

# Greek Wordnet Extension in the Domain of Psychology and Computer Science

Athanasios KREMIZIS<sup>1</sup>, Ioanna KONSTANTINIDI<sup>1</sup>, Marina PAPADAKI<sup>1</sup>, George KERAMIDAS<sup>1</sup>,  
Maria GRIGORIADOU<sup>2</sup>

*Abstract*— In this paper we present the investigative construction-methodology of a Greek dictionary in the domain of psychology (PsyGrTerms) and in the domain of computer science (CompSciTerms). We also present the construction methodology of PsychoGreekWordnet (PsyGrWn), a semantic Greek lexicon in the domain of psychology, stored in XML format. It contains synonym-sets of some of the most important Greek terms of contemporary psychology and their lexical semantic relations (Hypernymy, Holonymy), without recursive (cyclic) definitions or ambiguous definitions. Then we compare the searching results a) between English WordNet 2.0.c, Greek Wordnet 0.9.a and PsyGrWn, and b) between English WordNet 2.0.c, Greek Wordnet 0.9.a and CompSciTerms. Problematic issues and practical solutions adopted or proposed are described. Finally we discuss the foreseen advantages and improvements.

*Index Terms*—computational linguistics, machine-readable lexical database, monolingual semantic lexicon, WordNet, psychology domain, domain of computer science, jargon file.

## I. INTRODUCTION

WordNet® or Princeton WordNet [3] is a machine-readable lexical database of English organized by meanings. It was developed by the Princeton University. More specifically it is nouns, verbs, adjectives and adverbs grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. The synsets are interlinked by means of conceptual-semantic and lexical relations and form a network of meaningfully related terms and concepts. This structure is useful as a tool for computational linguistics and natural language processing (NLP).

Wordnet is any of the machine-readable lexical databases modeled after the PWN. Wordnets have been created for a lot of languages. EuroWordNet (EWN) [14][21] extended the PWN approach with the multi-lingual dimension adding an Inter-Lingual Index (ILI) to which all the monolingual wordnets for the languages represented in the project were aligned (based on PWN 1.5). EWN connected Dutch, Spanish, Italian, English, French, German, Czech, and Estonian. Balkanian WordNet (BalkaNet) [20], based on the PWN 2.0, extended the number of languages that the EWN contains connecting it with Bulgarian, Greek, Rumanian, Serbian, and Turkish.

<sup>1</sup> Interdisciplinary Program of Graduate Studies in Basic & Applied Cognitive Science, National and Kapodistrian University of Athens, Panepistimiopolis, Ilissia, Athens 15784, Hellas-Greece, [PsyGrWNGroup@C4U.gr](mailto:PsyGrWNGroup@C4U.gr)

<sup>2</sup> Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Panepistimiopolis, Ilissia, Athens 15784, Hellas-Greece, ([gregor@di.uoa.gr](mailto:gregor@di.uoa.gr)).

Greek Wordnet [1][16] is developed in the University of Patras from a team of linguists with the attendance of University of Athens. Statistical data of the Greek Wordnet are [4]: synsets 18.677, literals 24.811, literals/synset 1.33, Inter-Lingual Relations (ILR) 24.582, ILR per synset 1.33, nonlexicalized concepts 46, and definitions 18.649. The Global Wordnet project attempts to coordinate the production and linking of wordnets for all languages.[20]

VisDic [17][18] is a graphical application for viewing and editing machine-readable dictionaries structured in XML format. VisDic was primarily developed for browsing and editing wordnets (databases of meanings). It is the standard browser of the BalkaNet wordnets. The VisDic was developed by the Faculty of Informatics, Masaryk University Brno, Czech Republic (<http://nlp.fi.muni.cz/projects/visdic/>) and apart from the English language, supports also all the languages of BalKaNet (Greek, Bulgarian, Serbian, Rumanian and Turkish).

Domain relations may be attributed to each word in a wordnet in order to give us broader knowledge of the possible contexts for each wordnet entry. PWN [13] has domain relations that include psychology. Greek wordnet (GrWN) [15] has no domain relations, and the amount of terms that could be included in the domain of psychology and computer science is limited. The necessity of Greek wordnet extension becomes obvious if we take into account that domain information becomes crucial while we approach wordnet usage in information retrieval (IR) systems.[20]

In the following sections we describe the construction methodology of a monolingual (Greek) dictionary specialized in the domain of psychology (PsyGrTerms.txt) (Section II), the construction methodology of PsychoGreekWordnet (PsyGrWn) a semantic lexicon (wordnet) stored in XML format (Section III), the comparison of the searching results between PWN, GrWN and PsyGrWn (Section IV), the construction methodology of a Greek dictionary in the domain of computer science (CompScieTerms.txt), and the comparison of the searching results between PWN, GrWN and CompScieTerms.txt (Section V). In section VI we discuss the foreseen advantages and improvements.

## II. CONSTRUCTION-METHODOLOGY OF PSYGRTERMS TEXT FILE

In order to create a Greek dictionary in the domain of psychology, we find the most important English terms (nouns and verbs) from the domain of psychology, which derive from various

---

indexes of books of contemporary psychology. In other words, we choose to use the most frequently met English terms from the domain of psychology in contemporary bibliography. Then, we search these terms in PWN under the domain of psychology. Most of the terms were in other domains (e.g. pedagogy, factotum, physiology etc) but we used those that were hyponyms of the term *psychological feature* in the hierarchical tree of hypernymic relations. Then, we translate in Greek ([6], [12]) the terms, their definitions, synonyms and semantic relations (Hypernyms, Holonyms). Each definition is also examined for the existence of recursive (cyclic) definition. In case of existence, we continue the search of definition terms in PWN, until an accepted term (AT) with a noncircular definition is found. We also examine each definition for the existence of ambiguity. In case of existence, we continue the search of definition terms in PWN, until an AT with a more explicit definition is found. The process is repeated for 201 different synsets and this way we create the PsyGrTerms Text File (PsyGrTerms.txt), which contains the synsets and all their semantic relations (synonymy, hyponymy, hypernymy, meronymy, and holonymy) – in case they exist in PWN. The procedure of definition clarification results to the repetition of some synsets in the text file. E.g. the term *Αντιλαμβάνομαι* (in English: *to perceive*) is the AT in 8 different synsets that have cyclic definitions. This repetition adds an issue to be solved during the construction methodology of the PsyGrWn xml database.

### III. CONSTRUCTION-METHODOLOGY OF THE PSYGRWN XML DATABASE

The creation of PsyGrWn lexicon (PsyGrWn.xml) is modeled after PWN. A file in XML is created, with the use of VisDic, which at the end of below procedure contains synsets with their relations as found in PsyGrTerms Text File (PsyGrTerms.txt).

Using VisDic environment we create a new empty xml lexicon PsyGrWn.xml, and we do the following. If the term, that we want to add in the PsyGrWn.xml, has a circular or ambiguous definition in the PsyGrTerms.txt, we do not add it in the xml file but instead we add the AT. This is because we do not want to complicate the readers' comprehension of terms in the PsyGrWn. The disadvantage in this case is that the PsyGrWn.xml does not contain and does not present 201, but 175 different synsets of PsyGrTerms.txt. E.g. for the entry *Ακουστική αντίληψη* (in English: *auditory perception*) the PsyGrTerms.txt reports circularity, and has the entry *Αντιλαμβάνομαι* as not circular AT. Thus in the PsyGrWn we create the synset *Αντιλαμβάνομαι* and we ignore *Ακουστική αντίληψη*. In the process of eliminating recursive definitions the synset

*Αντιλαμβάνομαι* appears 8 times in the PsyGrTerms.txt as the AT with no cyclic definition. Thus, in order to add a term in PsyGrWn.xml, we seek the latter using VisDic's search box. If the term has been already added in the xml lexicon then it will be displayed in the results box, and we will seek the next term. If the term has not been added in the xml lexicon, then in VisDic's EDIT tab we select the command button NEW, we copy from PsyGrTerms.txt and paste the definition in Definition-box and the term with the synonyms in the Literal-boxes. We repeat previous procedure for all the terms with their definition and synonyms, without creating the relations that will be the last step. Prior to creation of relations, we have to add as new synsets the terms used as hypernyms/holonoms but have no definition in PsyGrTerms Text File. Thus, we search in the PsyGrWn.xml, for each of these terms. If the term has been already added in the xml lexicon then it will be displayed in the results box, and we will seek the next term. If the term has not been added in the xml lexicon, then we select the command button NEW and we copy/paste the term in the Literal box. We repeat previous procedure for all the terms used as hypernyms/holonoms in PsyGrTerms Text File. This way all related synsets are added in the PsyGrWn.xml file and we conclude to a total of 416 different sets of synonyms. Then, as last step we create the relations between these synsets. E.g.: From the PsyGrTerms.txt we have, for the term *ΑΠΝΗΣΗ* (in English: *denial*) *ΥΠΕΡΩΝΥΜΑ=ΜΗΧΑΝΙΣΜΟΣ ΑΜΥΝΑΣ, ΑΜΥΝΤΙΚΗ ΑΝΤΙΑΠΑΣΗ, ΑΜΥΝΑ* (in English: *hypernyms=defense mechanism, defense reaction, defence*)

In the related-synsets-search-box we type the related term e.g. *ΑΜΥΝΑ* (hypernym) and by pressing Enter key all previously added synsets that contain the term *ΑΜΥΝΑ* are presented in the same box. There we select the suitable synset *ΑΜΥΝΑ* and in the list box of type of relation, we select the corresponding relation from the list of relations (hypernym). Then we create relations (hypernym), with each one of the other 2 previously created synsets *ΜΗΧΑΝΙΣΜΟΣ ΑΜΥΝΑΣ*, and *ΑΜΥΝΤΙΚΗ ΑΝΤΙΑΠΑΣΗ*.

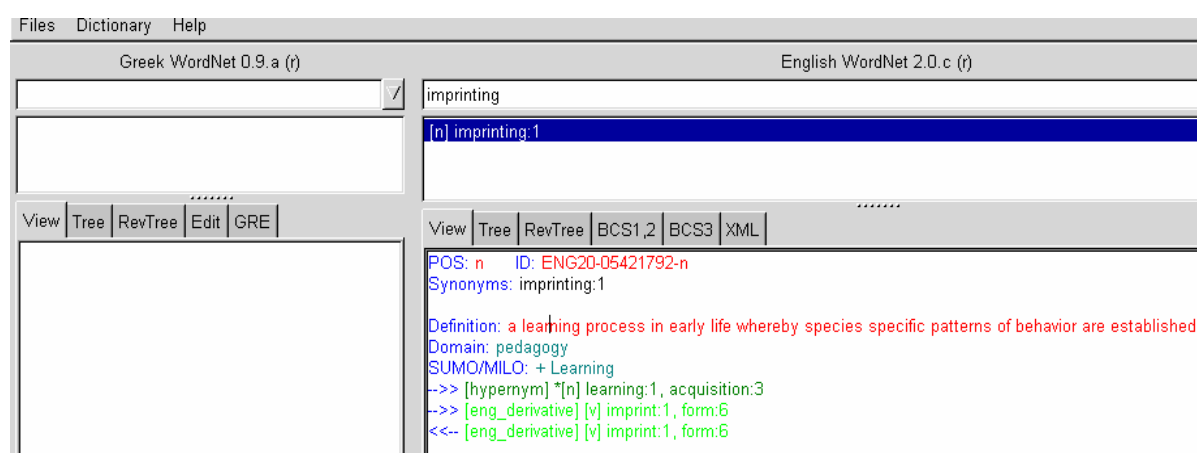
Repeating all above steps for all the terms of PsyGrTerms.txt we create the final PsyGrWn.xml that contains 416 different synsets of which 175 synsets with definitions, and. 67 different relations (hypernym, holo\_part).

#### IV. RESULTS

In order to compare GrWN and PsyGrWn, we count how many of the 175 synsets with definition included in PsyGrWn.xml exist in the hierarchical tree of hyponyms under the

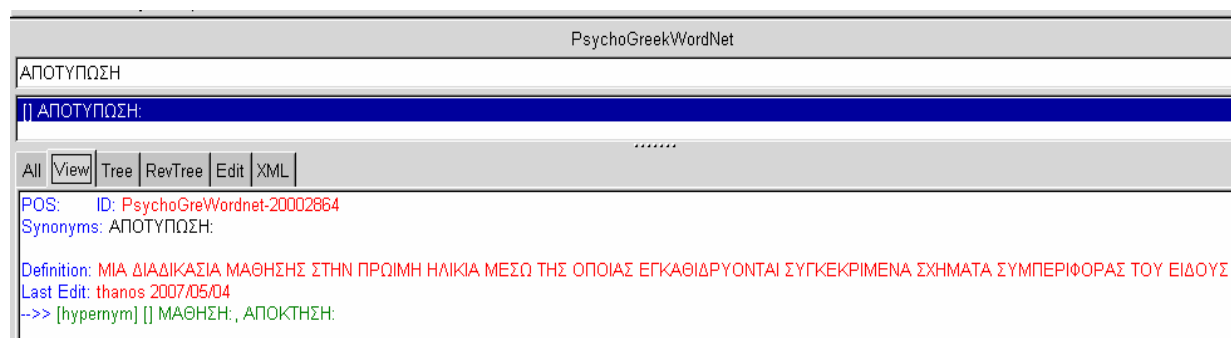
hypernym term *Ψυχολογικό Χαρακτηριστικό* (in English: *Psychological Feature*) of GrWN. We find 41 out of 175. We conclude that PsyGrWn.xml has 134 more, frequently met in contemporary bibliography, terms from the domain of psychology than GrWN.

In order to compare PWN and PsyGrWn, we do the following: Using the English-Greek dictionary of terms of psychology [5], we seek the English translation of the terms in PsyGrWn.xml and find the translation of 57 out of 175 synsets. For the comparison we select each of the 57 translated terms. E.g. we select *ΑΠΟΤΥΠΩΣΗ* - *imprinting*. With the use of VisDic, we seek the term *imprinting* in PWN (Fig. 1).



**Fig. 1.** Search results of term *imprinting* in the English WordNet 2.0.c (right frame) and the corresponding translation in GrWN (left frame, without results).

The search of term *ΑΠΟΤΥΠΩΣΗ* in the PsyGrWn has the results as they appear in Fig.2.



**Fig.2.** Search results of term *ΑΠΟΤΥΠΩΣΗ* in the PsyGrWn

The definition of the term *imprinting* in PWN is: "*a learning process in early life whereby species specific patterns of behavior are established*"

The definition of the corresponding term *ΑΠΟΤΥΠΩΣΗ* in the PsyGrWn is «*ΜΙΑ ΔΙΑΔΙΚΑΣΙΑ*

ΜΑΘΗΣΗΣ ΣΤΗΝ ΠΡΩΙΜΗ ΗΛΙΚΙΑ ΜΕΣΩ ΤΗΣ ΟΠΟΙΑΣ ΕΓΚΑΘΙΛΡΥΟΝΤΑΙ ΣΥΓΚΕΚΡΙΜΕΝΑ ΣΧΗΜΑΤΑ ΣΥΜΠΕΡΙΦΟΡΑΣ ΤΟΥ ΕΙΔΟΥΣ».

The two definitions are identical.

The hypernyms of the term *imprinting*, in PWN, are *learning*, *acquisition*. The hypernyms of the term ΑΠΟΤΥΠΩΣΗ in the PsyGrWn is ΜΑΘΗΣΗ, ΑΠΟΚΤΗΣΗ.

The hypernyms are respectively identical (*learning* = ΜΑΘΗΣΗ, *acquisition* = ΑΠΟΚΤΗΣΗ).

The comparison between hierarchical tree of hypernyms of the term *imprinting* in PWN (Fig. 3) and of the term ΑΠΟΤΥΠΩΣΗ in the PsyGrWn is (Fig. 4) is presented in **Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε..**

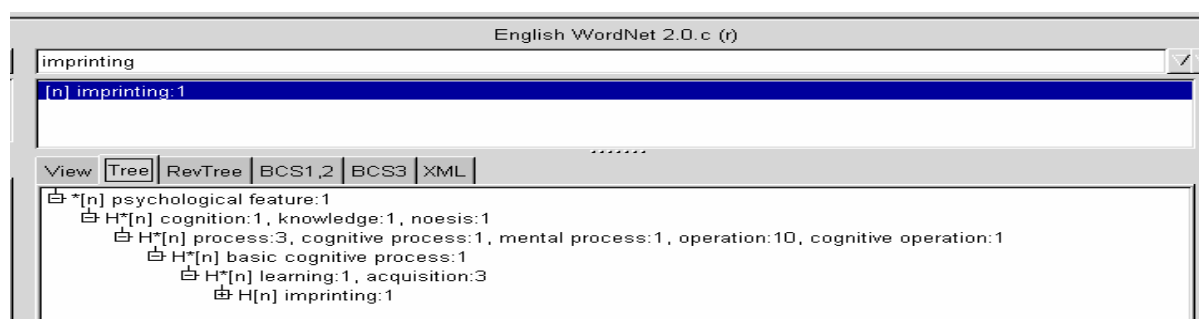


Fig. 3. The hierarchical tree of hypernyms of the term *imprinting* in PWN

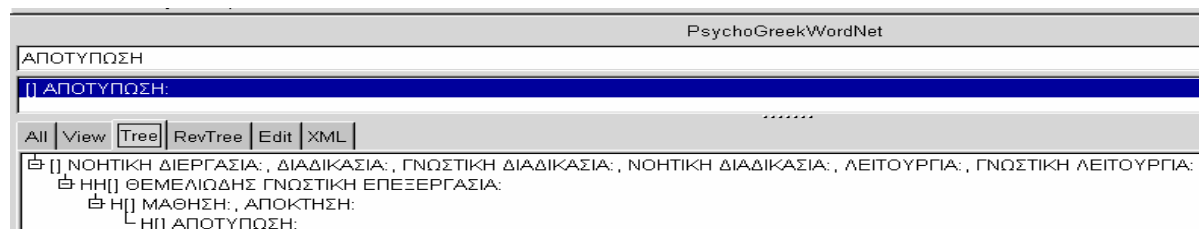


Fig. 4. The hierarchical tree of hypernyms of the term ΑΠΟΤΥΠΩΣΗ in the PsyGrWn

From the following table 1, of above hierarchical trees, appears that the hierarchy levels 3, 4, 5, 6 are identical, and 1, 2 are not.

Thus, is concluded that PsyGrWn needs more enrichment in the levels of relations.

**Table 1. Respected equivalences of hierarchical trees derived from PWN and PsyGrWn.**

Hierarchical Level	Hierarchical Tree derived from PWN	Hierarchical Tree derived from PsyGrWn
1.	psychological feature:1	
2.	cognition:1, knowledge:1, noesis:1	
3.	process:3, cognitive process:1, mental process:1, operation:10, cognitive operation:1	ΝΟΗΤΙΚΗ ΔΙΕΡΓΑΣΙΑ:, ΔΙΑΔΙΚΑΣΙΑ:, ΓΝΩΣΤΙΚΗ ΔΙΑΔΙΚΑΣΙΑ:, ΝΟΗΤΙΚΗ ΔΙΑΔΙΚΑΣΙΑ:, ΛΕΙΤΟΥΡΓΙΑ:, ΓΝΩΣΤΙΚΗ ΛΕΙΤΟΥΡΓΙΑ:
4.	basic cognitive process:1	ΘΕΜΕΛΙΩΔΗΣ ΓΝΩΣΤΙΚΗ ΕΠΕΞΕΡΓΑΣΙΑ:
5.	learning:1, acquisition:3	ΜΑΘΗΣΗ:, ΑΠΟΚΤΗΣΗ:
6.	<i>imprinting</i> :1	ΑΠΟΤΥΠΩΣΗ:

## V. CONSTRUCTION-METHODOLOGY OF COMPSCIE TERMS TEXT FILE

Jargon file [9] is a glossary that describes all senses of slang terms used in the domain of Computer Science. It was created from the compilation of terminology mainly used by the avid programmers of computer apparatus or networks of the American Universities and covers all the technical computing cultures. In order to extent GrWN in the domain of computer science we do the following: We extract 50 terms with their definitions in English, from Jargon file. The translation of these terms and definitions, in the Greek language, is subserved by the dictionary references [7][8][10][11]. This way we create the CompScieTerms Text File (CompScieTerms.txt). Then we search within PWN, the terms included in CompScieTerms.txt, specifying the domain of computer science (2 results), and the equivalents in GrWN (0 results). Only two terms were found in PWN from the domain of computer science. A) The term *Boot* with definition *Cause to load (an operating system) and start the initial processes*, with hypernym *resuscitate, revive* and no hyponym. The definition from jargon file is *to load and initialise the operating system*, which is similar to the definition found in PWN. B) The term *Down* with definition *Out*, with no hypernym and no hyponym. The definition from jargon file is *to stop functioning*. There is no equivalent in GrWN for either term. The work of GrWN extension in the domain of computer science is still under process.

## VI. CONCLUSIONS

The PsyGrWn that we have developed contains 416 different sets of synonyms from which 175 synsets have definitions, with 67 different relations (hypernym, holo\_part). There is a precise equivalence in the fields of definition and relations, between synsets in the PsyGrWn and PWN. GrWN has 41 equivalences out of 175 terms with definitions of PsyGrWn. PsyGrWn needs enrichment with the synsets that exist in the PWN in the hierarchy levels and enrichment with more synsets from psychology domain and their corresponding relations.

We only found 2 terms, out of 50 of the CompScieTerms.txt, in PWN in the domain of Computer Science. No equivalent was found for these terms in GrWN. Future plans are the translation of Jargon file to Greek, and the creation of a monolingual semantic lexicon in the domain of Computer Science in Greek.

Another future plan for both Greek wordnet extensions is the validation checking of inconsistencies which will be examined with a set of checks that validate wordnet data in the

XML format [2][16][19] and finally the merge with GrWN, so that the latter can be used also from the students of psychology and computer science, and those who deal with the psychology and computer science domain in general.

#### ACKNOWLEDGMENT

The work, described in this paper, was partially supported by a grant from the project “Language Engineering Tools in Learning Environment: Application, Research, Innovation (Let’s LEARN), which is funded by The General Secretariat for Research and Technology (Contract GSRT 092-g).

#### REFERENCES

- [1] BalkaNet project website (2007) <http://www.ceid.upatras.gr/Balkanet/>.
- [2] Budanitsky A., Hirst G. (2006) Evaluating WordNet-based Measures of Lexical Semantic Relatedness. Association for Computational Linguistics.
- [3] Christiane Fellbaum, ed., WordNet – An Electronic Lexical Database, The MIT Press, 1998, pp 4-10, 14, 24, 37-39.
- [4] Deliverable D6.2: The Final monolingual WordNets for each of the Balkan languages, Design and Development of a Multilingual Balkan Wordnet Balkanet, IST-2000-29388
- [5] Dictionary English-Greek of terms of psychology, The laboratory of Cognitive Science, Department of Philosophy and History of Science, National and Kapodistrian University of Athens
- [6] Dictionary, English-Greek “The Pocket Oxford Greek Dictionary” (2000)
- [7] Dictionary, Free on line of computing. FOLDOC (<http://foldoc.org/>).
- [8] Dictionary, Intertran (foreignworld) ( <http://web.tiscali.it/astatalista/en/main.html>).
- [9] Dictionary, The Jargon File, Version 2.9.10, 01 Jul 1992 as published in the page of “Project Guttenberg”
- [10] Dictionary, Ερμηνευτικό Λεξικό Όρων Πληροφορικής (Αγγλοελληνικό - Ελληνοαγγλικό Χρ. Κοΐλιας, Δημ Παναγιωτακος).
- [11] Dictionary, Λεξικό Πληροφορικής Β. Pfaffenberger Γκυορδας 1991
- [12] Dictionary,on-line <http://www.in.gr>
- [13] English WordNet 2.0.c
- [14] Eurowordnet project website, <http://www.illc.uva.nl/EuroWordNet/>.
- [15] Greek Wordnet 0.9.a
- [16] Grigoriadou M., Kormilakis H., Galiotou E., Stamou S., Papakitsos E. (2004) “The Software Infrastructure For The Development And Validation Of The Greek Wordnet”, Romanian Journal Of Information Science And Technology, 2004, Vol. 7, Numbers 1-2, P. 89-105.
- [17] Horak, A., SMRZ, P., (2004) VisDic - Wordnet Browsing and Editing Tool, in Proceedings of the Second International WordNet Conference - GWC 2004, Brno, Czech Republic. Masaryk University.
- [18] Horak, A., SMRZ, P.,(2004), New Features of Wordnet Editor VisDic. Romanian Journal of Information Science and Technology, Volume 7, Numbers 1 2, 2004, 201 213.
- [19] Smrž, P. (2004). Quality Control for Wordnet Development. Proceedings of the 2nd International WordNet Conference GWC 2004, Brno, Czech Republic, January 20-23, 2004
- [20] Tufis D., Cristea D., Stamou S. (2004).Balkanet: Aims, Methods, Results And Perspectives.A General Overview. Romanian Journal Of Information Science And Technology Volume 7, Numbers 1-2, 2004, 9-43.
- [21] Vossen P., EuroWordNet A Multilingual Database with Lexical Semantic Networks. Computers and the Humanities, Volume 32, Nos. 2-3, 1998. Kluwer Academic Publishers.