Memoirs of the Scientific Sections of the Romanian Academy
Tome XXXIV, 2011



# HYPONYMY PATTERNS IN ROMANIAN

VERGINICA BARBU MITITELU

*Research Institute for Artificial Intelligence, Romanian Academy, Bucharest, Romania*
*13, Calea 13 Septembrie, 050711 Bucharest, Romania.*
*E-mail: vergi@racai.ro.*

Hyponymy is a lexical-semantic relation that has been studied in Romanian linguistics only occasionally and almost exclusively in connection to other relations. However, for computational engineers it offers a very effective way of organizing the lexical material useful in many applications involving Natural Language Processing. In this paper we present two methods of identifying Romanian hyponymy patterns, their results and evaluation; we also envisage the applicability of these patterns and future work.

*Key words:* wordnet, semantic relations, hyponymy patterns.

## 1. INTRODUCTION

Lexical-semantic relations have been a subject of interest even since antique times. The scientific context in the second half of the last century fostered a rebirth of interest in this topic, due to the more and more prominent outline of a new domain of research, namely computational linguistics, and to its preoccupation for processing and generating texts in natural languages. Lexical-semantic relations intervene at many levels in the understanding and producing of natural language. Moreover, they represent a semantic factor, that, along others (frequency, the stylistic-functional factor, etymological and psychological factors) contribute to the organization of the huge quantity of words that make up the lexicon of a language [4].

Hyponymy is a paradigmatic relation corresponding to the class inclusion in logics [11], [6]. The word designating the extensionally more comprising class is called hypernym, and the one designating the included class is the hyponym. In theoretical linguistics this relation was defined, characterized [6], [13], types [6], [10], [22], and tests [7], [14] were proposed; it was compared and contrasted with similar relations such as meronymy and instance-of [10]. Most of the experiments in automatic extraction of relations from texts in Computational Linguistics prefer hyponymy.

In this paper we present two ways of identifying hyponymy patterns in Romanian, which can be useful for enriching the Romanian wordnet with new synsets. After a

description of the state of the art in this domain (section 2), we describe two ways of identifying hyponymy patterns in Romanian (subsections 3.1 and 3.2), compare the results of the two approaches (subsection 3.3), and present the results of the evaluation of these patterns on a domain corpus (section 4), followed by the presentation of the possible applications of this study (section 5) and the conclusions of the paper.

## 2. BACKGROUND

Studies in theoretical linguistics are characterized by a paradigmatic perspective on lexical-semantic relations. Their common semantic properties have been identified, the terms making up a paradigm have been enumerated (when their number was finite) with the help of componential analysis. The relations have been defined, characterized, classified according to various criteria. However, the semanticists have underlined the interdependency between the paradigmatic and syntagmatic perspectives in the semantic study of words [4].

Computational Linguistics and Artificial Intelligence are both interested in the syntagmatic aspect. Having access to huge texts, the researchers can extract different information from them, the relations between words being one of them. There are two main approaches to extraction of relations from corpora: the pattern-based approach, and the clustering approach. The former relies on the presupposition that, at the text level, one can identify lexical-syntactic structures specific to a certain semantic relation. Most of the experiments focused on hyponymy [1], [2], [9], [12], [15], [17], [18], etc., meronymy [3], [8], [17], and on other relations such as person – date of birth, invention – inventor, discoverer – discovery [19], etc. The clustering approach also focused on hyponymy [5], [16], [17]. This paper belongs to the first type of approach.

## 3. IDENTIFYING HYPONYMY PATTERNS IN ROMANIAN

Hyponymy patterns are lexical-syntactic patterns that allow for the co-occurrence, at short distance in text, of a hypernym and of one or more of its hyponyms either in the order HYPERNYM *hyponymy pattern* HYPONYM, or HYPONYM *hyponymy pattern* HYPERNYM.

The experiments in which hyponymy patterns for English have been identified are numerous (see references above). As far as we know, this is the first experiment addressing the Romanian language.

In the subsections below we present two ways in which we have identified hyponymy patterns for Romanian.

### 3.1. SEMI-AUTOMATIC IDENTIFICATION OF HYPONYMY PATTERNS FOR ROMANIAN

Using a Perl script, we extracted from a (segmented and lemmatized) Romanian corpus (of 881817 lexical units) the sentences containing nouns in (direct and indirect[i]) hyponymy relations. We disregarded the textual distance between the two nouns. In order to recognize the hyponym – hypernym pairs, we used the Romanian wordnet (RoWN) [21], that had, at the time of the experiment, 46269 synsets. We imposed no restriction on the distance between the nouns in the hyponymy tree in wordnet. The sentences thus extracted have been automatically grouped according to the similarity of the lexical material between the hyponym and its hypernym: if in *n* sentences there appear a hyponym and one of its hypernyms, and between them there is the same group of words, with identical lemmas, in the same order, then the *n* sentences are grouped together, irrespective of the hyponym – hypernym pair(s) they contain, as examples of the same hyponymy pattern. Thus, we identified the following patterns (in which GN stands for noun phrase, h for hyponym, and H for hypernym):

- *GN(H)* ***şi anume*** *GN(h)* "namely"*: Ea nu are nici_o bază de susţinere, numai o* ***idee****, şi anume o* ***idee*** *indestructibilă.*
- *GN(h)* ***fi un fel de*** *GN(H)* "be a kind of"*:* ***exprimare*** *e un_fel_de* ***pornografie***
- *GN(H)* ***care avea fi*** *GN(h)* "that have be"***: oameni*** *care au fost* ***participanţi***
- *GN(h)* ***şi (orice) alt*** *GN(H)* "and (any) (an)other"***: bani*** *şi alte* ***lucruri*** *suspecte;* ***taxele*** *şi orice alte* ***venituri*** *ale bugetului de stat*

[i] Direct hyponymy is established between two nodes linked by an edge: one of them is the mother-node and the other is the daughter-node. Indirect hyponymy is established, due to the transitivity of the hyponymy, between two nodes A and B that are not connected by any edge, but there is at least one node C between them so that A and C are in hyponymy relation and C and B are also in hyponymy relation. As hyponymy is a transitive relation, it results that A and B are, indirectly, in hyponymy relation.

•*GN(h)* ***şi (tot) celălalt*** *GN(H)* "and (all) the other(s)"***: Legile*** *şi toate celelalte* ***acte*** *normative*

•*GN(H),* ***mai_ales*** *GN(h)* "especially"***: delicvenţilor*** *de drept comun, mai_ales* ***gangsterilor***

•*GN(h)* ***deveni*** *GN(H)* "become"*: s-ar putea ca anumite* ***prevederi*** *să devină* ***subiect*** *de dispută*

•*GN(H)* ***nu ca un*** *GN(h)* "not as a"*: se poarte ca un* ***om****, nu ca un* ***primar***

•*GN(h)* ***fi considerat (ca)*** *GN(H)* "be considered (as)"*:* ***televiziunile*** *sunt considerate ca principale* ***instrumente*** *de luptă politică*

•*GN(H)* ***sine numi*** *GN(h)* "called"*: Aceşti* ***oameni*** *se numeau* ***capitalişti***

•*GN(H),* ***inclusiv*** *GN(h)* "inclusive"***:*** *toţi* ***oamenii*** *puterii, inclusiv* ***miniştrii***

•*GN(h)* ***fi un*** *GN(H)* "be a"***: Turcul*** *e un* ***om*** *sărman, are întotdeauna mustaţă, face bani din piatră seacă şi te întreabă dacă îţi place Turcia*

•*GN(h)* ***sau alt*** *GN(H)* "or (an)other"*: bănci* ***sau alte*** *instituţii de împrumut*

•*GN(H),* ***adică*** *GN(h)* "that is"*: revin la "acest* ***subiect*** *sensibil", adică la* ***cazul*** *Vântu - FNI*

•*GN(H),* ***ci (şi/doar) un*** *GN(h)* "but (also/only) a"***:*** *e un* ***om*** *cu idei, ci doar un* ***animal***

We tested these patterns on a journalistic corpus of 900000 lexical units. We automatically extracted from it those sentences in which the above patterns occur. We manually selected only those that also obey the syntactic criterion. The results are in Table 1. The "number of occurrences" column displays the total number of the respective pattern occurrences in the whole corpus. The "number of relevant occurrences" column displays the number of cases in which the positions of the two noun phrases are occupied by words in hyponymy relation. This relation can already be in the RoWN, can be valid, although not registered in RoWN, as this is incomplete, or can be a contextual hyponymy that need not be registered in the semantic network. This evaluation has been made manually, on the one hand, because of the incomplete character of the RoWN, and, on the other, in order not to lose the cases of contextual hyponymy, that would not have been recognized automatically. Moreover, as the corpus was not parsed, the manual evaluation also allowed for recognizing those cases when dependents of the hyponym and/or of the hypernym occur in the structure, without affecting its character: for instance, "revin la "acest **subiect** sensibil", adică la **cazul** Vântu - FNI", where the hypernym *subiect* "subject" is followed by a modifier (the

adjective *sensibil* "sensible") and the hyponym *cazul* "the case" is preceded by the preposition *la* "to".

The pattern precision was calculated as a percentage of the relevant number of occurrences of a pattern in its total number of occurrences. In this specific case, precision is equal to recall, so it also represents the accuracy of these patterns.

*Table 1*

Romanian hyponymy patterns evaluation results

| **No.** | **Pattern** | **Number of occurrences** | **Number of relevant occurrences** | **Accuracy (%)** |
|---|---|---|---|---|
| 1. | GN şi orice alt GN "and any other" | 2 | 2 | 100 |
| 2. | GN şi celălalt GN "and the other" | 4 | 4 | 100 |
| 3. | GN mai_ales GN "especially" | 1 | 1 | 100 |
| 4. | GN fi considerat GN "be considered" | 3 | 3 | 100 |
| 5. | GN sine numi GN "called" | 6 | 6 | 100 |
| 6. | GN fi un GN "be a" | 36 | 36 | 100 |
| 7. | GN sau alt GN "or (an)other" | 2 | 2 | 100 |
| 8. | GN ci (şi/doar) un GN "and (also/only) a" | 3 | 3 | 100 |
| 9. | GN deveni GN "become" | 15 | 14 | 93,3 |
| 10. | GN şi anume GN "namely" | 11 | 10 | 90,1 |
| 11. | GN şi alt GN "and (an)other" | 7 | 6 | 85,7 |
| 12. | GN inclusiv GN "inclusively" | 31 | 23 | 74,2 |
| 13. | GN adică GN "that is" | 8 | 5 | 62,5 |
| 14. | GN nu ca un GN "not as a" | 1 | 0 | 0 |

We mention that two patterns were not found in the test corpus: *GN fi un fel de GN* "be a kind of", *GN care avea fi GN* "that have be".

## 3.2. TRANSLATING ENGLISH HYPONYMY PATTERNS AND SEARCHING THE EQUIVALENTS IN A ROMANIAN CORPUS

In [2] we presented an experiment in which we identified hyponymy patterns for English. In Table 2 below we included the most relevant English hyponymy patterns and their accuracy (NP stands for noun phrase):

*Table 2*

English hyponymy patterns and their accuracy

| **No.** | **Pattern** | **Accuracy (%)** |
|---|---|---|
| 1. | NP other than NP | 100 |
| 2. | NP especially NP | 100 |
| 3. | NP principally NP | 100 |
| 4. | NP usually NP | 100 |

| | | |
|---|---|---|
| 5. | NP such as NP | 99,2 |
| 6. | NP in particular NP | 92,3 |
| 7. | NP e(.)g(.)NP | 91,4 |
| 8. | NP become NP | 91 |
| 9. | NP another NP | 87 |
| 10. | NP notably NP | 86,8 |
| 11. | NP particularly NP | 84,6 |
| 12. | NP except NP | 84,6 |
| 13. | NP called NP | 81,5 |
| 14. | NP like NP | 81,3 |
| 15. | NP including NP | 80,6 |
| 16. | NP mainly NP | 75 |
| 17. | NP mostly NP | 70,8 |
| 18. | NP i.e. NP | 65 |

We mention that, besides the patterns in the above table, we also identified others, for which we could not establish the accuracy, as they were not found in the test corpus: NP *be another* NP, NP *namely* NP, NP *and other* NP, NP *or other* NP, NP *a form of* NP, NP *or another* NP, NP *and similar* NP, NP *or similar* NP, NP *not least* NP, NP *but not* NP, NP *a kind of* NP, NP *like other* NP, NP *in common with other* NP, NP *and sometimes other* NP, NP *and many other* NP, NP *and in other* NP, NP *or any other* NP, NP *which be* NP, NP *for example* NP, NP *that is* NP, NP *apart from* NP, NP *even* NP, NP *be* NP, NP *for instance* NP, NP *as* NP, NP *either* NP, NP *as well as* NP. We translated these patterns and checked their occurrences in a Romanian corpus of 900000 lexical units. The evaluation method is identical to the one described above, for the automatically found Romanian patterns. The results are in Table 3.

*Table 3*

Romanian hyponymy patterns obtained by translating the English ones and their accuracy

| **No.** | **Pattern** | **Number of occurrences** | **Number of relevant occurrences** | **Accuracy (%)** |
|---|---|---|---|---|
| 1. | GN de exemplu GN “for example” | 1 | 1 | 100 |
| 2. | GN ca de pildă GN “for instance” | 1 | 1 | 100 |
| 3. | GN de pildă GN “for instance” | 2 | 2 | 100 |
| 4. | GN cum ar fi GN “such as” | 7 | 7 | 100 |
| 5. | GN, mai puţin GN “except” | 1 | 1 | 100 |
| 6. | GN de obicei GN “usually” | 1 | 1 | 100 |
| 7. | GN altul/alta/alţii/altele decât GN “(an)other than” | 2 | 2 | 100 |
| 8. | GN sau alt/altă/alţi/alte GN “or (an)other” | 9 | 9 | 100 |
| 9. | GN sau orice alt/altă/alţi/alte GN “or any other” | 2 | 2 | 100 |

| | | | | |
|---|---|---|---|---|
| 10. | GN şi anume GN “namely” | 2 | 2 | 100 |
| 11. | GN numit GN “called” | 51 | 50 | 98 |
| 12. | GN deveni GN “become” | 114 | 101 | 88,6 |
| 13. | GN mai ales GN “especially” | 8 | 7 | 87,5 |
| 14. | GN şi alt/altă/alţi/alte GN “and (an)other” | 53 | 44 | 83 |
| 15. | GN adică GN “that is” | 30 | 19 | 63,3 |
| 16. | GN cu excepţia GN “with the exception” | 16 | 10 | 62,5 |
| 17. | GN care fi GN “that be” | 8 | 5 | 62,5 |
| 18. | GN în special GN “especially” | 5 | 3 | 60 |
| 19. | GN inclusiv GN “inclusively” | 53 | 29 | 54,7 |
| 20. | GN afară de GN “besides” | 11 | 5 | 45,5 |
| 21. | GN precum GN “such as” | 6 | 2 | 33,3 |
| 22. | GN în afară de GN “besides” | 9 | 2 | 22,2 |
| 23. | GN chiar şi GN “even” | 29 | 6 | 20,7 |
| 24. | GN un fel de GN “a kind of” | 27 | 5 | 18,5 |
| 25. | GN alt/altă/alţi/alte GN “(an)other” | 8 | 1 | 12,5 |
| 26. | GN ca GN “as” | 700 | 40 | 5,7 |
| 27. | GN până şi GN “even” | 11 | 0 | 0 |
| 28. | GN dar nu GN “but not” | 2 | 0 | 0 |

### 3.3. COMPARISON BETWEEN THE RESULTS OF THE TWO METHODS

The quantitative difference between the translated and the semi-automatically identified patterns has three explanations:

- the quantitative difference between the two wordnets: the English one, Princeton WordNet 2.1, contains 117597 synsets, while the RoWN we used contains 46269 synsets;
- the quantitative difference between the corpora used: the fragments from the British National Corpus we used for identifying the English patterns have 1863 MB, while the Romanian one, only 25.6 MB;
- when we evaluated the translated patterns, we also accepted as relevant the cases of contextual hyponymy, which is not registered in wordnet and is established between words that, in the respective context, function as hyponym – hypernym pair, and also those that result from the metonymic interpretation of the context[ii] .

[ii] Hearst (1992) also remarks the existence of some hyponyms that result from metonymy, dependent on the context or on the perspective assumed on the matter. Considering that her aim was to enrich the

Comparing the set of Romanian patterns identified by translating the English ones (subsection 3.1.) with the set of those semi-automatically identified (subsection 3.2.), we notice the following:

- the intersection of the two sets, represented by the patterns identified by both methods, contains: *sau alt* "or (an)other", *şi anume* "namely", *sine numi* "called", *deveni* "become", *mai_ales* "especially", *şi alt* "and (an)other", *adică* "that is", *inclusiv* "inclusively". This means that almost 29% of the translated patterns have also been identified semi-automatically, and that 57% of the ones found semi-automatically were also identified by translating the English patterns;
- these common patterns are, as expected, among those with huge accuracy; in general, more than 80%, except for *adică* "that is" and *inclusiv* "inclusively". If we calculate a means of the precision for each pattern, we find that the pattern *sau alt* "or (an)other" has a maximal precision; the others have a mean precision above 90%: *numit/sine numi* "called" 99%, *şi anume* "namely" 95%, *mai_ales* "especially" 93,8%, *deveni* "become" 91%;
- the common patterns are not necessarily the most frequent in language, which was also to be expected. Their huge precision implies a reduced syntactic polysemy, so with a low frequency.

If we compare the English and the Romanian patterns with maximal precision, we notice that the intersection set contains three elements: *other than – altul decât*, *especially – mai_ales*, *usually – de obicei*.

70% of the total number of English patterns have equivalent Romanian hyponymy patterns. The other way round, 66.(6)% of the Romanian patterns have equivalent English hyponymy patterns. As one can notice, the percentages are quite close; this can

English wordnet, the problem is not trivial at all and urges for the analysis of the examples before introducing them in wordnet.

be a proof of the fact that different languages allow for the co-occurrence, in similar contexts, of word in hyponymy relations to each other.

## 4. EVALUATING THE ROMANIAN HYPONYMY PATTERNS AGAINST A DOMAIN CORPUS

We tested the hyponymy patterns for Romanian against a sub-corpus of the OPUS corpus (http://www.let.rug.nl/~tiedemann/OPUS/), that is against the EMEA (European Medicines Agency) documents containing 11,914,802 tokens. This sub-corpus has two main characteristics that influence the experiment results: abundance of repeated expressions and the specialized vocabulary [20].

The evaluation method applied here is the same as the one described above for the evaluation of the patterns identified via the two methods (see 3.1 and 3.2). The results are in Table 4.

*Table 4*

Accuracy of Romanian hyponymy patterns in a domain corpus

| Pattern | Accuracy (%) |
|---|---|
| GN *chiar şi* GN "even" | 100 |
| GN *de obicei* GN "usually" | 100 |
| GN, *ci (şi/doar)* GN "but (also/only)" | 100 |
| GN *în special* GN "especially" | 96.88 |
| GN *precum* GN "as" | 94.83 |
| GN *cum ar fi* GN "such as" | 93.75 |
| GN *(în) afară de* GN "besides" | 92.11 |
| GN *şi (orice) alt* GN "and (any) other" | 90.1 |
| GN *fi un* GN "be a" | 87.98 |
| GN *sau alt* GN "or (an)other" | 86.96 |
| GN *mai ales* GN "especially" | 85.71 |
| GN *alt decât* GN "other than" | 85.71 |
| GN *sine numi* GN "called" | 84 |
| GN *inclusiv* GN "inclusively" | 83.51 |
| GN *de exemplu* GN "for example" | 79.57 |
| GN *fi considerat* GN "be considered" | 79.17 |
| GN *care fi* GN "that be" | 74.12 |
| GN, *adică* GN "that is" | 66.66 |
| GN *cu excepţia* GN "except for" | 54.55 |
| GN *şi (tot) celălalt* GN "and (all) others" | 54.29 |

Comparing these results with the ones obtained against the journalistic corpora, we notice that, in most cases, the accuracy decreases. However, there are some patterns that have a higher accuracy in the domain corpus: GN, *adică* GN "that is", GN *care fi* GN

"that be", GN *în special* GN "especially", GN *inclusiv* GN "inclusively", GN *precum* GN "as", GN *(în) afară de* GN "besides". GN *de obicei* GN and GN *ci (şi/doar)* GN preserve their maximal accuracy.

## 5. APPLICATIONS OF THE HYPONYMY PATTERNS

RoWN is a useful linguistic resource in CL applications. The quality and quantity of its synsets influence the quality of the applications using RoWN. Thus, its enrichment is a priority. It has been developed manually so far, but we can add it automatically identified synsets, with the help of the hyponymy patterns presented above. A linguist needs to add these synsets a gloss and ensure their completeness.

Hyponymy patterns can be used, on the one hand, for the enrichment of the RoWN with both hyponyms and instances, and, on the other, both with words from the general and the specialized vocabularies.

In the work described here we did not distinguish between hyponymy and instance-of relations, as they are treated identically in the wordnet versions we used. However, these are distinct semantic relations, and the latest version of the Princeton WordNet 3.0 differentiates between them. Further study should check to what extent the patterns identified here have a different accuracy if we consider this distinction and how precise they are in the identification of instances. Named Entities Recognition tasks fully justifies the inclusion of instances in a linguistic resource such as a wordnet.

Most of the words forming the general vocabulary have already been included in RoWN. Our interest now can be its enrichment with terms from various domains. The experiment of testing the hyponymy patterns on a specialized corpus offers some data in this respect.

Form the lexical semantics perspective, experiments such as the one presented here can bring supplementary data, examples and refinements of the mostly paradigmatic analysis of hyponymy.

## 6. CONCLUSIONS

This paper is a contribution to the research in semantic relations extraction from corpora by means of lexical-syntactic patterns. We presented two methods of

identifying hyponymy patterns in Romanian. One method is semi-automatic, the other presupposes the translation of some English hyponymy patterns and the evaluation of their equivalents in Romanian texts. We calculated the accuracy of the patterns both in a journalistic and a domain corpus. The results prove that these patterns could be used for mining corpora in order to enrich the RoWN with new hyponyms and instances, and even for checking the completeness of the existent synsets.

**Acknowledgements.** We are grateful to the Romanian Ministry of Education, Research, Youth and Sport, that has financed the SIR-RESDEC project, in which this research was carried out.

REFERENCES

1. ALFONSECA E., MANANDHAR S., *Improving an Ontology Refinement Method with Hyponymy Patterns*, Third International Conference on Language Resources and Evaluation, Las Palmas, 2001, 235-239.
2. BARBU MITITELU V., *Hyponymy Patterns. Semi-automatic Extraction, Evaluation and Inter-lingual Comparison,* in *Text, Speech and Dialogue* (P. Sojka, A. Horak, I. Kopecek, P. Karel, eds.), Springer, 2008, 37-44.
3. BERLAND M., CHARNIAK E., *Finding parts in very large corpora*, Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics, 1999, 57-64.
4. BIDU-VRĂNCEANU A., FORĂSCU N., *Limba română contemporană. Lexicul*, Humanitas Educaţional, Bucureşti, 2005.
5. CARABALLO S., *Automatic construction of a hypernym-labeled noun hierarchy from text*, Proceedings of the 37th Annual Meeting of the Association for Computational Liguistics, 1999, 120-126.
6. CRUSE D. A., *Lexical Semantics*, Cambridge, CUP, 1986.
7. CRUSE D. A., *Meaning in Language. An Introduction to Semantics and Pragmatics*, Second ed., Oxford, OUP, 2004.
8. GÎRJU R., BADULESCU A., MOLDOVAN D., *Automatic Discovery of Part-Whole Relations*, Computational Linguistics, 2006, **32** (*1*), , 82-135.

9. HEARST M. A., *Automated Acquisition of Hyponyms from Large text Corpora*, Proceedings of the Fourteenth International Conference on Computational Linguistics, Nantes, 1992.
10. KLEIBER G., TAMBA I., *L'hyponymie revisité: inclusion et hiérarchie*, in *Langages 98: L'hyponymie et l'hyperonymie*, Larousse, 1990.
11. LYONS J., *Semantics*, vol. 1, Cambridge University Press, 1977.
12. MANN G. S., *Fine-Grained Proper Noun Ontologies for Question Answering*, in *COLING-02 on SEMANET: building and using semantic networks*, 2002, 1-7.
13. MURPHY M. L., *Semantic Relations and the Lexicon*, Cambridge, CUP, 2003.
14. NYCKEES V., *La sémantique*, Paris, Belin, 1998.
15. OAKES M. P., *Using Hearst's Rules for the Automatic Acquisition of Hyponyms for Mining a Pharmaceutical Corpus*, Proceedings of the Workshop Text Mining Research, Practice and Opportunities, Borovets, 2005, 63-67.
16. PANTEL P., RAVICHANDRAN D., *Automatically labeling semantic classes*, Proceedings of HLT/NAACL-04, Boston, 2004, 321-328.
17. PANTEL P., PENNACCHIOTTI M., *Espresso: Leveraging Generic Patterns for Automatically Harvesting Semantic Relations*, Proceedings of the Conference on Computational Linguistics / Association for Computational Linguistics (COLING/ACL-06), Sydney, 2006, 113-120.
18. PAŞCA M., *Acquisition of Categorized Named Entities for Web Search*, Proceedings of CIKM'04, Washington, 2004.
19. RAVICHANDRAN D., HOVY E., *Learning surface text patterns for a question answering system*, Proceedings of ACL-2002, Philadelphia, 2002, 41-47.
20. TIEDEMANN J., *News from OPUS – A collection of multilingual parallel corpora with tools and interfaces*, in *Recent Advances in Natural Language Processing: Selected Papers from RANLP 2007*, John Benjamins, 2009, 237-248.
21. TUFIŞ D., ION R., BOZIANU L, CEAUŞU A., ŞTEFĂNESCU D., *Romanian Wordnet: Current State, New Applications and Prospects*, Proceedings of 4th Global WordNet Conference, GWC-2008 (Attila Tanacs, Dora Csendes, Veronika Vincze, Christiane Fellbaum, Piek Vossen, eds.), Szeged, 2008, 441-452.

22. WIERZBICKA A., *Apples are not a "kind of fruit"*, American Ethnologist, 1984, *11,* 313-28.

Received August 16, 2010
Revised January 12, 2011