - a: неустановен:3 {неустановен}; неуточнен:1 {неуточн

category_domain bg - n: Неопределително/ местоимение:1 {неопред

1 неопределително местоимение прилагателно (поясняващо именна фраза или самостоятелно употребено), с което се определят неизвестни, непознати лица

2 неопределително местоимение съществително (заместващо именна фраза), с което се означават неизвестни лица

3 неопределително местоимение прилагателно (поясняващо именната фраза или самостоятелно употребено) за обозначаване на един или повече обекти, пре

ДУМА: ЛОКОМОТИВ

Chooser: /home/cvetana/chooser/p51.txt

File Pass Word Stats End

<Clipping="II_P_Az_maniaka.txt"> Момчето четеше на глас. Хемингуей не беше преценил тая работа. В този миг откъм се чу някакво свирене. Сакаш някой локомотив беше дерайлирал в съседната кърчина. Рукина се изправи и отиде до прозореца. Роклицката беше поленянала по облите ѝ форми и тя небрежно я пооправи. Отсреща, точно под прозореца на третия етаж, се мъдреше един надпис, изписан с бяла боя в неопределен шат. Девойката се вторачи като късоплед орел и някъде дълбоко в съзнанието ѝ нещо избухна и се разкъса на милиони парчета. Ей там, на някакви си 30 метра, върху поапената от родните мазица на блока, с огромни цветни букви пишеше: Обичам те. - Ало! Добър ден.

</Clipping>

<Clipping="II_P_Baven_tanc.txt"> Там имаше място само за легло, хладилник и - неговата гордост и радост - старо пиано "Стайнуей", на което се упражняваше денонощно през времето, когато не беше в компания. Една вечер там край на март, след особено тежък ден, Мали пристигна в квартирата на Себастиан. Щой ѝ отвори вратата, отвяре я лъмна вирината на застоял цигарен дим. - Здравей, мила. Визвай и слагай чайника. Мали

main tree synset

Synset: bg - n: локомотив:1 (локомотив)

□ Synset: bg - n: локомотив:1 (локомотив)

— POS n |L| ENG20-03547471-n BCS 2

□ literal локомотив:1 (локомотив)

— Definition самоходна транспортна машина на колела, която се движи по релсов път, предназначена ена за теглене на влакова композиция

□ holo_member bg - n: влак:1 (влак); композиция:3 (композиция); /влак:

□ hyponym bg - n: /парен/ локомотив:1 (парен локомотив)

□ hyponym bg - n: /самоходна/ машина:1 (самоходна машина); /самоход:

1 самоходна транспортна машина на колела, която се движи по релсов път, предназначена за теглене на влакова композиция

ДУМА: БЕШЕ

Chooser: /home/cvetana/chooser/p51.txt

File Pass Word Stats End

<Clipping="II_P_Az_maniaka.txt"> Момчето четеше на глас. Хемингуей не беше преценил тая работа. В този миг откъм се чу някакво свирене. Сакаш някой локомотив беше дерайлирал в съседната кърчина. Рукина се изправи и отиде до прозореца. Роклицката беше поленянала по облите ѝ форми и тя небрежно я пооправи. Отсреща, точно под прозореца на третия етаж, се мъдреше един надпис, изписан с бяла боя в неопределен шат. Девойката се вторачи като късоплед орел и някъде дълбоко в съзнанието ѝ нещо избухна и се разкъса на милиони парчета. Ей там, на някакви си 30 метра, върху поапената от родните мазица на блока, с огромни цветни букви пишеше: Обичам те. - Ало! Добър ден.

</Clipping>

<Clipping="II_P_Baven_tanc.txt"> Там имаше място само за легло, хладилник и - неговата гордост и радост - старо пиано "Стайнуей", на което се упражняваше денонощно през времето,

main tree synset

Synset: bg - v: съм:1 (съм)

□ Synset: bg - v: съм:1 (съм)

— POS v |L| BUL-174320789 BCS 6

□ literal съм:1 (съм)

— Definition спомогателен глагол, който е носител на граматичното значение в речките на анализните словоформи

□ category_domain bg - n: /спомогателен/ глагол:1 (спомогателен глагол)

□ usage Образът на автора бива пресъздаван с лекота.

1 спомогателен глагол, който е носител на граматичното значение в рамките на анализните словоформи

2 имам определено качество, характеристика, характеризирам се с качество, характеристика, белег

3 идентичен съм с някого или нещо друго, изразено със сказуемото определение

4 съществувам или в даден момент присъствам на определено място в пространството

5 копулативен глагол, с който фразата или изречението, съдържащи описание, оценка и под., се приписват на субекта'n/n/n/n'

6 еквивалентен или равнозначен съм на нещо или някого или имам определен статут, качества, особености'n

7 оказвам се, пребивавам (продължително) в определено положение или ситуация'n

8 представлявам съставна част от света, имам реално битие

9 работя на определено място с определена длъжност или имам определени функции

a намирам се в определено отношение спрямо други хора, неща или събития, обикновено включващо зависимост от тях или наличие на отговорност от

дума: дерайлирал

Chooser: /home/cvetana/chooser/p51.txt

File Pass Word Stats End

```

<Clipping="II.P_Az_maniaka.txt"> Момчето четеше на глас. Хемингуей не беше преценил тая работа. В този миг откъм се чу никакво смирено. Сякаш някой локомотив беше дерайлирал в съседната кръчма. Ружина се изправи и отиде до прозореца. Рокличката беше поленявала по облите ѝ форми и тя небрежно пооправи. Отстрема, точно под прозореца на третия етаж, се мъдреше един надпис, изписан с блясна боя в неопределен шрифт. Девојката се вторачи като кърселед орел и никакво дълбоко в съзнанието ѝ нешто избухна и се разкъса на милиони парчета. Ей там, на никакви си 30 метра, върху попанелата от горните мазилка на блока, с огромни цветни букви пишеше: Обичан те. - Ало! Добър ден.

</Clipping>

<Clipping="II.P_Baven_tanc.txt"> Там изаше няшто само за лепо, илашкинџи и - илювава гордост и радост - старо пиано "Сејкуей", на което се упражняваше ленивоносно през времето, когато не беше в колежа. Една вечер към края на март, след особено тежок ден, Нади пристигна в квартирата на Себастиан. Дон Ё отвори вратата, отвори ѝ лъсна вризмата на застоял цигарен дим - Здравей, мила. Влизай и спалай чайовка. Нади

1 напускан релсите по време на движение (за превозно средство)\n
  
```

main tree synset

гирам:1 (дерайлирам); излизам от релси:1 (излизам от релси); изляза от релси

- Synset: bg - v: дерайлирам:1 (дерайлирам); излизам от релси:1 (излизам от релси)
 - POS v [v] |L| [ENG20-01953370-v] BCS 0
 - literal дерайлирам:1 (дерайлирам)
 - literal излизам от релси:1 (излизам от релси)
 - literal изляза от релси:1 (изляза от релси)
 - Definition напускан релсите по време на движение (за превозно средство)
 - hypermym bg - v: вървя:9 (вървя); отивам:6 (отивам); отида:5 (отида);

дума: В

Chooser: /home/cvetana/chooser/p51.txt

File Pass Word Stats End

петите от собственика на клуба, подмигва и измъкна леквар, с почти десет иллади долара в джобовете на спортинга.

```

</Clipping>

<Clipping="II.P_Az_maniaka.txt"> Момчето четеше на глас. Хемингуей не беше преценил тая работа. В този миг откъм се чу никакво смирено. Сякаш някой локомотив беше дерайлирал в съседната кръчма. Ружина се изправи и отиде до прозореца. Рокличката беше поленявала по облите ѝ форми и тя небрежно пооправи. Отстрема, точно под прозореца на третия етаж, се мъдреше един надпис, изписан с блясна боя в неопределен шрифт. Девојката се вторачи като кърселед орел и никакво дълбоко в съзнанието ѝ нешто избухна и се разкъса на милиони парчета. Ей там, на никакви си 30 метра, върху попанелата от горните мазилка на блока, с огромни цветни букви пишеше: Обичан те. - Ало! Добър ден.

</Clipping>

<Clipping="II.P_Baven_tanc.txt"> Там изаше няшто само за лепо, илашкинџи и - илювава гордост и радост - старо пиано
  
```

main tree synset

Synset: bg - p: o:1 {o}; v:2 {v}; v&v:2 {v&v}

- Synset: bg - p: o:1 {o}; v:2 {v}; v&v:2 {v&v}
 - POS p [p] |L| [BUL-233974173] BCS 6
 - literal o:1 {o}
 - literal v:2 {v}
 - literal v&v:2 {v&v}
 - Definition предлог, означаващ място на повърхността контакт, допир
 - category_domain bg - n: предлог:1 (предлог)
 - usage Пъшката изтрака о камъняка.

r предлог за посочване на някого или нещо, за които е присъщо, характерно определено качество, черта, факт\и\n

s предлог за посочване на определена линия, курс на действие или поведение, които се следват\и\n

t предлог, означаващ място на повърхността контакт, допир\и\n

u предлог за въвеждане на темпо на движение\и\n

v предлог, изразяващ асоциация, свързване на образ, представа и под. с даден обект (предложното допълнение)\и\n

ДУМА: СЪСЕДНАТА

Chooser: /home/cvetana/chooser/p51.txt

File Pass Word Stats End

петите от собственика на клуба, полцията и измания лекар, с почти десет хиляди долара в джобовете на смокингa.

</Clipping>

<Clipping="II_P_Az_maniaka.txt"> Новичето четеше на глас. Хемингуей не беше преценил тая работа. В този миг откъм се чу някакво ширене. Сякаш някой локомотив беше дерайширал в съседната кръчма. Руфина се изправи и отиде до прозорца. Рокличката беше полетнала по облите й форми и тя небрежно я поправи. Отсреща, точно под прозорца на третия етаж, се издърже един нашин, изписан с бяла боя в неопределен шат. Девойката се вторачи като кясолед орел и наведе дълбоко в съзнанието й нещо избухна и се разкъса на милиони парчета. Ей там, на някакви си 30 метра, върху пощадената от горните назишка на блока, с огромни цветни букви пишеше: Обичам те. - Ало! Добър ден.

</Clipping>

<Clipping="II_P_Baven_tanc.txt"> Там имаше място само за нещо, ухажване и - човека попроси и влезе - сякаш плавно

main tree synset

Synset: bg - a: съседен:1 (съседен)

- Synset: bg - a: съседен:1 (съседен)
 - POS [a] [И] [ENG20-00421306-a BCS 4]
 - literal съседен:1 (съседен)
 - който се намира в непосредствена близост до някого или нещо, без място помежду им
 - Definition [И]
 - similar_to bg - a: близък:5 (близък; недалечен):1 (далечен)
 - usage в съседната стая
 - usage Бяхме в съседни стаи.

1 който се намира в непосредствена близост до някого или нещо, без място помежду им

2 който се намира, съществува близо до друг, особено от същия вид или подобен на него

ДУМА: КРЪЧМА

Chooser: /home/cvetana/chooser/p51.txt

File Pass Word Stats End

петите от собственика на клуба, полцията и измания лекар, с почти десет хиляди долара в джобовете на смокингa.

</Clipping>

<Clipping="II_P_Az_maniaka.txt"> Новичето четеше на глас. Хемингуей не беше преценил тая работа. В този миг откъм се чу някакво ширене. Сякаш някой локомотив беше дерайширал в съседната кръчма. Руфина се изправи и отиде до прозорца. Рокличката беше полетнала по облите й форми и тя небрежно я поправи. Отсреща, точно под прозорца на третия етаж, се издърже един нашин, изписан с бяла боя в неопределен шат. Девойката се вторачи като кясолед орел и наведе дълбоко в съзнанието й нещо избухна и се разкъса на милиони парчета. Ей там, на някакви си 30 метра, върху пощадената от горните назишка на блока, с огромни цветни букви пишеше: Обичам те. - Ало! Добър ден.

</Clipping>

<Clipping="II_P_Baven_tanc.txt"> Там имаше място само за нещо, ухажване и - човека попроси и влезе - сякаш плавно

main tree synset

Synset: bg - n: кръчма:1 (кръчма); пивница:1 (пивница)

- Synset: bg - n: кръчма:1 (кръчма); пивница:1 (пивница)
 - POS [n] [И] [ENG20-03686529-n BCS 3]
 - literal кръчма:1 (кръчма)
 - literal пивница:1 (пивница)
 - евтино питейно заведение с маси за сядане, в което се продават и сервират алкохолни напитки и лека храна (мезета)
 - Definition [И]
 - region_domain bg - n: Обединено кралство:1 {Обединено кралство}; Обединено кралство:1 {Обединено кралство}; Обединено кралство:1 {Обединено кралство}
 - hypertym bg - n: литеино/ заведение:1 (питейно заведение)
 - mero_part bg - n: бар:2 {бар}

1 евтино питейно заведение с маси за сядане, в което се продават и сервират алкохолни напитки и лека храна (мезета)

Семантично анотиран корпус на българския език

редактор и съставител, доц. д-р Светла Пенева Коева

Светла Пенева Коева, Борислав Красимиров Ризов, Екатерина Николова Търтоманова, Мария Андреева Тодорова, Светлозара Илиева Лесева, Христина Николова Кукова, Цветана Иванова Димитрова

Рецензенти: проф. д-рн Владко Данаилов Мурдаров, доц. д-р Гергана Кирилова Дачева

Технически редактор и графичен дизайн Ангелина Владиславова Аврамова

Художник на корицата Божидар Тодоров Чемиширов

Българска. Първо издание, 2010 г. Формат: 60/90/16, Печатни коли: 13