

## ANALIZA MARCERILOR AUXILIARI DE SENSURI DIN INTRĂRILE DE DICȚIONAR DLRO

Eugenia Burca  
Universitatea Agrară de Stat din Moldova  
[eugenia\\_burca@yahoo.com](mailto:eugenia_burca@yahoo.com)

**Abstract.** *This paper describes the process to analyze the sense of the auxiliary markers from the dictionary entry of ODRL (Ojogov's Dictionary of Russian Language) using the optimal parsing method of SCD (Segmentation-Cohesion-Dependency) configurations.*

**Cuvinte cheie:** *configurația SCD, marcheri de sens, marcheri de definiție, marcheri auxiliari.*

### I. Introducere

Resursele lingvistice cu o acoperire destul de largă reprezintă o componentă pe care se bazează diverse aplicații de limbaj natural. Completarea și extinderea resurselor rămâne în continuare o problemă actuală. Orice sursă în format electronic prezintă o valoare pentru tehnologizarea limbajului natural, problema constă însă nu doar în a avea acest tezaur, ci și în a-l putea utiliza eficient, or aici ne confruntăm cu problema automatizării adnotărilor, indexărilor etc. În studiul ce urmează ne vom axa pe problemele, ce țin de transformarea intrărilor de dicționare în resurse lingvistice indexabile, utilizând în acest scop configurații de tip SCD (Segmentare-Coeziune-Dependență), descrise în [1]. În conformitate cu tehnologia propusă de autorii acestei strategii, mai întâi de toate se studiază structura intrărilor în dicționar, apoi se determină marcherii de sens și de definiții, se stabilește dependența dintre marcheri și se efectuează clasificarea acestor structuri în conformitate cu ierarhia claselor de marcheri.

O configurație SCD are următoarele componente:

1. un set de clase de marcheri, ce identifică hotarele pentru diferite categorii lingvistice specifice;
2. un arbore care stabilește dependențele dintre clasele de marcheri;
3. un algoritm de parsare.

Sînt cunoscute configurațiile SCD pentru o serie de dicționare (DAR - Dicționarul tezaur al Limbii Române – formatul vechi, DLR - Dicționarul tezaur al Limbii Române – formatul nou, TLF - Le Trésor de la Langue Française, DWB - Deutsches Wörterbuch – GRIMM și GWB – GoetheWörterbuch [2], DLRLC – Dicționarul Limbii Ruse Literare Contemporane[3], DLRO – Dicționarul Limbii Ruse de S. Ojogov[4]).

Scopul lucrării noastre constă în analiza marcherilor de sens, stabiliți în [4] pentru cunoscutul dicționar DLRO [5].

### II. Clasele de marcheri DLRO, structura lor de dependențe

În intrările DLRO au fost depistați următorii marcheri de sens:

- § cifre romane (I., II., III.,...,etc.), ce reprezintă *nivelul superior* al ierarhiei de sensuri pentru orice intrare de dicționar;
- § cifre arabe urmate de punct (1., 2.,3.,... etc.), ce reprezintă *al doilea nivel* de sens;
- § două bare verticale paralele (||), care specifică *al treilea nivel* de sens și este folosit pentru enumerarea sensurilor morfologice ale unei intrări;

§ *romb gol* (◇), ce specifică *al patrulea nivel de sens* și este folosit pentru enumerarea sensurile “expresiilor fraziologice”;

§ *cifre arabe urmate de paranteze rotunde* ( **1**, **2**, **3**),... etc.), sînt folosiți, doar la *enumerarea* unui sens sau subsens.

Acești marcheri determină hotarele pentru diferite categorii lingvistice specifice.

Marcherii de definiție, ce stabilesc dependențele dintre clasele de marcheri, sînt folosiți pentru a crea un număr finit de modele. Aceste modele delimitează subsensurile informației lingvistice, prin trăsături morfologice, sintactice, semantice sau pragmatice și prin aspecte lexicale, etimologice, diacronice și prozodice. Au fost depistate următoarele tipuri de definiții:

§ *MorfDefs*, ce descriu categoriile morfologice la diferite niveluri ale arborelui de sensuri;

§ *RegDefs*, ce reprezintă cel mai frecvent instrument lingvistic folosit în DLRO pentru a descrie sensuri;

§ *BoldDefs*, care este folosit pentru a explica sensul unor “expresii fraziologice”;

§ *ItalDefs*, ce descriu, în general, *colocații*;

§ *SpecDefs*, ce descriu definiții scrise cu font *regular* și cuprinse, de obicei, între paranteze rotunde:

§ *DefExems*, ce are scopul evidențierea sensului altor definiții ce au sensuri mai generale decît ele (sensuri secundare și principale).

### III. Polivalența marcherilor: studiu de caz pentru marcherul de sens auxiliar „paranteze rotunde”

DLRO, fiind un dicționar explicativ, la prima vedere prezintă o structură cu mult mai simplă[4], decît dicționarele – tezaur, menționate mai sus. Cu toate acestea, chiar și în acest caz aparent simplu, ne confruntăm cu o serie de probleme dificile, una din ele fiind ambiguitatea unor marcheri. Din această categorie face parte marcherul „paranteze rotunde”, care, în funcție de context, are roluri diferite, prin urmare, necesită eforturi suplimentare pentru dezambiguizare.

Marcherul de sens auxiliar „paranteze rotunde” sînt utilizate în următoarele scopuri:

§ La specificarea *terminațiilor* (formelor hotărîte/nehotărîte) a *verbelor* ce participă la explicarea sensului cuvîntului rădăcină în cadrul marcherului de definiție *RegDefs*, de exemplu:

**АКЦЕНТИРОВАТЬ**, -рую, -руешь;-анный; *сов. и несов., что*. **1.** Поставить (ставить) акцент (в 1 знач.) на чем-н. (спец.). *А. текст*. **2. перен.** Выдвинуть (-игать) на первый план, подчеркнуть (-кивать) какую-н. мысль (книжн.)...

§ Poate fi utilizat în cadrul marcherului de definiție *MorfDefs* pentru a completa informația morfologică pentru cuvîntul lema, de exemplu:

**ВИД**, -а (-у), о виде, в виде, в виду, на виду...

**ВЕСЕЛИТЬ**, -лю, -лишь; *несов., кого-что*. Вызывать веселье, радость. *В. публику. Песня веселит душу*. ◇ **Веселящий газ** (спец.) - закись азота, вызывающая особое состояние опьянения. || *сов. развеселить*, -лю, -лишь; -ленный (-ен, -ена).

§ De asemenea, poate fi utilizat înainte de a include informația morfologică pentru sensul în marcherul de sens 1. și 2. utilizînd marcherul de definiție *RegDefs* în exemplul de mai jos:

**ВИД**, -а (-у), о виде, в виде, в виду, на виду, м. **1.** (предл. о виде, в виде). Внешность, видимый облик; состояние. *Внешний в. человека. Здоровый в. С видом знатока. На в. или с виду ему мало лет. В исправленном виде. В исправленном виде*. **2.** (предл. о виде, в виде). Местность, видимая взором. *В. на озеро. Комната с видам на море*.

- § La indicarea informației de stil a cuvîntului rădăcină într-o expresie (de exemplu: (спец.), (устар.), (книжн.) etc.) în cadrul marcherului de definiție *SpecDefs*;  
**ВЕРБАЛЬНЫЙ**, -ая, -ое; -лен, -льна (книжн.). Словесный, устный.  
**Вербальное заявление**. ◇ **Вербальная нота** (спец.) - дипломатическая нота без подписи, приравняемая к устному заявлению.
- § La explicarea sensului cuvîntului rădăcină dintr-o expresie-exemplu sau a expresiei în întregime în cadrul marcherului de definiție *DefExems*;  
**ЗЕЛЁНЫЙ**, -ая, -ое; зелен, зелена, зелено, зелены и зелены. 1. Цвета травы, листвы. 3. *свет* (в светофоре: разрешающий движение). 3. *огонек* (фонарик, горящий за стеклом незанятого такси). 3. *чай* (сорт чая)...
- § La explicarea sensului unei expresii-exemplu cu sens figurat după cuvîntul rezervat „перен.” în cadrul marcherului de definiție *DefExems*;  
**АСТРОНОМИЯ**, -и, ж. Наука о космических телах, образуемых ими системах и о Вселенной в целом. ... *Астрономическое число* (перен. : чрезвычайно большое)...
- § La indicarea părții de vorbire sau de propoziție a cuvîntului rădăcină într-o expresie exemplu;  
**ИЗВЕСТНЫЙ**, -ая, -ое. -тен, -тна. 1. Такой, о к-ром знают, имеют сведения. *Известное всем событие. Стало известно* (в знач. сказ.) *о чем-н...*  
**ВЕСЁЛЫЙ**, -ая, -ое; вёсел, -á, -о, -ы и (разг.) -ы. 1. Проникнутый весельем, полный веселья. В. *характер. В. взгляд. Мне весело* (в знач. сказ.). *Весело* (нареч.) *смеяться...*
- § În calitate de referințe la un sens principal a cuvîntului rădăcină în cadrul intrării curente în interiorul marcherului de definiție *RegDefs*;  
**ЖИТЬ**,... 3. Проводить жизнь в каком-н. месте, среди кого-н., обитать. Ж. *в Москве. Ж. с семьей.* 4. *чем и на что.* Поддерживать свое существование чем-н. Ж. *своим трудом. Ж. на литературный заработок.* ... 8. *в ком и кем.* Работать, проживая в доме нанимателя (устар.). Ж. *в дворниках. Ж. гувернером.* // *многokr.* **живать**, наст. не употр. (к 3, 4 и 8 знач.; разг.).
- § Sau referință la altă leamnă din dicționarul DLRO. De exemplu:  
**БЕРАНДА**, -ы, ж. То же, что терраса (в 1 знач.), *индикă că cuvîntul „БЕРАНДА” are același sens ca și cuvîntul „ТЕРРАСА” în sensul, indicat de marcherul de sens „1.” în cuvîntul „ТЕРРАСА”.*
- § În interiorul unor unități frazeologice din cadrul marcherului de definiție *BoldDefs* la specificarea unor forme de utilizare a expresiei date;  
**ВИД**,...◇ **В виде** *кого-чего, предлог с род. п.* - в качестве кого-чего-н., как кто-что-н., сходно с кем-чем-н... **Не показать** (**не подать**) **виду** - сохранить спокойный вид, не дать понять, заметить что-н. *Не показал виду, что обижен.*
- § La enumerarea unor sinonime ale cuvîntului utilizat în explicarea sensului cuvîntului rădăcină cu scopul de a reflecta varietatea lui de utilizare în cadrul marcherului de definiție *RegDefs*;  
**АРТИЛЕРИЯ**, -и, ж. 1. Огнестрельные орудия (пушки, гаубицы, минометы)...
- § Sau la indicarea frazelor sinonime cu frazele indicate în fața marcherului “()”:  
**ЖИЗНЬ**,...*Между жизнью и смертью* (в очень опасном для жизни положении). *Отдать ж. за кого-что-н.* (пожертвовать собой). *Положить (класть) ж. за кого-что-н.* (умереть, умирать за кого-что-н.).
- § Sau poate indica sinonimul unei fraze figurate de exemplu:  
*...Вся местность на виду* (хорошо видна)...

§ După o expresie-exemplu la specificarea utilizării ei cu sens figurat (так-же перен.) în cadrul marcherului de definiție *DefExems*;

**АРХИВ**, -а, м. **1.** Учреждение для хранения старых,старинных документов, документальных материалов. ... **2.** Отдел учреждения, где хранятся старые документы. *Сдать в а.* ( так-же перен. : упразднить, признать устаревшим, негодным)...

§ Poate complementa fraza, de exemplu:

...**Упустить из виду что** - забыть (обычно по невниманию, рассеянности).

#### IV. Concluzii

În această lucrare sînt expuse rezultatele cercetării marcherului de sens auxiliar „( )” ce se poate întâlni în intrările de dicționar DLRO[5], bazat pe configurație SCD [1]. În rezultatul acestor investigații s-au obținut: sensurile de utilizare ale acestui marcher în conținutul marcherilor de definiții, care vor permite să elaborăm procedura de dezambiguizare a lui în scopul construirii algoritmului de parsare.

#### V. Referințe

1. N. Curteanu, A. Moruz, D. Trandăbăț, C. Bolea, M. Spătaru, M. Husarciuc. 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, Editura Universității „Alexandru Ioan Cuza”, Iași, pp. 65-74.
2. Curteanu, N., A. Moruz, D. Trandabăț (2010): An Optimal and Portable Parsing Method for Romanian, French, and German Large Dictionaries. *CogAlex II - The Second Workshop on Cognitive Aspects of the Lexicon, COLING-2010, Beijing, China*, ISBN: 978-7-302-23456-2, pp. 38-47
3. Burca E., *Parsarea dicționarului DLRLC prin metoda bazată pe configurații Segmentare-Coeziune-Dependență*. *Studia Universitatis, Revistă științifică, USM, Seria “Științe exacte și economice”, nr.2(42), Chișinău, 2011, ISSN 1857-2073, pp. 71-81.*
4. Burca E., *Parsarea arborilor de sensuri din intrările de dicționar*. *Proceedings of the international conference „Information and Comunication Technologies – 2009”, ICT + I<sup>st</sup> edition, Chișinău, 2009, ISBN 978-9975-66-134-8, pp 38-41.*
5. С. И. Ожегов «Словарь Русского Языка»: 70000 слов/ Под ред.Н. Ю. Шведовой.- 22-е изд., стер.- М.: Рус. Яз., 1990.-921 с. – ISBN 5-200-01088-8.