# Romanian WordNet: Current State, New Applications and Prospects

Dan Tufiş, Radu Ion, Luigi Bozianu, Alexandru Ceauşu, and Dan Ştefănescu

Romanian Academy Research Institute for Artificial Intelligence
13, Calea 13 Septembrie, 050711, Bucharest 5, Romania
{tufis, radu, bozi, aceausu, danstef}@racai.ro

## 1 Introduction

The development of the Romanian WordNet began in 2001 within the framework of the European project BalkaNet which aimed at building core WordNets for 5 new Balkan languages: Bulgarian, Greek, Romanian, Serbian and Turkish. The philosophy of the BalkaNet architecture was similar to EuroWordNet [1, 2]. As in EuroWordNet, in BalkaNet the concepts considered highly relevant for the Balkan languages (and not only) were identified and called BalkaNet Base Concepts. These are classified in three increasing size sets (BCS1, BCS2 and BCS3). Altogether BCS1, BCS2 and BCS3 contain 8516 concepts that were lexicalized in each of the BalkaNet WordNets.

The monolingual WordNets had to have their synsets aligned to the translation equivalent synsets of the Princeton WordNet (PWN). The BCS1, BCS2 and BCS3 were adopted as core WordNets for several other WordNet projects such as Hungarian [3], Slovene [4], Arabic [5, 6], and many others.

At the end of the BalkaNet project (August 2004) the Romanian WordNet, contained almost 18,000 synsets, conceptually aligned to Princeton WordNet 2.0 and through it to the synsets of all the BalkaNet WordNets. In [7], a detailed account on the status of the core Ro-WordNet is given as well as on the tools we used for its development.

After the BalkaNet project ended, as many other project partners did, we continued to update the Romanian WordNet and here we describe its latest developments and a few of the projects in which Ro-WordNet, Princeton WordNet or some of its BalkaNet companions were of crucial importance.

## 2 The Ongoing Ro-WordNet Project and its Current Status

The Ro-WordNet is a continuous effort going on for 6 years now and likely to continue several years from now on. However, due to the development methodology adopted in BalkaNet project, the intermediate WordNets could be used in various other projects (word sense disambiguation, word alignment, bilingual lexical knowledge acquisition, multilingual collocation extraction, cross-lingual question answering, machine translation etc.).

Recently we started the development of an English-Romanian MT system for the legalese language of the type contained in JRC-Acquis multilingual parallel corpus [8] and of a cross-lingual question answering system in open domains [9, 10]. For these projects, heavily relying on the aligned Ro-En WordNets, we extracted a series of high frequency Romanian nouns and verbs not present in Ro-WordNet but occurring in JRC-Acquis corpus and in the Romanian pages of Wikipedia and proceeded at their incorporation in Ro-WordNet. The methodology and tools were essentially the same as described in [11], except that the dictionaries embedded into the WNBuilder and WNCorrect were significantly enlarged.

The two basic development principles of the BalkaNet methodology, that is Hierarchy Preservation Principle (HPP) and Conceptual Density Principle (CDP), were strictly observed. For the sake of self-containment, we restate them here.

**Hierarchy Preservation Principle**

> If in the hierarchy of the language L1 the synset $M_2$ is a hyponym of synset $M_1$
>
> ($M_2$ $H^m$ $M_1$) and the translation equivalents in L2 for $M_1$ and $M_2$ are $N_1$ and $N_2$ respectively, then in the hierarchy of the language L2 $N_2$ should be a hyponym of synset $N_1$ ($N_2$ $H^n$ $N_1$). Here $H^m$ and $H^n$ represent a chain of m and n hierarchical relations between the respective synsets (hyperonymy relations composition).

**Conceptual Density Principle (noun and verb synsets)**

> Once a nominal or verbal concept (i.e. an ILI concept that in PWN is realized as a synset of nouns or as a synset of verbs) was selected to be included in Ro-WordNet, all its direct and indirect ancestors (i.e. all ILI concepts corresponding to the PWN synsets, up to the top of the hierarchies) should be also included in Ro-WordNet.

By observing HPP, the lexicographers were relieved of the task to establish the semantic relations for the synsets of the Ro-WordNet. The hypernym relations as well as the other semantic relations were imported automatically from the PWN. The CDP compliance ensures that no dangling synsets, harmful in taxonomic reasoning, are created.

The tables below give a quantitative summary of the Romanian WordNet at the time of writing (September, 2007). As these statistics are changing every month, the updated information should be checked at http://nlp.racai.ro/Ro-wordnet.statistics. The Ro-wordnet is currently mapped on the various versions of Princeton WordNet: PWN1.7.1, PWN2.0 and PWN2.1. The mapping onto the last version PWN3.0 is also considered. However, all our current projects are based on the PWN2.0 mapping and in the following, if not stated otherwise, by PWN we will mean PWN2.0.

**Table 1.** POS distribution of the synsets in the Romanian WordNet.

| Noun synsets | Verb synsets | Adj. synsets | Adv. synsets | Total |
|---|---|---|---|---|
| 33151 | 8929 | 851 | 834 | 43765 |

**Table 2**. Internal relations used in the Romanian WordNet.

| | | | |
|---|---|---|---|
| hypernym | 42794 | category_domain | 2668 |
| near_antonym | 2438 | also_see | 586 |
| holo_part | 3531 | subevent | 335 |
| similar_to | 899 | holo_portion | 327 |
| verb_group | 1404 | causes | 171 |
| holo_member | 1300 | be_in_state | 570 |
| DOMAINS classes | 165 | SUMO&MILO categories | 1836 |
| objective synsets | 34164 | subjective synsets | 9601 |

As one can see from Table 2, the synsets in Ro-WordNet have attached, via PWN, DOMAINS-3.1 [12], SUMO&MILO [13, 14] and SentiWordNet [15] labels.

The DOMAINS labeling (http://wndomains.itc.it/) uses Dewey Decimal Classification codes and the 115425 PWN synsets are classified into 168 distinct classes (domains).

The SUMO&MILO upper and mid level ontology is the largest freely available (http://www.ontologyportal.org/) ontology today. It is accompanied by more than 20 domain ontologies and altogether they contain about 20,000 concepts and 60,000 axioms. They are formally defined and do not depend on a particular application. Its attractiveness for the NLP community comes from the fact that SUMO, MILO and associated domain ontologies were mapped onto Princeton WordNet. SUMO and MILO contain 1107 and respectively 1582 concepts. Out of these, 844 SUMO concepts and 1582 MILO concepts were used to label almost all the synsets in PWN. Additionally, 215 concepts from some specific domain ontology were used to label the rest of synsets in PWN (instances).

The SentiWordnet [15] adds subjectivity annotations to the PWN synsets. Their basic assumptions were that words have graded polarities along the Subjective-Objective (SO) & Positive-Negative (PN) orthogonal axes and that the SO and PN polarities depend on the various senses of a given word (context). The word senses in a synset are associated with a triple P (positive subjectivity), N (negative subjectivity) and O (objective) so that the values of these attributes sum up to 1. For instance, the sense 2 of the word *nightmare* (a terrifying or deeply upsetting dream) is marked-up by the values P:0.0, N:0,25 and O:0.75, signifying that the word denotes to a large extent an objective thing with a definite negative subjective polarity.

Due to the BalkaNet methodology adopted for the monolingual WordNets development, most of the DOMAINS, SUMO and MILO conceptual labels in PWN are represented in our Ro-WordNet (see Table 3).

**Table 3.** The ontological labeling (DOMAINS, SUMO, MILO, etc.) in RO-WordNet vs. PWN.

| LABELS | PWN | Ro-WordNet |
|---|---|---|
| DOMAINS-3.1 | 168 | 165 |
| SUMO | 844 | 781 |
| MILO | 949 | 882 |
| Domain ontologies | 215 | 173 |

The BalkaNet compliant XML encoding of a synset, including the new subjectivity annotations is exemplified in Figure 1.

```
<SYNSET>
   <ID>ENG20-05435872-n</ID>
   <POS>n</POS>
   <SYNONYM>
      <LITERAL>nightmare<SENSE>2</SENSE></LITERAL>
   </SYNONYM>
   <ILR><TYPE>hypernym</TYPE>ENG20-05435381-n</ILR>
   <DEF>a terrifying or deeply upsetting dream</DEF>
   <SUMO>PsychologicalProcess<TYPE>+</TYPE></SUMO>
   <DOMAIN>factotum</DOMAIN>
   <SENTIWN><P>0.0</P><N>0.25</N><O>0.75</O></SENTIWN>
</SYNSET>
<SYNSET>
   <ID>ENG20-05435872-n</ID>
   <POS>n</POS>
   <SYNONYM>
      <LITERAL>co•mar<SENSE>1</SENSE></LITERAL>
   </SYNONYM>
   <DEF>Vis urât, cu senza•ii de ap•sare •i de în•bu•ire</DEF>
   <ILR>ENG20-05435381-n<TYPE>hypernym</TYPE></ILR>
   <DOMAIN>factotum</DOMAIN>
   <SUMO>PsychologicalProcess<TYPE>+</TYPE></SUMO>
   <SENTIWN><P>0.0</P><N>0.25</N><O>0.75</O></SENTIWN>
</SYNSET>
```

**Fig. 1.** Encoding of two EQ-synonyms synsets in PWN and Ro-WordNet.

The visualization of the synsets in Figure 1, by means of the VISDIC editor[1] [16], is shown in Figure 2.

[1] http://nlp.fi.muni.cz/projekty/visdic/

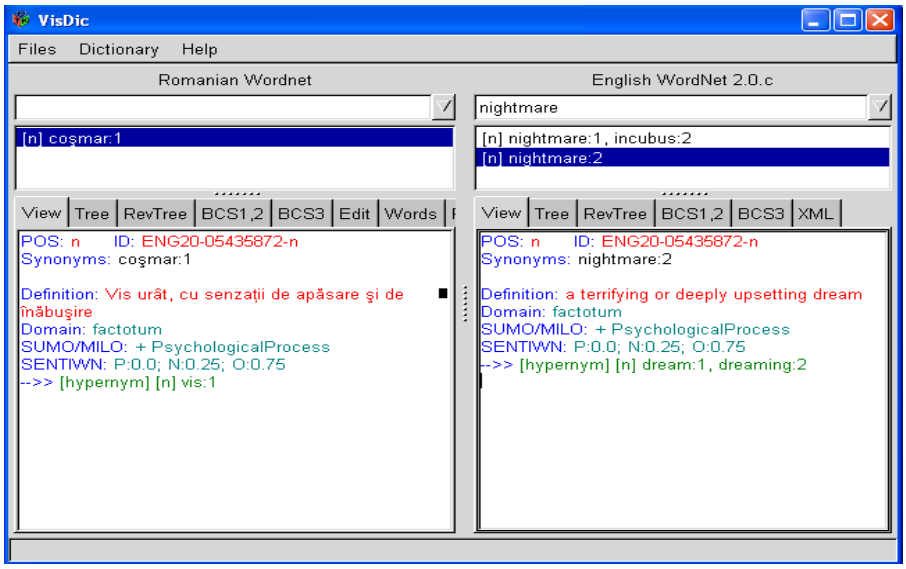


**Fig. 2.** VISDIC synchronized view of PWN and Ro-WordNet.

The Ro-WordNet can be browsed by a web interface implemented in our language web services platform (see Figure 3). Although currently only browsing is implemented, Ro-WordNet web service will, later on, include search facilities accessible via standard web services technologies (SOAP/WSDL/UDDI) such as distance between two word senses, translation equivalents for one or more senses, semantically related word-senses, etc.

## 3 Recent applications of the Ro-WordNet

In previous papers [17, 18] we demonstrated that difficult processes such as word sense disambiguation and word alignment of parallel corpora can reach very high accuracy if one has at his/her disposal aligned WordNets. Various other researchers showed the invaluable support of aligned WordNets in improving the quality of machine translation. In this section we will discuss some new applications of Ro-En pair of WordNets, the performances of which strongly argue for the need to keep-up the Ro-WordNet development endeavor.

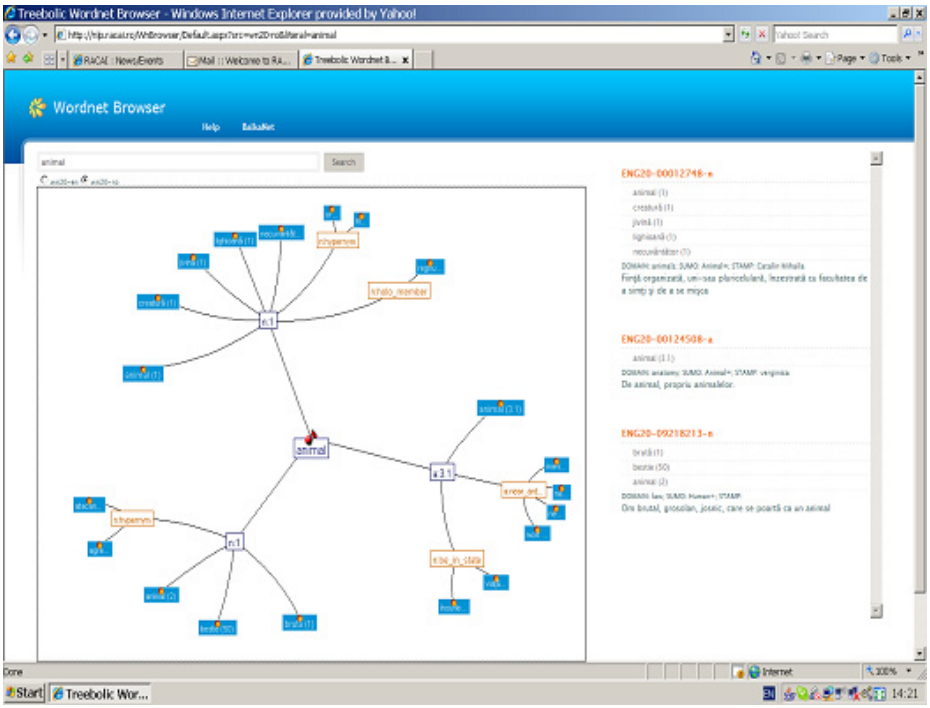


**Fig. 3.** Web interface to Ro-WordNet browser.

### 3.1 WordNet as an important resource to monolingual WSD

The WordNet concept has practically revolutionized the way a WSD application is thought. The explicit semantic structure of WordNet enables WSD application writers to use the semantic relations between synsets as a way of primitive reasoning when establishing senses of the words in a text. The presence of a particular semantic relation called hypernymy has also provided the much-expected mechanism of generalization of words' senses allowing for deploying of machine learning methods to WSD. It can be safely stated that WordNet has pushed forward the very nature of WSD algorithms in the direction of true semantic processing.

In [19] we presented an unsupervised WSD algorithm whose disambiguation philosophy is entirely based on the WordNet architecture. The idea of the algorithm is that of combining the paradigmatic information provided by the WordNet with the contextual information of the word in both the training and the disambiguation phases. The context of the word is given by its dependency relations with the neighboring words (which are not necessarily adjacent). In [20] we introduced the concept of *meaning attraction model* as theoretical basis for our monolingual WSD algorithm.

In the training phase, we estimate the measure of the meaning attraction between dependency-related words of a sentence. Given two dependency-related words, $W_a$

and $W_b$, each with its associated WordNet synset identifiers[2], the meaning attraction between synset id *i* of word $W_a$ and synset id j of word $W_b$ is a function of the frequency counts between pairs <*i*,*j*>, <*i*,*> and <*,*j*> collected from the entire training corpus. As meaning attraction functions, we chose DICE, Log Likelihood and Pointwise Mutual Information which can be all computed given the pair frequencies described above. Consider for instance the examples in Figure 4.

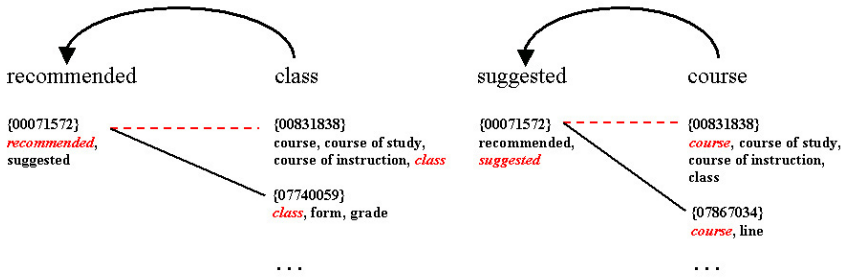


**Fig. 4.** Two examples of dependency pairs with the relevant information for learning.

Both "recommended" and "suggested" are in the same synset with the id 00071572. Also "class" and "course" are in the same synset with the id 00831838. This means that the pair <00071572,00831838> receives count 2 from these two examples as opposed to any other pair from the cartesian products which is seen only once. This fact translates to a preference for the meaning association "*mentioned as worthy of acceptance*" and "*education imparted in a series of lessons or class meetings*" which may not be correct in all contexts but it is a part of the natural learning bias of the training algorithm.

The synonymy lexical-semantic relation is just the first way of generalization in the learning phase. Another one, more powerful is given the by the semantic relations graph encoded in WordNet. In Figure 4 we have simply used the synsets' ids for computing frequencies. But their number is far too big to give us reliable counts. So, we are making use of the hypernym hierarchies for nouns and verbs to generalize the meanings that are learned but without introducing ambiguities. So, for a given synset id we select the uppermost hypernym that only subsumes one meaning of the word.

This WSD algorithm has recently participated in the 4th Semantic Evaluation Forum, SEMEVAL 2007 on the English All-Words Coarse and Fine-Grained tasks where it attained the top performance among the unsupervised systems.

Because it is language independent, it also has been applied with encouraging results to Romanian using the Romanian WordNet. The test corpus was the Romanian SemCor, a controlled translation of the English version of the corpus. The test set comprised of 48392 meaning annotated content word occurrences and for different meaning attraction functions and combinations of results using them, the best F-measure was 59.269%.

[2] By synset identifier, we understand the offset of the synset in the WordNet database. Knowing this ID and the word, we can extract the sense number of that word in the respective synset.

### 3.2 Romanian WordNet and Cross-Language QA

Romanian WordNet and its translation equivalence with the Princeton WordNet have been used as a general-purpose translation lexicon in the CLEF 2006 Romanian to English question answering track [9]. The task required asking questions in Romanian and finding the answer from an English text collection. For this task, the question analysis (focus/topic identification, answer type, keywords detection, query formulation, etc.) was made in Romanian and the rest of the process (text searching and answer extraction) was made in English.

Our approach was to generate the query for the text-searching engine in Romanian and then to translate every key element of the query (topic, focus, keywords) into English without modifying the query. Since we don't have a Romanian to English translation system and because neither the question nor the text collection were word sense disambiguated, for every key element of the query, we selected all the synonyms in which it appeared from the Romanian WordNet. Then, for every synonym of the latter list, we extracted all English literals of the corresponding English synset making a list of all possible translation equivalents for the source Romanian word. Finally, we ordered this list by the frequency of its elements computed from the English text collection and selected the first 3 elements as translation equivalents of the Romanian word. While this translation method does not assure a correct translation of each source Romanian word of the initial question, it is good enough for the search engine to return a set of documents in which the correct answer would be eventually identified. The evaluation of the recall for the IR part of the QA system [10] was close to 80%, but its major drawback was not in the translation part but in the identification of the keywords subject to translation.

### 3.3 Machine Translation Development Kit

The aligned Ro-En WordNets have been incorporated into our MT development kit, which comprises tokenization, tagging, chunking, dependency linking, word alignment and WSD based on the word alignment and the respective WordNets. The interface of the MTKit platform allows for editing the word alignment, word sense disambiguation, importing annotations from one language to the other and a friendly visualization of all the preprocessing steps in both languages. Figure 5 illustrates a snapshot from the MTKit interface. One can see the word alignment of translation unit (no. 16) from a document (nou-jrc42002595) contained in the JRC-Acquis multilingual parallel corpus. The right-hand windows display the morpho-lexical information attached to selected word from the central window (journal). The upper-right window displays the POS-tag, the lemma, the orthographic form and the WordNet sense number. The windows below it display the WordNet relevant information as well as the SUMO&MILO label pertaining to the corresponding sense number. The lowest-right window displays the appropriate WordNet gloss and SUMO documentation.

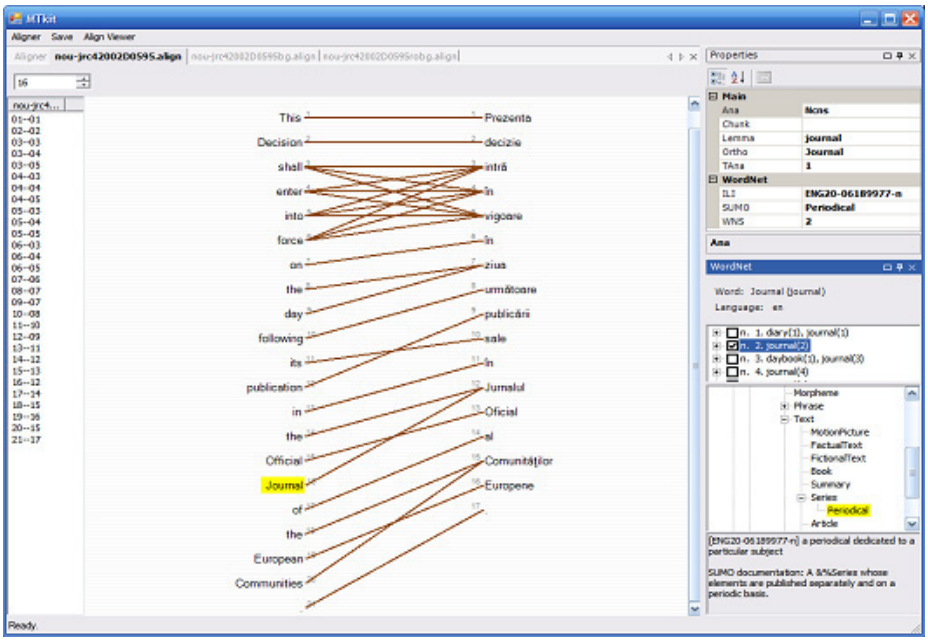


**Fig. 5.** MTKit interface.

### 3.4 Opinion analysis

One of the hottest research topics nowadays is related to subjectivity web mining with many applications in opinionated question answering, product review analysis, personal and institutional decision making, etc. Recent release of SentiWordNet [15] allowed for automatic import of the subjectivity annotations from PWN into any WordNet aligned with it. Thus, it became possible to develop subjectivity analysis programs for various languages equipped with a WordNet aligned to PWN.

We made some preliminary experiments with a naive opinion sentence classifier [21]. It simply sums up the O, P and N scores for each word in a sentence. For the words in the chunks immediately following a valence shifter, until the next valence shifter, the O, P and N scores are modified so that the new values are the following: $O_{new}$=1-$O_{old}$, $P_{new}$= $P_{old}$ $O_{old}$/($P_{old}$+$N_{old}$) and $N_{new}$= $N_{old}$ $O_{old}$ /($P_{old}$+$N_{old}$). Taking advantage of the *1984* Romanian-English parallel corpus which is word aligned and word sense disambiguated in both languages, we applied our naive opinion sentence classifier on the English original sentences and their Romanian translations and OpinionFinder [22] on the English original sentences. Since the WordNet opinion annotations are the same in PWN and Ro-WordNet aligned synsets it was obvious that our opinion classifier would give similar results for the two languages. So, in the end we compared the classifications made by our opinion classifier and OpinionFinder for a few English sentences. From the total number of 6411 sentences in *1984* corpus there were selected 954 for which both internal classifiers of Opinion Finder agreed in

judging the respective sentences as being subjective (see for details [22]) as in the sentence below:

*<MPQASENT autoclass1="subj" autoclass2="subj" diff="30.8">The stuff was like nitric_acid , and moreover , in swallowing it one had the sensation of being hit on the back of the head with a rubber club .</MPQASENT>*

We manually analyzed the 20 top-certainty sentences from the 954 selected ones, extracted the valence shifters they contained and dry-run the naive classifier described above, using the subjectivity values from SentiWordNet. When the O value for a sentence was smaller then 0.5 we arbitrary decided that it was subjective. All the 20 sentences were thus classified as subjective. For the same sentence[3], chunked and WSDed as below, the naive opinion classifier computed the following scores: P:0.063; N:0.563; O:0.375.

*[The stuff(1)] was(1) like (3) [nitric_acid(1)] , and moreover(1) , [in swallowing (1) it] one had (1) the sensation (1) of [being hit (4)] [on the back_of_the_head (1)] [with a rubber(1) club(3)].*

Whether the threshold value or the final P, N and O values might be debatable, the main idea here is that one can use the SentiWordNet annotation of the synsets in a WordNet for language L, aligned to PWN, for dwelling on subjectivity mining in arbitrary texts in language L.

## 4 Conclusions and further work

The development of Ro-WordNet is a continuous project, trying to keep up with the new updates of the Princeton WordNet. The increase in its coverage is steady (approximately 10,000 synsets per year for the last three years) with the choice for the new synsets imposed by the applications built on the basis of Ro-WordNet. Since PWN was aimed to cover general language, it is very likely that specific domain applications would require terms not covered by Princeton WordNet. In such cases, if available, several multilingual thesauri (EUROVOC - http://europa.eu/eurovoc/, IATE - http://iate.europa.eu/iatediff/about_IATE.html, etc.) can complement the use of WordNets. Besides further augmenting the Ro-WordNet, we plan the development of an environment where various multilingual aligned lexical resources (WordNets, framenets, thesauri, parallel corpora) could be used in a consistent but transparent way for a multitude of multilingual applications.

[3] The underlined words represent valence shifters, and the square parentheses delimit chunks as determined by our chunker; the numbers following the words represent their PWN sense number.

## Acknowledgements

The work reported here was supported by the Romanian Academy program "Multilingual Acquisition and Use of Lexical Knowledge", the ROTEL project (CEEX No. 29-E136-2005) and by the SIR-RESDEC project (PNCDI2, 4th Programme, No. D1.1-0.0.7), the last two granted by the National Authority for Scientific Research. We are grateful to many colleagues who contributed or continue to contribute the development of Ro-WordNet with special mentions for Cătălin Mihăilă, Margareta Manu Magda and Verginica Mititelu.

## References

1. Vossen, P. (ed.): A Multilingual Database with Lexical Semantic Networks. Kluwer Academic Publishers, Dordrecht (1998)
2. Rodriguez, H., Climent, S., Vossen, P., Bloksma, L., Peters, W., Alonge, A., Bertagna, F., Roventini, A.: The Top-Down Strategy for Building EuroWordNet: Vocabulary Coverage, Base Concepts and Top Ontology. J. Computers and the Humanities 32 (2-3), 117-152 (1998)
3. Miháltz, M., Prószéky, G.: Results and evaluation of Hungarian nominal wordnet v1.0. In: Proceedings of the Second International Wordnet Conference (GWC 2004), pp. 175–180. Masaryk University, Brno (2003)
4. Erjavec, T., Fišer, D.: Language Resources and Evaluations, LREC 2006 22–28 May 2006. Genoa, Italy (2006)
5. Black, W., Elkateb, S., Rodriguez, H., Alkhalifa, M., Vossen, V., Pease, A., Fellbaum, C.: Introducing the Arabic WordNet Project. In: Sojka, P., Choi, K.S., Fellbaum, C., Vossen, P. (eds.) Proceedings of the third Global Wordnet Conference, Jeju Island, 2006, pp. 295–299 (2006)
6. Elkateb, S., Black, W, Rodriguez, H, Alkhalifa, M., Vossen, P., Pease, A., Fellbaum, C.: Building a WordNet for Arabic: In Proceedings of the Fifth International Conference on Langauge Resources and Evaluation. Genoa, Italy. (2006)
7. Tufiş D., Cristea, D., Stamou, S.: BalkaNet: Aims, Methods, Results and Perspectives: A General Overview. J. Romanian Journal on Information Science and Technology, Special Issue on BalkaNet, Romanian Academy, 7(2-3) (2004a)
8. Steinberger, R., Pouliquen, B., Widiger, A., Ignat, C., Erjavec, T., Tufiş, D.: The JRC-Acquis: A multilingual aligned parallel corpus with 20+ languages. In: Proceedings of the 5th LREC Conference, Genoa, Italy, 22-28 May, 2006, pp. 2142-2147, ISBN 2-9517408-2-4, EAN 9782951740822 (2006)
9. Puşcasu, G., Iftene, A., Pistol, I., Trandabăţ, D., Tufiş, D., Ceauşu, A., Ştefănescu, D., Ion, R., Orăşan, C., Dornescu, I., Moruz, A., Cristea, D.: Developing a Question Answering System for the Romanian-English Track at CLEF 2006. In: Peters, C., Clough, P., Gey, F.C., Karlgren, J., Magnini, B., Oard, D.W., de Rijke, M., Stempfhuber, M. (eds.) LNCS Lecture Notes in Computer Science, ISBN: 978-3-540-74998-1, pp. 385–394. Springer-Verlag (2007)
10. Tufiş, D, Ştefănescu, D., Ion, R., Ceauşu, A.: RACAI's Question Answering System at QA@CLEF 2007. CLEF2007 Workshop, p. 15., September, 2007. Budapest, Hungary (2007)
11. Tufiş, D., Barbu, E., Mititelu, V., Ion, R., Bozianu, L.: The Romanian Wordnet. J. Romanian Journal on Information Science and Technology, Special Issue on BalkaNet, Romanian Academy, 7(2-3) (2004b)

12. Bentivogli, L, Forner, P., Magnini, B., Pianta, E.: Revising WordNet Domains Hierarchy: Semantics, Coverage, and Balancing. In: Proceedings of COLING 2004 Workshop on "Multilingual Linguistic Resources", pp. 101–108. Geneva, Switzerland, August 28, 2004 (2004)
13. Niles, I., Pease, A. Towards a Standard Upper Ontology. In: Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001). Ogunquit, Maine, October 17–19, 2001 (2001)
14. Niles, I. Pease, A.: Linking Lexicons and Ontologies: Mapping WordNet to the Suggested Upper Model Ontology. In: Proceedings of the 2003 International Conference on Information and Knowledge Engineering. Las Vegas, USA (2003)
15. Esuli, A., Sebastiani, F.: SentiWordNet: A publicly Available Lexical Resourced for Opinion Mining. LREC2006 22 - 28 May 2006. Genoa, Italy (2006)
16. Horák, A., Smrž, P.: New Features of Wordnet Editor VisDic. J. Romanian Journal of Information Science and Technology 7(2-3) (2004)
17. Tufiş, D., Ion, R., Ide, N.: Fine-Grained Word Sense Disambiguation Based on Parallel Corpora, Word Alignment, Word Clustering and Aligned Wordnets. In: Proceedings of the 20th International Conference on Computational Linguistics, COLING2004, pp. 1312–1318. Geneva (2004d)
18. Tufiş, D., Ion, R., Ceauşu, Al., Ştefănescu, D.: Combined Aligners. In: Proceeding of the ACL2005 Workshop on "Building and Using Parallel Corpora: Data-driven Machine Translation and Beyond", pp. 107–110. Ann Arbor, Michigan, June, 2005 (2005)
19. Ion, R.: Word Sense Disambiguation Methods Applied to English and Romanian. (in Romanian). PhD thesis. Romanian Academy, Bucharest (2007)
20. Ion, R., Tufiş, D.: Meaning Affinity Models. In: Proceedings of the 4th International Workshop on Semantic Evaluations, SemEval-2007, p. 6. Prague, Czech Republic, June 23–24 2007, ACL 2007 (2007)
21. Tufiş, D., Ion, R.: Cross lingual and cross cultural textual encoding of opinions and sentiments. Tutorial at Eurolan 2007: "Semantics, Opinion and Sentiment in Text" Iaşi, July 23–August 3, 2007 (2007)
22. Wilson, T., Hoffmann, P., Somasundaran, S., Kessler, J., Wiebe, J., Choi, Y., Cardie, C., Riloff, E., Patwardhan, S.: OpinionFinder: A system for subjectivity analysis. In: Proceedings of HLT/EMNLP 2005 Demonstration Abstracts, pp. 34–35. Vancouver, October 2005 (2005)
23. Fellbaum C. (ed.): WordNet: An Electronic Lexical Database. MIT Press (1998)
24. Magnini, B., Cavaglià, G.: Integrating Subject Field Codes into WordNet. In: Gavrilidou, M., Crayannis, G., Markantonatu, S., Piperidis, S., Stainhaouer, G. (eds.): Proceedings of LREC-2000, Second International Conference on Language Resources and Evaluation, pp. 1413–1418. Athens, Greece, 31 May–2 June, 2000 (2000)
25. Miller, G.A., Beckwidth, R., Fellbaum, C., Gross, D., Miller, K.J.: Introduction to WordNet: An On-Line Lexical Database. J. International Journal of Lexicography 3(4), 235–244 (1990)
26. Tufiş, D., Cristea, D.: Methodological issues in building the Romanian Wordnet and consistency checks in Balkanet. In: Proceedings of LREC2002 Workshop on Wordnet Structures and Standardisation, pp. 35–41. Las Palmas, Spain (2002)
27. Tufiş, D., Ion, R., Barbu, E., Mititelu, V.: Cross-Lingual Validation of Wordnets. In: Proceedings of the 2nd International Wordnet Conference, pp. 332–340. Brno (2004c)
28. Tufiş, D., Mititelu, V., Bozianu, L., Mihăilă, C.: Romanian WordNet: New Developments and Applications. In: Proceedings of the 3rd Conference of the Global WordNet Association, pp. 337–344. Seogwipo, Jeju, Republic of Korea, January 22–26, 2006, ISBN 80-210-3915-9 (2006)
29. Tufiş, D., Barbu, E.: A Methodology and Associated Tools for Building Interlingual Wordnets. In: Proceedings of the 4th LREC Conference, pp. 1067–1070. Lisbon (2004)