News about the Romanian Wordnet

Verginica Barbu Mititelu
RACAI
13, Calea 13 Septembrie
Bucharest 050711, Romania
vergi@racai.ro

Ștefan Daniel Dumitrescu
RACAI
13, Calea 13 Septembrie
Bucharest 050711, Romania
sdumitrescu@racai.ro

Dan Tufiș
RACAI
13, Calea 13 Septembrie
Bucharest 050711, Romania
tufis@racai.ro

Abstract

There are more than 60 wordnets worldwide; the Romanian wordnet is among those that are maintained and further developed. Begun within the BalkaNet project and further enriched in various (application oriented) projects, it was used in word sense disambiguation, machine translation and question answering with promising results. We present here the latest qualitative and quantitative improvements of our lexical resource, special attention being paid to derivational relations, the latest statistics, as well as the development of an Application Programming Interface, meant to facilitate work with the wordnet, both for its further development purposes and for its use in applications. In the context of creating a common European research infrastructure network, our wordnet is licensed through META-SHARE, being freely available for scientific purposes.

1 Introduction

The development of the Romanian wordnet (RoWN henceforth) started within BalkaNet project[1]. Afterwards, it has been developed and maintained within several projects by the Natural Language Processing (NLP) group of the Romanian Academy Research Institute for Artificial Intelligence (RACAI): ROTEL[2], STAR[3], SIR-RESDEC[4], ACCURAT[5], METANET4U[6], the Romanian Academy research plan.

Within BalkaNet a core of 18000 synsets was created. They were aligned to the Princeton WordNet (PWN) versions available throughout time, respectively version 2.0 at the end of the project. Among those synsets there were more than 400 that lexicalize concepts specific to the Balkan area. These were implemented in all six languages of the project (Bulgarian, Czech, Greek, Romanian, Serbian, Turkish) and were linked to hypernym synsets, already existing in PWN, so they were not left dangling in the network.

RoWN contains words belonging both to the general vocabulary and to various domains of activity. Throughout time, we aimed at a complete coverage of the basic common sets from EuroWordNet[7], of the 1984 corpus[8], of the newspaper articles corpus NAACL2003[9], of the Acquis Communautaire corpus and the Eurovoc thesaurus[10], of VerbNet 3.1[11], and as much as possible from the ROWikipedia lexical stock.

Two basic development principles have always been followed: the Hierarchy Preservation

[1] http://www.dblab.upatras.gr/balkanet
[2] http://www.ai.ici.ro/rotel_eng/index.htm
[3] http://www.racai.ro/star
[4] http://www2.racai.ro/sir-resdec
[5] http://www.accurat-project.eu/, http://valhalla.racai.ro/accurat/index.php?page=despre
[6] http://www.racai.ro/metanet4u-racai
[7] http://www.illc.uva.nl/EuroWordNet
[8] http://nl.ijs.si/ME/Vault/CD/docs/1984.html
[9] http://ws.racai.ro:9191/repository/browse/the-naacl-2003-english-romanian-corpus/da86dc2efb6811e2a8ad00237df3e35886f019db7a16437f801cba30dd6ab209
[10] http://optima.jrc.it/Acquis/JRC-Acquis.3.0/doc/README_Acquis-Communautaire-corpus_JRC.html
[11] http://verbs.colorado.edu/~mpalmer/projects/verbnet.html

Principle (HPP) (according to which the hierarchical structure of the concepts in a wordnet is the same irrespective of the natural language for which the wordnet is developed) and the Conceptual Density Principle (which ensures that once a concept is selected to be implemented, all its ancestors up to the unique beginners are also selected, thus preventing the existence of dangling nodes) (Tufiş et al., 2004). The former principle was the assumption behind our development methodology, namely the expand method. The latter ensured the lack of dangling nodes in the nouns and verbs hierarchies. As a consequence of the way we chose to create our language resource, the lexical density has never been our preoccupation, thus there are many words that do not occur in as many synsets as how many meanings they have. Nevertheless, we do not exclude such an objective from our further developments.

At present, RoWN is aligned to PWN version 3.0. Details about the way we performed the alignment from PWN 2.0 to PWN 3.0 and about the way we solved the encountered problems (the n:1 or 1:n matches between synsets in the two versions) are presented in Tufiş et al. (2013).

RoWN is licensed through META-SHARE[12] (). It is free for academic research, but restricted for commercial use.

In this paper we present the latest qualitative and quantitative improvements of our lexical resource, the latest statistics (Section 3), special attention being paid to derivational relations (Section 4), as well as the development of an Application Programming Interface, meant to facilitate work with the wordnet, both for its further development purposes and for its use in applications (Section 5). Our intentions for further development are included in the Conclusions section. Before proceeding, we enumerate the applications in which our team used RoWN and which, throughout time, influenced our decisions about the concepts to be further implemented in the network.

2 Uses of RoWN

Ion and Tufiş (2009) and Ion and Ştefănescu (2011) describe word sense disambiguation (WSD) methods that make use of wordnets: the former is set in a multilingual environment and the WSD is done with the help of aligned wordnets. The latter is set in a monolingual environment and the WSD is done with the help of the lexical chains established between the co-occurring words in the text, chains whose length is calculated in the wordnet. The assumption is that the shorter the lexical chain, the more similar the words. The length of the lexical chain depends on the number of relations marked in the network. The results in the multilingual environment are reported as better than those in the monolingual one.

For a Question Answering (QA) system, RoWN was used for expanding the query introduced by the user (Ion et al., 2008) with words semantically related (i.e., synonyms, hypo- and hyperonyms) to the ones it contained. Moreover, RoWN was also used in the last phase, that of ranking the found results by calculating the semantic distance, again as a lexical chain, between the words introduced by the user and those occurring in the text. It was noticed that the relations with the greatest contribution at calculating the score are hyponymy and derivational relations.

Aligned wordnets are valuable sources of cross-language equivalents, especially multiword terms, in machine translation.

3 Latest Quantitative Developments

Lately our efforts of implementing new synsets aimed at a complete coverage of VerbNet 3.1, with the prospect of creating a syntactic parser for Romanian.

The up-to-date statistics about RoWN are presented in Table 1 and 2 below. In the former, PoS stands for part of speech, S for synset, L for literal, UL for unique literals and NL for nonlexicalized synsets. Obeying the HPP stated above implies the transfer of the hierarchies from PWN into RoWN. The lack of perfect equivalences among languages is widely known; nevertheless, we chose to disregard it. Moreover, there are lexical gaps in all languages. We call them nonlexicalized concepts and represent them as empty synsets. For example, for the PWN verbal synset {zip_up:1} (gloss: close with a zipper) there is no literal in the corresponding Romanian synset. However, such synsets do not lack relations: the corresponding ones from PWN are transferred into RoWN.

[12] http://ws.racai.ro:9191/repository/browse/18

PoS	S	L	UL	NL
Nouns	41063	56532	52009	1839
Verbs	10397	16484	14210	759
Adjective	4822	8203	7407	79
Adverbs	3066	4019	3248	110
TOTAL	59348	85238	75656	2787

Table 1: Statistics about synsets and literals in RoWN.

Relation	Number
hypo/hyperonymy	48316
instance_hypo/hyperonymy	3889
antonym	4131
similar_to	4838
verb_group	1530
member_holonym	2047
part_holonym	5573
substance_holonym	410
also_see	1333
attribute	958
cause	196
entailment	371

Table 2. Semantic relations in RoWN.

It is worth noticing that antonymy, which is a lexical relation in PWN, is represented as a semantic one in RoWN. The conceptual opposition between the synsets is more useful in various applications than the mere antonymy between two literals.

With the exception of *attribute* relation, all the others enumerated in Table 2 link synsets with literals of the same part of speech. A path between two words of a different part of speech, about which any speaker would say they are related, although not impossible to find, would be too long, thus providing wrong information about the similarity between those words.

4 Derivational Relations

Using RoWN in applications, as presented above, showed unnatural lexical chains, such as one of the possible chains between *inventator* "inventor" and *inventa* "to invent":

inventator(1.1) *instance_hyponym* James_Watt(x)
James_Watt(x) *instance_hypernym* inginer(1.1)
inginer(1.1) *hyponym* inginer_software(1)
inginer_software(1) *domain_member_TOPIC* ştiinţa_calculatoarelor(x)
ştiinţa_calculatoarelor(x) *domain_TOPIC* programa(3)
programa(3) *hyponym* crea_mental(1)
crea_mental(1) *hypernym* **inventa**(1)

The strangeness of this example results from the intricate path from *inventator* to *inventa*, uncommon for whatever speaker of Romanian: *inventator – James Watt – inginer* "engineer" – *inginer software* "software engineer" – *programa* "to program" – *crea mental* "to create by mental act" – *inventa*.

Faced with a number of such cases, we decided to implement derivational relations into our wordnet.

This type of relations exists in other wordnets as well: the Turkish WordNet (Bilgin et al., 2004), PWN (Fellbaum et al., 2007), the Czech WordNet (Pala and Hlaváčková, 2007), the Polish WordNet (Piasecki et al., 2012), the Estonian one (Kahusk, et al., 2010). Given the language-specific character of such relations, each team undertook their own strategy for finding the relations in their wordnet. However, there are teams that transferred the derivational relations in PWN and then validated them: this is the case for the Bulgarian WordNet (Koeva, 2008), the Serbian (Koeva et al., 2008) and the Finnish one (Lindén and Niemi, 2013).

Whereas most of the undertakings above aimed at expanding the network with new synsets derivationally linked with the literals already in the wordnet, we were interested in adding such relations between literals that are in the synsets. No extension was intended, at least for the moment.

We discuss below some theoretical aspects of derivational relations and the significance of their representation in a wordnet and then present the methodology we adopted for identifying and annotating them in RoWN.

4.1 Pre-requisites

Derivation is one means of creating new words in a language from existing morphemes, i.e. the smallest units of a language that have their own meaning. It ensures both formal and semantic relatedness between the root and the derived word: the formal relatedness is ensured by the fact that the root and the derived word contain (almost) the same string of letters that represent the root, while the semantic relatedness is ensured by the compositionality of meaning of the derived word: its meaning is a sum of the meaning of the root and the meaning of the affix(es). Thus, the Romanian words *alerga* "run" and *alergător* "runner" are derivationally related: the

latter is obtained from the former by adding the suffix *-ător* (after removing *-a*, the infinitive suffix) and it means "the one who runs". However, derivational relations cannot be established for all meanings of these words: when considered with their proper meaning, they are related, but when *alerga* is considered with its figurative meaning "to try hard to get something", it does not establish a derivational relation with *alergător*, as it has not developed any related figurative meaning.

In the derivation process only one affix of a type is added. So, a prefix and a suffix can be added to a root in the same derivation step, but never two suffixes or/and two prefixes. If a word contains more than two affixes of the same type, then they were attached in different steps in the derivation.

4.2 Identifying derivational relations between literals in RoWN

Having available a list of (492) Romanian affixes and the list of (31872) simple literals in RoWN, we searched for pairs of literals ($literal_1$ and $literal_2$) such that $literal_1$ +/- affix(es) = $literal_2$. The "+" version covers progressive derivation, while the "-" version covers backformation. We allowed for at most 2 affixes, but of different types, as discussed above. The results are presented in Table 3:

Derivation type	Number of derived words	Percent
Prefixation	2862	17.43
Suffixation	13556	82.57
TOTAL	16418	

Table 3. Derivational relations between simple literals in RoWN.

The percents are reasonable: it is a well-known fact that prefixation is weakly productive in Romanian, unlike suffixation.

We subjected the found pairs to an automatic and then a manual validation. For the former, we enriched the list of affixes with information about the part of speech of the words to which they can attach and of the words they help create. The list is available at www.racai.ro/~vergi under Research. For example, the suffix *-a* can be attached to nouns or to adjectives to create verbs:

-a n>v a>v

Examples include: *buton* ("button") + *-a* > *butona* ("to channel-surf"), *scurt* ("short") + *-a* > *scurta* ("to shorten").

Afterwards we proceeded to a manual validation of the whole number of pairs. The results are presented in Table 4: for each type of derivation (DT) (prefixation P or suffixation S), from the found pairs (column 2) we present the number of those passing the automatic validation (AV) in column 3 and then of those that passed the manual validation (MV) in column 4; the last column presents the percent of manually validated pairs for each derivation type.

DT	Found	AV	MV	%
P	2862	2621	1990	69.53
S	13556	8345	8452	62.35
TOTAL	16418	10966	10442	

Table 4. Validated pairs.

Examples of pairs that passed the automatic validation but not the manual one include: *prinde* "to catch" – *surprinde* "to surprise", *abate* "to deviate" – *abator* "slaughter house".

4.3 Sense level annotation

Having already established that derivational relations need to be marked at the word sense level, not for all senses of the words in a pair, the next necessary step is to calculate the Cartesian product of the sets of synsets in which the members of the validated pairs occur. Thus, for the 10442 pairs of literals resulted after manual validation, we calculated the Cartesian product for each pair, obtaining a total of 101729 pairs of synsets. They display formal relatedness and, in order to mark a derivational relation for them, it is also necessary to subject them to a semantic evaluation. A linguist goes through them and whenever semantic similarity is noticed, the pair is labeled with one of the 57 semantic labels we established: 16 for prefixed words (together, subsumption, opposition, mero, eliminate, iterative, through, repeat, imply, similitude, instead, aug, before, anti, out, back) and 41 for suffixed ones (subsumption, member_holo, member_mero, substance_holo, substance_mero, ingredient_holo, holonym, part, agent, result, location, of_origin, job, state, period, undergoer, instrument, sound, cause, container, vehicle, body_part, material, destination, gender, wife, dim, aug, object_made_by, subject_to, by_means_of, clothes, event, abstract, colour, tax, make_become, make_acquire, manner, similitude, related).

The most frequently attached semantic labels are: for prefixed words: opposition (*neesenţial* "unessential" – *essential* "essential") (792), subsumption (*subclasă* "subclass" – *clasă* "class") (363), repeat (*reaprinde* "reignite" – *aprinde* "ignite") (305); for suffixed words: related (*călduros* "warm" – *căldură* "warmth") (1294), event (*împărtăşanie* "communion" – *împărtăşi* "commune") (699), abstract (*cerinţă* "requirement" – *cere* "require") (490), manner (*primejdios* "dangerous" - *primejdie* "danger") (436), agent (*linguşitor* "adulator" – *linguşi* "adulate") (394). At the end of the article, in the Annex, containing Table7 and Table 8, we present the semantic labels and their frequencies for prefixed and, respectively, suffixed words, accompanied by examples.

4.4 Statistics about derivational relations

Going through 55849 such pairs of synsets, we obtained the results in Table 5.

	Prefixed	**Suffixed**	**TOTAL**
Pairs subject to validation	30132	25717	55849
Validated pairs	3145	13916	17061
Percent	10.43	89.64	**30.55**

Table 5. Semantically annotated pairs.

The aim of marking these derivational relations was to increase the number of links between synsets, especially between synsets of different parts of speech. For the validated pairs we included in Table 6 statistics about the derivational relations involving words of the same and of different part of speech. It is obvious that, on the whole, adding derivational relations to a wordnet increases the number of cross-part of speech (PoS) relations.

	Same PoS %	Cross PoS %
Prefixed	97	3
Suffixed	15	85
TOTAL	38	**62**

Table 6. Distribution of derivational relation on PoS.

5 RoWordNetLib

We have built an Application Programming Interface (API) for RoWN, called RoWordNetLib, meant as a tool to aid quick implementations of RoWN into both research-oriented and industry applications. When designing it, we envisaged a tool that should be easy to use, easy to extend and that would offer a sufficiently large array of functionalities. The chosen programming language is Java.

The main functionalities that RoWordNetLib provides are:

- Input/Output for working with XML-based RoWN files;
- Methods for working with the semantic network itself (RoWordNet objects containing RoWN);
- Set operations for working with multiple RoWordNet objects (reunion, intersection, complement, difference, merge, etc.);
- Basic Word Sense Disambiguation (WSD) algorithms;
- Similarity Metrics (both distance-based and semantic).

The API's uses can be classified as (1) internal – it helps to facilitate the continuous work of enriching RoWN and (2) external – to quicken the development of Romanian-enabled smart applications. By providing set operations like difference, intersection or reunion on RoWordNet objects, more people can work in parallel on RoWN and then easily join their versions into a single wordnet, thus easing its development. Externally, wordnets are successfully used to perform word sense disambiguation, information retrieval, information extraction, machine translation, automatic text classification and summarization.

RoWordNetLib is structured into several packages, each with its assigned functionality. The main packages are: 'data', 'io', 'op' and 'wsd'.

The 'data' package contains the data structures RoWordNetLib uses internally. Its structure is simple, following the way the data is naturally structured in a wordnet: a RoWordNet object contains an array of Synset objects which are indexed by the synset ID for retrieval speed. Each Synset object contains a number of primitive types as well as an array of Literal objects and an array of Relation objects. A Literal object contains a word and an associated sense. A Relation object contains a relation (string) that points to a target synset (defined as an ID), as well as optionally having a source and target literal for

cases where the relation is not between synsets but between two synsets' particular literals.

The 'io' package provides input and output functions. The most important I/O function is reading and writing RoWordNet objects in their native XML format.

The 'op' package provides different operational tools: (1) set operation methods for joining, intersecting, complementing, etc., multiple RoWordNet objects; (2) through the BFWalk class, the ability to perform a breadth-first walk through the RoWN semantic network; (3) a number of distance-based and semantic similarity measures (Resnik, 1995) for measuring the closeness of concepts (lexicalized by literals in synsets).

The 'wsd' package implements two Word Sense Disambiguation algorithms: Lesk (1986) and an adapted version of Lesk. They are used to obtain information content values for synsets in RoWN given an arbitrary Romanian text as the input corpus, which is further used to enable the semantic similarity measures.

6 Conclusions and Further Work

RoWN is a valuable resource for the Romanian language and the NLP group of RACAI uses it in most of their applications. We presented here our latest qualitative and quantitative achievements.

Further enrichment of RoWN is a constant preoccupation of our team. It follows all the time the other interests of the group. For instance, the last set of implemented synsets was made up of verbs exclusively, given our present interest to cover VerbNet 3.1, with the prospect of creating a parser for Romanian.

Increasing the density of relations between synsets in order to make RoWN more effective in applications was obtained by adding derivational relations. Although they are relations between literals, the semantic labels we attached to them can be viewed as a link between the synsets to which the respective literals belong. After finishing the semantic annotation of the derivative pairs, we could try to expand the network with automatically derived words. For Romanian an experiment of automatically deriving words is reported by Petic (2011), who used very productive and reliable affixes. With the list of affixes and their combination possibilities (available at www.racai.ro/~vergi under Research) that we have created, we can dare test new cases of automatic derivation for Romanian.

Reference

Orhan Bilgin, Özlem Çetinoglu, and Kemal Oflazer. 2004. Morphosemantic relations in and across wordnets: A study based on Turkish. P. Sojka, K. Pala, P. Smrz, C. Fellbaum, P. Vossen (Eds.), *Proceedings of GWC*.

Christiane Fellbaum, Anne Osherson, and Peter E. Clark. 2007. Putting Semantics into WordNet's "Morphosemantic" Links. *Proceedings of the 3rd Language and Technology Conference*.

Radu Ion and Dan Ştefănescu. 2011. Unsupervised Word Sense Disambiguation with Lexical Chains and Graph-Based Context Formalization. Zygmunt Vetulani (ed.): *LTC 2009*, Lecture Notes in Artificial Intelligence, 6562/2011: 435—443.

Radu Ion and Dan Tufiş. 2009. Multilingual versus Monolingual Word Sense Disambiguation. *International Journal of Speech Technology*; 12 (2-3):113-124.

Radu Ion, Dan Ştefănescu, Alexandru Ceauşu, and Dan Tufiş. 2008. RACAI's QA System at the Romanian-Romanian Multiple Language Question Answering (QA@CLEF2008) Main Task. Carol Peters et al. (eds.) *Working Notes for the CLEF 2008 Workshop*: 10.

Neeme Kahusk, Kadri Kerner, and Kadri Vider. 2010. Enriching Estonian WordNet with Derivations and Semantic Relations. *Proceeding of the 2010 conference on Human Language Technologies – The Baltic Perspective*:195-200.

Svetla Koeva. 2008. Derivational and Morphosemantic Relations in Bulgarian Wordnet. *Intelligent Information Systems*; XVI:359-369.

Svetla Koeva, Cvetana Krstev, and Duško Vitas. 2008. Morpho-semantic Relations in Wordnet – A Case Study for two Slavic Languages. *Proceedings of the Fourth Global WordNet Conference:*239-254.

Michael Lesk. 1986. Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone. *5th SIGDOC*:24-26.

Krister Lindén and Jyrki Niemi. 2013. Is it possible to create a very large wordnet in 100 days? An evaluation. *Language Resources and Evaluation*, http://link.springer.com/article/10.1007%2Fs10579-013-9245-0.

Karel Pala and Dana Hlaváčková, D. 2007. Derivational relations in Czech Wordnet. *Proceedings of the Workshop on Balto-Slavonic Natural Language Processing*: 75-81.

Mircea Petic. 2011. Generative mechanisms of Romanian derivational morphology. *Memoirs of the Scientific Section of the Romanian Academy*. Series IV, Tome XXXIV:21-30.

Maciej Piasecki, Radoslaw Ramocki, and Marek Maziarz. 2012. Recognition of Polish Derivational Relations Based on Supervised Learning Scheme. *Proceedings of LREC 2012*: 916-922.

Philip Resnik. 1995. Using information content to evaluate semantic similarity in a taxonomy. *14th International Joint Conference on Artificial Intelligence*.

Dan Tufiș, Dan Cristea, and Sofia Stamou. 2004. BalkaNet: Aims, Methods, Results and Perspectives. A General Overview. *Romanian Journal on Information Science and Technology*; 7:9-34.

Dan Tufiș, Verginica Barbu Mititelu, Dan Ștefănescu, Radu Ion. 2013. The Romanian Wordnet in a Nutshell. *Language Resources and Evaluation*, http://link.springer.com/article/10.1007%2Fs10579-013-9230-7.

Annex

Label	**Occurrences**	**Example**
BACK	2	reflux “low tide” – flux “high tide”
TOGETHER	29	întreţese “interweave” – ţese “weave”
AUG	5	supraabundenţă “overabundance” – abundenţă “abundance”
OUT	1	epidermal “epidermis” – derma “dermis”
SIMILITUDE	61	reţine “withhold” – tine “hold”
IMPLY	26	desconsidera “disconsider” – considera “consider”
THROUGH	5	răzbate “get through” – bate “beat”
MERO	17	suprafaţă “surface” – faţă “face”
BEFORE	14	preambalare “prepacking” – ambalare “packing”
OPPOSITION	792	neesenţial “unessential” – esenţial “essential”
REPEAT	305	reaprinde “reignite” – aprinde “ignite”
SUBSUMPTION	363	subclasă “subclass” – clasă “class”
ANTI	10	anticolinesterază “anticholinesterase” – colinesterază “cholinesterase”
INSTEAD	6	vicepreşedinte “vicepresident” – preşedinte “president”
ITERATIVE	2	răsfoi “thumb through” – foaie “leaf”
ELIMINATE	9	deşela “override” – şale “loin”

Tabel 7. Semantic labels for prefixed words and their frequency in RoWN.

Label	**Occurrences**	**Example**
RELATED	1294	călduros “warm” – căldură “warmth”
SOUND	163	bufneală “plunk” – bufni “to plunk”
STATE	284	îndoială “doubt” – îndoi “to doubt”
DESTINATION	5	patentant “patentee” – patenta “to patent”
AUG	1	grăsan “big fat person” – gras “fat”
SIMILITUDE	115	încărcătură “loading” – încărcare “loading”
PERIOD	43	bătrâneţe “old age” – bătrân “old”
JOB	179	semănător “sower” – semăna “sow”
PART	12	optime “eighth” – opt “eight”
MEMBER_MERO	17	orăşean “town dweller” – oraş

		"town"
BY_MEANS_OF	104	oprelişte "obstructor" – opri "obstruct"
CAUSE	19	umezeală "dampness" – umezi "to damp"
MEMBER_HOLO	37	soldăţime "soldiery" – soldat "soldier"
RESULT	227	tencuială "plastering" – tencui "plaster"
SUBJECT_TO	19	chinui "to anguish" – chin "anguish"
ABSTRACT	490	cerinţă "requirement" – cere "require"
SUBSUMPTION	42	căpetenie "headman" – cap "head"
OF_ORIGIN	29	sătean "villager" – sat "village"
EVENT	699	împărtăşanie "communion" – împărtăşi "to commune"
INSTRUMENT	84	ondulator "crimper" – ondula "to crimp"
INGREDIENT_HOLO	1	sticlărie "glass work" – sticlă "glass"
TIME	1	cătănie "period of military service" – cătană "serviceman"
MANNER	436	primejdios "dangerous" – primejdie "danger"
MAKE_ACQUIRE	110	îndigui "to dam" – dig "dam"
CONTAINER	17	afişier "board" – afiş "poster"
HOLONYM	26	pieptar "vest" – piept "breast of a garment"
DIM	50	căsuţă "little house" – casă
		"house"
OBJECT_MADE_BY	50	chinezărie "Chinese work" – chinez "Chinese"
CLOTHES	1	pieptar "vest" – piept "breast"
SUBSTANCE_HOLO	2	cerat "waxy" – ceară "wax"
AGENT	394	linguşitor "adulator" – linguşi "adulate"
LOCATION	87	cărămidărie "brickyard" – cărămidă "brick"
MATERIAL	4	îndulcitor "sweetener" – îndulci "sweeten"
UNDERGOER	47	setos "thirsty" – sete "thirst"
COLOUR	19	cenuşiu "ashen" – cenuşă "ash"
GENDER	13	călugăriţă "nun" – călugăr "monk"
SUBSTANCE_MERO	1	ricină "ricin" – ricin "castor oil plant"
MAKE_BECOME	89	caricaturize "to caricature" – caricatură "caricature"

Tabel 8. Semantic labels for suffixed words and their frequency in RoWN.