

Parsing the Dictionary of Modern Literary Russian Language with the Method of SCD Configurations. The Lexicographic Modeling

Neculai Curteanu, Svetlana Cojocaru, Eugenia Burcă

Abstract

This paper extends the experience of parsing other five, sensibly different, Romanian, French, and German largest dictionaries, to *DMLRL* (Dictionary of Modern Literary Russian Language) [18], using the optimal and portable parsing method of SCD (Segmentation-Cohesion-Dependency) configurations [7], [11], [15]. The purpose of the present paper is to elaborate the lexicographic modeling of *DMLRL*, which necessarily precedes the sense tree parsing dictionary entries. The following *three* SCD configurations are described: the *first one* has to separate the lexicographic segments in a *DMLRL* entry, the *second* SCD-configuration concentrates on the SCD marker classes and their hypergraph hierarchy for *DMLRL* primary and secondary senses, while the *third* SCD configuration hands down the same modeling process to the atomic sense definitions and their examples-to-definitions. The dependency hypergraph of the third SCD configuration, interconnected to the one of the second SCD configuration, is specified completely at the atomic sense level for the first time, exceeding the SCD configuration modeling for other five dictionaries [15], [14]. Numerous examples from *DMLRL* and comparison to *DLR-DAR* Romanian thesaurus-dictionary support the proposed *DMLRL* lexicographic modeling.

Keywords: new approach to dictionary entry parsing; the parsing method of SCD configurations; parsing the largest Romanian, German, French, and Russian dictionaries; lexicographic modeling.

1 Dictionary Entry Parsing with SCD Configurations

The general idea behind parsing a thesaurus or dictionary can be reduced to transforming a raw text entry into an indexable linguistic resource. The typical representation of the parsing result of a dictionary entry is its *sense tree structure*.

The aim of this paper is to prepare the **DMLRL** (Dictionary of Modern Literary Russian Language) [18] for parsing with the *method* of SCD (Segmentation-Cohesion-Dependency) *configurations* [7], [15], starting with its necessary *lexicographic modeling* [16], [1]. We rely on the experience of modeling and parsing very efficiently other *five* largest, complex, and sensibly different thesaurus-dictionaries [11], [15]: **DLR** (The Romanian Thesaurus – new format) [2], [11], **DAR** (The Romanian Thesaurus – old format) [28], **TLF** (Le Trésor de la Langue Française) [23], **DWB** (Deutsches Wörterbuch – GRIMM) [17], **GWB** (Göthe-Wörterbuch) [17].

An SCD *configuration* has the following computational components [7], [15], [27]:

- A set of *marker classes*: a *marker* is a boundary for a specific linguistic category;
- A *hypergraph-like hierarchy* that establishes the dependencies among the marker classes;
- A *searching (parsing) algorithm*. The parsing algorithm is designed to perform the following actions: recognize the markers within the text, identify the text structures / spans they bound, and classify these structures according to the pre-assigned hierarchy of marker classes. The algorithm is applied to a *specific* SCD configuration of marker classes and hierarchy, strictly depending on the *semantics* standing behind *that* SCD configuration. Such a semantics involves specific markers, marked categories, and their (partial ordering) hierarchies to be applied along the corresponding text span (or lexicographic segment) to be parsed.

The developed parsing strategy merges the following three SCD configurations: *the first one* has to separate the main *lexicographic segments* [22 :2] of a thesaurus entry; *the second* SCD-configuration should parse each *entry segment*, concentrating on the sense description segment and its sense-tree extraction [11], [12], [14], [15], [16]. This partic-

ularly important SCD-configuration, which obtains the entry *sense tree* exclusively from the sense *marker sequences*, coincides with the DSSD algorithm in [11]. This algorithm (and SCD-configuration) continues with a *third* SCD-configuration that parses each node in the generated sense-tree for obtaining the atomic definitions / senses (*i.e.* finest-grained meanings) of the entry, according to the lexical-semantics modeling of the thesaurus-dictionary; *e.g.* for **DLR-DAR** [11], [14], [15], for **DMLRL** [16], [1], and also the remarks concerning the new types of definitions / senses, definition examples, etc.), *i.e.* their lexicographic types and dependencies.

Parsing with SCD configurations means a good (sometimes, thorough) measure of prerequisite semantic modeling of the text, establishing of the marker classes for syntactic-discursive structures driven by certain precise semantics, determining the dependency hypergraph(s) for the considered marker classes, recognition of the markers in the text, and extraction of the *marker sequences* (only). In such a concrete SCD configuration, parsing means to compute the dependency relations between (among) the markers in the marker sequences of the text, according to the dependencies incorporated *ab initio* in the pre-established *dependency hypergraphs* for the *marker classes* of the configuration.

To notice the important computational characteristic that parsing with SCD configurations is a *procedural-oriented* tool and a completely *formal grammar-free* one, the latter device being proved to be cumbersome and inefficient when applied to free, general, or specialized (such as dictionary entry) kinds of natural language texts.

2 Homonymic Entries in DMLRL

The homonymic entries in **DMLRL** (**D**ictionary of **M**odern **L**iterary **R**ussian **L**anguage) are discriminated by indexing each of the homonyms with Arabic numerals followed by dot, all in *Arial font*, *Regular* and *Bold* format. These indexes are positioned in front of each homonym-word lemma, enumerating increasingly all the homonyms of the same word-lemma in the dictionary. An example of *four* homonymic entries

of the word "БЫЧОК" is present in **DMLRL** [18 :860-861].

The first two of these entries may cause the same possible ambiguity between the second sense of "БЫЧОК" *first entry*, introduced by the sense marker "2." (font Arial; correct font: Times New Roman), and the index of the second homonymic entry, starting with the similar but slightly different marker "2." (Times New Roman; correct font: Arial). If the parsing program works properly and associates unequivocally the homonymy index to the **DMLRL** entry lemma (which is written with bolded capital Cyrillic letters), then there should not appear ambiguities when discriminating and parsing the lexically independent homonymic entries in **DMLRL**. The (ambiguity-introducing) original example is [18 :861]:

1. 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. *Лоси сбрасывают рога; старые самцы в декабре — январе, молодые бычки — в конце февраля — в марте.* Формоз. Спутн. Следопыта. ◇ В сравн. [Сергей] *выслушивал предложения молча, насупившись, склонив, как бычок, голову и напряженив шею.* Первенц. Дир. Томилин. ◇ Смотреть, глядеть и т. п. бычком. *Смотреть хмуро, исподлобья.— Ну, а парнишку-то!.. сажай и его! Что, смотрю, он у тебя таким бычком глядит, слова не скажет.* Григор. Рыбаки. *Лешка смотрел на него [мир] не как прежде — широко открытыми, ясными серыми глазами, — а бычком, исподлобья и ожидал от него одних неприятностей.* Дубов, Горе одному. ◇ Пить бычком. См. Пить.~ **Сказка про белого бычка.** См. Сказка.

2. Перен. Разг. О молодом упрямом человеке (обычно в функции сказуемого).— *Эх ты, бычок несуразный .. грохотал Сиволап шатающемуся Кромулину.* Леон. Конец мелк. чел. *Всех широкоцев обозвал он кротами, а Яшку — бычком, бездельником-буяном.* Панфер. Бруски. ~ **Быть бычку на веревочке.** См. Быть.

— Поликарпов, 1704: быччк; Нордстет, 1780: бычок.

2. БЫЧОК, чка, м. **1.** Рыба отряда окунеобразных.

Черноморские, каспийские бычки. Я и механик удили с палубы рыбу и нам попадались очень крупные, толстоголовые бычки. Чех. Остр. Сахалин. Они поймали .. одну горбушу и двух бычков-подкаменщиков с пестрой окраской и оранжевой каймой на темно-оливковом спинном плавнике. Арсен. Дерсу Узала. До чего ж прозрачна байкаль

3 The Main Lexicographic Segments in DMLRL

In [11], [12], [14], [15] there have been analyzed the first and second SCD configurations of the following five thesaurus-dictionaries: **DLR** (The Romanian Thesaurus – new format) [2], [11], **DAR** (The Romanian Thesaurus – old format) [28], [13], [15], **TLF** (Le Trésor de la Langue Française) [23], [15], **DWB** (Deutsches Wörterbuch – GRIMM) [17], [15], and **GWB** (Göthe-Wörterbuch) [17], [15]. This knowhow is applied and extended in this paper to **DMLRL** (Dictionary of Modern Literary Russian Language) [18].

The *first SCD configuration* has to recognize the lexicographic segments of a **DMLRL** entry. **DMLRL** comprises (at least) five types of lexicographic packages / segments: **(1)** a *morpho-lexical* package / segment, **(2)** the *sense description* segment, **(3)** a *TildaDef* package or segment of definitions (see subsection 3.3), **(4)** the *morpho-syntactic variant* segment, and **(5)** the *etymology* segment of the word-lemma. The morpho-lexical definition package is obligatorily present at the beginning of each entry, immediately after the word-lemma. The morpho-lexical package may occur also at the sense lower-levels of the entry sense tree. The *TildaDef* package can be attributed not only to any (sub)sense description level of the entry but also to the root-sense (zero-level sense hierarchy). When this package / segment begins at *new paragraph* (*NewPrg* typographic marker), the *TildaDef* package is assigned to the entry root-sense. The same *NewPrg* lexicographic marker is met in **DAR** thesaurus-dictionary [15], [14] (see subsections 3.2, 5.2).

In the following subsections, some examples of **DMLRL** lexicographic segments are given. We notice that the structure of lexico-

graphic segments for large thesaurus-dictionaries, recognized within the *first SCD configuration*, is linear and simple, in general [14], [15]. However, remarkable exceptions are the oldest dictionaries studied, namely the German **DWB** [17] and the Romanian **DAR** [28], [15], whose design began in 19-th century for both.

3.1 The Morpho-Lexical Segment of a DMLRL Entry

The entry **БЫТЬ** is enlightening for the morpho-lexical segment: this lexicographic package / segment covers the first rows, from the word-lemma until the first primary sense introduced by the marker "I.", in bold [18 :856]. With this marker begins the most important lexicographic segment of **DMLRL** entries (and in any dictionary), *viz.* the segment containing the lexical-semantics descriptions of entry senses, called the *sense description segment*.

БЫТЬ, *наст.* не употр. кроме *3 л. ед. е с т ь* и (*устар.*) *3 л. мн. с у т ь*, *буд.* б у д у, б у д е ш ь, *прош.* б ы л, б ы л а, б ы л о (с отрицанием: н е б ы л, н е б ы л а, н е б ы л о), *повел.* б у д ь (т е), *прич. действ. прош.* б ы в ш и й, *деепр.* б у д у ч и, *несов., неперех.*

I. Как самостоятельный глагол означает: **1.** Существовать. *Не говори с тоской: их нет; Но с благодарностью: были.* Жук. Воспоминание. *В дымном зареве вставал рассвет. Город был. Сегодня нет его.* Сурк. Город О. *Прошлое человечества — драгоценная сокровищница неисчислимых богатств .. Эти богатства были, то есть существовали когда-то реально во времени.* Шагинян, Воскреш. из мертвых. ◇ Жил-б ы л, жила-б ы л а и т. п. жил да б ы л. *Нар.-поэт.* Употр. как начало повествования, сказки и т. п. *Жила-была вдова, Тому лет восемь, бедная старушка.* Пушкин. Домик в Коломне. *Во время оно жил да был В Москве боярин Михаил, Прозваньем Орша.* Лерм. Боярин Орша. ◇ О каком-л. времени, периоде, поре и т. п. *Была осень. Был полдень. Была та смутная пора, Когда Россия молодая, В бореньях силы напрягая, Мужала с гением Петра.* Пушкин. Полтава.

Other examples of **DMLRL** morpho-lexical segments are the following (shaded part) ones [18 :781]:

БРОШЮРОВАТЬ, р у ю, р у е ш ь, прич. страд, прош. б р о ш ю р о в а н н ы й, а я, о е, несов. перех. Сшивать, скреплять отпечатанные листы в книгу или брошюру соответственно нумерации. При издании журнала мне доверялось только брошюровать тираж, написанный под копирку в несколько экземпляров. А. Гусев, От Эльбр. до Антаркт.

— С иным (*устар.*) напис. и произнош.: б р о ш и р о в а т ь.— Даль: брошировать; Ушаков, 1934: брошюровать.— От франц. brocher — сшивать листы книги.

БРОСАТЬ, а ю, а е ш ь, несов.; **бросить**, б р о ш у, б р о с и ш ь, прич. страд, прош. б р о ш е н н ы й, а я, о е, сов.; перех. и неперех.

1. Перех. Резким движением, взмахом заставлять перемещаться в воздухе в каком-л. направлении копн, что-л.; кидать (в 1 знач.). *Бросить камень, палку.*

Within *MorfDefs* of the morpho-lexical segment, several *SpecDefs*, *SpSpecDefs*, *LexVarDefs* etc. may be inserted (see subsection 5.2).

3.2 The Sense-Description Segment

The investigation of the lexicographic segment devoted to the lexical-semantic sense description is focused in Sections 4 and 5 below, which contain interesting examples for sense and definition description markers, together with their dependencies, represented as procedural, interconnected hypergraphs. The complete analysis of this segment constitutes the *second* and *third SCD configurations*, which is naturally the most important enterprise for the lexicographic modeling of the dictionary entry parsing process.

3.3 The *TildaDef* Package / Segment of Definitions

The *TildaDef* definition is introduced by the **DMLRL**-specific marker *tilda* "˜", being written in *bolded* and *italics* Times New Roman font,

экипажем. Левченко, Капли воен. грозы. ~ **Бросать/бросить деньги на ветер.** См. Д е н ь г и. **Бросать оружие.** Сдаваться, отказываться воевать. [Суконщики] *последние бросили оружие и уступили превосходной силе.* Пушк. Ист. Пугачева. **Оторви да брось.** См. О т р ы в а т ь. **Поднять да бросить.** См. П о д н и м а т ь. **Хоть брось.** О ком-, чем-л. плохом, никуда не годном, не поддающемся исправлению и т. п. [Хлестова:] *А ты, мой батюшка, неисцелим, хоть брось. Изволил вовремя явиться!* Гриб. Горе от ума.— *Ну, что там? — Ось сломалась.— Барин для порядка ее потрогал. — Да, хоть брось.* Тург. Помещик.

4. *Перех.* Класть что-л. небрежно, не на свое место; кидать (в 3 знач.).— *Ты, Оксана, смотри, не бросай так ключа от твоей комнаты; не пропало бы что у тебя.* Данил. Беглые в Новороссии
... ..

or to the root-sense, as in **БЫТЬ** [18 :857]:

5. *Разг.* Употр. в формах будущего времени в знач. связи настоящего времени. ◇ При выяснении происхождения, родства, имени, социального положения и т. п.— *Гостит у нас .. Иван Иванович Мизинчиков, тебе будет двоюродный брат, кажется.* Дост. Село Степанчиково.. — *Ты сам откуда же будешь? — Мы рязанские.* Сераф. В пути.— *Тетья,— окликает ее [женщину] Вика храбро,— вы будете художница?* Лидина, Леванги.

~ **Будет и на нашей, моей** и т. п. **улице праздник.** См. П р а з д н и к. **Будь здоров.** См. З д о р о в ы й. **Будь не во гнев; не во гнев будь сказано.** См. Г н е в. **Будь спокоен, будьте спокойны.** См. С п о к о й н ы й. **Будь то..., или...; будет ли то..., или...** Употр. для выражения предположения при перечислении, сопоставлении и т. п. чего-л. *Начиная работать над каким-нибудь портретом, будь то изображение самое известное и документальное, или, наоборот, потерявшее свое имя, всегда можно ожидать любых осложнений.* Немилова, Загад.стар.картин. *Но всякая материальная сила, будь то сила мужского тела или же сила машины, нуждается еще в духовном водительстве.* Горьш.

Водопад. **Будь (ты, он, она) (трижды) проклят, проклята.**
См. П р о к л и н а т ь

The lexicographic *Tilda* package / segment illustrated above contains:

(a) Several *TildaDef* headers, namely: “*Будет и на нашей, моей и т. п. улице праздник.*”; “*Будь здоров.*”; “*Будь не во гнев; не во гнев будь сказано.*”; “*Будь спокоен, будьте спокойны.*”; “*Будь то..., или...; будет ли то..., или...*”;

(b) Several *RefDefs* (*Reference Definitions*), namely: “См. П р а з д н и к.”; “См. З д о р о в ы й.”; “См. Г н е в.”; etc., completing the *TildaDefs*;

(c) A general form of the *TildaDef* shape, made up of a *TildaDef* header, followed by a *RegDef* and two *DefExems* (quoted text and its *sigle* – i.e. its bibliographic source reference(s); the term *sigle* is assumed from French): “*Будь то..., или...; будет ли то..., или...* Употр. для выражения предположения при перечислении, сопоставлении и т. п. чего-л. *Начиная работать над каким-нибудь портретом, будь то изображение самое известное и документальное, или, наоборот, потерявшее свое имя, всегда можно ожидать любых осложнений.* Немилова, Загад. стар. картин. *Но всякая материальная сила, будь то сила мужского тела или же сила машины, нуждается еще в духовном водительстве.* Горьш. Водопад.” . . .

(d) Other *TildaDefs* of the package; see the hypergraph in Fig. 2, showing the (sequences and) dependencies for the atomic definitions and examples-to-definitions within the *lexical-semantics projection* of the primary and secondary senses into atomic senses of **DMLRL**.

A *special situation*, interesting from several points of view, demonstrates the following entry [18]:

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 энергически принялся за очистку авгиевых конюшен попечительства над училищем. Гл. Усп. Бог грехам терпит.

This *TildaDef* package can express, *by itself* (even with a single component definition), the lexical-semantic sense contents of **DMLRL** entry. The entry **АВГИЕВЫ** above, whose sense is defined basically through a *TildaDef* definition header, is further refined by *literal enumeration*. A similar type of entry sense definition can also be met in **DLR-DAR**, where entries can be described exclusively through a *BoldDef* or *ItalDef* definition [10], [11], [12], [13]. Equally, the literal enumeration may refine such zero-level definitions of atomic kind. Thus, in case of **DMLRL** dictionary, *RegDef*, *TildaDef*, and *RefDef* are *autonomous definitions*, in the sense described in subsection 5.2, initially proposed for **DLR-DAR** dictionaries [12], [14]. Section 5 comes into details on the atomic definitions / senses and types of examples-to-definitions.

3.4 The Morpho-Syntactic Variant Segment

The *Morpho-Syntactic Variant Segment* describes an independent subentry, associated as a syntactic variant to the basic entry. Typical examples are *adverbial* forms (**По-бычьи**) associated to certain *adjectives* (**Бычий**), as in the following sample [18 :860, 772]. The *Morpho-Syntactic Variant Segment* is located between the *sense description* segment, possibly ended with a *Tilda package* (or segment), and the *etymology* segment of **DMLRL** entry.

БЫЧАЧИЙ, ья, ье. *Разг.* 1. Относящ. к быку, быкам (1. Бык в 1 знач.),...

2. Свойственный быку; такой, как у быка.

По-бычьи, нареч. То же, что по-бычьи. *Гараська по бычачьи мотнул головой.* Аник. Гараська-диктатор.

— Слов. XI—XVII вв.: бычатий; Вейсманн, 1731, с. 454: бычачий; Росс. Целлариус 1771,

БЫЧИЙ, б е, ь я. **1.** Относящ. к быку, быкам (1. Бык в 1 знач.),

3. В составных народных названиях растений. *Бычья трава. Бычий ноготок.*

По-бычи, нареч. Как бык (1. Бык в 1 знач.), подобно быку. *Старший сержант по... ..*

— Срезневский: б ы ч и й; Вейсманн, 1731, с. 609: б ы ч и й; Нордстет, 1780: б ы ч и й; БАС 1948: п о-б ы ч ь и.

3.5 The Etymology Segment

The etymological description segment (shaded part), which always ends a **DMLRL** entry, is illustrated here also on the entry **БЫТЬ** [18 :856]. The **DMLRL** etymology segment is always introduced by the specific *etymological-dash*, *NewPrg* (*New Paragraph*) marked, and written with (two points) smaller font than the (Times New Roman) common text font measure of **DMLRL** dictionary entries.

... .. *Чтоб тебе, ему и т. п. пусто было.* См. П у с т о.
Чтобы духу твоего, вашего и т. п. не было. См. Д у х.
Чтобы неповадно было. См. Н е п о в а д н о. (**Я**) н е
я буду, если не... Употр. для выражения твердой уверенности или решительного намерения.— А где ж она, родительница-то? али спряталась? Не я б у д у, если не сидит где-нибудь там, за ширмами. Дост. Село Степанчиково.. — Я не я б у д у, если не окажется [в рапорте], что мы вырвались после отчаянной борьбы против охраны. Мстислав. Грач — птица весенняя.

— В иной (*разг.*) форме: *деепр. б ы в ш и*; в иной (*устар.*) форме: *деепр. б ы в*.— Срезневский: б ы т и; Берында, 1627: б ы с т ь; Вейсманн, 1731, с. 351: да б ы т ь так, не б у д е т, не б ы т ь удаче; Лекс. 1762: б ы т ь.

4 The Dependency Hypergraph at Sense Marker Classes in DMLRL

4.1 Primary and Secondary Senses in DMLRL. Examples, Dependencies

The *primary sense* markers in **DMLRL** pointed out so far by the lexicographic analysis are: **(1)** capital Roman numerals followed by a dot (**I.**, **II.**, **III.**,...etc.), in bold (*LatCapNumb_Mark*), and **(2)** Arabic numerals followed by a dot (**1.**, **2.**, **3.**,... etc.), in bold (*ArabNumb_Mark*). The markers of these classes are positioned at the beginning of the text row, in fact, at new paragraph (*NewPrg* marker), except for the *first sense markers* (**I.**, **1.**), which usually does not occur at new paragraph.

The sense markers of the class denoting Roman capital numerals followed by a dot (**I.**, **II.**, **III.**,...etc. or simply, *LatCapLett_Enum*) represent the top of the sense hierarchy in **DMLRL**. These markers establish the lexicographic limits for the *most general senses* of the word-lemma. To notice that they are the equivalent of the marker class denoted by bolded Latin capital letters **A.**, **B.**, etc. (abbreviated as *LatCapLett_Enum*) in **DLR** [9], [11], [12].

The sense marker class of *Arabic numerals* followed by dot, point (**1.**, **2.**, **3.**,... etc.), in bold (*ArabNumb_Enum*) stands for the second level of primary sense representation in **DMLRL**. The place of these two sense marker classes is displayed within the hypergraph of Fig. 1 below. The sense marker classes *LatCapNumb_Enum* and *ArabNumb_Enum* are considered to be the set of **DMLRL** *primary senses*, similarly to **DLR-DAR** lexicographic modeling [9], [11], [15].

We placed the *two-oblique-bars* "//" sense marker, which is specific to **DMLRL**, on the *third level* of the hierarchical dependency structure of **DMLRL** senses (Fig. 1). In the same time, the sense marker "//" is considered to be the first element of the two-markers set {//, ◇} denoting the *secondary senses* in **DMLRL**. The sense marked by "//" is in lexical-semantics subordination to (or subsumed by) any other primary sense marked by an element in the marker classes {*LatCapNumb_Enum*, *ArabNumb_Enum*}, when they exist in the entry text.

Otherwise (when a primary super-ordinated sense lacks), the secondary sense marker “//” may occur immediately under the topmost level of the **DMLRL** sense hierarchy. The marker “//” is embodied explicitly into the entry text, even for the case when this level has only one element of this type. For instance [18]:

АБРИКОСОВЫЙ, а я, о е. 1. Относящийся к абрикосу, абрикосам (в 1 знач.). ◇ А б р и к о с о в о е дерево. То же, что абрикос. // Состоящий из абрикосов. *Абрикосовый сад.*

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

We notice that *RegDefs* in the // -marked subsenses to the primary senses in the above entry are refined by the so-called *DictExem*, i.e. examples-to-definitions given by the **DMLRL** authors. Usually, *DictExems* are separated from *DefExems* that follows through the **DMLRL**-specific marker *traverse* “□”. See subsection 5.1-(ii) for further discussion on “□” marker, the first sense description “1.” of entry **ВЕДУЩИЙ** that follows (and Fig. 2). In this entry, the secondary sense “//” is refined through literal enumeration. In analogy with the **DLR** hypergraph of sense dependencies, we associate the **DMLRL** “//” marker with the **DLR** “◆” sense marker: they are both secondary sense markers and subsume the similar secondary sense marker denoted in both dictionaries by the *empty-diamond* “◇” (see below) [11], [15], [18].

ВЕДУЩИЙ, а я, е е. 1. Идущий впереди; головной. *Ведущий самолет. Каждый из ведущих броненосцев больше всего осыпался неприятельскими снарядами.* Нов.-Прибой, Цусима. // В знач. *сущ.* В е д у щ и й, е г о, м, В е д у щ а я, е й, 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.

Actually, the second marker in the *secondary sense* marker set used by **DMLRL** is the "horizontal-empty-diamond", which will be replaced in the **DMLRL** entry text, for the ease of graphical representation, lexical-semantic role, and uniformity, with the **DLR-DAR** sense marker "◇", *i.e.* the "vertical-empty-diamond" or, simply, the *empty-diamond* "◇" marker. The lexical-semantic sense defined by the "◇" marker is subsumed (thus subordinated) by the **DMLRL** sense defined with the sense marker "//".

We associate the secondary sense marker set {◆, ◇} in **DLR-DAR** with the corresponding set of markers {//, ◇} in **DMLRL**, relying on the following facts supported by the current stage of our investigation: (a) ◆ subsumes ◇, thus ◆ subordinates ◇ in **DLR-DAR** (actually, these relations refer to the entry senses introduced by these markers) [11]. (b) Similarly, // subsumes ◇, thus // subordinates ◇ in **DMLRL**. (c) The senses introduced by these markers are considered to be *secondary senses*, each pair in its corresponding dictionary, because of the high similarity of their lexical-semantic description refinement (a concept which we called *lexical-semantic granularity* of dictionary entry senses) [11], [14]. (d) Another argument for the proposed relationship is that these sense markers behave likewise when related to the sense refinement technique of *literal enumeration*: both markers in the above pairs of secondary sense markers, for the dictionaries **DLR-DAR** and **DMLRL**, are interleaving with the literal enumeration, recursively calling each other on several (but finite number of) levels. Typical examples are the entry "CAL" in **DAR**, demonstrated in [15], [14], and the entry "БЫ" [18 :844] in subsection 4.3 below. (e) Finally, preserving similar measures of lexical-semantic granularities in the thesaurus-dictionaries **DLR-DAR** and **DMLRL**, the primary senses (**A.**, **B.**, ...; **I.**, **II.**, ...; **1.**, **2.**, ...) in **DLR-DAR** and (**I.**, **II.**, ...; **1.**, **2.**, ...) in **DMLRL** do not interleave with the literal enumeration(s), while the secondary senses, {◆, ◇} in **DLR-DAR** and {//, ◇} in **DMLRL** do, as noticed in (d) above.

If there is no other higher-level sense marker, the "◇" marker may occur immediately below the root-sense of the entry sense tree, as in the following example [18 :781]:

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

Besides, the sense derived immediately from the *empty-diamond* marker "◇" can be refined by literal enumeration, as in the example "БРАТЬ" below, for the subsense no. "14." (the shaded part) [18 :742].

... ..

14. *Перех.* С некоторыми существительными (с предлогами и без предлогов) обозначает: производить какое-л. действие в соответствии со значением существительного. *Брать на буксир*. ◇ **Брать на прицел, на мушку**. *Целясь, готовиться к стрельбе. Завтра они будут ползти по окопам, закладывать мины, брать на мушку фрица.* Эренб. Буря. *Глядит Громак и молвит: — Есть! Заметил вражью точку, Берет тот кустик на прицел, Припав к ружью наводчик.* Твард. Ив. Громак. ◇ **Брать на учет**. а) Заносить в списки лиц, входящих в состав чего-л. (обычно в официальной речи). *Лейтенант брал Бондарева на учет.* Родичев, *Стоял старик на обочине.* б) Устанавливать наличие, количество кого-, чего-л. [*Двенадцать комсомольцев*] *ушли..брать на учет богатства, которые надо будет вывозить из мест затопления.* Песков, *Счастье перв. тропы.* в) Принимать во внимание, учитывать что-л. *Те, кто работал с ним в лаборатории, удивлялись тщательности его экспериментов — он брал на учет все мелочи, исключал возможность малейшей ошибки.* Гранин, Вар. второй. ◇ **Брать под вопрос**. См. *Вопрос*.

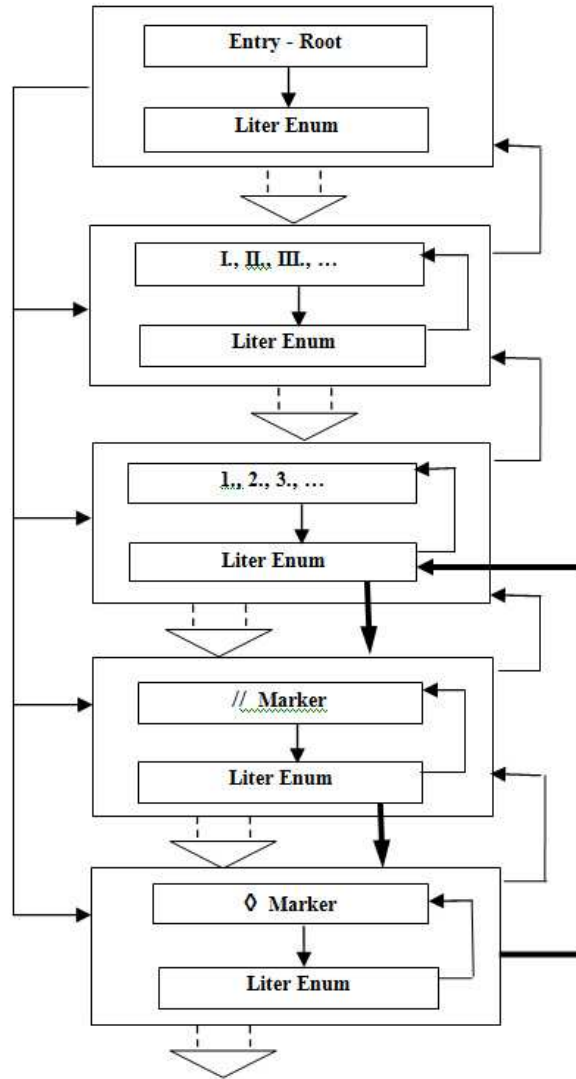


Figure 1. The Dependency Hypergraph at Sense Marker Classes in DMLRL

4.2 The Literal Enumeration and Its Recursive Dependency with DMLRL Sense Markers

The problem of literal enumeration in **DMLRL** is, for the moment, the most challenging one concerning the sense dependencies introduced by **DMLRL** marker classes. This is because one may find entry samples that display a recursion between the *literal enumeration* and the *secondary senses* “//” and “◇” (at least these markers). This level of recursion can be raised towards the higher (primary) senses, or may step down to the atomic senses / definitions. The solution of reducing these recursions to a finite number of cycles, and disambiguation of the cyclic application of secondary sense markers and of the literal enumeration should be consistent with the possible extension of the literal enumeration recursion to the higher or lower levels on the **DMLRL** hypergraph of marker class dependencies, pre-established for **DMLRL** (Fig. 1).

The following lexicographic sense description levels in **DMLRL** are specifiable through literal enumeration:

(1) **Refinement of the primary senses**; for this situation we deliver examples concerning the lexical-semantics refinement introduced by Arabic numerals (*ArabNumb_Enum*), but not the senses marked with Roman numerals (*LatCapNumb_Enum* class). A logical explanation would be that, for the lexical-semantics *granularity measure* of senses introduced by the *LatCapNumb_Enum* markers, the literal enumeration should not be an adequate refinement tool but rather the immediately lower, still primary or secondary levels of sense specification, managed by the *ArabNumb_Enum* marker class, “//” and “◇” markers.

(2) **Refinement of the secondary senses** “//” and “◇” by literal enumeration; for instance, in the entry **ВЕДУЩИЙ**, subsection 4.1.

(3) **Refinement of atomic senses / definitions** by literal enumeration. We have the **DMLRL** entry example of **БЫВШИЙ** [18 :846], where the sense introduced by “◇”, which details the entry root-sense definition, is refined by literal enumeration. Another example is the entry **АВГИЕВЫ** (subsection 3.3 above), whose root-sense is described by a *TildaDef* atomic definition, refined at its turn through

literal enumeration.

4.3 Which Sense Levels Could Refine the Literal Enumeration?

We are interested now in *the reverse situation*: which are the sense levels that could refine the lexical-semantics sense / definition(s) of a *letter* marker (or several, for instance) belonging to the sense refinement procedure of *literal enumeration*? The most interesting case we met (until now) is the entry "БЫ" [18 :844], under the primary sense no. "3." This subsense begins to be refined through literal enumeration, the first sense marker letter "а)" being further detailed with the following sequence of secondary sense markers \diamond , //, \diamond , \diamond . This marker sequence is followed by literal enumeration second letter "б)", further refined by the sequence \diamond , \diamond , \diamond of (secondary) sense markers. The next letter-marker is "в)", with some atomic definitions, followed by the letter marker "г)", which is specified by two secondary subsenses: \diamond , \diamond . The entry excerpt of "БЫ" [18 :844] is illustrative:

... ..

2. В придаточной части сложного предложения обозначает действие, обуславливающее собой то, о чем сообщается в главной части. *Когда б разбойника облавою не взяли, То многие еще бы пострадали.* Михалк. Бешен, пес

3. Обозначает различные оттенки желаемости действия; **а)** Собственно желаемость. *Учился бы сын. Были бы дети здоровы.* \diamond Если бы, когда бы, хоть бы и т. п. *О, если бы когда-нибудь Сбылась поэта сновиденья!* Пушкин. Посл. к Юдину. [Николка:] *Хоть бы дивизион наш был скорее готов.* Булгаков, Дни Турб. \diamond С неопр. ф. глаг. *Полететь бы пташечке К синю морю; Убежать бы молодцу в лес дремучий.* Дельв. Пела, пела пташечка.. [Настя:] *Ах, тетенька, голубок! Вот бы поймать!* А. Остр. Не было ни гроша... — *Жара, дедушка Лодыжкин .. Нет никакого терпения! Искупаться бы!* Купр. Бел. пудель. // Употр. для выражения опасения по поводу какого-л. нежелательного действия (с отрицанием). *Не заболел бы он.* \diamond С неопр. ф. глаг., имеющей

перед собой отрицание. — *Гляди, — говорю, — бабочка, не кусать бы тебе локтя! Так-таки оно все на мое вышло.* Леск. Воительница. ◇ Только бы (б) не. — *По мне жена как хочешь одевайся, .. только б не каждый месяц заказывала себе новые платья, а прежние бросала новешенькие.* Пушкин. Арап Петра Великого. [Варя:] *Не опоздать бы только к поезду.* Чех. Вишневый сад. **б) Пожелание.** **Условие я бы предпочел не подписывать.** Л. Толстой. Письмо А. Ф. Марксу, 27 марта 1899. ◇ **С неопр. ф. глаг.** *Поохотиться бы по настоящему, на коня бы денег добыть, — мечтал старик.* Г. Марков, Строговы. ◇ **В сочетании с предикативными наречиями со знач. долженствования, необходимости, возможности.** [Алеша Бровкин] *сверкнул глазами и понесся .. по гнилым полам приказной избы. Вслед ему косились плешивые понышники: “Потише бы надо, бесстрашной, здесь не конюшня”.* А. Н. Толстой. Петр I. ◇ Только бы (б), лишь бы, Употр. со знач. желательности действия. [Скалозуб:] *Мне только бы досталось в генералы.* Гриб. Горе от ума. **в) Желание-просьба, совет или предложение (обычно при мест. 2л.).** [Марина:] *И чего засуетился? Сидел бы:* Чех. Дядя Ваня. — *Пошел бы ты к ним счетоводом, полковник.* Павлен. Счастье. — *Ты бы, Сережа, все-таки поговорил с Лидией:* Пришв. Кац. цепь. **г) Желательность целесообразного и полезного действия.** ◇ **С неопр. ф. глаг.** *Вам бы вступить за Павла-то!* — воскликнула мать, вставая. — *Ведь он ради всех пошел.* М. Горький, Мать. ◇ **С неопр. ф. глаг., имеющей перед собой отрицание.** [Лиза:] *А вам, искателям невест, Не нежиться и не зевать бы.* Гриб, Горе от ума.

~ *Во что бы то ни стало.* См. Стать. *Как бы не так.* См. Как. *Кто бы ни был, что бы ни было, как бы то ни было.* См. БЫТЬ. *Хоть бы хны.* См. ХОТЬ. *Хоть бы что.* См. ХОТЬ.

— Срезневский: бы; Лекс. 1762: бы.

This example shows that literal enumeration can be further refined through secondary subsenses introduced by the sense markers ”//” and ”◇”. In the previous examples, we have seen that both primary senses (demonstrated for those defined by the marker class *ArabNumb_Enum*,

at this time) and secondary senses can be refined through literal enumeration. We did not (and didn't expect to) find the situation when the literal enumeration to be refined through primary sense marker. Since we have the concrete situation when secondary senses are detailed through literal enumeration, and the reverse holds too (at least for the example above), the two processes are calling each other for a finite (times of) recursion calls that are sequencing the two procedures of lexical-semantics particularization. The problem is to stipulate an *explicit criterion* for stopping effectively the mutual calling of the two refinement processes (through secondary senses and literal enumeration) in a finite number of steps.

Figure 1 provides the scheme of dependencies between the sense marker classes, in **DMLRL**, for the primary and secondary senses, possibly refined through the lexicographic device of *literal enumeration*. The hypergraph of dependencies at the classes of sense markers in **DMLRL** displays in Fig. 1 the *finite recursion* between the blocks of secondary sense markers, // and \diamond , and the literal enumeration: usually, each of the two secondary sense markers may call the literal enumeration, but the **DMLRL** dependency hypergraph specifies that the reverse is also true, *i.e.* the literal enumeration may call, at its turn, each of the secondary sense markers! The *direct calls* made from the marker sub-blocks to the other marker class blocks are put on view with *bolded arrows*.

The procedure, called the "*enumeration closing condition*" for the literal or numeral enumeration, is explained in the sequel. The programming solution for a deterministic condition of a *finite* number of cycles, when mutual calls of (the mentioned) marker classes are performed, is to check the following *closing condition*: for getting out of the (literal or numeral) enumeration (or, in other words, to terminate the enumeration procedure), after the last letter (number) closing the enumeration list, the sense level description is raised at least *one unit higher* than any of the marker levels used as subordinated sense markers under the (literal or numeral) items in the enumeration list.

More precisely, for instance, if secondary sense markers were used under a certain letter of a literal enumeration, and *after* the last letter

in the enumeration it is used a primary sense marker (thus higher with at least one unit in comparison to secondary markers), then the literal enumeration cycle at hand *can be closed* (one may not continue the literal enumeration refinement with the next letter in the alphabetic order).

We remind that we met a somehow similar (but more complex) problem for modeling the thesaurus-dictionary **DAR**, where the literal enumeration and the sense refinement introduced by the *NewPrg* (New Paragraph) typographic marker defining new senses (in various contexts) were calling each other [13], [14], [15]. The solution was there to introduce a special, *numeral enumeration* (with Roman small numerals, *LatSmaNumb_Enum*) for the sense markers *NewPrg*, then to close the finite mutual calls relying on the *enumeration closing condition* applied to several levels of sense description:

- the *literal enumeration* closing condition when this enumeration is developed *inside* a sense defined by a *single NewPrg* marker;
- the *numeral enumeration* closing condition when several *NewPrg* markers, encoded with the implicit, small Latin numbers *LatSmaNumb_Enum*, are developed within a sense described by a *single* small Latin letter of a literal enumeration *LatSmaLett_Enum*;
- once again the *literal enumeration* closing condition, when this enumeration is developed within a single, primary or secondary, regent sense.

Thus, we have here a *double enumeration*, a literal and a numeral one (the latter, generated by *NewPrg* markers), interleaving each other, each one with its enumeration closing condition. The entry **CAL** in **DAR** thesaurus-dictionary illustrates the exposed situation [15], [14], [28].

5 Atomic Sense Definitions and Example-To-Definitions in DMLRL: Their Dependency Hypergraph

5.1 DMLRL Specific Markers

(i). **The *tilda* "∼" marker.** The role of this DMLRL-specific sense marker is to introduce a package of at least one definition of *TildaDef* type, with the aim of detailing the meaning of the sense definitions. The *TildaDef* package can be initiated at any level on the sense tree of a DMLRL entry, including the root-sense level of the word-lemma. *TildaDef*, *RefDef*, together with the *RegDef* most common device of sense description, provides the set of *autonomous definitions* in DMLRL (see the taxonomy in subsection 5.2). Subsection 3.3 describes in detail the role of DMLRL-specific "∼" marker.

(ii). **The *traverse* "□" marker.** In DMLRL, this marker has several functions at the level of *atomic definitions* [11], [12], [14]:
 (1) The "traverse" sense marker is used to separate the author's *example text* (called *DictExem* in DMLRL) from the *quoted text example* that follows (denoted *DefExem*, as in DLR-DAR), both (possibly) preceded by *specifying definitions* (*SpecDefs*, *SpSpecDefs*, or other ones).
 (2) The *traverse* marker "□" is also employed in DMLRL for displaying certain grammatical forms of the word-lemma. See also subsection 5.1, (E8: *DictExem*). Examples are [18 :780]:

БРОСОК, с к а, 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. □ **Б р о с к о м**, в знач. нареч. *Правой рукой он [рыбак] брал лежащую на парاپете полубесформенную массу осьминога и резким броском кидал ее на камни парапета.*

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. В и о л о н ч е л ь, я, м. Мы присутствуем при последних усилиях борьбы виолончеля за свое самостоятельное существование. Чайков. Третья неделя концерта сезона.

(iii). **The asterisk “*” marker.** The task of this DMLRL-specific marker is to introduce a citation containing the use of the entry word-lemma with its *figurative meaning*. E.g. [18 :772-773]:

БРОСАТЬ, а ю, а е ш ь, несов.; **бросить**, б р о ш у, б р о с и ш ь, прич. страд, прош. б р о ш е н н ы й, а я, о е, сов.; перех. и неперех. 1. Перех. Резким движением, взмахом заставлять перемещаться в воздухе в каком-л. направлении копна, что-л.; кидать (в 1 знач.). *Бросить камень, палку.* [Чацкий:] *Кричали женщины: ура! И в воздух чепчики бросали!* Гриб. Горе от ума. *Иногда аппетит [Прасковьи Павловны] даже совсем пропадал, и она с досадой бросала на стол вилку и ножик.* Салт. Сатиры в прозе. [Доктор] *бросал мне стул, который я должна была поймать за ножки и бросить обратно.* Кавер. Два капит. *Войдя в избу, Михаил поставил на пол плетеную из бересты корзину, .. бросил к кровати мешок с валенками.* Ф. Абрам. Две зимы и три лета. **Море глухо шумело, бросая на песчаную отмель гряды пенящихся волн.* Мам.-Сиб, Вокруг раки, куста. *Ветер бросал горсти листьев на стол, на койку, на прл.* Паустов. Желт. цвет.

(iv). **The one-oblique-bar “/” marker.** This marker joins pairs of paradigmatic alternatives for the basic form of the entry word-lemma. E.g. [18 :773]:

... .. // В спортивной борьбе — вынуждать противника падать, касаясь лопатками ковра, земли. *Борьба велась без приза, по просьбе дирекции, и Арбузов два раза бросал англичанина, почти шутя, редкими и эффектными трюками, которые он не рискнул бы употребить в состязании с мало-мальски опасным борцом.* Купр. В цирке. ~ **Бросать/бросить грязь, грязью** в кого-л. См. Грязь. **Бросать/бросить** кого-, что-л. *за борт.* См. 1. Борт. *Жребий брошен.* См. Жребий. **Бросать/бросить**

камень, *камнем* в кого-л. См. Камень. **Бросать/бросить**
камешки в чей-л. *огород*. См. Камешек. **Бросать/бросить**
перчатку. См. Перчатка. **Бросать/бросить** что-л. *на чашу*
весов. См. Чаша.

2. *Перех.* Разводя руки, пальцы, выпускать, переставать
 держать что-л.

5.2 Atomic Definitions and Examples-to-Definitions in DMLRL: Taxonomies, Sequencing, and Dependencies

The definition types received specific functional roles in describing the meanings under **DLR** primary and secondary senses [11]. For the atomic senses / definitions, two taxonomies have been proposed in [12], [14], [15], to be used not only for **DLR-DAR** but also for **TLF**, **DWB**, **GWB**. Adapted and applied here to the **DMLRL** dictionary, the first taxonomy contains the following classes:

(**obli**) *obligatory definitions*, which are the *MorfDefs* and, for each **DMLRL** entry, *one* of the following three definitions, *RegDef*, *TildaDef*, (not exclusively when *RegDef* is present), or *RefDef*. The meaning of obligatory definitions is that there are no entries to have no *MorfDef*, and (at least) *one* of the *RegDef*, *TildaDef*, or *RefDef* definitions.

(**opti**) *optional definitions / examples-to-definition(s)* in **DMLRL**: *SpecDef*, *SpSpecDef*, *TildaDef* (when a *RegDef* is present), *RefDef*, *LexVarDef*, *DictExem*, and *DefExem*, whose presence is optional, as modifiers for an obligatory sense / definition.

A complementary taxonomy classifies **DMLRL** sense definitions and examples-to-definitions in:

(**auto**) *autonomous definitions*: *RegDef*, *TildaDef*, *RefDef*, and *LexVarDef*, meaning that these definitions have a stand-alone role in introducing **DMLRL** senses;

(**cont**) *contingent definitions / examples-to-definitions*: *MorfDef*, *SpecDef*, *SpSpecDef*, *LexVarDef*, *TildaDef* (when a *RegDef* is present), *DictExem* and *DefExem*, which do not have an independent, self-

determining meaning, but (possibly) playing the role of adjuncts, *i.e.* modifiers to some other definitions (including themselves).

MorfDef is obligatory at the root level of any **DLR** entry (*except* when the entry is defined by *RefDef*), being inherited (by default, when not present) on the lower levels of the entry sense tree. *MorfDef* is both an obligatory (at the root level) and also a *contingent* definition, when placed in front of an *autonomous* definition.

MorfDefs, *SpecDefs*, *SpSpecDefs*, *LexVarDefs*, *DictExems* and *DefExems* are *contingent* definitions since they cannot define a (sub)sense in an autonomous manner but they serve as auxiliary adjuncts to modify, to complete either autonomous definitions or other contingent definitions.

The lexicographic modeling of **DMLRL** for the parsing method of SCD *configurations* has to reveal at the beginning the entry segments (the *first* SCD configuration), the main segment of sense description being refined by primary and secondary senses, with their markers and dependencies, and their (possible) recursive relationship to literal enumeration (the *second* SCD configuration), whose image is the hypergraph in Fig. 1). The final level of lexical-semantics refinement is represented by the *third* SCD configuration, consisting of atomic sense definitions, examples to **DMLRL** autonomous definitions, their specific (sometimes, complex) markers, their sequencing and dependencies, their (autonomous / contingent and / or obligatory / optional) lexical-semantic role within a **DMLRL** entry. The third SCD configuration of **DMLRL** is illustrated in Fig. 2 below, a marker class dependency hypergraph, interconnected with that one in Fig. 1, and established for *the first time* at this level of specification for atomic sense definitions, among the studied large dictionaries **DLR**, **DAR**, **TLF**, **DWB**, **GWB** [15].

Trying to keep as close and unitary as possible to the already existing lexicographic SCD modeling of the atomic definitions and examples for **DLR-DAR**, **TLF**, **DWB-GWB**, we outline the following **DMLRL** atomic senses definitions, examples-to-definitions, their markers and dependencies [10], [11], [13], [15], [1], [16]. Each atomic sense definition is classified accordingly to the taxonomies proposed

above in this subsection (based on [12], [14], [15]). We found (until the current stage of **DMLRL** lexicographic investigation) that it is necessary to operate with the following **DMLRL** atomic sense definitions and examples-to-definitions:

(D1) *MorfDef* (*Morphologic Definition*); *Obligatory* and *Contingent* definition. When non-present, it should be inherited from a regent or a higher-level sense. It is written with Times New Roman, Italics font.

(D2) *SpecDef* (*Specification Definition*); *Contingent* and *Optional* definition. This is a *modifying* type definition applied in a cyclic or recursive manner to an autonomous definition. It is written with Times New Roman, Italic font. The expressions representing *SpecDefs* are usually abbreviated, reserved words.

(D3) *SpSpecDef* (*Spaced Specification Definition*); Similar to *SpecDef* but written with spaced-characters. *Internal reference* (inside the same entry), *external reference* (to another **DMLRL** entry), *morphological suffixes* or lexical variants are written, in certain contexts, with spaced-characters. See also *RefDefs*.

(D4) *RegDef* (*Regular Definition*); *Autonomous* and *Obligatory* definition. It is written with Times New Roman, Regular font. This is the basic tool to describe the semantic lexical-meaning of an entry sense / subsense in **DMLRL** (and in the largest majority of other dictionaries).

(D5) *TildaDef* (*Tilda-marker Definition*); *Autonomous* and *Optional* definition. Its description is enclosed in subsection 3.3.

(D6) *RefDef* (*Reference Definition*); *Autonomous* and *Optional* definition. *RefDefs* are *external references*, frequently met as constitutive part of the *TildaDef* definition package, or *internal references* to an entry sense (including the root-sense) inside which such a reference is used. We notice that all *RefDefs* are *SpSpecDefs* but the reverse is not true. See (D2) and (D3) below for typical examples.

(D7) *LexVarDef* (*Lexical-Variant Definition*); *Contingent* and *Optional* definition, used to provide lexical variation(s) to the entry-word. It is written with *bolded font*, and met within a *MorfDef*, when the meaning of the lexical variant is the same as that of the word-lemma.

(E8) *DictExem* (*Dictionary authors' Example*); *Contingent* and *Optional* example. This type of examples is given by the **DMLRL** dictionary authors to support the refinement of semantic explanations to autonomous definitions assigned to entry senses. *DictExems* usually follow an autonomous definition and are separated from *DefExems* by the traverse “□” **DMLRL** specific marker (see also subsection 5.1-(iii)).

(E9) *DefExem* (*Definition Example*); *Contingent* and *Optional* example. It is very similar to *DefExem* in **DLR-DAR** dictionaries [11], [15]. This type of examples represents quotations, text excerpts from bibliographic sources, with the role of refining and completing the meanings of autonomous definition(s) assigned to a (sub)sense of an entry. To each *DefExem* is associated a *sigle*, *i.e.* the *reference* of *DefExem* citation excerpt to its *bibliographic source(s)* or authorship.

The following further specifications and exemplifications concerning the above **DMLRL** atomic senses / definitions and their markers [1], [16] are considered. The relevant text of **DMLRL** definitions or examples-to-definitions at hand is highlighted in gray.

(D1: *MorfDef*) The *morphologic definitions* *MorfDefs* can form even a morpho-lexical package / segment in a **DMLRL** entry (see subsection 3.1) and describe the morphological categories at different levels of the entry sense tree. In general, the first element of a dictionary entry is a *MorfDef* that specifies certain syntactic categories, each one with its characteristic morphological-syntactic features. For **DMLRL**, the part-of-speech category of the word-lemma is *not present explicitly* but deduced from and described by its specific linguistic features; *e.g.*, “м.” (masculine), “Перен.” (figurative) for nouns; ”аю, аешь, *несов.*; **бросить**, брошу, бросишь, *прич. страд, прош.* брошенный, ая, ое, *сов.*; *перех.* и *неперех.*” contains flexional forms, reference to a sibling form (**бросить**, in bold and distinct font), transitive (*перех.*) and intransitive (*неперех.*) information for verbs etc. If *MorfDef* is missing at a (sub)sense level, then it is *inherited* implicitly from the regent or higher-level sense endowed with a *MorfDef* definition. Often, *MorfDef* is followed by *SpecDefs*, even when, by inheritance, it is missing overtly (as in the examples that follow).

АВАНС, а, м. **1.** Деньги, а также продукты, товары, выдаваемые а счет 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. ...

БРОСАТЬ, а ю, а е ш ь, *несов.;* **бросить**, б р о ш у, б р о с и ш ь, *прич. страд, прош. б р о ш е н н ы й, а я, о е, сов.; перех. и неперех.*

1. *Перех.* Резким движением, взмахом заставляя перемещаться в воздухе в каком-л. направлении копн, что-л.; кидать (в 1 знач.). *Бросить камень, палку.* [Чацкий:] *Кричали женщины: ура! И в воздух чепчики бросали!* Гриб. Горе от ума.

(D2: *SpecDef*) The *specification definitions SpecDefs* are various types of linguistic information (morphologic, syntactic, semantic, pragmatic, discursive, stylistic etc.) which refine and concentrate the meaning of the word, phrase, or text definition at hand. *SpecDefs* are written in Times New Roman, Italic font, many of them are *abbreviations* and *reserved* words (“*Снев.*”, “*Перен.*”, “*Разг.*” etc.), or parenthesized descriptions specifying different contexts of use within **DMLRL** senses. Functionally working as modifier expressions to be applied to the sense-subsense described, *SpecDefs* are both *contingent* and *optional* definitions. *SpecDefs* are present at any level of the entry sense tree. They are frequently enclosed in and associated with *MorfDefs*. Examples of *SpecDefs* (and also *SpSpecDefs*):

... ..

11. *Перех. и неперех. Разг.* Достигать.цели, добиваться успеха посредством чего-л. О нем [Прокофий] *стали говорить тогда:— Новый-то круто берет, а!., новый-то что удумал.*

4. *Только Зл. Разг.* Ловиться на удочку (о рыбе). *Ходил рыбачить на озеро. Плотва хорошо бралась, только успевай червя насаживать.* Горыш. Тридц. лет спустя.

БРОСАТЬ, а ю, а е ш ь, *несов.;* **бросить**, б р о ш у, б р о с и ш ь, *прич. страд, прош. б р о ш е н н ы й, а я, о е, сов.; перех. и неперех.*

1. **Перех.** Резким движением, взмахом заставляя перемещаться в воздухе в каком-л. направлении копн, что-л.; кидать (в 1 знач.). *Бросить камень, палку.* □ [Чацкий:] *Кричали женщины: ура! И в воздух чепчики бросали!*

БРАТЬ, беру, берешь, **прош.** брал, ла, ло, **несов., перех. и неперех.**, (**сов.** брать). 1. Захватывать рукой, руками; принимать в руки. *Брать ложку. Брать со стола книгу. Откинув локонь от милого чела, Сама из рук моих свирель она брала.* Пушкин. Муза. [Доктор] *брал его руку, отсчитывал пульс.* Горбат. Мое поколение. ◇ **Б р а т ь** чем-л. [Сахар] *приходилось брать щипчиками.* В. Катаев, Хуторок в степи. ◇ **Б р а т ь** руками, в руки что-л. *Он тянулся за дудкой, брал ее дрожащими руками и прикладывал к губам.* Корол. Слеп, музыкант.

БРАТСТВО, а, **ср.** 1. Содружество, единение, союз, основанные на общности целей, взглядов, принципов и т. п. [Пьер] *твердо верил в возможность братства людей, соединенных с целью поддерживать друг друга на пути добродетели.* Л. Толст. Война и мир. // **Собир.** Люди, объединенные общей целью, общим делом ит. п. ◇ **Б р а т с т в о** какое-л., кого-л. *Студенческое братство,* □ *Газетное братство распадалось на целый ряд категорий: передовики, фельетонисты, хроникеры, заведующие отделами вообще.* Мам.-Сиб. Черты из жизни Пепко.

(D3: **SpSpecDef**) The *spaced-specification definitions* (*SpSpecDefs*) are used, in general, either to specify morphological / lexical forms and variants, or to *internally* (inside the same entry) and *externally* (to another **DMLRL** entry or entry sense) refer a **DMLRL** entry sense / subsense. *SpSpecDefs* may occur not only in the sense description segment but also into the morphological, *TildaDef*, and etymological description segments. It is important to mention that a *SpSpecDef* expression in **DMLRL** is rather distinct from that defined in **DLR-DAR** [11], [14], [15]. *SpSpecDefs* in **DMLRL** are employed to describe the following situations [1]:

- (i) Collocations of the word-lemma in various expressions. *E.g.:*

АБРИКОСОВЫЙ, а я, о е. 1. Относящийся к абрикосу, абрикосам (в 1 знач.). ◇ **А б р и к о с о в о е** дерево. То же, что абрикос.

(ii) Lexical and syntactic-phrase variants of the entry-word, as in:

АВАНС, а, м. 1. Деньги, а также продукты, товары, выдаваемые а счѣт предстоящих платежей. *Получать аванс...* □ **А в а н с о м**, в знач. нареч. Вперѣд, заранее. [На дачу] *пошли деньги взятые авансом у издателя*. В. Андреева, Дом на Чѣрн. Речке.

(iii) *Internal references* (inside the same entry) and *external references* (to another **DMLRL** entry), thus *RefDefs*, are also *SpSpecDefs* in **DMLRL**, i.e. Times New Roman, regular, spaced-character written. The example that follows contains (grayed) external *RefDefs* (ending the previous entry that precedes “**БРАТЬСЯ**”), morphological derivations, and internal *RefDefs* (inside the “**БРАТЬСЯ**” entry).

... См. **2. М у ш к а** *Братъ* кого-, что-л. *на прицел*. См. **Прицел**. *Братъ* кого-л. *на пушку*. См. **1. П у ш к а**. *Братъ* кого-, что-л. *под обстрел*. См. **Обстрел**.

– Срезневский: б р а т и; Поликарпов, 1704: б е р у; Вейсманн, 1731, с 154: б р а т и денги; Росс Целлариус 1771, с 9: б е р у, б р а т ь.

БРАТЬСЯ, б е р у с ь, б е р е ш ь с я, прои. б р а л с я, б р а л а с ь, б р а л о с ь и б р а л о с ь несов. (сов. в з я т ь с я). 1. Захватывать что-л., хвататься за что-л., касаться чего-л. Руками. ◇ **Б р а т ь с я** за что-л. *Братъся за поручни*. *Братъся за голову, за подбородок*. □ *Он умолкал, иногда надолго*. *Справляясь с волнением, он крепко брался за спинку стула*. Кавер. Откр. книга. ◇ **Б р а т ь с я** руками, пальцами

(iv) Flexion suffixes of **DMLRL** entry-word, usually met in the morphological segment. *E.g.*:

ВИРТУОЗНЫЙ, а я, о е; з е н, з н а, з н о. Относящ. к виртуозу, свойственный ему. *Виртуозное исполнение*.

(v) The use of the spaced-character entry-word (lemma or derivations) within specific phrases, as in:

БЫЧИЙ, *ь е, ь я*. **1.** **2.** ◇ **БЫЧИЙ** глаз. *Разг.* Болезненное растяжение и выпячивание глазного яблока. ◇ **БЫЧЬЕ** сердце. *Разг.* Болезненно увеличенное (в размерах и по массе) сердце.

(D4: *RegDef*) *RegDef* is an *autonomous* and *obligatory* definition, written with Times New Roman, Regular font. *RegDef* is the standard device to describe the semantic lexical-meaning of an entry sense / subsense in **DMLRL** (and in many other thesaurus-dictionaries, including **DLR-DAR**). Sample of (grayed) *RegDef*:

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

(D5: *TildaDef*) The *TildaDef* package / segment of definitions is described in subsection 3.3.

(D6: *RefDef*) *RefDefs* are autonomous, *external references*, frequently met as constitutive parts of the *TildaDef* definition (or package), or *internal references* to an entry sense (including the root-sense) inside which such a reference is used. *RefDefs* are written with Times New Roman, regular, spaced-characters, thus they all are *SpSpecDefs*; the reverse is not true. (D2) and (D3) contain instances of *RefDefs*. In the example (D3: *SpSpecDef*)-(iii) given above, the first four grey fields are external *RefDefs* and the last three ones are internal *RefDefs*. The autonomous role of external *RefDefs* is shown in the following examples [18 :771]:

БРОНХ. См. **Б р ó н х и**.

... ..

БРОНХИÓЛА. См. **Б р о н х и ó л ы**.

(D7: *LexVarDef*) The *LexVarDef* (*Lexical-Variant Definition*) is a *contingent* and *optional* definition, used to provide lexical variation(s) to the entry word-lemma. It is written with small, regular characters, *bolded font*, and met within a *MorfDef* (when the meaning of the lexical variant is the same as that of the word-lemma; *e.g.* the pairs in the fol-

lowing examples: **ВИЛЯТЬ-вильнуть**, **БРОСАТЬ-бросить** [18 :860, 772].

ВИЛЯТЬ, я ю, я е ш ь, *несов.*; **вильнуть**,
БРОСАТЬ, а ю, а е ш ь, *несов.*; **бросить**, б р о ш у, б р о с и ш ь,
прич. страд, прош. б р о ш е н н ы й, а я, о е, сов.; перех. и неперех.
1. Перех. Резким движением, взмахом заставляя перемещаться в
воздухе

(E8: *DictExem*) *DictExems* are examples given by **DMLRL** dictionary to support the refinement of semantic explanations to autonomous definitions assigned to the entry senses. *DictExems* usually follow an autonomous definition and are separated from the other *DefExems* by the *traverse* “□” **DMLRL** specific marker (subsection 5.1-(ii) describes the *traverse* marker role). The difference between a *DictExem* and a *DefExem* is that the former do not bear a *sigle*, *i.e.* the reference to the bibliographic source of the example-to-definitions (this one is just the dictionary authorship), while *DefExem* has to provide its bibliographic source, *viz.* its *sigle(s)*. When both *DictExems* and *DefExems* are present, the former are located always as the first ones, followed by the *traverse* “□” marker, which signals the end of *DictExems* sequence and the beginning of the *DefExems* block. Numerous samples of *DictExems* and *DefExems* are already shown in the paper.

(E9: *DefExem*) The role and structure of a *DefExem* (*Definition Example*) is to support and refine a lexical-semantics sense definition, already outlined in (E8: *DictExem*) above. *DefExem* in **DMLRL** is actually the same *DefExem* example-to-definition that is working for **DLR-DAR** dictionaries [11], [15].

6 Conclusions

The special *features* of the *new parsing method with SCD configurations* (SCD-*configs*) are: • The SCD-*configs* method for dictionary entry parsing is based on sense marker classes, their lexical-semantics dependency (*i.e.* sense structure subsumption), and *procedural hypergraphs* reflecting the sense marker class sequencing and dependencies

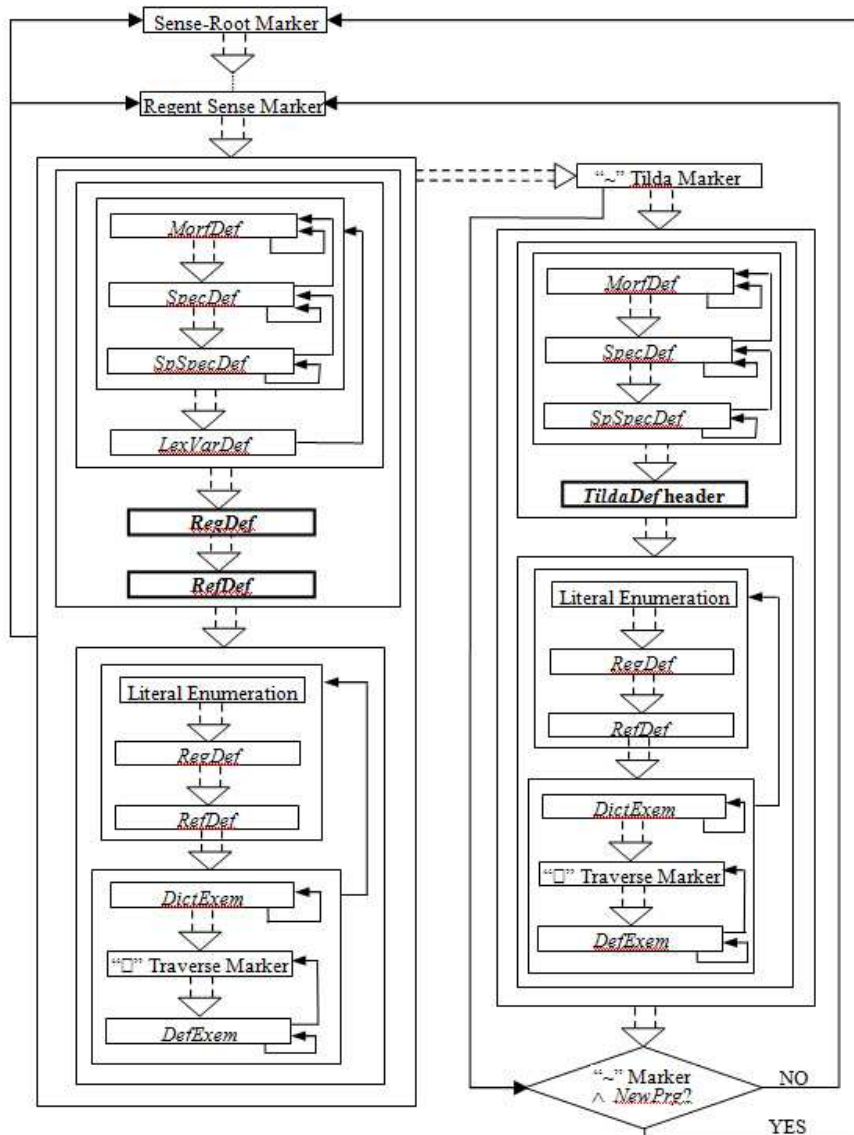


Figure 2. *RegDef* block and *TildaDef* block sequences and dependencies for **DMLRL** atomic sense definitions and examples-to-definitions

for each SCD configuration [11], [15]. • SCD-*configs* is a completely *formal grammar-free* approach which involves simple, efficient (weeks-time adaptable), thus portable modeling and programs [15]. • The method of SCD-*configs* for dictionary entry parsing is derived from the more general SCD linguistic theory and parsing strategy for natural language free text [7], [5], [3], [4]. • The main drawback of the currently existing parsing methods for dictionary entry parsing is that the sense tree construction of each entry is recursively embedded and mixed within the definition parsing procedures [6]. • To overcome this essential problem, the SCD-*configs* separate and run sequentially, on independent levels (*viz.* SCD configurations), the processes of *lexicographic segment recognition*, *sense tree extraction* (for entry senses defined by explicit marker classes), and *atomic definition parsing*. • This makes the whole *dictionary entry parsing* process with SCD-*configs* to be *optimal* [15], [11].

The main results of this paper consist in identification and behavior description of the three SCD configurations that are specific to **DMLRL** dictionary: SCD-*config1* shows the linear sequence of **DMLRL** lexicographic segments, while SCD-*config2* deals with sense marker classes associated to the primary and secondary senses in **DMLRL** and to their dependencies, displayed as the hypergraph in Fig. 1. Already pointed out in subsection 4.3, the solution to the problem of recursive calls between the secondary senses ($//$ and \diamond) and the refinement procedure of literal enumeration is the *enumeration closing condition*. The SCD-*config3* is represented in Fig. 2 as the hypergraph of the atomic sense / definition markers in **DMLRL** and *interconnected* with the hypergraph in Fig. 1. That one gives the dependency relationships among the higher-order sense marker classes, handing down from the root-sense, through primary and secondary senses, continued with the dependency hypergraph for the lower and atomic senses / definitions, represented in Fig. 2. When structurally accomplished, **DMLRL** lower-level senses are raising up, called by higher-level sense markers, until the structure of the entry sense tree is completed.

We provide in this paper the atomic definitions and examples-to-definitions that contribute to sense construction, their obligatory, au-

tonomous, contingent and / or optional functional role, described with their marker class sequences and dependencies. The type of dependency hypergraph in Fig. 2 is displayed for the *first time*, at this level of lexical-semantics specification, among the other similar dictionaries investigated for lexicographic modeling and parsing [15]. The **DMLRL** lexicographic segments, along with the higher-level marker class dependencies and hypergraph behavior in Fig. 1, procedurally interconnected with the hypergraph in Fig. 2, represent the complete lexicographic modeling of the three SCD configurations, which can ensure a high-performance parsing process of **DMLRL** dictionary, as proved for similar or more complex thesaurus-dictionaries [14], [15].

References

- [1] Burcă, Eugenia (2011): *Parsing the Dictionary of Modern Literary Russian Language (DMLRL) using the Method of Segmentation-Cohesion-Dependency Configurations*, Institute of Mathematics and Computer Science, Chişinău, Rep. of Moldova, 12 p. (in Romanian, draft paper).
- [2] Cristea, D., Răschip, M., Forăscu, C., Haja, G., Florescu, C., Aldea, B., Dănilă, E. (2007): *The Digital Form of the Thesaurus Dictionary of the Romanian Language*. In Proceedings of the 4th International IEEE Conference SpeD 2007.
- [3] Curteanu, Neculai (1994): *From Morphology to Discourse Through Marker Structures in the SCD Parsing Strategy. A Marker Hierarchy-Based Approach*. Language and Cybernetics, INTERK-IBERNETIK'93, Akademia Libroservo, Prague, Czech Republic, pp. 61–73.
- [4] Curteanu, Neculai, G. Holban (1996): *The SCD Linguistic Strategy Applied to the Analysis and Generation of Romanian*. In the volume "Language and Technology", (Ed. Dan Tufiş), The Romanian Academy Editorial House, Bucharest, pp. 169-176 (In Romanian).

- [5] Curteanu, N., D. Gălea, C. Butnariu, C. Bolea (2004): *Marcu's Clause-like Discourse Segmentation Algorithm and SCD Clause Segmentation-based Parsing*, Proceedings ECIT-2004 Conference, pp. 59–86, Iași, România.
- [6] Curteanu, N., E. Amihăesei (2004): *Grammar-based Java Parsers for DEX and DTLR Romanian Dictionaries*. ECIT-2004 Conference, Iasi, Romania.
- [7] Curteanu, N. (2006): *Local and Global Parsing with Functional FXbar Theory and SCD Linguistic Strategy*. (I.+II.), Computer Science Journal of Moldova, Academy of Science of Moldova, Vol. 14 no. 1 (40):74–102 and no. 2 (41):155–182.
- [8] Curteanu, N., D. Trandabăț, G. Pavel, C. Vereștiuc, C. Bolea (2007): *eDTLR Project – The Romanian Thesaurus-Dictionary in Electronic Format*. Research Report to the PNCDI II Project No. 91_013/18.09.2007, Stage on 2007 (in Romanian).
- [9] Curteanu, N., G. Pavel, C. Vereștiuc, D. Trandabăț (2008): *eDTLR Parsing with Lexicographic Grammars in the JavaCC Framework. The Current Stage, Problems, and Development Solutions*. In Proceedings of the Workshop on Linguistic Resources and Instrument for Romanian Language Processing – ConsILR-2007, (Ed. I. Pistol, D. Cristea, D. Tufiș), The "Al.I. Cuza" University Editorial House, Iași, ISSN: 1843-911X, pp. 87–96 (in Romanian).
- [10] Curteanu, N., D. Trandabăț, A. Moruz, C. Bolea, M. Husarciuc (2008): *Parsing the Romanian Language Thesaurus Dictionary (new format) at Sense Trees and Definitions, with the Method of SCD Configurations*. Research Report to the Grant Project PNCDI 2, Nr. 91_013/18.09.2007, Stage on 2008 (in Romanian).
- [11] Curteanu, N., Moruz, A., Trandabăț, D. (2008): *Extracting Sense Trees from the Romanian Thesaurus by Sense Segmentation & Dependency Parsing*, Proceedings of CogAlex-I Workshop, COLING 2008, Manchester, United Kingdom, pp. 55–63, ISBN 978-1-905593-56-9.

- [12] Curteanu, N., A. Moruz, D. Trandabăț, Cecilia Bolea, Mădălina Spătaru, Maria Husarciuc (2009). *Sense Tree Parsing and Definition Segmentation in the eDTLR Thesaurus-Dictionary eDTLR*, In Proceedings of the Workshop on Linguistic Resources and Instrument for Romanian Language Processing – ConsILR-2008, (Ed. D. Trandabăț, D. Cristea, D. Tufiș), Editura Univ. “Al.I. Cuza” Iași, ISSN: 1843-911X, pp. 65–74 (in Romanian).
- [13] Curteanu, N., A. Moruz, D. Trandabăț, C. Bolea (2009): *Parsing the Romanian Academy Thesaurus Dictionary (old format) and Romanian Language Thesaurus Dictionary (new format) at Sense Trees and Definitions, with the Method of SCD Configurations*. Research Report to the Grant Project PNCDI 2, Nr. 91_013/18.09.2007, Stage on 2009 (in Romanian).
- [14] Curteanu, N., A. Moruz, D. Trandabăț (2010): *Comparative Parsing of the Romanian, French, and German Thesaurus Dictionaries*. In Proceedings of the Workshop on Linguistic Resources and Instrument for Romanian Language Processing, (Ed. A. Iftene, H.N. Teodorescu, D. Cristea, D. Tufiș), Editura Univ. “Al.I. Cuza” Iași, ISSN: 1843-911X, pp. 113–122 (in Romanian).
- [15] Curteanu, N., Trandabăț, D., Moruz, A. (2010): *An Optimal and Portable Parsing Method for Romanian, French, and German Large Dictionaries*, Proceedings of COGALEX-II Workshop, COLING-2010, Beijing, China, August 2010, pp. 38–47.
- [16] Curteanu, Neculai (2011): *The SCD-based Lexicographic Modeling of DMLRL – Dictionary of Modern Literary Russian Language*, Research Report, Institute of Computer Science, Romanian Academy, Iasi Branch, June 2011 (in Romanian).
- [17] Das Woerterbuch-Netz (2010): <http://germazope.uni-trier.de/Projects/WBB/woerterbuecher/>
- [18] Dictionary of Modern Literary Russian Language (20 volumes – 1994): Editorial House: M.: Russkii Iazyk; Second edition, revised

- and supplemented, 864 p; 1991 – 1994. ISBN: 5-200-01068-3 (in Russian).
- [19] DLR revision committee. (1952). *Coding rules for DLR* (in Romanian). Romanian Academy, Institute of Philology, Bucharest.
- [20] Erjavec, T, Evans, R., Ide, N., Kilgariff A., (2000): The CONCEDE Model for Lexical Databases. Research Report on TEI-CONCEDE LDB Project, Univ. of Ljubljana, Slovenia.
- [21] Hauser, R., Storrer, A. (1993): *Dictionary Entry Parsing Using the LexParse System*. Lexikographica 9 (1993), 174–219.
- [22] Kammerer, M. (2000): *Wörterbuchparsing Grundsätzliche Überlegungen und ein Kurzbericht über praktische Erfahrungen*, <http://www.matthias-kammerer.de/content/WBParsing.pdf>
- [23] Le Trésor de la Langue Française informatisé (2010). <http://atilf.atilf.fr/tlf.htm>
- [24] Lemnitzer, L., Kunze, C. (2005): *Dictionary Entry Parsing*, ESS-LLI 2005.
- [25] Marcu, Daniel. 1997: *The Rhetorical Parsing, Summarization, and Generation of Natural Language Texts*, Ph.D. Thesis, Univ. of Toronto, Canada, pp. 331.
- [26] Neff, M., Boguraev, B. (1989) *Dictionaries, Dictionary Grammars and Dictionary Entry Parsing*, Proc. of the 27th annual meeting on Association for Computational Linguistics Vancouver, British Columbia, Canada Pages: 91 – 101.
- [27] ORDA License Copyright Registration (2011). Owner: Curteanu, Neculai. Title: *The SCD (Segmentation-Cohesion-Dependency) Lexicographic Modeling and Parsing Strategy for Natural Language Text of Some Romanian, French, German, and Russian Thesaurus Dictionaries*, ORDA – The Romanian Copyright Office, RNO Registration No. 9134 / 25.07.2011.

- [28] Puşcariu, Sextil *et al.* (1906): Dictionary of the Romanian Language (Dictionary of the Romanian Academy – **DAR**), Bucharest, Edition 1940 (old format).
- [29] Tiktin, H. (1989): *Rumänisch-deutsches Wörterbuch*, 2., überarbeitete und ergänzte Auflage von Paul Miron. [Band I–III]. Otto Harrassowitz. Wiesbaden. I: 1986; II: 1988; III: 1989.
- [30] Tiktin, H. (2005): *Rumänisch-deutsches Wörterbuch*, 3., überarbeitete und ergänzte Auflage von Paul Miron und Elsa Lüder. Band I–III. Cluj-Napoca, Clusium. I: 2000; II: 2003; III: 2005.
- [31] Tufiş, D., Rotaru, G., Barbu, A.M. (1999): *Data Sampling, Lemma Selection and a Core Explanatory Dictionary of Romanian*. Proc. of the 5th International Workshop on Computational Lexicography COMPLEX, Pecs, Hungary, pp. 219–228, 1999.
- [32] Tufiş, Dan (2001): From Machine Readable Dictionaries to Lexical Databases, RACAI, Romanian Academy, Bucharest, Romania.
- [33] XCES TEI Standard, Variant P5 (2007): <http://www.tei-c.org/Guidelines/P5/>

Neculai Curteanu, Svetlana Cojocaru, Eugenia Burcă, Received March 30, 2012

Neculai Curteanu,
Institute of Computer Science, Romanian Academy, Iaşi Branch,
Str. Gh. Asachi, Nr. 3,
700483 Iaşi, România
E-mail: ncurteanu@yahoo.com

Svetlana Cojocaru, Eugenia Burcă,
Institute of Mathematics and Computer Science, Academy of Sciences of Moldova,
Str. Academiei nr. 5, Chişinău,
MD 2028, R. Moldova
E-mails: Svetlana.Cojocaru@math.md, eugenia_burca@yahoo.com