Toward the Soundness of Sense Structure Definitions in Thesaurus-Dictionaries. Parsing Problems and Solutions*

Neculai Curteanu, Alex Moruz

Abstract

In this paper we point out some difficult problems of thesaurus-dictionary entry parsing, relying on the parsing technology of SCD (Segmentation-Cohesion-Dependency) configurations, successfully applied on six largest thesauri – Romanian (2), French, German (2), and Russian. **Challenging Problems: (a)** Intricate and / or recursive structures of the lexicographic segments met in the entries of certain thesauri; **(b)** Cyclicity (recursive) calls of some sense marker classes on marker sequences; **(c)** Establishing the hypergraph-driven dependencies between all the atomic and non-atomic sense definitions. Classical approach to solve these parsing problems is hard mainly because of depth-first search of sense definitions and markers, the substantial complexity of entries, and the sense tree dynamic construction embodied within these parsers. **SCD-based Parsing Solutions: (a)** The SCD parsing method is a procedural tool, completely formal grammar-free, handling the recursive structure of the lexicographic segments by procedural non-recursive calls performed on the SCD parsing configurations of the entry structure. **(b)** For dealing with cyclicity (recursive) calls between secondary sense markers and the sense enumeration markers, we proposed the Enumeration Closing Condition, sometimes coupled with New_Paragraphs typographic markers

transformed into numeral sense enumeration. **(c)** These problems, their lexicographic modeling and parsing solutions are addressed to both dictionary parser programmers to experience the SCD-based parsing method, as well as to lexicographers and thesauri designers for tailoring balanced lexical-semantics granularities and sounder sense tree definitions of the dictionary entries.

Keywords: dictionary entry parsing; parsing method of SCD configurations; recursive lexicographic segments; recursive calls of sense markers; Enumeration Closing Condition; soundness of sense structure definitions.

1 Thesaurus-Dictionary Parsing with SCD Configurations

This section goal is two-fold: to briefly introduce the *parsing method* of SCD (Segmentation-Cohesion-Dependency) *configurations*, which was applied to parse six largest Romanian, French, German, and Russian dictionaries [7], [4], [3], [5], [6], and to outline the issue of the present paper.

The *parsing method* of SCD *configurations* consists in applying *breadth-first* (completed with depth-first, stack-type) searching algorithms for the recognition and establishing the dependencies between the sense marker classes of dictionary entries [4], [3], [5], [6], [7]. In general, an SCD *configuration* (hereafter, SCD*config*) has the following computational components: • A set of *marker classes*: a *marker* is a boundary for a specific linguistic category; • A *hypergraph-like hierarchy* that pre-establishes the dependencies among the marker classes; • A *searching* (*parsing*) *algorithm*.

When applied to dictionary entry parsing, the method of SCD configurations merges the following sequence of (at least) three specific configurations (*i.e.* lexical-semantics sense levels): **(a)** The *first one*, abbreviated hereafter SCD*config*1, performs the segmentation and dependencies for the *lexicographic segments* [11 :2], [10] of each dictionary entry [4], [5], [7]. **(b)** Stepping down into the lexicographic segments of a thesaurus-dictionary entry, the *second* SCD *configuration* (SCD-

276

*config*2) usually parses the *sense description* segment, extracting its *sense tree* structure [4], [3], [5], [7]. Actually, the SCD*config*2 parses the entry sense definitions of larger lexical-semantics granularity in the sense description segment: primary, secondary, and literal / numeral enumeration senses. **(c)** The *third* SCD *configuration* (henceforth SCD*config*3) continues to refine the sense definitions of SCD*config*2, parsing each node in the generated sense-tree for obtaining the atomic definitions / senses (*i.e.* finest-grained meanings) of the dictionary entry.

We experienced the method of SCD *configurations* for *modeling* and *parsing*, with outstanding results (over 90% accuracy), on *six* largest, complex, and sensibly different thesaurus-dictionaries for *Romanian*: **DLR** (The Romanian Thesaurus – new format) [3], [4], [7], and **DAR** (The Romanian Thesaurus – old format) [4], [7], [16]; for *French*: **TLF** (Le Trésor de la Langue Française) [4], [7], [12]; for *German*: **DWB** (Deutsches Wörterbuch – GRIMM) [4], [7], [8], and **GWB** (Göthe-Wörterbuch) [4], [7], [8]; and for *Russian* – **DMLRL** (Dictionary of the Modern Literary Russian Language) [5], [6], [7].

The paper is organized as follows: *Section* 2 discusses the problems met in SCD*config*1 for recognizing the intricate or recursive structure of the *lexicographic segments* in German **DWB**, Romanian **DAR**, and French **TLF** thesauri. *Section* 3 examines situations of cyclicity (recursive) calls that may occur between secondary sense markers and sense enumeration(s) in **DAR**, **DMLRL**, and **DLR**, the transformation of the typographic *New_Paragraphs* into sense enumeration markers (*e.g.* in **DLR**, **DAR**, and **DMLRL**), and the solution provided by the *Enumeration Closing Condition* when recursive calls occur [5], [6], [4], [7]. *Section* 4 points out few examples of (atomic) definition parsing problems in **DLR**, **TLF**, and **DMLRL** [5], [4], [7]. *Section* 5 outlines the impact of the discussed parsing problems and solutions on both the robust parser construction and the soundness of lexicographic design for the largest thesaurus-dictionaries, obtained within the optimal and portable framework of SCD configurations.

277

2 Parsing the Lexicographic Segments on SCD-*Config*1

2.1 Intricate Lexicographic Segments in German DWB

The German **DWB** (Deutsches Wörterbuch – GRIMM) entries comprise a complex structure of the lexicographic segments, which provide a non-uniform and non-unitary composition [8]. A special feature is that **DWB** (Deutsches Wörterbuch) and **GWB** (Göthe-Wörterbuch) [8] lexicographic segments are composed of two parts: a first (optional) *root-sense* subsegment, and the *body* subsegment, which contains the explicit sense markers, easily recognizable. For **DWB**, the parsing of lexicographic segments is not at all a comfortable task since they are defined by three distinct means, displaying a rather intricate structure:

(A) *After* the *root-sense* of a **DWB** entry, or *after* the *root-sense* of a lexicographic segment, (a list of) italicized-and-spaced key-words are placed to constitute the *label* of the lexicographic segment *that follows*. Samples of such key-word labels for **DWB** lexicographic segments are: "*Form, Ausbildung und Ursprung*", "*Formen*", "*Ableitungen*", "*Verwandtschaft*", "*Verwandtschaft und Form*", "*Formelles und Etymologisches*", "*Gebrauch*", "*Herkunft*", "*Grammatisches*", etc., or, for **DWB** (most important) sense-description segment: "*Bedeutung und Gebrauch*" (or just "*Bedeutung*"). In the example below, they are marked in 25% grey.

Example 2.1.1. **GRUND**, *m.*, *dialektisch auch f. gemeingerm. wort; fraglich ist das geschlecht von got.* *grundus *in* grunduwaddjus, *vgl.* afgrundiþa; *sonst meist masc.: ahd.* grunt, crunt; *mhd.* grunt; *as.* grund; *mnd.* grunt *meist f., selten m.; mnl.* gront *meist m., selten f.; ndl.* grond; *afries.* grund, grond; *ofries.* grund; *wfries.* groun, grùwn; *ags.* grund; *engl.* ground; *anord.* grunnr *m.*, grund *f.; dän.* grund *comm. gen.; schwed.* grund; *als dem german. entlehnt gelten lit.* gruntas *m., preusz.* gruntan acc. m., grunte *f., lett.* grunts *m.*, grunte *f., poln. russ. slov. nlaus.* grunt *m.* *f o r m   u n d   h e r k u n f t.*

1) *für das verständnis der vorgeschichte des wortes ist die* z w i e g e s c h l e c h t i g k e i t

278

... ...

H. V. SACHSENHEIM *spiegel* 177, 30;

die neuen grundt zu der kirchen *zimm. chron.*[2] 2, 539, 36; du findest noch vil gar alter meür und grunt und thürn SIGMUND MEISTERLIN *in städtechron.* 3, 51, 14. *auszerschwäb. im obd. nur selten:* mosige grunde SEBIZ *feldbau* (1579) 149. *anders, als rein graphische erschei-nung versteht sich das fehlen des umlautzeichens in md. texten; häufig z. b. bei* LUTHER: grebt die grunde 1, 148; drey starcke grund 6, 290. *b e d e u t u n g.* *die bedeutungsgeschichte des wortes läszt sich schwer aufbauen, weil ihre wesentlichsten etappen in vorgeschichtliche zeit fallen. die auch auszerdeutsch altbezeugten verwendungen im sinne von 'tiefe' (s. u.* I) *und im sinne ron 'erde' (*II) *stellen offenbar die beiden cardinalen bedeutungssträng dar. aber auch die bedeutung 'tal-, wiesengrund' (*III), *anscheinend auf der*

... ...

... ...hat *(s. u.* II A 1 a). *nach* JAC. GRIMM *liegt der unterschied darin, 'dasz* gr. *mehr nach innen geht,* boden *die ober-fläche bezeichnet' (th.* 2, 211). *das trifft mehrfach zu; doch erschöpft diese unterscheidung einer mehr räumlichen und mehr flächenhaften vorstellung die sache nicht.*

I. grund *bezeichnet die feste untere begrenzung eines dinges.*

A. *grund von gewässern; seit ältester zeit belegbar: profundum (sc. mare)* crunt *ahd. gl.* 1, 232, 18; latid thea odra *(fisch)* eft an gr. faran *Hel.* 2633.

1) *am häufigsten vom meer (in übereinstimmung mit dem anord. gebrauch):*

... ...

In Ex. 2.1.1 above, these notions are illustrated as follows: between the entry lemma **GRUND** and the label "*f o r m u n d h e r k u n f t*", it spans the *root-sense subsegment* of the first lexicographic segment for the entry "**GRUND**". The key-words "*f o r m u n d h e r k u n f t*" rep-resent the first label for the *first segment* of the lemma, described with several sense markers, among which the first one is "**1)**". The segment "*f o r m u n d h e r k u n f t*" ends when the label "*b e d e u t u n g*" occurs for the next lexicographic segment. Between this label and the

279

effective description of the segment senses, which begins with the sense markers "**I.**" ... "**A.**" ... etc., it spans the *root-sense* of the segment labeled with "*b e d e u t u n g*". Thus each lexicographic segment in **DWB** may contain, optionally, in a "preamble", the root-sense (subsegment) description of that segment. The key-words (or a list of key-words) placed at the end of a segment correspond to (and represent) the label of the lexicographic segment *that follows*.

(**B**) *The second* way to specify the lexicographic segments in **DWB** is expressed as follows: *after* the primary sense markers, there are specified those key-words representing *the label* of the lexicographic segment that follows. The example 2.1.2 is enlightening:

Example 2.1.2. GEBEN, *dare.*

I. *Formen, ableitungen, verwandtschaft.*

1) *es ist ein allgemein, aber ausschlieszlich germanisches wort: goth.* giban (*praet.* gaf), *ahd.*

... ...

II. *Bedeutung und gebrauch.*

1) geben *und* nehmen, *die beiden sich ergänzenden gegenstücke, verdienen die erste...*

... ...

The entry **GEBEN** of **DWB** has the *Latin definition* "*dare*", which is at the same time the root-sense of the entry. The first segment (which begins with the marker "**I.**") is labeled with "*Formen, ableitungen, verwandtschaft*", while the the second segment (which begins with the marker "**II.**") has the label "*Bedeutung und gebrauch*". This is the proper *sense description segment* of the lemma **GEBEN** from **DWB**, actually.

(**C**) The *third* (and most frequent) way to identify the lexical description segment(s) of a **DWB** entry is simply the lack of a segment label at the beginning of the sense description segment. By default, after the entry *root-sense segment* (which can be reduced to the Latin definition, *i.e.* the translation of the German word-lemma), the sense-description segment comes without any "*Bedeutung*" label, introducing explicit sense markers and definitions.

Example 2.1.3. BESUCHEN, *ahd.* pisuochan (GRAFF 6, 84),

mhd. besuochen, *nnl.* bezoeken, *schw.* besöka, *dän.* besöge.

1) *den jägern,* das wild besuchen, *aufspüren.*

2) einen ort besuchen, *mhd.* einen turnei besuochen. *Engelh.* 2359; *nhd.* die kirchen, spielhäuser, theater besuchen, *franz.* fréquenter; das sie dein haus und deiner unterthanen . . .

. . . .

While the lexicographic segment structure is not easy to be obtained for **DWB** (SCD-*config*1), as shown in this subsection, the dependency hypergraph for the sense description segment (SCD*config*2), represented in [4 :Fig. 6], looks more feasible when the former task has been achieved.

2.2 Recursive Structure of Lexicographic Segments in DAR

We present here the recursive configuration for two lexicographic segments in **DAR** (the old format of **DLR**): the *French* and *Nest* segments.

The *French* segment [4], [7] "looks" like the *sense description* segment, while the *Nest* (Romanian "*cuib*") segment delivers, at smaller dimensions, a similar (thus *recursive*) lexicographic structure as that of **DAR** general entry.

Example 2.2.1. The entry **LĂMURÍ** [Eng: *elucidate, explane, clear up*] in **DAR**, followed by the *French* segment, the *sense description SenseSeg* segment, and a *Nest* segment (the segment and sense markers are highlighted in 25% grey:

LĂMURÍ vb. IV^a. 1°. *Purifier, raffiner.* 2°. *Préciser; fixer; éclairer; s'éclairer, s'élucider.* 3. *Expliquer.* 4°. *Distinguer, apercevoir.*

1°. T r a n s. (Despre metale, etc.) A curăţi prin foc de corpurile necurate; p. g e n e r. a c u r ă ţ i, a l i m p e z i, a p u r i f i c a. *Ca aurul în ulcea i-au lămurit.* MINEIUL (1776) 154^2/1. *În cuptoriul înfrânării ţi-ai lămurit trupul.* ib. 45^1/$_2$. *Argintarul lucrează argintul lămurindu-l prin foc cu plumb, care trage arama.* I. IONESCU, M. 714.

. . . .

[Şi: **lămurá** † vb. I^a. *Hierul* (= fierul) *ce lămura* [făurarul]. herodot, 28. || A d j e c t i v e: **lămurít** (cu negativul **nelămurit**), **-ă** = curăţit, limpezit, purificat; clarificat, desluşit, limpezit, explicat, clar, limpede. (Ad 1°) *Argintul lămuritu iaste cuvântul lu Dumnezeu.* coresi, ev. 318/$_5$; cf. dosofteiu, ps. 38. *Tăia iarăşi bani de argint lămurit.* herodot, 262. *Argintul cel cu foc lămurit.* biblia (1688) 372$_2$. *Laptele cel lămurit.* mineiul (1776) ... ...

... ...

... ...*Să-şi facă o idee lămurită de sine însuşi.* marcovici, c. 11/$_1$. *Adevăruri lămurite.* i. ionescu, c. vi. *Să-i dea mai lămurit răspuns.* c. negruzzi, i 197. *Hotărirea împărătesei era lămurită.* ispirescu, l. 307; – (în poezia populară cu caracter mistic) **lămurát, -ă.** *Să rămână curat, Lămurat, Cum Dumnezeu l-o dat.* marian, d. 34, 39, 125; – **lămuritór,-oáre** adj. = curăţitor, limpezitor, purificator; care lă- mureşte, care desluşeşte, care clarifică. *Dovezi lămuritoare.* donici, f. 44. *Lămuritoare cuvinte de dreptate.* c. negruzzi, II 297. | A b s t r a c t: **lămuríre** s. f. = acţiunea de a lămuri; limpezire, curăţire, purificare; claritate, desluşire, explicaţiune. **Cu lămurire** loc. adv. = în mod lă- murit, clar, limpede. *Urmează a se face socotealile tovărăşiei cu multă lămurire.* pravila (1814) 87. *Am văzut cu lămurire.* uricariul, i 216/$_2$. *Acest adevăr rămâne cu lămurirea cuvenită.* i. ionescu, c. 243. *Trebue să dăm mai întâi o lămurire despre acest rege.* c. negruzzi, i. 177. *Să aibă la cine alerga la lămuriri, când lecţia ar fi fost prea grea.* g. vifor, luc. iv 309. (Învechit) *Lămurire a socotelelor* = lichidare. pontbriant, barcianu. *Despre Bârlad... iarăşi avem preţioase lămuriri.* bogdan, c. m. 2.].

2.3 Recursive Configuration of Lexicographic Segments in TLF

Example 2.3.1. "**Rem.**", "**Étymol. et Hist.**", and "**DÉR.**" lexico- graphic segments in the **TLF** entry **ÉLÉPHANT**. Along with lexical- semantics *sense trees* (with primary, secondary, and enumeration- described subsenses) inside several lexicographic segments, see also the **Rem.** segment inside the last **DÉR.** segment!

. . . .

Rem. On rencontre ds la docum. **a)** *Éléphantarque,* subst. masc., antiq. Chef d'une compagnie de soldats montés sur des éléphants. *Deux armées entières : trente mille hommes d'un côté, onze mille de l'autre, sans compter les éléphants avec leurs éléphantarques* (FLAUB., *Corresp.,* 1860, p. 384). **b)** *Éléphante,* subst. fém. rare. Femelle de l'éléphant. *Emploi métaph.* Femme lourde qui manque de souplesse (*cf.* HUYSMANS, *Art mod.,* 1883, p. 133). **c)** *Éléphas,* subst. masc. Nom scientifique de l'éléphant. *L'*"*Elephas meridionalis*", *comme d'ailleurs la plupart des éléphants qui se baladaient autrefois en Europe, n'avait pas de fourrure* (FARGUE. *Piéton Paris,* 1939, p. 129).

Prononc. et Orth. : [e l e f ä]. Ds *Ac.* dep. 1694. **Étymol. et Hist. 1.** 1121 *elefant* (*Ph. Thaon Best.,* 1416 ds T.-L. : une beste truvum qu'**elefan** apelum); **2.** 1825 p. ext. " personne à la démarche lourde et peu gracieuse " (BRILLAT-SAV., *Physiol. goût,* p. 227); **3.** 1560 *elephant de mer* (PARÉ, éd. Malgaigne, *Discours de la licorne,* III, chap. XI, p. 502). Empr. au lat. *elephantus* " éléphant ", en a. fr. on rencontre plus souvent la forme *olifant**. **Fréq. abs. littér. :** 926. **Fréq. rel. littér. :** XIXe s. : a) 1 789, b) 2 429; XXe s. : a) 678, b) 701.

DÉR. 1. Éléphanteau, subst. masc. Petit de l'éléphant; jeune éléphant. *Des éléphanteaux se séchant au soleil* (GREEN, *Journal,* 1938, p. 144). – [e l e f ä t o] – 1re attest. XVIe s. (Ant. du Pinet ds DELB. *Rec.* ds *DG*); de *éléphant,* suff. *-eau**. – Fréq. abs. littér. : 1. **2. Éléphantesque**, adj. Comparable à l'éléphant; qui est, en poids et en taille, supérieur à la moyenne. Synon. *énorme, gigantesque, gros, monumental. C'est une dame* [*la comtesse Fontaine*] *aux proportions éléphantesques, dans la fleur de la soixantaine* (COPPÉE, *Toute une jeun.,* 1890, p. 220). – [e l e f ä t ɛ s k] – 1re attest. 1890 *id.*; de *éléphant,* suff. *-esque**. **3. Éléphantin, ine**, adj. **a)** Relatif à l'éléphant; qui rappelle l'éléphant. *L'épiderme éléphantin des mendiants* (HUYSMANS, *Là-bas,* t. 2, 1891, p. 20). *Belle autrefois* [*Taitou*], *de cette beauté grasse que recherchent les Orientaux, mais devenue avec le temps d'une corpulence éléphantine* (THARAUD, *Passant Éthiopie,* 1936, p. 110). . . . .

283

... ...

L'énorme Suédoise beauté éléphantique (SIMONIN, BAZIN, *Voilà taxi!* 1935, p. 141). Qui est atteint d'éléphantiasis. Synon. *éléphantiasique, éléphantiaque.* Attesté ds LITTRÉ, *Ac. Compl.* 1842, BESCH. 1845, *Lar. 19e − 20e* et QUILLET 1965. **Rem.** Certains dict. attestent l'emploi subst. dans le sens de " éléphantiasique, éléphantiaque ". – Dernière transcr. ds LITTRÉ : é-lé-fan-ti-k'. – 1res attest. **a)** XVe s. subst. (*Valenciennes,* ap. La Fons. ds GDF.), **b)** adj. " d'éléphant " 1506-1516 (FOSSETIER, *Chron. Marg.,* ms. Bruxelles, 10512, IX, II, 5 ds GDF. *Compl.*); *de éléphant,* suff. *-ique*.*
BBG. – GILI GAYA (S.). Miscelánea. *Revista de Filologia española.* 1949, t. 33, pp. 145-146. – GOTTSCH. Redens. 1930, p. 42, 121. – GRIMAUD (F.). Pt gloss. du jeu de boules. *Vie Lang.* 1968, p. 194. – ROG. 1965, p. 42, 178, 180. – ROMMEL 1954, p. 98. – SPITZER (L.). Über einige Wörter der Liebessprache. Leipzig, 1918, p. 56. – VAGANAY (H.). Qq. mots peu connus. *In :* [*Mél. Chabaneau (C.)*]. *Rom. Forsch.* 1907, t. 23, p. 226 (*s.v. éléphantin*).

Example 2.3.2. Highly refined description of the sense tree for the "**Étymol. et Hist.**" lexicographic segment in the **TLF** entry **VENIR**.

... ...

Prononc. et Orth.: [v ə n : R], *(il) vient* [-v j ɛ̃]. Att. ds *Ac.* dep. 1694. Conjug. ind. prés.: *je viens, tu viens, il vient, nous venons, vous venez, ils viennent*; imp.: *je venais*; passé simple: *je vins*; fut.: *je viendrai*; passé composé: *je suis venu*; plus-que-parfait: *j'étais venu*; passé ant.: *je fus venu*; futur ant.: *je serai venu,* cond.: *je viendrais*; cond. passé: *je serais venu*; subj. prés.: *que je vienne*; imp.: *que je vinsse* ; passé *que je fus venu*; plus-que-parfait: *que je fusse venu*: impér.: *viens, venons, venez*; passé: *sois venu, soyons venu, soyez venu*; inf. prés.: *venir*; passé: *être venu*; part. prés.: *venant*; passé: *venu, -ue; étant venu.* **Étymol. et Hist. A. 1.** *Venir a* + subst. marquant le terme du mouvement **a)** *ca* 880 " se déplacer pour arriver près du point de référence " (*Eulalie,* 28 ds HENRY *Chrestomathie,* p. 3); *ca* 1050 *en venir* " *id.* " (*Alexis,* éd. Chr. Storey, 113); spéc. 1690 " atteindre un certain point " (FUR.); 1842 mar. (*Ac. Compl.:* **Venir** au vent [...]. **Venir** à bâbord ou à tribord); **b)** 1176-81 fig. *venir à* + subst. ab-

284

str. " apparaître dans l'esprit, être conçu " (CHRÉTIEN DE TROYES, *Charrete*, éd. M. Roques, 495); **2.** *venir de* + subst. indiquant l'origine du mouvement **a)** *ca* 1050 " arriver en provenance de " (*Alexis*, 251); **b)** *ca* 1170 fig. " provenir, découler de " (CHRÉTIEN DE TROYES, *Erec*, éd. M. Roques, 4392); spéc. *ca* 1250 " descendre (de quelqu'un) " (*Grant mal fist Adam*, I, 28 ds T.-L.); 1606 " dériver (d'un mot) " (NICOT, *s.v. bohourd*); **c)** loc. 1176-81 *don vos vient?* (CHRÉTIEN DE TROYES, *Charrete*, 137); 1580 *d'où venoit celà* (MONTAIGNE, *Essais*, I, 20, éd. P. Villey et V.-L. Saulnier, p. 96); 1664 *d'où vient que* (MOLIÈRE, *Tartuffe*, I, 1); **3. a)** *ca* 1050 *venir* sans compl. de lieu (*Alexis*, 467); *ca* 1050 *faire venir qqn* " lui demander de venir " (*ibid.*, 335); 1539 *venir au secours* (EST.); . . . . . .

. . . .

D. Avec l'inf. *venir* servant de simple auxil. **1.** fin Xe s. *venir* + inf. " faire en sorte de " (*Passion*, 407); **2.** *ca* 1050 *venir a* surtout à la 3e pers. + inf. " se trouver en train de " (*Alexis*, 47); **3.** *ca* 1225 *venir de* + inf. " avoir juste fini de " (GAUTIER DE COINCI, *Mir.*, éd. V. Fr. Koenig, I Mir 12, 44). Du lat. *venire* " venir ", " arriver, se présenter ", " parvenir à ", " venir à quelque chose, venir dans tel ou tel état " et " en venir à ". **Fréq. abs. littér.:** 98 961. **Fréq. rel. littér.:** XIXe s.: a) 142 843, b) 153 800; XXe s.: a) 144 519, b) 129 650. **Bbg.** BAMBECK (M.). Galloromanische Lexikalia aus volksprachlichen mittelalterlichen Urkunden. *Mél. Gamillscheg (E.)* 1968, p. 69. – DABÈNE (L.). *Aller* et *venir*: de la ling. à la didact. *Mél. Pottier (B.)* 1988, pp. 217–224. – DEJAY (D.). Les Rel. actancielles appréhendées à travers un corpus de verbes fr. Thèse, Nancy, 1986, pp. 37–42. . . . .

It is clear that any dictionary parser should recognize first (explicitly expressed or by default) the lexicographic segments within the first SCD parsing configuration.

3 Parsing Problems at the Level of Primary and Secondary Sense Definitions on the SCD-*Config*2

3.1 Cyclicity Calls between Secondary Sense Markers and Literal Enumeration in DMLRL

Example 3.1.1. It is common in **DMLRL** that (primary and) secondary senses to be refined by literal enumeration. For the reverse, atypical and uncommon situation, where the literal enumeration is further refined through secondary sense markers // and ◇, the most interesting case we met in **DMLRL** is the entry **БЫ** [9 :844], under the primary sense no. "**3.**".

. . . .

2. В придаточной части сложного предложения обозначает действие, обусловливающее 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.* ◇ Если бы, когда бы, хоть бы и т. п. *О, если бы когда-нибудь Сбылись поэта сновиденья!* Пушк. Посл. к Юдину. [Николка:] *Хоть бы дивизион наш был скорее готов.* Булгаков, Дни Турб. ◇ 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.* ◇ 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.* Пушк. Арап Петра Вел. [Варя:] *Не опоздать бы только к поезду.* Чех.

Вишн. сад. **б)** Пожелание. *Условие я бы предпочел не подписывать.* Л. Толст. 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 повытчики: "Потише бы надо, бесстрашной, здесь не конюшня".* А. Н. Толст. 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. Кащ. цепь. **г)** Желаемость целесообразного и полезного действия. ◇ С неопр. ф, глаг. *Вам бы вступиться 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.

The parsing result of this part of **БЫ** entry is the following:
\<entry\>
\<list\>БЫ 1. ◇ ◇ ◇ ◇ ◇ 2. 3. a) ◇ ◇ // ◇ ◇ б) ◇ ◇ ◇ в) г) ◇ ◇ n-23\</list\>
\<sense value="БЫ"class="0"\>
\<definition\> (сокращенно \<b\>Б\</b\>), частица. В сочетании с глаголами в форме прошедшего времени образует сослагательное наклонение. \</definition\>
\<sense value="1."class="4"\>
. . . .

<sense value="3."class="4">

<definition> Обозначает различные оттенки желаемости действия; </definition>

<sense value="а)"class="5">

<definition> Собственно желаемость. Учился бы сын. Были бы дети здоровы. </definition>

<sense value="◇"class="8">

<definition> Если <spaced> б ы </spaced>, когда <spaced> б ы </spaced>, хоть <spaced> б ы </spaced><spaced> и </spaced> т. п. О, если бы когда-нибудь Сбылись поэта сновиденья! Пушк. Посл. к Юдину. [Николка:] Хоть бы дивизион наш был скорее готов. Булгаков, Дни Турб. </definition>

< /sense>

<sense value="◇"class="8">

<definition> С неопр. ф. глаг. Полететь бы пташечке К синю морю; Убежать бы молодцу в лес дремучий. Дельв. Пела, пела пташечка.. ... ... </definition>

< /sense>

<sense value="//"class="6">

<definition> Употр. для выражения опасения по поводу ... ... </definition>

<sense value="◇"class="8">

<definition> 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. </definition>

< /sense>

<sense value="◇"class="8">

<definition> Только <spaced> б ы</spaced> (б) не. - По мне жена как хочешь одевайся, .. только б не каждый месяц ... ... </definition>

< /sense>

< /sense>

< /sense>

<sense value="б)"class="5">

<definition> Пожелание. Условие я бы предпочел не подписывать. Л. Толст. Письмо А. Ф. Марксу, 27 марта 1899. </definition>

<sense value="◊"class="8">

<definition> С неопр. ф. глаг. Поохотиться бы по-настоящему, на коня бы денег добыть, - мечтал старик. Г. Марков, Строговы. </definition>

< /sense>

<sense value="◊"class="8">

<definition> В сочетании с предикативными наречиями со знач. долженствования, необходимости, возможности. ... ... </definition>

< /sense>

<sense value="◊"class="8">

<definition> Только <spaced> б ы</spaced> (б), лишь бы, Употр. со знач. желательности действия. [Скалозуб:] Мне только бы досталось в генералы. Гриб. Горе от ума. </definition>

< /sense>

< /sense>

<sense value="в)"class="5">

<definition> Желание-просьба, совет или предложение.... ... ... </definition>

... ...

< /sense>

< /sense>

< /sense>

<EtymologicalPart>

<p> – Срезневский: <spaced> б ы</spaced>; Лекс. 1762: <spaced> б ы</spaced>.</p>

< /EtymologicalPart>

< /entry>

The *Enumeration Closing Condition* (ECC) represents a deterministic, computational constraint devoted to check the sound termination (*i.e.* in a deterministic, finite number of steps) of the literal or numeral enumeration marker list, when higher-level sense markers break into this list. When this happens, contextual look-ahead verifications are needed

to obtain the correct closing of the enumeration list. More precisely, ECC means that whether after a certain (let us say, *current*) *letter* in the sense enumeration marker list occur higher-level sense markers (on the dependency hypergraph), then one should look ahead in the sense marker sequence until the *next* letter of the same enumeration type occurs. If such a letter does exist and follows *monotonously* (in the alphabetic order) the current one in the enumeration list, then the enumeration should continue. Otherwise, *i.e.* the letter does not exist or it begins another enumeration, of the same or another kind as the current one, then the ECC holds and the current literal enumeration must be closed. For instance, in the Romanian **DLR**, with the filled and empty diamonds ♦, ◇ as secondary sense markers, the enumeration list **a) b) c)** ◇ ♦ ◇ ◇♦ ◇ **d)...** should continue, while the marker sequence **a) b) c)** ◇ ♦ ◇ ◇♦ ◇ **a)...** should close the first literal enumeration (see also [5], [4], [6], [7]).

The same is true if non-enumerable sense markers (such as ♦, ◇) are replaced by another enumeration of sense markers, be it of numeral or another literal type. Two different enumerations, a standard, *literal* one, and a *numeral* one coming from transforming the *New_ Paragraphs* into sense markers, are illustrated by the entry **CAL** of the Romanian **DAR** thesaurus.

3.2 Cyclicity Calls between Secondary Sense Markers, Literal Enumeration, and *New_ Paragraphs* in DAR and DLR

Example 3.2.1. [7 :Chap. 9] In the **DAR** entry of the preposition **DE** (En: *of, by, for, to, from...*, Fr: *de*) we encounter the situation of the *NewPrg* (*New_ Paragraph*) use as *numeral* enumeration, pursued or not by another sense marker: *NewPrg* introduces component subsenses in the (Romanian) *RomSeg* segment, which follows the (French) *FreSeg* segment.

    <FreSeg>
    NewPrg **DE** prep. A. I. 1°. a). Marque le lieu d'où part une action...

    ... ...

NewPrg F. Elément de nombreaux mots composés.

< /FreSeg>

<RomSeg>

NewPrg De neaccentuat în frază şi proclitic, formează o singură unitate fonetică. . .

NewPrg Substantivul în legătură cu *de* rămâne de obiceiu nearticulat, dacă nu e urmat de un atribut al său. . .

NewPrg Cuvântul de sub regimul lui *de* are de cele mai multe. . .

. . . .

{*RomSeg* contains 14 paragraphs introduced by *NewPrg*, followed by *RomSeg* and *SenseSeg*. Hence:}

. . . .

< /RomSeg>

<SenseSeg>

NewPrg A. Construcţia prepoziţională are funcţiunea sintactică. . .

NewPrg I. Ca determinare privitoare la spaţiu sau la timp.

NewPrg 1^0. Complemente circumstanţiale de loc.

NewPrg a) Complementul circumstanţial de loc răspunde la întrebarea u n d e?...

. . .

< /SenseSeg>

Example 3.2.2. [7 :Chap. 9] The illustrative example of entry **CAL** from **DAR** is important and rather complex, showing the use of *NewPrg* markers as sense *numeral* enumeration, interleaving with the already existing sense *literal* enumeration.

NewPrg **CAL** s.m. *Cheval.*

NewPrg 1°. Numele generic al speţei cavaline; s p e c. individ masculin...

. . .

NewPrg Adecă amù cailoru zăbalele în gură lă. . .

. . . {a large block of definitions and *DefExems* of the entry **CAL**}

NewPrg În compoziţii:

NewPrg a.) (Entom.) **Cal-de-apă** = o specie a c a l u l u i - d r a c u l u i, numită. . .

. . .

291

NewPrg **Calul-dracului** = a.) insectă cu corpul lung... | (De aici) Babă rea... ; –b.) = **cal-de-apă**...

...

NewPrg **Calul-popii** = a.) c a l u l-d r a c u l u i...; –b.) = cal-de-apã... Insectă lungă şi cu aripile pătate...

NewPrg **Cal-turtit** = c a l u l-d r a c u l u i...

NewPrg b.) (Zool.; la românii din A.-U.) **Cal-de apă** s. (după germ. Nilpferd) –**cal-de-Nil** = h i p o p o t a m LB., BARCIANU ...

...

NewPrg **Cal-de-mare** = *hyppocampus brevirostris...*

...

NewPrg 2°. P. a n a l. (Mor.) *Caii* cu spetezele ţin coşul şi alcătuesc...

...

The sense dependency subtree between the sense markers ”1°.” and ”2°.” looks as follows (Fig. 1. below):
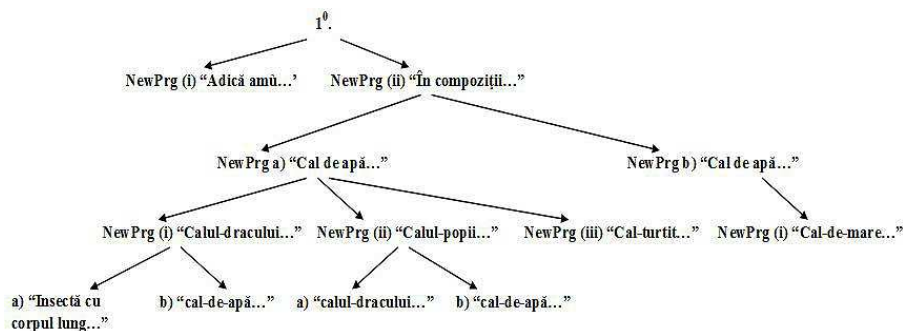

Figure 1. Partial sense dependency subtree of the **CAL** entry in **DAR**

A good exercise of solving this problem is to parse correctly the entry **CAL** in **DAR**, partially shown below. This complete representation extends a slightly less refined output obtained by the automatic SCD-based parser.

<entry>
<sense value="CAL" class="0">

<definition> s.m. <i>Cheval.</i></definition>
<sense value="1°." class="12">
<definition> Numele generic al speţei cavaline; s p e c. individ masculin... M â n z u l dacă nu se ţine de prasilă</definition>

... ...
<sense value="NewPrg" class="1°.+i.">
<definition><i>Adecă amù cailoru zăbalele în gură lă </i>[= le]<i>băgăm COD. VOR. 122/13.</i>
<i>Nu fireţi</i>...</definition>
< /sense>
<sense value="NewPrg" class="1˚.+ii.">
<definition>În compoziţii:</definition>
< /sense>
<sense value="a.)" class="16">
<definition> (entom.) Cal-de-apă = o specie a c a l u l u i-d r a c u l u i, numită şì c ă l u ţ - d e - a p ă, c a l u l - d r a c u l u i, c a l u l - p o p i i, c ă l u ţ, p ă u n i ţ ă, p i ţ i n g ă u l - d r a c u l u i, s c ă l u ş - d e - a p ă, ţ â n ţ a r - d e -apă (<i>Calopteryx splendens</i>). MARIAN, INS. 559-560, cfr. H. XI 195.</definition>
< /sense>
<sense value="NewPrg" class="a+i.">
<definition>Calul-dracului =
<sense value="a.)" class="16"> <definition> insectă cu corpul lung şi turtit, de coloare galbenă închisă, cu aripile lungi şi late, şi străvezii ca o păioară. Zboară foarte iute, mai ales pe de-asupra apelor. Se mai numeşte: c a l u l - p o p i i, c a l-t u r t i t, c o b i l i ţ ă, c ă l u g ă r i ţ ă (H. x 355) (<i>Libellula depressa</i>). MARIAN, INS. 558 ž. u., „un fel de ţânţar mare" H. IX 52. Cfr. H. I 59, IV 54, V 116, IX 437, 473, x 259, XII 27, 374.</definition>
< /sense>
<sense value="#" class="22"> <definition> <i>A fi ca calul-dracului</i>, se zice de un om neastâmpărat. marian, ins. 565. </definition>
< /sense>

<sense value="|" class="20"> <definition> (De aici) Babă rea, cfr. n e a g a r e a. Cfr. coşbuc, b. 92. <i>Baba asta (vrăjitoare) erà calul-dracului<i>: afurisită şi rea. PAMFILE, J. I, cfr. ZANNE, P. II 3;– </definition>

< /sense>

<sense value="b.)" class="16"> <definition> c a l-d e-a p ã. MARIAN, INS. 559. </definition>

< /sense>

< /definition>

< /sense>

<sense value="NewPrg" class="a+ii.">

<definition>Calul-popii =

<sense value="a.)" class="16">

<definition> c a l u l - d r a c u l u i. MARIAN, INS. 558; </definition>

< /sense>

<sense value="b.)" class="16">

<definition> c a l-d e-a p ă. id. ib. 559. Insectă lungă şi cu aripile pătate, având ochii mari. H. VII 481; cfr. H. I 59, II 307, 227, 117, V 280, X 151, 355, 498, XII 226, 429, XIV 350, 397, 467. </definition>

< /sense>

< /definition>

< /sense>

<sense value="NewPrg" class="a+iii.">

<definition> Cal-turtit = c a l u l - d r a c u l u i. MARIAN, INS. 558.

< /definition>

< /sense>

<sense value="NewPrg" class="b+i."="a+iv.">

<sense value="b.)" class="16">

<definition> (Zool.; la Românii din A.-U.) Cal-de apă s. (după germ. Nilpferd) -de-Nil = h i p o p o t a m LB., BARCIANU. </definition>

< /sense>

<sense value="NewPrg" class="1°.+iii."="b+ii.">

<definition> Cal-de-mare= <i>hyppocampus breviro-stris</i>. BARCIANU. <i>Cai-de-mare, albi ca spuma</i>, EMI-NESCU, p. 114. </definition>

< /sense>

< /sense>

<sense value="2°." class="12">

<definition> P. a n a l. (Mor.) <i>Caii</i> cu spetezele ţin coşul şi alcătuesc... </definition>

<sense value="||" class="20">

<definition> (Dulgh.) S c a u n u l cu cleştele de strâns... </definition>

< /sense>

< /sense>

.. ..

<sense value="4°." class="12">

<definition> (Cor.) Numele unui danţ ţărănesc...</definition>

< /sense>

< /sense>

< /entry>

The partial sense marker sequence in the above representation is the following: ... ... 1°. i. ii. a.) BoldDefMark i. BoldDefMark a.) # | b.) ii. BoldDefMark a.) b.) iii. BoldDefMark i. b.) BoldDefMark ii. BoldDefMark 2°.|| ... . We remark the distinct role of *NewPrg* typographic-type sense marker in the context of subsequences New-Prg DefMark Enum and NewPrg Enum DefMark: the first sequence introduces lower, local level dependencies, while the second one defines higher level ones, all depending on the look-ahead sense markers. The subsequence contextual analysis and two passages along the whole sense marker sequence provide the correct sense dependencies.

Such an approach would be rather difficult to be implemented within the classical, formal grammar-based grammars, since it works depth-first search on *all* the dictionary forms, definition bodies, and sense markers, while ECC and the emphasized contextual analyses on the

marker subsequences are performed on the *bare sequence* of the *extracted sense markers* from the entry. Dependency structures such as in the entry **CAL** of **DAR** represent, in our evaluation, lexicographic mistakes or inadequacies at the dictionary design stage; parsing it correctly with the method of SCD configurations is both a technical challenge and also a warning for more sound and careful sense structure constructions in the greatest thesaurus-dictionaries.

Example 3.2.3. While the secondary sense markers are naturally refined through literal enumeration in **DLR** thesaurus, we found yet the reverse, atypical situation, *e.g.* for the entries **DOAR**, **DOÁSCĂ** (fragment below), and especially **LUMÍNĂ** (fragment below), where the recursive calls for literal enumeration is mixing with secondary sense markers. The first literal *enumeration* is notably further marked by another, *numeral* enumeration, introduced by the *NewPrg* (*New_Paragraph*) markers.

DOÁSCĂ s. f. **1.** Nume dat unor scânduri, unor bucăţi de lemn sau unor obiecte făcute din acestea:

a) (Popular) Scândură (**1**). *Strunga de muls e închisă cu o doscă, scândură, până se pun la muls păcurarii.* DR. II, 336. *Şi-a lăsat abatajul nearmat … şi coperişul fără doasce.* DAVIDOGLU, M. 70. *Îl pun caşul undeva pe-o doscă.* Com. din LUGAŞU DE JOS – ALEŞD, cf. ALR I 1 853/61, 65, 80, 107. ♦ *Gard de doşte* = gard de scânduri. Cf. ALR II/I h 267/64, ALRM II/I h 359/64. … … … … … … … … …… … … … … …… … …

g) (Învechit) Copertă de carte, confecţionată din lemn şi învelită în piele. *Mi se încredinţase un dulap nou-nouţ … Era încărcat cu fel de fel de bucoavne vechi, cu doascele de lemn.* CIAUŞANU, R. SCUT. 55, cf. ARH. FOLK. VII, 121. ◇ Loc. adv. **Din doască-n doască** = în întregime, de la un capăt la altul. *Secretarul întreprinderii luă traducerea şi o citi din doască-n doască.* AGÎRBICEANU, A. 53.

2. (Regional) Perete subţire (Bonţ – Gherla). Cf. PAŞCA, GL. **3.** (Regional) Vas făcut din coajă de dovleac. *Sus pe corlată … trei doaşte de dovlete.* PLOPŞOR, C. 39. … … … … … …… … … …

LUMÍNĂ s.f. **A.** (Predomină sensul concret de radiaţie; în opoziţie

cu î n t u n e r i c)

I. (Adesea cu determinări calificative) Radiaţie care face corpurile vizibile.

1. (Ca atribut al universului, al naturii ambiante; componentă a lumii înconjurătoare) *Lăudaţil toate stealele şi ... ... gonească Cât va fi câmp de gonit Şi lumină de zărit".* ALECSANDRI, O. I, 8.

a) (Ca radiaţie solară, element al peisajului diurn) *Voi întoarce lumira soarelui de cătră voi, de va fi întunrearecu* (a. 1600). CUV. D. BĂTR. II, 49/9. *Lumina soarelui face dzua.* PRAV. 141. ... ... *Deopotrivă se găseşte-n toate Amestecată umbră şi lumină.* ISANOS, V. 281. ♦ L o c. a d j. **De lumină = a)** luminos, sclipitor; s p e c. (despre ochi) strălucitor. *Deunăzi ... mă simţii cufundat ca într-un nor întunecos ... Ancuţo! tu ai prefăcut acel nor în soare de lumină! Tu ai deşteptat în sufletu-mi o viaţă necunoscută!* ODOBESCU, S. I, 143. ... ... *Ochi de lumină avea fiul lui Ieronim, privirea lui în noapte fulgera.* ROMÂNIA LITERARĂ, 1970, nr. 93, 17/3 **; b)** (despre un spaţiu, un loc) în care pătrunde lumina **(A I 1),** plin de lumină *Acest loc ... era pe atunci, în 1650, un ochi de lumină în mijlocul marelui codru al Căpoteştilor.* IORGA, C. I. II, 5 **; c)** (despre plante) care trăieşte la lumină **(A I 1).** *După o fază de 2-3 ani cu floră de buruieni de lumină, urmează faza de fâneaţă cu ierburi cu rizomi.* CHIRIŢĂ, P. 71. ♦ L o c. a d v. **Pe** (sau, rar, **la) lumină** = în timpul zilei **(I 2),** de ... ... ARHIVA R. I, 87/20. *A înviat din morţi ..., Lumina ducându-o Celor din morminte!* EMINESCU, O. IV, 359. *Zâmbetul sfânt al martirului care-ntrevede ... lumina vieţii eterne.* CARAGIALE, O. II, 64. (Contextul aduce sensul figurat privind viaţa interioară a individului) *Cine va îmbla zioa nu se va poticni ...; iară cine va îmbla noapte poticni-se-va, că lumină nu iaste întru el.* CORESI, EV. 95.

b) (Ca radiaţie reflectată de lună; element al peisajului nocturn) *Luna, ... fire are lumina ce iase den ea să turbure udăturile trupului.* CORESI, EV. 81. ... ...

297

Mare şi minunată este lucrarea luminii lunii asupra feţii pământului şi a sănătăţii locuitorilor lui. EPISCUPESCU, PRACTICA, 335/2. . . .
. . . .

The discussion and solution is similar as that for the entry **CAL** from **DAR**.

Example 3.2.4. In this **DLR** entry, the ♦ secondary sense is inserted within the literal enumeration and, irregularly, subordinated to it!

LÚBENE s.m. (Munt.) Numele dat unor plante din familia cucurbitaceelor: **a)** (şi în sintagma *lubene turcesc,* H II 326, ALR I 855/725, ib. 856/725, 730, 735, 740) dovleac *(Cucurbita maxima).* ♦ *Lubene scoromic =* pepene galben *(Cucurbita melo).* Cf. ALR I 857/740. *Era nouă morţi. Şedea ca lubenii.* GEORGESCU-TISTU, B. 35. Cf. ALR I 856/710, 725, 730, 735, 740, ALR SN I h 198/723, ALRM SN I h 137/723; **b)** dovleac, bostan *(Cucurbita pepo).* Cf. DDRF, SCRIBAN, D., ALR I 855/710, 725, 730, 735, 740. Cf. H II 79, 326, XI 321. *.Albina zbărrr! dup-o floare de lubene, unde se pitise ca s-audă ce va zice.* POP., ap. HEM 1 650

4 Parsing the Atomic Sense Definitions on SCD-*Config*3

The complete parsing of atomic definitions of a dictionary entry relies essentially on the pre-established *dependency hypergraph* of the SCD-*Config*3, as that in [5 :Fig. 2, p. 75], connected to the hypergraph(s) on SCD*Config*2. In this section we point out only few problems that may generate unsound dependencies within the sense trees of the parsed entries on the SCD*Config*3 level: **(1)** Reliable recognition of the atomic sense definitions, including context-depending ones (*e.g. TildaDef* in **DMLRL** [5 :48], *BoldDef* and *ItalDef* in **DLR**, **DAR** [4], [3], [7]); **(2)** Cycling calls between atomic sense definitions and literal enumeration, marked or not by *NewPrg*; **(3)** New kinds, non-standard types of sense definitions and examples-to-definitions; **(4)** Various situations

of *definition inheritance*, either explicit ones (*e.g.* with the *inheritance-dash* marker) as in **TLF** or **GWB**, or by implicit (non-marked) definition inheritance, as frequently occur in **DLR** or **DAR**, along with the sense dependencies they generate.

Remark 4.1. Atomic definitions *BoldDef* and *ItalDef* in **DLR-DAR** may often be refined through literal enumeration. Since the reverse situation is also frequent, when met together they may cause dependency assignment disagreements, as illustrated in examples 3.2.3 and 3.2.4 above.

Example 4.2. Here it is a sample of 'new' atomic definition, something between *ItalDef* and *DefExem* (excerpt from **LÍMBĂ** in **DLR**). Another (this time, very useful) case: *Indexed DefExem* (excerpt from **BRAVE** in **TLF**) [4], [7]. "Unknown" definition species may always be invented, either useful or not, but they may involve recognition problems in the parsing process.

. . . .

Limba oase n-are (= poţi spune cuiva ceva, îl poţi sfătui, ştiind însă că nu va lua în seamă, nu se va conforma spuselor tale). I. CR. IV, 22. *Limba oase n-are, dar oase sfarmă* (= cu cuvântul mari lucruri săvârşim). I. GOLESCU, ap. ZANNE, P. II, 217, PANN, P. V. I, 21. *Limba izbeşte în dintele ce te doare* (= te defaimă unde îţi pasă). I. GOLESCU, ap. ZANNE, P. II, 223. *Toată pasărea pe limba ei piere* (= într-un fel sau altul, fiecare suportă consecinţele vorbelor, ale faptelor proprii). PANN, P. V. I, 25, NEGRUZZI, S. I, 247, LĂCUSTEANU, A. 127, ODOBESCU, S. III, 10, CREANGĂ, ...

. . . .

A. — **1.** Homme courageux qui ne craint pas les dangers ou les entreprises difficiles, qui les a affrontés. *Il n'y a pas d'heures pour les braves* (VERLAINE, *Œuvres posthumes,* t. 1, Souvenirs, 1896, p. 206) :

- 11. . . . . tu es sûr du cœur et du bras de ce gladiateur? Il faut *un* **brave** pour défaire Sigognac, lequel, je l'avoue, bien que je le haïsse, n'est point *lâche,* puisqu'il a bien osé se mesurer contre moi-même. T. GAUTIER, *Le Capitaine Fracasse,* 1863, p. 347. . . . .

Example 4.3. When explicitly marked (as in **TLF**, **GWB**), the sense *definition inheritance* means to establish the correct mother-node in the sense tree from where the definition should be handed down. When inheritance is 'marked' by the *lack of definition* (as in **DLR**), the work on the entry sense tree is more complex and challenging. This is an exacting topic.

... ...

3. Titlu purtat de conducătorii Ţărilor Române; persoană care avea acest titlu; domnitor (**1**), vodă, voievod (**3**), (învechit) gospodar, vlădică, biruitor. V. principe[1] (**1**), prinţ. *La putenciosul domnu Pătru-Vodă amu fost de multe ori* (a. 1593). DOC. î. (XVI), 181. ... ...

... ...

♦ (Atribuind calitatea ca un adjectiv) *Un părinte domnu să aşaze pe un fiiu al său în scaonul părintescu.* GHEORGACHI, CER. (1762), 271. ◇ Spec. Conducător al unui principat sau al unui cnezat; principe[1] (**1**), prinţ, cneaz. Cf. MARDARIE, L. 159/14. *Domnilor de Ardeal dzicem crai ungureşti.* M. COSTIN, O. 43. ... ...

5   Toward the Soundness of Sense Structures in Thesauri

This paper discussed a series of parsing problems and solutions in the context of parsing *six* very large and sensibly different dictionaries of *four* European languages. The typical parsing problems presented are related to the cyclicity (recursive) calls of sense markers on the parsing layers of *three* SCD *configurations*. Working on modules (SCD configurations), reducing the parsing problems (almost only) to sense *marker sequence* analysis, transforming the typographical *New_Paragraphs* into sense *numeral enumeration*, which interleaves with literal enumeration and other sense marker classes, employing the *Enumeration Closing Condition* to check the sound and deterministic (and possibly multiple) use of the sense enumeration device represent the solutions and novelty contributions of the present paper. They are addressing both the dictionary parser designers and thesauri lexicographers, since almost all

the raised problems can be seen as irregularities and / or inadequacies of the sense structure definitions, affecting their lexical-semantic soundness.

References

[1] N. Curteanu, E. Amihăesei (2004). *Grammar-based Java Parsers for DEX and DTLR Romanian Dictionaries.* ECIT-2004 Conference, Iasi, Romania.

[2] N. Curteanu (2006). *Local and Global Parsing with Functional* FX-*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, http://www.math.md/files/csjm/v14-n2/v14-n2-(pp155-182).pdf.

[3] N. Curteanu, A. Moruz, D. Trandabăţ (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, http://aclweb.org/anthology/W/W08/W08-1908.pdf.

[4] N. Curteanu, D. Trandabăţ, A. Moruz (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, http://www.aclweb.org/anthology-new/W/W10/W10-3407.pdf.

[5] N. Curteanu, S. Cojocaru, E. Burcă (2012). *Parsing the Dictionary of Modern Literary Russian Language with the Method of SCD Configurations. The Lexicographic Modeling.* Computer Science Journal of Moldova, Academy of Sciences of Moldova, Vol. 20, No.1(58), pp. 42–81, http://www.math.md/files/csjm/v20-n1/v20-n1-(pp42-82).pdf.

[6] N. Curteanu, S. Cojocaru, A. Moruz (2012). *Lexicographic Modeling and Parsing Experiments for the Dictionary of Modern Literary Russian Language*, ConsILR-2012, Bucharest, The Editorial House of "Al. I. Cuza" University, Iaşi, pp. 189–198.

[7] N. Curteanu. (2012). *The Segmentation-Cohesion-Dependency Parsing Strategy and Linguistic Theory*, TehnoPress, Iaşi, România, xix + 420 p., ISBN: 987-973-702-928-7.

[8] Das Woerterbuch-Netz (2010). http://germazope.uni-trier.de/Projects/WBB/woerterbuecher/.

[9] Dictionary of Modern Literary Russian Language (20 volumes – 1994). M.: Russian language; Second edition, revised and supplemented, 864 p.; 1991 – 1994. ISBN: 5-200-01068-3 (in Russian).

[10] R. Hauser, A. Storrer (1993). *Dictionary Entry Parsing Using the LexParse System.* Lexikographica 9 (1993), pp.174–219.

[11] M. Kammerer (2000). *Wöterbuchparsing Grundsätzliche Überlegungen und ein Kurzbericht über praktische Erfahrungen*, http://www.matthias-kammerer.de/content/WBParsing.pdf.

[12] Le Trésor de la Langue Française informatisé (2010). http://atilf.atilf.fr/tlf.htm.

[13] L. Lemnitzer, C. Kunze (2005). *Dictionary Entry Parsing*, ESSLLI 2005.

[14] C. Mărănduc (2010). *Dictionary of expressions, locutions, and phrases*, Corint Editorial House, Bucharest, 560 p., ISBN 973-135-570-2 (in Romanian).

[15] M. Neff, B. Boguraev (1989). *Dictionaries, Dictionary Grammars and Dictionary Entry Parsing*, Proc. of the 27th annual meeting on Association for Computational Linguistics Vancouver, British Columbia, Canada Pages: pp. 91 – 101.

[16] S. Puşcariu, *et al.* (1906). Dictionary of the Romanian Language
(Dictionary of the Romanian Academy – **DAR**), Bucharest, Edition 1940 (old format).

[17] D. Tufiş (2001). *From Machine Readable Dictionaries to Lexical Databases*, RACAI, Romanian Academy, Bucharest, Romania.

[18] XCES TEI Standard, Variant P5 (2007).
http://www.tei-c.org/Guidelines/P5/

Neculai Curteanu, Alex Moruz   Received June 27, 2012

Neculai Curteanu

Institute of Computer Science,
Romanian Academy, Iaşi Branch
Str. Gh. Asachi, Nr. 3,
700483 Iaşi, România
E–mails: *ncurteanu@yahoo.com*,
         nicu.curteanu@iit.academiaromana-is.ro

Alex Moruz

Institute of Computer Science,
Romanian Academy, Iaşi Branch,
Faculty of Computer Science,
"Al. I. Cuza" University of Iaşi,
E–mails: *alex.moruz@gmail.com*,
         alex.moruz@iit.academiaromana-is.ro