

PARSAREA DICȚIONARULUI LIMBII RUSE LITERARE CONTEMPORANE PRIN METODA BAZATĂ PE CONFIGURAȚII SEGMENTARE-COEZIUNE-DEPENDENȚĂ

Eugenia BURCA

Institutul de Matematică și Informatică al AȘM

The basic idea of this paper is to transform an entry that contains non-processed text, into index linguistic resource, using a parsing method of Segmentation-Cohesion-Dependency (SCD) for the Contemporary Russian Literary Language Dictionary. The first SCD configurations are applied on a dictionary entry to extract the tree of its senses/subsenses and the second one are used to parse its semantic content of sense definitions.

Introducere

Orice aplicație ce ține de procesarea limbajului natural are nevoie, de regulă, de un lexicon de cuvinte, care conține informație morfologică, sintactică și semantică. Pentru a obține informația semantică, este necesar a prelucra dicționarele explicative existente. Ideea generală ce stă la baza acestei lucrări constă în elaborarea unui mecanism de transformare a unui dicționar, pentru a-l obține în format electronic sub formă de resursă lingvistică indexabilă. În acest scop, este necesar ca pentru toate intrările din dicționar să se obțină arbori de sensuri și de definiții. Arborii de sensuri pentru fiecare intrare din dicționar identifică hotarele pentru diferite categorii lingvistice specifice. La rândul lor, acești arbori sunt divizați în sensuri principale și în sensuri secundare, respectându-se o ierarhie introdusă de marcherii de sensuri. Arborii de definiții stabilesc dependențele dintre clasele de marcheri de definiții. Pentru a obține arborele de sensuri, adică ierarhia sensurilor desemnate de o intrare din dicționar, a fost utilizat un parser cu configurații de tip SCD (Segmentare-Coeziune-Dependență – Segmentation-Cohesion-Dependency) [1,2]. Parsarea unui dicționar constă în recunoașterea marcherilor de sens, în identificarea structurilor dintre doi marcheri și în clasificarea acestor structuri în conformitate cu ierarhia prestabilită a claselor de marcheri de sens.

Vom construi arborii de sensuri și de definiții cercetând marcherii de sensuri și de definiții din intrările Dicționarului Limbii Ruse Literare Contemporane (DLRLC) [3]. Intrările DLRLC sunt alcătuite din mai multe părți componente, fiecare din ele furnizează diferite tipuri de informații despre intrarea respectivă. Aceste intrări vor fi studiate utilizând noțiunea de configurație SCD.

Conform [4], o configurație SCD are următoarele componente:

- ✓ un set de clase de marcheri de sens ce identifică hotarele pentru diferite categorii lingvistice specifice;
- ✓ o ierarhie de tip arbore care stabilește dependențele dintre clasele de marcheri;
- ✓ un algoritm de parsare.

Componentele intrărilor din DLRLC sunt:

- ✓ informația despre forma cuvântului respectiv (ortografie, pronunție etc.);
- ✓ informația gramaticală (partea de vorbire, subcategorizare gramaticală etc.);
- ✓ definițiile sensurilor cuvântului de bază;
- ✓ exemple de utilizare a cuvântului de bază: expresii-frazeologice sau citate din diferite surse literare;
- ✓ definițiile subsensurilor cuvântului-intrare utilizat în expresiile exemple;
- ✓ referințe către alte intrări;
- ✓ etimologia cuvântului de bază etc.

1. Clasele marcherilor de sens din DLRLC și structura lor de dependență

Unii marcheri de sens depistați pentru intrările din DLRLC coincid cu cei pentru intrările din Dicționarul Tezaur al Limbii Române (DTLR) [4]. Ei își păstrează forma și, parțial, funcțiile, dar apar și marcheri diferiți cu funcții noi. Acești marcheri formează un set de clase de marcheri (marcheri de sens) ce identifică hotarele pentru diferite categorii lingvistice specifice.

Inițial vom evidenția două categorii ale marcherilor de sens: principali și secundari. Marcherii de sens principali în intrările DLRLC sunt următorii: *cifre romane urmate de punct (I, II, III,...etc.)*, *cifre arabe urmate de punct (1., 2., 3,... etc.)*, *două bare oblice (//)*. În calitate de marcheri de sens secundari servesc: *rombul gol (◇)*, *traversa (⇒)*, *tilda (~)*, *enumerare literală (a), б), в),... etc.)*.

Vom analiza în continuare poziția ierarhică a marcherilor enumerați. Vom începe cu cei principali.

➤ Marcherii de tip *cifre romane urmate de punct (I., II., III.,...etc.)* reprezintă *nivelul superior* al ierarhiei de sensuri pentru orice intrare de dicționar. Acești marcheri desemnează cele mai *generale sensuri* ale cuvântului. (De menționat că acesta este echivalentul marcherului de tip *majuscule A., B., etc.* din DTLR [4]). De exemplu:

БЫТЬ, наст. не употр. кроме 3л. ед. есть и (устар.) 3л. мн. суть, буд. буду, будешь, прош. был, была, было (с отрицанием: не был, не была, не было), повел. будь(те), прич. действ. прош. бывший, деенр. будучи, несов., неперех. **I.** Как самостоятельный глагол означает: **1.** Существовать. *He говори тоской: их нет; Но с благодарностью: были.* Жук. Воспоминание.... **II.** В знач. вспомогательного глагола или связки. **1.** Употр. в составном именном сказуемом (в настоящем времени обычно опускается). *В то время был еще жених Ее супруг.* Пушкин. Е. О....

Marcherul de tip *cifre romane urmate de punct* nu este reprezentat în mod explicit, dacă acest nivel are doar un element de acest tip.

➤ Marcherii de tip *cifre arabe urmate de punct (1., 2., 3.,... etc.)* reprezintă *al doilea nivel* de sens. Acest nivel este înglobat de un marcher de tip *cifră romană*, dacă acesta există. În lipsa *cifrei romane*, marcherul de tip *cifră arabă urmată de punct* apare pe nivelul cel mai de sus al ierarhiei de sensuri. (De menționat că marcherul specificat mai sus este echivalentul marcherului de tip *cifre romane I., II., etc.* din DTLR [4]). De exemplu:

АБАК, а, м., **АБАКА**, и, ж. **1.** В архитектуре – плита над основной частью капители. *В дорическом ордере Греции колонны не имеют баз, капитель состоит из круглой подушки – эхина и прикрывающей его квадратной плиты – абаки.* Бартенев, От пирамид до современных зданий. **2.** Только м. Доска для арифметических вычислений (первоначально у древних греков, римлян).

Marcherul nu este reprezentat în mod explicit, dacă intrarea lexicală are doar un sens pentru acest nivel al analizei.

➤ Marcherul *două bare oblice (//)* specifică *al treilea nivel* de sens și este folosit pentru a evidenția nuanțele de *sens principal* ale unei intrări din dicționar. Acest nivel este înglobat de un marcher de tip *cifră arabă* sau *cifră romană*, dacă acestea există. În lipsa acestora, marcherul *două bare oblice* apare pe nivelul cel mai de sus al ierarhiei de sensuri. Marcherul este reprezentat în mod explicit, chiar dacă acest nivel are doar un element de acest tip. De exemplu:

АБРИКОСОВЫЙ, ая, ое. **1.** Относящийся к абрикосу, абрикосам (в 1 знач.). ◊ Абрикосовое дерево. То же, что абрикос. // Состоящий из абрикосов. *Абрикосовый сад.*

2. Относящийся к абрикосу, абрикосам (во 2 знач.) *Абрикосовая косточка.* // Приготовленный из абрикосов, с абрикосами. *Абрикосовый сироп. Абрикосовое варенье.*

ВЕДУЩИЙ, ая, ее. **1.** Идущий впереди; головной. *Ведущий самолет.* = *Каждый из ведущих броненосцев больше всего осыпался неприятельскими снарядами.* Нов.-Прибой, Цусима. // В знач. суц. Ведущий, его, м, Ведущая, ей, ж. **а)** Тот, кто ведет, возглавляя какую-л. группу. *В тайге заблудиться легко, если к тому же окажется самонадеянным и не очень опытным ведущим.* Ворон. Волев. прием. *Последнее время Драченко ходит у нас в качестве разведчика. А теперь думаем посылать его ведущим.* Кудреватых, Стр. нашей жизни. **б)** Летчик, летящий на головном самолете, направляющий действия своего ведомого. *За те немногие минуты, что они провели в воздухе, Петров сумел оценить уверенную и поистине мастерскую манеру полета своего ведущего.* Б. Полев. Пов. о наст. чел. ...

Marcherii de *sens principal* în intrările **DLRLC** pot fi succedați de următorii marcheri de *sens secundar*:

➤ Marcherul *romb gol (◊)* specifică *al patrulea nivel* de sens și este folosit la enumerarea *sensurilor secundare* ale unei intrări din DLRLC. Marcherul este utilizat:

(1) la evidențierea compatibilității lexicale a „expresiilor frazeologice”. De exemplu:

АБИССИНСКИЙ, ая, ое. Относящ. к абиссинцам, принадлежащий, свойственный им. ◊ Абиссинское нагорье. *Устар.* Эфиопское нагорье.

(2) la indicarea caracteristicilor de utilizare a cuvântului, inclusiv prin comparații. De exemplu:

ВИНОГРАД, а и у, м **1.** ... **2.** Ягоды такого растения....◊ В сравн. [Мы] неизменно возвращались из леса с полными коробами черники, малины .. или крупного, как виноград, гонобобеля. Полторацк. Дарю вам Мещеру.

(3) la indicarea expresivității de utilizare a cuvântului (în aceste cazuri marcherul se plasează înaintea cuvintelor rezervate, de exemplu, *Ирон., Спец., Шутл., Устар.* etc.). De exemplu:

АВГУР, а м. **1.** В древнем Риме – жрец, толкующий волю богов и предсказывающий будущее по крику и полету птиц. *В важнейших случаях вопрос решали куры, Чьих предсказаний образцы Сводили*

в формулы жрецы, Куриной мудрости разгадчики, авгуры. Д. Бедный, Курология. ◇ Устар. Как авгуры (смотря, глядя и т. п. друг на друга, смеяться, улыбаться и т. п.). Выражение связано с тем, что авгуры, по сообщению Цицерона, без улыбки не могли смотреть друг на друга, так как сами не верили в свои гадания. *Взглянув друг на друга, потом, как Цицероновы авгуры, Мы расмеялись тишком.* Пушкин. Е. О. (Вар.)...

Dacă nu există nici un *marcher de sens superior*, *marcherul de tip romb gol* apare pe nivelul cel mai de sus al arborelui de sensuri. *Marcherul de tip romb gol* este reprezentat în mod explicit, chiar dacă acest nivel are doar un element de acest tip. De exemplu:

БРОШЮРНЫЙ, ая, ое. Относящ. к брошюре, брошюрам, связанный с их производством. *Брошюрное шитье.* ◇ Б р ошюрная литература. Устар. Литература, издаваемая в виде брошюр.

➤ *Marcherul traversa* (=) specifică *al cincilea nivel de sens* și este utilizat la:

(1) separarea textului autorului de citata ce urmează. De exemplu:

ВИРТУОЗНЫЙ, ая, ое; зен, зна, зно. Относящ. к виртуозу, свойственный ему. *Виртуозное исполнение. = Посмотреть неповторимую по блеску виртуозную технику юдинских операций стремились не только молодые, но и умудренные опытом хирурги.* И. Грекова, Странная вера доктора Швейдера.

(2) la afișarea anumitelor forme gramaticale ale cuvântului. De exemplu:

БРОСОК, ска, м. 1. Резкий взмах руки (рук), благодаря которому перемещается в воздухе что-л., находившееся в ней (в них). *Граната, разорвавшись при броске, оторвала мальчику правую кисть.* Коптяева, Дружба. = Броском, в знач. нареч. *Правой рукой он [рыбак] брал лежащую на паранете полубесформенную массу осьминога и резким броском кидал её на камни паранета.*

(3) la evidențierea variantelor învechite, depășite, ale intrării din dicționar. De exemplu:

ВИОЛОНЧЕЛЬ, и ж. Смычковый четырехструнный инструмент, средний по регистру и размерам между скрипкой и контрабасом. *Партии альты и виолончели были в руках учителей музыкальной школы.* Федин, Братья. = Устар. Виолончель, я, м. *Мы присутствуем при последних усилиях борьбы виолончеля за свое самостоятельное существование.* Чайков. Третья неделя концертн. сезона.

Acest nivel poate fi înglobat de orice *marcher de sens principal*, dacă acesta există; în caz contrar, *marcherul de tip traversă* apare pe nivelul cel mai de sus al arborelui de sensuri, fiind reprezentat în mod explicit, chiar dacă acest nivel are doar un element de acest tip.

➤ *Marcherul tilda* (~) specifică *al șaselea nivel de sens* și este utilizat înaintea primului aforism sau a expresiei idiomatice, întâlnite într-o intrare din dicționar. Acest nivel poate fi înglobat de orice *marcher de sens principal* sau secundar, dacă acesta există. De exemplu:

БЫТЬ, ...I. ...II. ...5. Разг. Употр. в формах будущего времени в знач. связи настоящего времени. ... ~ Будет и на нашей, моей и т. п. *улице праздник.* См. Праздник. *Будь здоров.* См. Здоровый. *Будь не во гневе; не во гневе будь сказано.* См. Гнев. *Будь спокоен, будьте спокойны.* См. Спокойный. *Будь то..., или...; будет ли то..., или...* Употр. для выражения предположения при перечислении, сопоставлении и т. п. чего-л.

Dacă nu există nici un *marcher de sens principal*, atunci *marcherul de tip tilda* apare pe nivelul superior al arborelui de sensuri. De exemplu:

АВГИЕВЫ. ~ *Авгиевы конюшни* (чего-л.). а) Об очень загрязненном, захламленном месте, помещении. *Письменный стол наш представляет авгиевы конюшни, и только теперь я мог обрести клочок бумаги.* Мусорг. Письмо В.В. Стасову, 31 марта 1972. б) О чем-л. находящемся в крайне запущенном состоянии; о беспорядке, неразберихе где-л. — *Говорят, ревизор энергетически принялся за очистку авгиевых конюшен попечительства над училищем.* Гл. Усп. Бог грехам терпит.

➤ *Marcherii enumerare literală* – de tip *litere chirilice, după care urmează paranteză rotundă (а), б), в),... etc.)* – nu specifică o clasă distinctă de *marcheri de sens*, ci sunt folosiți doar la *enumerarea* unui sub-sens. De exemplu:

БЫТЬ, ... I. ...II. ... 2. Употр. в сложном сказуемом. а) С кратким прилагательным *должен.* — *Наши отношения должны быть такие, какие они всегда были.* Л. Толст. Анна Карен. *Учитель должен быть артист, художник, горячо влюбленный в свое дело.* М. Горький, А. П. Чехов. [Лиза] *спешила угадать мысль, которая должна была быть самой главной.* Федин, Перв. радости. *Я должен был сказать жестко и непреклонно: — Нет, я не согласен.* Кузьмин, Круг царя Соломона. б) С предикативами (можно, надо, нужно и т. п.). *Быть можно дельным человеком И думать о красе ногтей.* Пушкин. Е. О. *Надо было видеть, с какой радостью наш кинооператор выскочил из дома, едва услышав гул машины.* Ефрем. Дорога ветров, ...

Marcherii de acest tip nu reprezintă un nivel anume în ierarhia claselor de marcheri, deoarece aparțin nivelului de sens al părintelui.

➤ Marcherul liniuță (“-”) este utilizat la specificarea etimologiei cuvântului de bază interpretat. De exemplu:

БРОШЮРОВАЛЬНЫЙ, а я, ое. Относящ. к брошированию, предназначенный для него.

– С иным (*устар.*) напис. и произнош.: брошировальный. – ТЭ: брошировальные машины; Ушаков, 1934: брошюровальный.

Analizând astfel subordonarea marcherilor (atât a celor principali, cât și a celor secundari), putem construi un graf, care stabilește ierarhia lor.

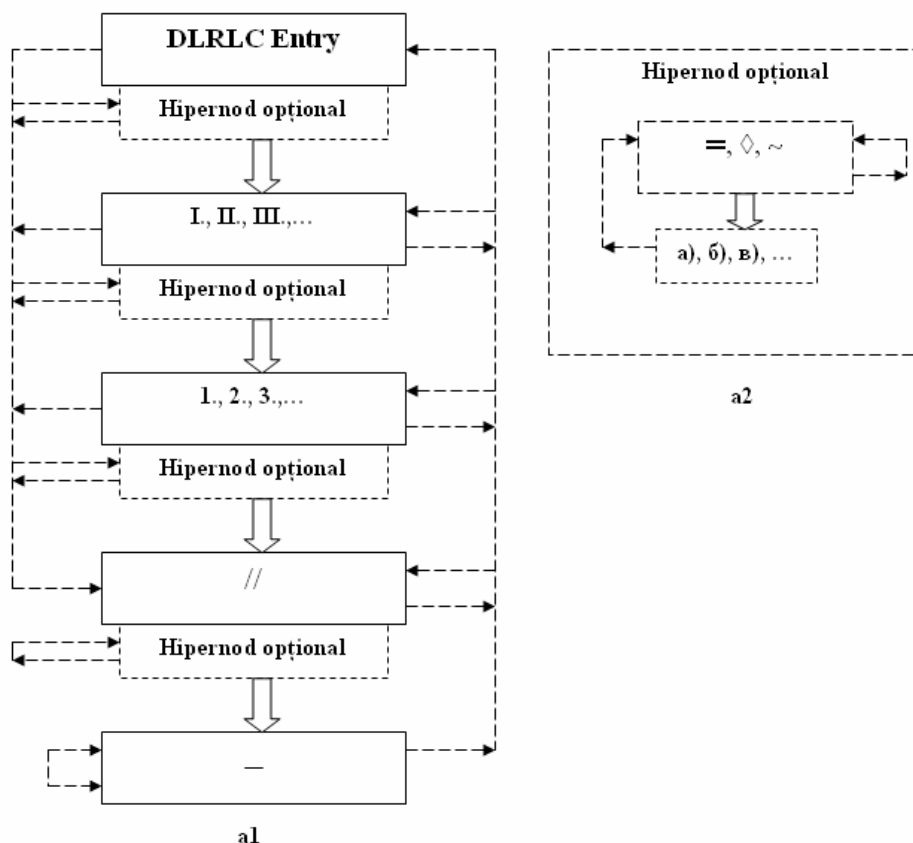


Fig.1. Ierarhia marcherilor de sensuri pentru intrările din DLRLC.

Figura 1 a1 prezintă ierarhia mulțimilor de marcheri de sensuri din DLRLC. Săgețile cu linie întreruptă arată că orice nivel de sens este opțional, iar săgețile continue indică ierarhia mulțimilor de marcheri de sensuri. Figura 1 a2 reprezintă structura hipernodului opțional. Datorită caracteristicilor sale specifice, marcherii *traversa*, *romb gol*, *tilda* și *enumerarea literală* (incluse în hipernodul opțional din Figura 1 a2) sunt reprezentați pe niveluri atașate nivelului asociat în ierarhia marcherilor de sensuri.

Comparând marcherii de sensuri pentru DLRLC și DTLR, putem constata că marcherii *cifre romane urmate de punct* (I., II., III.,...etc.) și *cifre arabe urmate de punct* (1., 2., 3.,... etc.) coincid. Marcherii: *două bare oblice* (//), *traversa* (=) și *tilda* (~) se utilizează numai în DLRLC, iar marcherii: *rombul gol* (◇) și *enumerare literală* (a), б), в),... etc.) își păstrează parțial forma și funcțiile.

2. Marcherii de definiții din DLRLC

La analiza conținutului semantic al sensurilor în DLRLC cu a doua configurație SCD, bazată pe un set de marcheri de definiții (ce desemnează o ierarhie de tip arbore care stabilește dependențele dintre clasele de marcheri), s-a constatat că o serie de marcheri păstrează același mod de prezentare și aceeași semnificație ca și pentru DTLR, dar apar și sensuri noi, specifice pentru DLRLC. Astfel, pentru parsarea DLRLC este necesar a opera cu următorii marcheri de definiții:

- ✓ *MorfDef* – definiții morfologice;
- ✓ *RegDef* – definiții scrise cu font *regular*;
- ✓ *BoldItalDef* – definiții scrise cu *bold - italic*;
- ✓ *BoldRegDef* – definiții scrise cu *bold - regular*;
- ✓ *SpecDef* – definiții ce conțin specificații;
- ✓ *SpSpecDef* – definiții ce conțin anumite specificații, scrise cu litere spațiate;
- ✓ *ExemDef* – exemple la definiții, cu rolul de a întregi înțelesurile unei definiții.

Vom menționa că marcherii *BoldItalDef* și *BoldRegDef* sunt specifici DLRLC și nu sunt prezenți în DTLR.

➤ Definițiile morfologice *MorfDef* descriu categoriile morfologice la diferite niveluri ale arborelui de sensuri. Primul element într-o intrare de dicționar este un *MorfDef*, care specifică toate categoriile morfologice posibile pentru cuvântul intrare. Dacă definiția unui sens nou nu conține un *MorfDef*, acesta se moștenește de la primul sens care are un *MorfDef*.

În continuare, exemplele de definiții din DLRLC sunt evidențiate cu gri:

АВАНС, а, м. 1. Деньги, а также продукты, товары, выдаваемые в счёт предстоящих платежей. *Получать аванс...* 2. *Перен.* О том, что заранее дано или обещано и что необходимо оправдать, подтвердить в будущем,... 3. *Только мн. Перен. Устар.* О знаках внимания, поведении, вселяющих надежды на расположение, симпатию и т.п.

➤ Definițiile *RegDef* (cel mai des întâlnite) reprezintă explicația cuvântului-intrare sau a unor sintagme care îl conțin, reprezentând descrierea standard a sensurilor în majoritatea dicționarilor. De exemplu:

БЫТОПИСАНИЕ, я, ср. 1. *Устар.* Историческое описание, история. *Он рыться не имел охоты В хронологической пыли Бытописания земли.* Пушкин. Е. О.

➤ O definiție de tip *BoldItalDef* este folosită cu scopul de a explica sensul unor aforisme, „expresii frazeologice”. Expresia este scrisă cu litere Bold (îngroșate) – Italic (cursiv), urmată de un *RegDef*. De exemplu:

АВГИЕВЫ. ~ *Авгиевы конюшни* (чего-л.). а) Об очень загрязненном, захлавленном месте, помещении. *Письменный стол наш представляет авгиевы конюшни, и только теперь я мог обрести клочок бумаги.* Мусорг. Письмо В.В. Стасову, 31 марта 1972. б) О чем-л. находящемся в крайне запущенном состоянии; о беспорядке, неразберихе где-л. – *Говорят, ревизор энергически принялся за очистку авгиевых конюшен попечительства над училищем.* Гл. Усп. Бог грехам терпит.

➤ O definiție de tip *BoldRegDef* este folosită cu scopul de a arăta altă categorie morfologică a verbelor. Aceste definiții sunt scrise cu Bold - Regular. De exemplu:

ВИЛЯТЬ, яю, яешь, *несов.*; *вильнуть*,...

➤ Definițiile ce conțin specificații, *SpecDef*, sunt scrise cu font Italic. Multe dintre ele sunt abrevieri: „*Спец.*”, „*Перен.*”, „*Разг.*” etc. sau expresii rezervate scrise între paranteze rotunde care specifică diferite contexte de utilizare ale intrării DLRLC. *SpecDef*-urile sunt folosite la orice nivel în arborele de sensuri. De exemplu:

(1) pe nivelul superior al cuvântului-intrare din dicționar. De exemplu:

БРОШЮРКА, и, род. мн. рок, ж. *Разг.* То же, что брошюра. [В лавочке] *пахло новыми тесовыми полками и лежащими на них новыми книжками и брошюрками.* Бунин, Освобожд. Толстого. *Подарили мне небольшую брошюрку о проблемах образования в Японии.* Ю. Семенов, На «козле» за волком.

(2) pe nivelul unui sens principal. De exemplu:

БЫЧОК, чка, м. 1. *Разг.* Уменьш.-ласк. к бык (1. Бык в 1 знач.); молодой бык. *В стайке у Кузнецовых рос бычок, низколобий, красный, с рожками, похожими на шишки.* Задорн. Амур-Батюшка. 2. *Перен. Разг.* О молодом упрямом человеке (обычно в функции сказуемого). – *Эх ты, бычок несуразный ... грохотал Сиволап шатающемуся Кромину.* Леон. Конец мелк. чел.

(3) pe nivelul unui sens secundar. De exemplu:

БЫЧИЙ, ье, ья. 1. ... 2. Свойственный быку; такой, как у быка. [Попова:] *Если у вас здоровые кулаки и бычье горло, то, думаете, я боюсь вас?* Чех. Медведь. *Человек в шинельке несоображающим бычьим взглядом уставился в лицо Лихарева.* Леон. Конец мелк. чел. ♦ Бычий глаз. *Разг.* Болезненное растяжение и выпячивание глазного яблока. ♦ Бычье сердце. *Разг.* Болезненно увеличенное (в размерах и по массе) сердце.

➤ *SpSpecDef* – definiție ce conține cuvântul-intrare scris cu litere spațiate utilizat în diferite expresii.

O definiție *SpSpecDef* urmează după marcherul de sens „*romb gol*”, „*traversa*” sau „*tilda*” după care, de obicei, urmează un *RegDef*. S-ar putea spune că o definiție *SpSpecDef* specificată aici este diferită de cea

definită pentru o intrare DTLR și, în același timp, se poate adăuga că ea este asemănătoare cu o definiție *ItalDef*, definită pentru o intrare DTLR [4], care în DLRLC lipsește. O definiție *SpSpecDef* se utilizează cu scopul de a descrie:

(1) cologații ale cuvântului-intrare în diverse expresii. De exemplu:

АБРИКОСОВЫЙ, ая, ое. 1. Относящийся к абрикосу, абрикосам (в 1 знач.). ◇ **Абрикосовое** дерево. То же, что абрикос.

(2) compatibilitatea lexicală și sintactică a cuvântului. De exemplu:

АВАНС, а, м. 1. Деньги, а также продукты, товары, выдаваемые в счёт предстоящих платежей. *Получать аванс... = Авансом*, в знач. нареч. Вперёд, заранее. [На дачу] *пошли деньги взятые авансом у издателя*. В. Андреева, Дом на Чёрн. Речке.

(3) referințe la alte cuvinte din dicționar ce fac parte din aceleași sau diferite familii lexicale. De exemplu:

БЫТЬ,... I. ...II. ...5. ...~ *Будет и на нашей, моей* и т. п. *улице праздник*. См. Праздник. *Будь здоров*. См. Здоровый. *Будь не во гнев; не во гнев будь сказано*. См. Гнев. *Будь спокоен, будьте спокойны*. См. Спокойный.

(4) terminațiile flexiunilor cuvântului-intrare ce se conțin în rădăcina arborelui de sensuri. De exemplu:

ВИРТУОЗНЫЙ, ая, ое; зен, зна, зно. Относящ. к виртуозу, свойственный ему. *Виртуозное исполнение*.

(5) specificul utilizării cuvântului, inclusiv în comparații; utilizarea expresivă a cuvântului:

БЫЧИЙ, бе, бя. 1. ... 2. ... ◇ **Бычий** глаз. *Разг.* Болезненное растяжение и выпячивание глазного яблока. ◇ **Бычье** сердце. *Разг.* Болезненно увеличенное (в размерах и по массе) сердце.

➤ Definițiile *ExemDef* sunt utilizate cu scopul de a întregi o definiție prin specificarea unor exemple aduse de către autor sau citate din careva opere literare. *ExemDef*-urile au rolul de a rafina sensul definițiilor secundare și principale. De exemplu:

АБРАЗИВНЫЙ, ая, ое. Относящ. к абразиву, абразивам, являющийся абразивом. *Абразивные материалы. Абразивный брусок*.

3. Algoritm de parsare

Pentru a obține arborele de sensuri (de definiții), adică ierarhia sensurilor (definițiilor) desemnate de o intrare din dicționar, a fost utilizat un parser construit în baza configurației SCD (Segmentation-Cohesion-Dependency) [1,2]. Parsarea unui dicționar constă în recunoașterea marcherilor de sensuri și de definiții, în identificarea structurilor dintre doi marcheri și în clasificarea acestor structuri în conformitate cu ierarhia prestabilită a claselor de marcheri, descrise în Figura 1.

Concluzii

În această lucrare au fost prezentați marcherii de sensuri și de definiții, care pot forma ierarhii de sensuri și de definiții pentru fiecare intrare din DLRLC, bazate pe configurații SCD. Rezultatele comparației efectuate între marcherii de sensuri și de definiții din [1] cu cei analogi din DLRLC [3] ne-au ajutat să facem o analiză detaliată a tuturor marcherilor, a referințelor interne și externe ale unei intrări, să verificăm acuratețea parserului bazat pe configurații SCD, descris în [4], și, în ultimă instanță, să elaborăm un instrument care ar permite automatizarea construirii resurselor lingvistice indexabile.

Referințe:

1. Curteanu N. Augmented X'-schemes, Proceedings of the 12th conference on Computational linguistics, p.130-132, August 22-27, 1988, Budapest, Hungry [doi>10.3115/991635.991662].
2. Curteanu N. (2006): Local and Global Parsing with Functional (F)X-bar Theory and SCD Linguistic Strategy. (I.+II.), Computer Science Journal of Moldova, Academy of Science of Moldova, vol. 14, no.1 (40), p.74-102 and no.2 (41), p.155-182.
3. Словарь современного русского литературного языка. В 20-ти томах. - Москва: Русский язык; Издание 2-е, перераб. и доп., 864 страницы; 1991 - 1994 г. ISBN: 5-200-01068-3.
4. Curteanu N., Moruz A., Trandăbăț D., Bolea C., Spătaru M., Husarciuc M. Parsarea arborilor de sensuri și segmentarea la definiții în Dicționarul Tezaur eDTLR. – În: Lucrările atelierului Resurse lingvistice și instrumente pentru prelucrarea limbii române. Iași, 19-21 noiembrie 2008. - Iași: Editura Universității „Alexandru Ioan Cuza”, p.65-74.

Prezentat la 31.01.2011