

Classification of Specification Definition Markers for DLRLC Entries

E. Burca

Abstract: This paper presents the results of specification definitions markers analysis, which were detected after application of the method of parsing based on Segmentation-Cohesion-Dependency (SCD) *configurations* for efficient transformation of the entries from Contemporary Russian Literary Language Dictionary (DLRLC) into an linguistic indexed resources.

Keywords: Segmentation – Cohesion - Dependency *configurations*, sense and definition markers, specification definitions markers, linguistic resources, parsing, dictionary entry, sense tree.

1 Introduction

The strategy behind parsing method of dictionary entries based on SCD *configurations* is to recognize sense markers (a *marker* is a boundary for a specific linguistic category), the structures identifying between two markers and classification of these structures in accordance with predetermined hierarchy of classes of sense markers [1, 2].

The main sense markers detected for entries of DLRLC [3] by the first SCD *configuration* are: *Roman numerals followed by a period points* (I., II., III etc.), *Arabic numerals followed by point* (1., 2., 3 etc.) and as secondary sense markers serve: *two slash (/), empty diamond (◇), cross (≡), tilde (~), literal enumeration (а), б), в),... etc.)*. [4].

During the semantic content analysis of the DLRLC entries with second SCD *configuration* based on a set of definitions markers (which mean a tree hierarchy that establishes the dependencies between classes of markers) there are identified the following definitions markers: *MorfDef* (morphological definitions); *RegDef* (definitions written with regular font); *BoldItalDef* (definitions written in bold – italic fonts), *BoldRegDef*

(definitions written in bold – regular fonts); *SpecDef* (definitions containing specifications); *SpSpecDef* (definitions that contain certain specifications, written with spaced letters); *ExemDef* (examples with the role to complete a definition meanings) [4]. In further the definitions containing specifications (*SpecDef*) will be explored in more detail.

2 Specifications definitions for the DLRLC entries

Definitions containing specifications *SpecDef*, are written in Italic font, specify different contexts of DLRLC entries usage. They represent some reserved words (abbreviations) or expressions obtained by concatenating two or more reserved words from a strictly limited list. *SpecDef*-s are used at any level of senses tree.

After the performed analysis we concluded, that depending on the sense, determined by a marker *SpecDef*, the list of reserved words can be divided into four distinct classes:

- *Class I* - *SpecDef* markers, which carry a morpho-syntactic information of the word;
- *Class II* - *SpecDef* markers, which contain a style characteristic of the word;
- *Class III* - *SpecDef* markers, which include a information about the expressivity of word pronunciation;
- *Class IV* - *SpecDef* markers, which contain a semantic information about words.

SpecDef markers of *Class I* are represented by reserved words, such as: союз, противит., частица, приставка, нареч., м., ж., ср., мн., ед., сов., несов. etc., that follows after the word-entry or after any main sense markers. For example:

БРАУНИНГ, а, **м.** Небольшой автоматический пистолет особой системы...

БРАТЬ, беру, берёшь, *прош.* брал, ла, ло, *несов.*, *перех.* и *неперех.*, (*сов.* взять). 1. Захватывать рукой, руками; принимать в руки...

The expressions formed by concatenating of two or more reserved words (*в знач. сущ.*, *в знач. нареч.*, *в знач. межд.*, *прич. в знач. прил.*, *сов.* и *несов.* etc.) are used in cases when the word is used in another part of speech (usually one of these expressions is indicated after markers "="). For example:

АВАНС, а, м. 1. Деньги, а также продукты, товары, выдаваемые а счёт предстоящих платежей. Получать аванс... = *А в а н с о м*, *в знач. нареч.* Вперёд, заранее.

SpecDef markers of *Class II*, which contain a characteristic style of speech, are represented by reserved words: *разг.*, *устар.*, *устаревающее*, *спец.*, *шутл.*, *прост.*, *обл.*, *поэт.* etc. In same mode the concatenating of the two markers is admitted: e.g. *устар.* и *поэт.*, *устар.* и *разг.* etc. The respective markers indicate the membership of a word or sense of certain speech areas. For example:

АБИССИНСКИЙ, ая, ое. Относящ. к абиссинцам, принадлежащий, свойственный им. ◇ *А б и с с и н с к о е* нагорье. *Устар.* Эфиопское нагорье...

БРАТУШКА, и, *род. мн.* шек, м. *Обл.* То же, что брат

Class III. SpecDef markers, which carry the information about the expressivity of word pronunciation, are determined by the reserved words *вопр.*, *усилит.*, *побудит.*, *пренебр.*, *ирон.* etc. For example:

БАБА, ...4. *Перен. Прост. Пренебр.* О робком, нерешительном мужчине, юноше...

Class IV. SpecDef markers, which include a semantic information about words, are represented by reserved *собир.* and *в знач. собир.* If words have collective sense before described sense is used *собир.* logo. For example:

БРАТСТВО, а, ср. ... // *Собир.* Люди, объединенные общей целью, общим делом...

In case when the entry word have singular and plural numbers, but the entry word is used in singular number with collective sense, then after the marker "=" is used the expression *в знач. собир.* For example:

СОБОЛЬ,...1. Хищный пушной зверек...*Охота на соболей.* = **B**
знач. собир. Водятся здесь и соболя и лисица...

3 Conclusion

This paper presents the results of specification definitions markers analysis, which were detected after application of the parsing method based on SCD *configurations* for efficient transformation the DLRLC entries in linguistic indexed resources.

The effectuated analysis of specification definitions markers permit us with more accuracy and completeness to determine more deeply lexicographic information about DLRLC entries.

References

- [1] Curteanu, N. *Augmented X'-schemes, Proceedings of the 12th conference on Computational linguistics*, p. 130-132, August 22-27, 1988, Budapest, Hungary [two > 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), pp.74-102 and no. 2 (41), pp.155-182.
- [3] *Contemporary Russian Literary Language Dictionary in 20 volumes*. Publisher: Moscow: Russian language, 2nd edition, revised. and add. 864 pages; 1991 - 1994 г. ISBN: 5-200-01068-3
- [4] Burca, E. *Parsarea dicționarului DLRLC prin metoda bazată pe configurații Segmentare-Coeziune-Dependență*, Studia Universitatis, Revistă științifică a Universității de Stat din Moldova, Seria "Științe exacte și economice", 2011, nr.2(42) Chișinău, ISSN 1857-2073, pp.71-81.

E. Burca

Institute of Mathematics and Computer Science of ASM

State Agricultural University of Moldova

E-mail: eugenia_burca@yahoo.com