

# Evaluating the Word Sense Disambiguation Accuracy with Three Different Sense Inventories

Dan Tufiş<sup>1,2</sup>, Radu Ion<sup>1</sup>

<sup>1</sup>Institute for Artificial Intelligence, 13, Calea 13 Septembrie, 050711, Bucharest 5, Romania

<sup>2</sup>Faculty of Informatics, University "A.I. Cuza", 16, Gral. Berthelot, Iaşi, 6600, Romania

**Abstract.** Comparing performances of word sense disambiguation systems is a very difficult evaluation task when different sense inventories are used and, even more difficult when the sense distinctions are not of the same granularity. The paper substantiates this statement by briefly presenting a system for word sense disambiguation (WSD) based on parallel corpora. The method relies on word alignment, word clustering and is supported by a lexical ontology made of aligned wordnets for the languages in the corpora. The wordnets are aligned to the Princeton Wordnet, according to the principles established by EuroWordNet. The evaluation of the WSD system was performed on the same data, using three different granularity sense inventories.

## 1 Introduction

Most difficult problems in natural language processing stem from the inherent ambiguous nature of the human languages. Ambiguity is present at all levels of traditional structuring of a language system (phonology, morphology, lexicon, syntax, semantics) and not dealing with it at the proper level, exponentially increases the complexity of the problem solving. For instance, it is instructive to observe that many text books, published in the 70's and 80's and even more recently, when dealing with parsing ambiguity, exemplify the phenomenon mainly by drawing on the morpho-lexical ambiguities and less on genuine structural syntactic ambiguity (such as PP-attachment). The early approach where the morpho-lexical ambiguities, as detected by a morpho-lexical analyzer, were left to the syntactic parser became obsolete in the '90, when the part-of-speech (PoS) tagging technology matured and turned into a standard preprocessing phase. Currently, the state of the art taggers (combining various models, strategies and processing tiers) ensure no less than 97-98% accuracy in the process of morpho-lexical full disambiguation. For such taggers a 2-best tagging<sup>1</sup> is practically 100% accurate.

One step further to the morpho-lexical disambiguation is the word sense disambiguation (WSD) process. WSD tagging is the great challenge and it is very likely that when this technology will reach the same performance as PoS tagging, the

---

<sup>1</sup> In k-best tagging, instead of assigning each word exactly one tag (the most probable in the given context), it is allowed to have occasionally at most k-best tags attached to a word and if the correct tag is among the k-best tags, the annotation is considered to be correct.

market will see a boom of robust and trustful applications empowered by natural language interaction.

Informally, the WSD problem can be stated as being able to associate to an ambiguous word ( $w$ ) in a text or discourse, the sense ( $s_k$ ) which is distinguishable from other senses ( $s_1, \dots, s_{k-1}, s_{k+1}, \dots, s_n$ ) prescribed for that word by a reference semantic lexicon. One such semantic lexicon (actually a lexical ontology) is Princeton WordNet [1] version 2.0<sup>2</sup> (henceforth PWN). PWN is a very fine grained semantic lexicon currently containing 203,147 sense distinctions, clustered in 115,424 equivalence classes (synsets). Out of the 145,627 distinct words, 119,528 have only one single sense. However, the remaining 26,099 words are those that one would frequently meet in a regular text and their ambiguity ranges from 2 senses up to 36. Several authors considered that sense granularity in PWN is too fine-grained for the computer use, arguing that even for a human (native speaker of English) the sense differences of some words are very hard to be reliably (and systematically) distinguished. There are several attempts to group the senses of the words in PWN in coarser grained senses – *hyper-senses* – so that clear-cut distinction among them is always possible for humans and (especially) computers. There are two hyper-sense inventories we will refer to in this paper since they were used in the BalkaNet project [2]. A comprehensive review of the WSD state-of the art at the end of 90's can be found in [3]. Stevenson and Wilks [4] review several WSD systems that combined various knowledge sources to improve the disambiguation accuracy and also address the issue of different granularities of the sense inventories. SENSEVAL series of evaluation competitions on WSD (three editions since 1998, with the fourth one approaching, see <http://www.cs.unt.edu/~rada/senseval>) is a very good information source on learning how WSD evolved in the last 6-7 years and where is it nowadays.

We describe a multilingual environment, containing several monolingual wordnets, aligned to PWN used as an interlingual index (ILI). The word-sense disambiguation multilingual method combines word alignment technologies, translation equivalents clustering and synset interlingual equivalence [9, 10]. Irrespective of the languages in the multilingual documents, the words of interest are disambiguated by using the same sense-inventory labels. The aligned wordnets were constructed in the context of the European project BalkaNet. The consortium developed monolingual wordnets for five Balkan languages (Bulgarian, Greek, Romanian, Serbian, and Turkish) and extended the Czech wordnet initially developed in the EuroWordNet project [5]. The wordnets are aligned to PWN, taken as an interlingual index, following the principles established by the EuroWordNet consortium. The version of the PWN used as ILI is an enhanced XML version where each synset is mapped onto one or more SUMO [6] conceptual categories and also is classified under one of the IRST domains [7]. In the present version of the BalkaNet ILI there are used 2066 SUMO distinct categories and 163 domain labels. Therefore, for our WSD experiments we had at our disposal three sense inventories, with very different granularities: PWN senses, SUMO categories and IRST Domains.

---

<sup>2</sup> <http://www.cogsci.princeton.edu/~wn/>

## 2 Multilingual corpus and the outline of the WSD procedure

The basic text we use for the WSD task is Orwell's novel "Ninety Eighty Four" [8] and its translations in several languages (Bulgarian, Czech, Estonian, Greek, Hungarian, Romanian, Serbian, Slovene, and Turkish). All the translations were sentence-aligned to English, POS tagged and lemmatized. The alignments are to a large extent (more than 93%) 1-1, that is, more often than not, one sentence in English is translated as one sentence in the other languages.

Although we showed elsewhere [9] that using more languages and (especially) more wordnets the accuracy of the WSD task is increasing<sup>3</sup>, one might object that this is a hard condition to meet for most of the languages. If tools needed for the preprocessing phases [10, 11] (tokenisation, POS-tagging, lemmatization) are easy to find for almost any natural language, reliable aligned wordnets are much rarer. Therefore, for the WSD experiments described and evaluated in this paper we used only the integral English-Romanian bitext and only the alignment between PWN and the Romanian wordnet. However, we should make it clear that, since our word alignment techniques do not use the wordnets, it is always possible to transfer the sense labels decided for a pair of language (here Romanian and English) to all the other languages in the parallel corpus for which a wordnet is not available. Yet, there is no guarantee that the annotation in other languages of the parallel corpus, as resulted from the above mentioned transfer, is as accurate as in the two source languages that generated it. The evaluation has to be done by native speakers of the annotation importing languages.

The word sense disambiguation method described here has four steps:

- a) word alignment of the parallel corpus and translation pairs extraction; this step results in different numbers of correct translation pairs (CT), wrong translation pairs (WT) and word occurrences without identified translations (NT); many NTs are either hapax legomena words or word occurrences that were not translated in the other language;
- b) wordnet-based sense disambiguation of the translation pairs found (CT+WT) in step a); this step results in sense-assigned words (SAWO) for CT and (very likely) sense-unassigned word occurrences (SUWO) for WT.
- c) word sense clustering for NT and SUWO; this phase takes advantage of the sense assignment in step b).
- d) generating the XML disambiguated parallel corpus with every content word (in both languages) annotated with a single sense label. Sense label inventory can be one of the three available in the BalkaNet lexical ontology: PWN unique synset identifiers, SUMO conceptual categories and IRST-Domains.

---

<sup>3</sup> During the final evaluation of the BalkaNet wordnets, we run a smaller scale experiment where we used 4 wordnets with an average F-measure improvement of almost 3% over the results reported in [9]. Although the better results could be partly justified by a smaller number of words and occurrences, the main improvement came from two extra wordnets we used in the experiment.

### 3 Word Alignment and Translation Lexicon Extraction

Word alignment is a hard NLP problem which can be simply stated as follows: given  $\langle T_{L1} T_{L2} \rangle$  a pair of reciprocal translation texts, in languages L1 and L2, the word  $W_{L1}$  occurring in  $T_{L1}$  is said to be aligned to the word  $W_{L2}$  occurring in  $T_{L2}$  if the two words, in their contexts, represent reciprocal translations. Our approach is a statistical one and comes under the “*hypothesis-testing*” model (e.g., [11], [12]). In order to reduce the search space and to filter out significant information noise, the context is usually reduced to the level of sentence. Therefore, a parallel text  $\langle T_{L1} T_{L2} \rangle$  can be represented as a sequence of pairs of one or more sentences in language L1 ( $S_{L1}^1 S_{L1}^2 \dots S_{L1}^k$ ) and one or more sentences in language L2 ( $S_{L2}^1 S_{L2}^2 \dots S_{L2}^m$ ) so that the two ordered sets of sentences represent reciprocal translations. Such a pair is called a translation alignment unit (or translation unit). The word alignment of a bitext is an explicit representation of the pairs of words  $\langle W_{L1} W_{L2} \rangle$  (called translation equivalence pairs) co-occurring in the same translation units and representing mutual translations. The general word alignment problem includes the cases where words in one part of the bitext are not translated in the other part (these are called *null alignments*) and also the cases where multiple words in one part of the bitext are translated as one or more words in the other part (these are called expression alignments). The word alignment problem specification does not impose any restriction on the part of speech (POS) of the words making a translation equivalence pair, since cross-POS translations are rather frequent. However, for the aligned wordnet-based word sense disambiguation we discarded translation pairs that did not preserve the POS (and obviously null alignments). Removing duplicate pairs  $\langle W_{L1} W_{L2} \rangle$  one gets a translation lexicon for the given corpus.

Although far from being perfect, the accuracy of word alignments and of the translation lexicons extracted from parallel corpora is rapidly improving. In the shared task evaluation of different word aligners, organized on the occasion of the NAACL2003, for the Romanian-English track, our winning system [13] produced relevant translation lexicons with an aggregated F-measure as high as 84.26%. Meanwhile, the word-aligner was further improved so that the current performances (on the same data) are about 3% better on all scores in word alignment and about 5% better in wordnet-relevant dictionaries (containing only translation equivalents of the same POS), thus approaching 90%.

### 4 Wordnet-based Sense Disambiguation

Once the translation equivalents were extracted, for any translation pair  $\langle W_{L1} W_{L2} \rangle$  and two aligned wordnets, the algorithm performs the following operations:

1. extract the interlingual (ILI) codes for the synsets that contain  $W_{L1}^i$  and  $W_{L2}^j$  respectively, to yield two lists of ILI codes,  $L_{ILI}^1(W_{L1}^i)$  and  $L_{ILI}^2(W_{L2}^j)$
2. identify one ILI code common to the intersection  $L_{ILI}^1(W_{L1}^i) \cap L_{ILI}^2(W_{L2}^j)$  or a pair of ILI codes  $ILI_1 \in L_{ILI}^1(W_{L1}^i)$  and  $ILI_2 \in L_{ILI}^2(W_{L2}^j)$ , so that  $ILI_1$  and  $ILI_2$  are the *most similar* ILI codes (defined below) among the candidate pairs  $(L_{ILI}^1(W_{L1}^i) \otimes L_{ILI}^2(W_{L2}^j))$  [ $\otimes$  = Cartesian product]

The rationale for these operations derives from the common intuition which says that if the lexical item  $W_{L1}^i$  in the first language is found to be translated in the second language by  $W_{L2}^j$ , then it is reasonable to expect that at least one synset which the lemma of  $W_{L1}^i$  belongs to, and at least one synset which the lemma of  $W_{L2}^j$  belongs to, would be aligned to the same interlingual record or to two interlingual records semantically closely related.

Ideally step 2 above should identify one ILI concept lexicalized by  $W_{L1}$  in language  $L1$  and by  $W_{L2}$  in language  $L2$ . However, due to various reasons, the wordnets alignment might reveal not the same ILI concept, but two concepts which are semantically close enough to license the translation equivalence of  $W_{L1}$  and  $W_{L2}$ . This can be easily generalized to more than two languages. Our measure of interlingual concepts semantic similarity is based on PWN structure. We compute semantic-similarity<sup>4</sup> score by formula:  $SYM(ILI_1, ILI_2) = \frac{1}{1+k}$  where  $k$  is the number

of links from  $ILI_1$  to  $ILI_2$  or from both  $ILI_1$  and  $ILI_2$  to the nearest common ancestor. The semantic similarity score is 1 when the two concepts are identical, 0.33 for two sister concepts, and 0.5 for mother/daughter, whole/part, or concepts related by a single link. Based on empirical studies, we decided to set the significance threshold of the semantic similarity score to 0.33. In case of ties, the pair corresponding to the most frequent sense of the target word in the current bitext pair is selected.

## 5 Word Sense Clustering Based on the Translation Lexicons

To perform the clustering, we derive for each target word  $i$  occurring  $m$  times in the corpus a set of  $m$  binary vectors  $VECT(TW^i)$ . The number of cells in  $VECT(TW^i)$  is equal to the number of distinct translations of the word  $i$  in the other language (called source language) of the parallel corpus. The  $k^{\text{th}}$   $VECT(TW^i)$  specifies which of the possible translations of  $TW^i$  was actually used in the source language as an equivalent for the  $k^{\text{th}}$  occurrence of  $TW^i$ . All positions in the  $k^{\text{th}}$   $VECT(TW^i)$  are set to 0 except at most one bit identifying the word used (if any) as translation equivalent for the target word  $i$ . The  $m$  vectors  $VECT(TW^i)$  are processed by an agglomerative algorithm based on Stolcke's Cluster2.9 [16] which produces clusters of similar vectors. Such a cluster would identify the occurrences of the target word that are likely to have been used with the same meaning. The fundamental assumption of this algorithm is that if two or more instances of the same target word are identically translated in the source language it is plausible that their meaning is the same. The likelihood is increased as the number of source languages is larger and their types are more diversified [17]. This is because the chance of preserving common sense ambiguities in the translation equivalents significantly decreases when more diversified languages are considered.

One big problem for the clustering algorithms in general and for agglomerative ones in particular is that the number of classes should be known in advance in order to obtain meaningful results. With respect to the word sense clustering this would mean

<sup>4</sup> Other approaches to similarity measures are described in [15].

knowing in advance for every word in a text how many of its possible senses are actually used in the given corpus. We use the results of the previous phase (word sense disambiguation based on the aligned wordnets) which generally successfully covers more than 75% of the text. For the words the occurrences of which were disambiguated by this phase we consider that any other sense-unassigned occurrence was used with one of the previously seen senses, and thus we provide the algorithm with the number of classes. For all the words for which none of its occurrences was previously disambiguated (the majority of these words occurred only once) we automatically assign the first sense number in PWN. The rationale for this heuristics is that PWN senses are numbered according to their frequencies (first sense is the most frequent). This back-off mechanism is justified when the texts do not belong to specialized registers. PWN is a general semantic lexicon and the statistics on senses were drawn from a balanced corpus. For a specialized text, a more successful heuristics would be to take advantage of a prior classification of the text according to the IRST-domains [7] and then to consider the most frequent sense with the same domain label as the one of the text.

## 6 Generation the WSD-annotation in the parallel corpus

The structure of the automatically generated WSD-annotated corpus (Figure 1) is a simplified version of the XCES-ANA<sup>5</sup> format [18] with the additional attributes *sn* (sense number), *oc* (ontological category) and *dom* (domain) for the <w> tag.

**Figure1.** A sample of the WSD corpus encoding

```
<body> . . .
  <tu id="Ozz20">
    <seg lang="en">
      <s id="Oen.1.1.4.9">
        <w lemma="the" ana="Dd">
The</w>
          <w lemma="patrol" ana="Ncnp" sn="1" oc="SecurityUnit"
            dom="military">
patrols</w>
            <w lemma="do" ana="Vais">
did</w>
              <w lemma="not" ana="Rmp" sn="1" oc="not" dom="factotum">
not</w>
                <w lemma="matter" ana="Vmn" sn="1"
                  oc="SubjectiveAssesmentAttribute" dom="factotum">
matter</w><c>
, </c>
          <w lemma="however" ana="Rmp" sn="1"
            oc="SubjectiveAssesmentAttribute|PastFn" dom="factotum">
however</w><c>
.</c></s></seg>
      <seg lang="ro">
```

<sup>5</sup> <http://www.cs.vassar.edu/XCES/>

```

    <s id="Oro.1.2.5.9">
      <w lemma="și" ana=Crssp>
și</w>
      <w lemma="totuși" ana="Rgp" sn="1"
        oc="SubjectiveAssesmentAttribute|PastFn" dom="factotum">
totuși</w><c>
      ,</c>
      <w lemma="patrulă" ana="Ncfpry" sn="1.1.x"
        oc="SecurityUnit" dom="military">
patrulele</w>
      <w lemma="nu" ana="Qz" sn="1.x" oc="not" dom="factotum">
nu</w>
      <w lemma="conta" ana="Vmii3p" sn="2.x"
        oc="SubjectiveAssesmentAttribute" dom="factotum">
contau</w><c>
    .</c></s></seg>
  </tu> . . .
</body>

```

The values of these attributes (defined for words belonging to the content words) have the following meanings:

- *sn* specifies the sense label for the current word as described in the wordnet of the respective language.
- *oc* represents the SUMO ontological concept(s) on which the wordnet sense of the current word is mapped on.

*dom* identifies the IRST domain under which the wordnet sense of the current word is clustered.

The use of all the three additional attributes is the default, but the user may specify one or two attributes to be generated in the WSD annotated parallel corpus.

## 7 Evaluation

The BalkaNet version of the “1984” corpus is encoded as a sequence of uniquely identified *translation units* (TU, see Figure 1). For the evaluation purposes, we selected a set of fairly frequent English words (123 nouns and 88 verbs) the meanings of which were also encoded in the Romanian wordnet. The selection considered only polysemous words (at least two senses per part of speech) since the POS-ambiguous words are irrelevant as this distinction is solved with high accuracy (more than 99%) by our tiered-tagger [14]. All the occurrences of the target words were disambiguated by three independent experts who negotiated the disagreements and thus created a gold-standard annotation for the evaluation of precision and recall of the WSD algorithm.

The same targeted words were automatically disambiguated both with and without the back-off clustering algorithm. For the basic wordnet-based WSD we used the PWN and the Romanian wordnet. For the back-off clustering we extracted an English-Romanian translation equivalence dictionary based on which we computed the initial clustering vectors for all occurrences of the target words.

Out of the 211 target words, with 1998 occurrences the system could not make a decision for 27 words (12.79%) with 51 occurrences (2.55%). Half of these words (14) occurred only once and neither the wordnet-based step nor the clustering back-off could do anything. Other 13 occurrences, were not translated by the same part of speech, were wrongly translated by the human translator or not translated at all. Applying the simple heuristics (SH) that says that any unlabelled target occurrence receives its most frequent sense, 36 out of 51 of them got a correct sense-tag. Therefore, since all the target occurrence words received a sense annotation, the recall and precision have the same values. The table below summarizes the results.

**Table 1.** WSD precision, recall and F-measure for the algorithm based on aligned wordnets (AWN), for AWN with clustering (AWN+C) and for AWN+C and the simple heuristics (AWN+C+SH).

WSD System	Precision	Recall	F-measure
AWN	83.27%	60.06%	69.78%
AWN+C	76.27%	74.32%	75.28%
AWN+C+SH	76.12%	76.12%	76.12%

The results in Table 1 show the best Precision for the simplest algorithm (AWN) as this one deals only with the occurrences for which correct translation pairs (CT) have been found. The occurrences for which wrong translation pairs (WT) have been found and those for which no translation has been identified are ignored. This explains the low Recall score. The back-off mechanisms (clustering and the simple heuristics) bring a major improvement of the Recall scores and thus, the value of the overall F-measure score is improved. The major sources of further improvements of the WSD performance are the following: a better accuracy of the word aligner, a larger Romanian wordnet and a cleaner interlingual alignment of the synsets.

We are working on a combined word aligner incorporating our previous algorithm [14] and a new one based on GIZA++[19]<sup>6</sup>. Since the two aligners have similar accuracy, but the errors done by any of them is not a proper subset of the errors done by the other, we are confident that the combination of the two will provide a better alignment than any of them alone. As far as the improvement of the Romanian wordnet, this is a continuous project and since the BalkaNet project ended the number of synsets is 22% larger and several interlingual mapping errors have been corrected.

Once the values for the *sn* attributes established, the values for the *oc* and *dom* attributes are deterministically appended to the <w> tag annotation.<sup>7</sup> The Table 2 shows a great variation in terms of Precision, Recall and F-measure when different granularity sense inventories are considered for the WSD problem. Therefore, it is important to make the right choice on the sense inventory to be used with respect to a given application. In case of a document classification problem, it is very likely that the IRST domain labels (or a similar granularity sense inventory) would suffice. The

<sup>6</sup> <http://www-i6.informatik.rwth-aachen.de/Colleagues/och/software/GIZA++.html>

<sup>7</sup> One should note that the actual WSD task is done in terms of PWN senses. The values for *oc* and *dom* attributes, corresponding to each PWN synset, are statically compiled of-line.

rationale is that IRST domains are directly derived from the Universal Decimal Classification as used by most libraries and librarians. The SUMO sense labeling will be definitely more useful in an ontology based intelligent system interacting through a natural language interface. Finally, the most refined sense inventory of PWN will be extremely useful in Natural Language Understanding Systems, which would require a deep processing. Also, such a fine inventory would be highly beneficial in lexicographic and lexicological studies.

**Table 2.** Evaluation of the WSD (AWN+C+SH) in terms of three different sense inventories.

<b>Sense Inventory</b>	<b>Precision</b>	<b>Recall</b>	<b>F-measure</b>
PWN 115424 categories	76.12%	76.12%	76.12%
SUMO 2066 categories	82.64%	82.64%	82.64%
DOMAINS 163 categories	91.90%	91.90%	91.90%

Similar findings on sense granularity for the WSD task are discussed in [4] where for some coarser grained inventories even higher precisions are reported. However, we are not aware of better results in WSD exercises where the PWN sense inventory was used. The major explanation for this is that unlike the majority work in WSD that is based on monolingual environments, we use for the definition of sense contexts the cross-lingual translations of the occurrences of the target words. The way one word in context is translated into one or more other languages is a very accurate and highly discriminative knowledge source for the decision making.

## 8 Conclusions

The results in Table 2 show that although we used the same WSD algorithm on the same text, the performance scores (precision, recall, f-measure) significantly varied, with more than 15% difference between the best (DOMAINS) and the worst (PWN) f-measures. This is not surprising, but it shows that it is extremely difficult to objectively compare and rate WSD systems working with different sense inventories.

The potential drawback of this approach is that it relies on the existence of parallel data and at least two aligned wordnets which might not be available yet. Nevertheless, parallel resources are becoming increasingly available, in particular on the World Wide Web, and aligned wordnets are being produced for more and more languages. In the near future it should be possible to apply our and similar methods to large amounts of parallel data and a wide spectrum of languages.

## References

1. Fellbaum, Ch. (ed.) WordNet: An Electronic Lexical Database, MIT Press (1998).

2. Tufiş, D., Cristea, D., Stamou, S., BalkaNet: Aims, Methods, Results and Perspectives. A General Overview. In: D. Tufiş (ed): Special Issue on BalkaNet. Romanian Journal on Science and Technology of Information, Vol. 7 no. 3-4 (2004) 9-44.
3. Ide, N., Veronis, J., Introduction to the special issue on word sense disambiguation. The state of the art. Computational Linguistics, Vol. 27, no. 3, (2001) 1-40.
4. Stevenson, M., Wilks, Y., The interaction of Knowledge Sources in Word Sense Disambiguation. Computational Linguistics, Vo. 24, no. 1, (1998) 321-350.
5. Peters, W., Vossen, P., Diez-Orzas, P., Adriaens, G., Cross-Linguistic Alignment of wordnets with an Inter-Lingual-Index. In P. Vossen (Ed.): Special Issue on EuroWordNet. Computers and the Humanities, Vol. 32, no.2-3 (1998) 221-251.
6. Niles, I., and Pease, A., Towards a Standard Upper Ontology. In Proceedings of the 2<sup>nd</sup> International Conference on Formal Ontology in Information Systems (FOIS-2001), Ogunquit, Maine, (2001) 17-19.
7. Magnini B. Cavaglià G., Integrating Subject Field Codes into WordNet. In Proceedings of LREC2000, Athens, Greece (2000) 1413-1418.
8. Erjavec, T, MULTEXT-East Version 3: Multilingual Morphosyntactic Specifications, Lexicons and Corpora, in Proceedings of LREC2004, Lisbon (2004) 1535-1538.
9. 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 20<sup>th</sup> International Conference on Computational Linguistics, COLING2004, Geneva, (2004) 1312-1318.
10. Gale, W.A. and Church, K.W. (1991). Identifying word correspondences in parallel texts. Fourth DARPA Workshop on Speech and Natural Language. Asilomar, CA, pp. 152-157.
11. Smadja, F., McKeown, K.R., and Hatzivassiloglou, V. (1996). Translating collocations for bilingual lexicons: A statistical approach. *Computational linguistics*, 22(1):1-38.
12. Tufiş, D., A cheap and fast way to build useful translation lexicons. In Proceedings of the 19th International Conference on Computational Linguistics, COLING 2002, Taipei (2002) 1030-1036.
13. Tufiş, D., Barbu, A., M., Ion, R. A word-alignment system with limited language resources. In Proceedings of the NAACL 2003 Workshop on Building and Using Parallel Texts; Romanian-English Shared Task, Edmonton (2003) 36-39.
14. Tufiş, D., Tiered Tagging and Combined Classifiers, in F. Jelinek, E. Nöth (eds) Text, Speech and Dialogue, Lecture Notes in Artificial Intelligence, Vol. 1692. Springer-Verlag, Berlin Heidelberg New-York (1999) 28-33.
15. A. Budanitsky and G. Hirst, Semantic distance in WordNet: An experimental, application-oriented evaluation of five measures. Proceedings of the Workshop on WordNet and Other Lexical Resources, Second meeting of the NAACL, Pittsburgh, June, (2001) 29-34.
16. Stolcke, A. Cluster 2.9. <http://www.icsi.berkeley.edu/ftp/global/pub/ai/stolcke/software/cluster-2.9.tar.Z> (1996).
17. Ide, N., Erjavec, T., Tufiş, D.: „Sense Discrimination with Parallel Corpora” in Proceedings of the SIGLEX Workshop on Word Sense Disambiguation: Recent Successes and Future Directions. ACL2002, July Philadelphia 2002, pp. 56-60
18. Ide, N., Bonhomme, P., Romary, L., XCES: An XML-based Standard for Linguistic Corpora. In Proceedings of LREC2000, Athens, Greece (2000) 825-30.
19. Och, F., J., Ney, H., Improved Statistical Alignment Models, *Proceedings of ACL2000*, Hong Kong, China, 440-447, 2000.