

Translation-oriented Word Sense Induction Based on Parallel Corpora

Marianna Apidianaki

LaTTiCe, University Paris 7, CNRS
ENS-1 rue Maurice Arnoux, F-92120, Montrouge
marianna@linguist.jussieu.fr

Abstract

Word Sense Disambiguation (WSD) is an intermediate task that serves as a means to an end defined by the application in which it is to be used. However, different applications have varying disambiguation needs which should have an impact on the choice of the method and of the sense inventory used. The tendency towards application-oriented WSD becomes more and more evident, mostly because of the inadequacy of predefined sense inventories and the inefficacy of application-independent methods in accomplishing specific tasks. In this article, we present a data-driven method of sense induction, which combines contextual and translation information coming from a bilingual parallel training corpus. It consists of an unsupervised method that clusters semantically similar translation equivalents of source language (SL) polysemous words. The created clusters are projected on the SL words revealing their sense distinctions. Clustered equivalents describing a sense of a polysemous word can be considered as more or less commutable translations for an instance of the word carrying this sense. The resulting sense clusters can thus be used for WSD and sense annotation, as well as for lexical selection in translation applications.

1. Introduction

The granularity of sense distinctions varies considerably among resources and a unique response concerning their number is difficult to be found (Kilgarriff, 1997). Both linguistic and extra-linguistic factors have a bearing on the definition of senses: linguistic factors are related to different theoretical semantic hypotheses that may be adopted during the construction of a resource, while extra-linguistic ones concern its envisaged uses. In a NLP context, sense inventories are needed for WSD and semantic annotation. These tasks being “intermediate” (Wilks & Stevenson, 1996), they are essential for achieving final goals, highly dependent on the envisaged application.

The efficient use of predefined semantic resources for WSD in particular applications is often hampered by the high granularity, the great number and the striking similarity of the senses described therein (Ide *et al.*, 2001; Edmonds & Kilgarriff, 2002; Ng *et al.*, 2003). Besides the complexity of processing in the case of very fine sense distinctions, there is also a risk of information loss, when a forced choice among closely related senses has to be made while relations between senses are not taken into account (Dolan, 1994). The high granularity of senses described in monolingual resources poses problems for establishing sense correspondences in a bilingual context as well (Miháľtz, 2005; Specia *et al.*, 2006).

Even the need of such distinctions in precise applications is often being doubted, prompting the development of methods that attempt to reduce the granularity found in predefined resources, by clustering senses in order to propose coarser sense distinctions (Dolan, *ibid.*, Peters *et al.*, 1998; Mihalcea & Moldovan, 2001; Navigli, 2006). These observations have also fostered the development of application-oriented WSD methods, taking into consideration the particular needs of final applications.

Moreover, supervised WSD techniques are subject to a serious limitation, the well-known ‘knowledge acquisition bottleneck’ (Resnik, 2004). Although these techniques perform best in public evaluations (Agirre & Soroa, 2007), existing hand-tagged corpora allow for a

small improvement over the simple most frequent sense heuristic (Snyder & Palmer, 2004). Inventories needed for supervised WSD may change from one domain to the other, as well as the distribution of senses, and additional hand-tagging of corpora is required. Unsupervised word sense induction and discrimination methods induce word senses directly from corpora, often using clustering techniques which group together similar instances of words. In this case, WSD can be done comparing a new instance of a polysemous word with the induced clusters (representing senses) and selecting one of them as its sense.

The method proposed in this article combines contextual and translation information coming from both language sides of a parallel corpus in order to identify the senses of SL polysemous words. The induced senses can be used for establishing sense correspondences between these words and their translation equivalents (EQVs) in the corpus. The proposed sense distinctions and correspondences are adequate for semantic processing in translation applications. More precisely, they can be used for disambiguation of new occurrences of polysemous words and for selection of semantically correct translation equivalents during lexical selection in Machine Translation (MT).

2. Theoretical Assumptions

The theoretical assumptions underlying our method are the following:

- (a) the contextual (distributional) hypothesis of meaning (Harris, 1954; Firth, 1957), according to which the meaning of words corresponds to their use in texts
- (b) the contextual hypothesis of semantic similarity (Miller & Charles, 1991), according to which context similarity of words reflects their semantic similarity
- (c) the assumption of a semantic correspondence between SL words and their EQVs in real texts.

These assumptions permit the emission of another one, which justifies the combination of contextual and translation information extracted from a parallel corpus:

(d) information coming from the contexts of a SL word when translated with a precise EQV, may shed light on the senses carried by the EQV; furthermore, the similarity of the SL word's contexts reveals the semantic similarity of its EQVs.

According to assumption (a), the analysis of the lexical context surrounding a word in texts can reveal its meaning. A high degree of context similarity shows the word's semantic homogeneity, while context dissimilarity indicates the existence of sense distinctions. Lexical context constitutes thus a valuable source of semantic information, exploited in various sense induction (Schütze, 1998; Pantel & Lin, 2002; Véronis, 2004; Purandare & Pedersen, 2004) and WSD methods (Lesk, 1986; Brown *et al.*, 1991; Kaji & Morimoto, 2002). According to assumption (c), in the case of a word correspondence in a parallel corpus, the senses carried by a SL word and its EQV are considered to be similar. Hence, different EQVs are translating the different senses of a polysemous SL word in the target language (TL), senses also reflected in the SL contexts.

Before sense identification, translation correspondences extracted from a parallel corpus are situated at the word level and polysemous words are associated with numerous EQVs. Our objective is the refinement of these relations and the establishment of correspondences at a higher level of analysis. The originality of our sense induction approach consists in the projection of cooccurrence information from one side of the bitext to the other using as a "bridge" the translation relations extracted from texts, without recourse to predefined lexical resources. The proposed method is totally data-driven and its core component is an unsupervised clustering algorithm which does not necessitate annotated data.

3. Description of the Method

3.1. Context in a Bilingual Framework

In monolingual contextual methods of sense induction, information used for clustering comes from the context of the occurrences of a polysemous word and the resulting clusters illustrate its different senses. The context used for clustering may be perceived differently when more languages are involved. For instance, in the work of Ide *et al.* (2001) and Tufiş *et al.* (2004), occurrences of polysemous words are described by context vectors representing their translations in six different languages found in parallel corpora. The senses of these words are identified by clustering the corresponding context vectors. A similar conception of context is found in the work of van der Plas & Tiedemann (2006), where the alignment contexts of a word constitute the features used for creating the corresponding vector.

In this work, information from the context surrounding a word in texts (following the "traditional" conception of context) is combined with translation information found in the results of a word alignment procedure. This set of information forms the input of the sense induction method.

3.2. Semantic Clustering in a Bilingual Framework

3.2.1. Training corpus

The training corpus used in this work is an English-Greek bitext of approximately 4 000 000 words aligned at the sentence level, lemmatized and part-of-speech (POS) tagged (Gavriliidou *et al.*, 2004). The sentence alignment results consist of "translation units" composed of a SL and a TL segment, each of which contains up to 2 sentences being in a translation relation¹.

3.2.2. Bilingual lexicon building

The training corpus has been word aligned, at the levels of tokens and types (Simard & Langlais, 2003). Here we use the results of the alignment of word types, their quality being clearly superior to that of the alignment of tokens; this difference confirms the beneficial impact of lemmatization on this kind of processing in the case of a morphologically rich language like Greek (Nießen & Ney, 2004). Two bilingual lexicons were built from these results, one for each translation direction (English-Greek/Greek-English); in these lexicons, words of each language are associated to their translation EQVs in the corpus.

As we are interested in correspondences between words of the two languages belonging to the same grammatical category, the lexicons have been filtered by POS-tag (so that SL nouns be aligned to TL nouns, verbs to verbs, etc.). This processing filtered out much of the noise present in the lexicon. An intersection filter has also been used in order to eliminate the remaining noise, keeping only word associations found in the lexicons of both translation directions.

The sense induction method was developed using the results of a manual alignment procedure (Apidianaki, 2007) and then applied to the automatically generated translation lexicons. Here, we present the results of the method for a sample of the words in the English-Greek lexicon. The lexicon entries used are given in Table 1; numbers in parenthesis show the frequency of use of each EQV as translation of the polysemous SL word in the training corpus.

SL word	EQVs
structure	δομή(272), διάρθρωση(32), κατασκευή(27)
guidance	προσανατολισμός(107), καθοδήγηση(34), συμβουλή(7)
survey	έρευνα(146), δημοσκόπηση(7), επισκόπηση(7)
power	αρμοδιότητα(117), εξουσία(113), δύναμη(71), ισχύς(50)
trade	εμπόριο(184), συναλλαγή (53), επάγγελμα(11)

Table 1. Sample of the English-Greek lexicon

¹ SL segments containing 0 sentences correspond to "additions" in translation while empty TL segments correspond to "omissions". A correspondence between 2 sentences of each language permits capturing crossing correspondences.

3.2.3. Sub-corpora building

A sub-corpus is created from the training corpus for each SL word (w), consisting of the translation units where it occurs in the SL segment. This sub-corpus is subsequently filtered on the basis of each EQV of the word, present in the TL segments. In this way, several “translation units sets”, described as ‘ w_EQV ’, are created, containing those units where w is translated by each one of the EQVs. For instance, filtering the sub-corpus of the word *structure*, we obtain three translation units sets, corresponding to its EQVs in the corpus: the first set can be described as ‘*structure_δομή*’; the second as ‘*structure_διάθροση*’ and the third as ‘*structure_κατασκευή*’.

3.2.4. Source language contexts of the EQVs

A “SL context” is created for each EQV of w from the corresponding translation units set (w_EQV). This context is composed by the lemmas of the content words (nouns, adjectives and verbs) surrounding w in the SL segments of ‘ w_EQV ’ and occurring more than once, as described in Figure 1. For instance, the SL context of the EQV *δομή* is composed by the content words found in the English context of *structure* whenever it is translated by this particular EQV in the corpus.

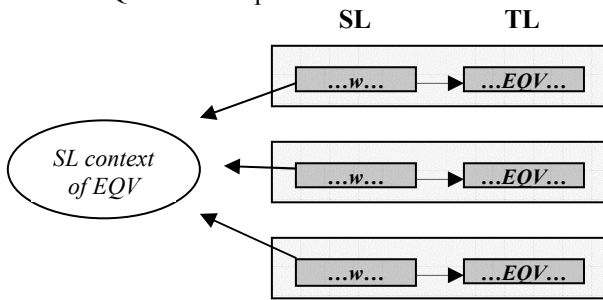


Figure 1. SL context of EQV in the ‘ w_EQV ’ translation units set

A frequency list of the retained context features is then generated. The frequency lists created for each of the EQVs of w form the input of a semantic similarity calculation method.

3.2.5. Context similarity calculation

Following our initial assumption (d), which concerns the possibility of using SL context information for the semantic analysis of the EQVs, the similarity of SL contexts corresponding to different EQVs indicates the degree of their semantic similarity.

The semantic calculation performed does not operate on the individual contexts of the occurrences of a SL word, but on the sets of “SL contexts” corresponding to its EQVs, obtained in the way described in the previous paragraph. Similarity estimations do not concern thus particular SL word occurrences but pairs of translation EQVs and are done using SL context features. Using these “extended” contexts as input of the similarity calculation method significantly reduces the impact of data sparseness on the results.

3.2.6. The similarity measure

The measure used for calculating similarity is a variation of the weighted Jaccard coefficient (Grefenstette,

1994). This weighted measure permits the definition of the relevance of each context feature for the estimation of the EQVs’ similarity. The input of the similarity calculation for two EQVs consists of their frequency lists as well as of those generated for the other EQVs of the SL word. The score attributed to a pair of EQVs indicates their degree of similarity.

Three weights are calculated for each context feature (j) of an EQV (i): first, a *global weight* (gw) is attributed to each j on the basis of its dispersion in the sub-corpus of the SL word and of its frequency of cooccurrence with the word when translated with each of the EQVs (i).

$$gw(\text{feature}_j) = 1 - \frac{\sum_{i=1}^{nbr_i} p_{ij} \log(p_{ij})}{nrels}$$

The gw of a feature depends on the number of EQVs with which it is related (in the SL word sub-corpus) and on its probability of occurrence with each one of the EQVs.

$$p_{ij} = \frac{\text{absolute frequency of feature}_j \text{ with EQV}_i}{\text{total number of features for EQV}_i}$$

$$nrels = \text{total number of relations extracted for } j$$

Then the *local weight* (lw) of a feature with a particular EQV is calculated, on the basis of its frequency of cooccurrence with the EQV in question.

$$lw(EQV_i, \text{feature}_j) = \log(\text{frequency of feature}_j \text{ with EQV}_i)$$

Finally, a feature’s *total weight* (w) relevant to one EQV corresponds to the product of its global weight and its local weight with this particular EQV.

$$w = gw * lw$$

The Weighted Jaccard (WJ) coefficient of two EQVs m and n is given by the following formula:

$$WJ(EQV_m, EQV_n) = \frac{\sum_1^{nbr_j} \min(w(EQV_m, \text{feature}_j), w(EQV_n, \text{feature}_j))}{\sum_1^{nbr_j} \max(w(EQV_m, \text{feature}_j), w(EQV_n, \text{feature}_j))}$$

The results of the similarity calculation are exploited by a clustering algorithm, which groups semantically similar EQVs.

3.2.7. Implementation details: dynamic programming

The input of the clustering algorithm consists in the set of EQVs of a SL word and the output consists in clusters of EQVs illustrating the senses of the word. Possible clustering solutions being numerous, but only one being optimal, clustering can be expressed in terms of a combinatorial optimization problem. This problem is resolved here using a dynamic programming technique: the construction of the optimal sense clusters containing the most similar EQVs constitutes the ‘global problem’, perceived as composed by a group of ‘sub-problems’, which concern the similarity estimation of each pair of

EQVs. This similarity is described by the score attributed to the pair by the similarity calculation method.

3.2.8. Properties of the Clustering Algorithm

- **Distance measure**

The similarity calculation results constitute the “distance measure” that conditions the EQVs’ grouping: two EQVs are clustered if their similarity score exceeds a certain threshold, defined locally for each SL word as the average of the similarity scores attributed to all the pairs of its EQVs. EQVs having a “significant” semantic relation are those having a score exceeding this threshold.

- **Clustering termination condition**

The resulting clusters could be described in graph theory terms as “complete graphs”, given that all their elements have to be linked to each other. The clustering procedure ceases when this condition is met while no more EQVs may enter a cluster without violating it.

- **Possibility of creation of overlapping clusters**

The algorithm allows for the creation of overlapping clusters. This property of the algorithm is in accord with the nature of the task at hand: the resulting clusters describe senses of the polysemous SL word and it is possible that one EQV (found in the intersection of clusters) translates more than one of its senses. This property of the algorithm is more obvious when the method is applied to manually extracted translation data where bigger clusters (containing more EQVs) are more often constructed. The reason for that is that the recall (which corresponds to the number of EQVs found for a word in the bilingual lexicon to the whole number of EQVs translating the word in the training corpus) is more limited in the automatically generated translation lexicon than in the manually generated ones.

3.3. Sense Induction by Inter-lingual Projection of Clustering Information

Clustered EQVs are supposed to translate the same sense of the SL word, contrary to EQVs of different clusters, which translate different senses. In a contextual approach to semantic similarity (assumption (b)), similar words are considered to be more or less commutable in the contexts revealing their relation (Miller & Charles, 1991). Consequently, we suppose that clustered EQVs can be more or less commutable as translations of the SL word when found in contexts close to the ones that induce their similarity.

The clusters formed are projected on the SL word allowing for the identification of its senses. Each sense induced in this way can be described by the elements of the corresponding cluster. The senses identified for the sample of polysemous words studied here are given in Table 2; we also include a short description of each sense.

SL word	Identified Senses	Sense description
structure	{διάρθρωση, δομή}	arrangement
	{κατασκευή}	construction
guidance	{προσανατολισμός, καθοδήγηση}	orientation
	{συμβουλή}	advice
survey	{δημοσκόπηση, έρευνα}	poll
	{επισκόπηση}	resume
power	{δύναμη}	force
	{αρμοδιότητα, εξουσία}	authority
	{ισχύς}	(electric) load
trade	{συναλλαγή, εμπόριο}	transaction, commerce
	{επάγγελμα}	job

Table 2. Senses of the SL polysemous words

3.4. Using sense clusters for WSD and annotation

The resulting sense clusters can be used for WSD and sense annotation of new instances of the polysemous SL words. The information gathered during training can be used by unsupervised WSD methods in order to select one of the senses for labeling a new instance of a polysemous word. The need for hand-tagged data for WSD is thus eliminated.

Using translation EQVs for WSD brings it closer to the Senseval multilingual tasks (Chklovski *et al.*, 2004), where the sense inventories used represent semantic distinctions performed in other languages. In these tasks, the existence of a biunivocal relation between an EQV and a sense is assumed and no distinction is made between semantically related and unrelated EQVs. Consequently, semantically similar and distant EQVs are considered as indicators of equivalent sense distinctions. On the contrary, in the clustering results, semantically similar EQVs are grouped together and so the identified sense distinctions are coarser.

Furthermore, using the results of this method for semantic annotation overcomes the need of a predefined sense inventory. This renders sense annotation possible for languages for which parallel corpora are available but good quality sense inventories are not.

4. Evaluation

We evaluate the impact of exploiting the semantic information acquired by the sense induction method on the results of a WSD task.

4.1. Test corpus

The corpus used for evaluation is different from the training one: it consists of the English-Greek part of the sentence-aligned first version of the EUROPARL corpus, which contains 623 604 sentence pairs (Koehn, 2005). As in the case of the training corpus, we extract translation units consisting of a SL and a TL segment, forming a “test sub-corpus” for each SL word. In this sub-corpus, the word appears in the English side (segment) of the translation units, while one of its EQVs is found in the Greek side². This EQV is considered as the “reference translation” that will be used for evaluation. Both parts of

² We don’t take into consideration translation units containing EQVs of the SL word not found in the corresponding lexicon entry; the reason is that, as these EQVs were not considered during training, no information relative to them is available.

the corpus have been lemmatized and POS-tagged³ (Schmid, 1994).

4.2. Exploiting the induced senses for WSD

The WSD method used exploits the sense inventory built by the sense induction method described in the previous sections. Sense clusters are characterized by the SL context features that revealed the similarity of the EQVs they contain. Clusters containing one EQV are characterized by the EQV's most pertinent context features. The comparison of this information acquired during training with the context of new SL words instances allows for their disambiguation.

WSD predictions may concern clusters of one or more translation EQVs. In the case where a one element cluster is selected (i.e. the sense chosen is described by only one EQV), this EQV can be considered as the most adequate translation of the new SL word instance. In the case of a cluster of more than one EQV, they can all be considered as (more or less) good translations. Hence, exploiting cluster information permits to the WSD method to take advantage of paradigmatic information relative to the EQVs' semantic similarity that enriches the correspondences between the items of the two languages.

4.3. Evaluation of the WSD method exploiting sense clusters

The WSD results are evaluated using "recall" and "precision": recall is defined as the ratio of correctly disambiguated instances to the total number of new instances of the polysemous word in the test corpus, while precision corresponds to the ratio of correctly disambiguated instances to the number of sense predictions made by the system. We consider as correct the prediction of a sense cluster containing the EQV that translates the new SL word instance in the test corpus (reference translation).

The results are compared with a baseline, which consists in the selection of the most frequent EQV for all the instances of the polysemous word. Hence, the baseline corresponds to both precision and recall, as WSD predictions are made for all test instances. The results obtained for the words studied here are presented in Table 3 (expressed in percentages).

In parenthesis we give the number of occurrences of the polysemous word that have been evaluated and also their distribution according to the reference translations. The most frequent EQV of each word in the training corpus, which serves for calculating the baseline, is given in bold.

SL word	Baseline	Recall	Precision
structure (2156) (δομή: 1649, διάρθρωση: 492, κατασκευή: 15)	76.48	88.91	90.39
guidance (143) (προσανατολισμός: 76, καθοδήγηση: 60, συμβουλή: 7)	53.14	86.71	87.32
survey (231) (έρευνα: 185, δημοσκοπήση: 36, επισκόπηση: 10)	80.08	85.71	86.46
power (5502) (εξουσία: 2764, αρμοδιότητα: 1464, δύναμη: 967, ισχύς: 307)	26.6	70.01	71.50
trade (4973) (εμπόριο: 4063, συναλλαγή: 883, επάγγελμα: 27)	81.7	97.18	99.07
TOTAL	63.6	85.7	86.95

Table 3. Evaluation results

The prediction and recall scores of the WSD method using cluster information clearly overcome the baseline scores for all SL words. It is interesting to note that unsupervised systems in Senseval-3 hardly reach the reported baseline, while best performing systems achieve a 65-70% score, due mainly to the fine granularity of the WordNet senses used (Snyder & Palmer, 2004). Our results are explained by the coarser granularity of the sense inventory exploited for WSD, which contains the senses proposed by our sense induction method.

In almost all cases, the most frequent EQV in the training corpus is also the most frequent reference translation in the evaluation corpus. This is not the case only for *power*, which explains its low baseline score.

5. Perspectives

The sense attributed by the WSD method to a new instance of a polysemous word may consist in a cluster of more than one EQV. In a Machine Aided Translation context, the EQVs contained in the cluster could constitute suggestions of multiple semantically pertinent translations at the word level, from which the translator could select the most adequate for translating the source word.

In an automatic framework, a cluster containing more than one EQV should be filtered out automatically. This could be done using a "lexical selection" method. The aim of this method would consist in deciding which of the semantically similar clustered EQVs would be more appropriate in the new TL context. In an experimental framework, this method would exploit the TL context provided by the parallel test corpus (Vickrey *et al.*, 2005), whereas in a real MT system, TL context would consist in the translations of the rest of the input sentence, depending on the adopted translation approach.

The TL information required for this filtering could be acquired during training from the TL contexts of the EQVs. These contexts would be analyzed and the features retained for each EQV would be weighted in the same way as the SL context features (cf. paragraph 3.2.6.). The features retained for each of the clustered EQVs could then be compared with the new TL context, so that the most appropriate translation of the new SL word instance can be selected.

Such a lexical selection method could complement the results of the WSD method, in cases where the attributed senses are described by clusters containing more than one

³ In order to tag and lemmatize the Greek part of the test corpus, the TreeTagger was trained on the Greek part of our training corpus.

EQV. In a preliminary version of this work, these two methods were merged. However, we have decided to separate them in order to be able to exploit the results of the WSD method, considering that they could be useful in tasks such as semantic annotation.

6. Conclusion

In this paper we have presented a data-driven sense induction method that exploits contextual and translation information extracted from a parallel aligned bilingual corpus. Sense clustering is performed using the results of a semantic similarity calculation concerning the EQVs of a polysemous word. Similarity is estimated using extended contexts corresponding to each EQV of the word, which reduces the data sparseness effect. The method being totally statistical, it can be used for sense induction from various corpora and for different languages. The only prerequisite is a large parallel corpus having undergone a number of preprocessing steps (lemmatization, POS-tagging, sentence and word alignment).

Senses proposed for a SL word are described using its clustered translation EQVs, taking into consideration their similarity relations. This clustering makes possible the suggestion of coarser sense distinctions than in the case of establishment of biunivocal relations between EQVs and senses. The results of the sense induction method, which consist in sense correspondences between words of two languages, can be used for WSD and lexical selection in translation applications.

Acknowledgments

We would like to thank Philippe Langlais and his colleagues at the RALI laboratory (University of Montreal) for aligning our parallel training corpus at the word level.

References

- Agirre, E. & Soroa, A. (2007). Semeval-2007 Task 02 : Evaluating Word Sense Induction and Discrimination Systems. In Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007), Association for Computational Linguistics, 23-24 June, Prague, Czech Republic, pp. 7-12.
- Apidianaki, M. (2007). Repérage de sens et désambiguïsation dans un contexte bilingue. In Proceedings of the 14th Traitement Automatique des Langues Naturelles conference (TALN 2007), Toulouse, France, June 5-8, Vol.1, pp. 207-216.
- Brown, P. F., Della Pietra, S. A., Della Pietra, V. J. & Mercer, R. L. (1991). A statistical approach to sense disambiguation in machine translation. In Proceedings of the Speech and Natural Language Workshop, Pacific Grove, CA, pp. 146-151.
- Chklovski, T., Mihalcea, R., Pedersen, T. & Purandare, A. (2004). The senseval-3 multilingual English-Hindi lexical sample task. In Proceedings of Senseval-3, Third International Workshop on Evaluating Word Sense Disambiguation Systems, Barcelona, Spain, July, pp. 5-8.
- Dolan, W. B. (1994). Word Sense Ambiguation: Clustering Related Senses. In Proceedings of the 15th International Conference on Computational Linguistics (COLING), Kyoto, Japan, 5-9 August, pp. 712-716.
- Edmonds, P. & Kilgarriff, A. (2002). Introduction to the special issue on evaluating word sense disambiguation systems. *Natural Language Engineering* 8(4), Cambridge University Press, pp. 279-291.
- Firth, J. R. (1957). *Papers in Linguistics, 1934-1951*. London/New York: Oxford University Press.
- Gavriliadou, M., Labropoulou, P., Desipri, E., Giouli, V., Antonopoulos, V. & Piperidis, S. (2004). Building parallel corpora for eContent professionals. In Proceedings of MLR 2004, PostCOLING Workshop on Multilingual Linguistic Resources, Geneva.
- Grefenstette, G. (1994). *Explorations in Automatic Thesaurus Discovery*. Boston/Dordrecht/London: Kluwer Academic Publishers.
- Harris, Z. (1954). Distributional structure. *Word*, 10, pp. 146-162.
- Ide, N., Erjavec, T. & Tufiş, D. (2001). Automatic sense tagging using parallel corpora. In Proceedings of the 6th Natural Language Processing Pacific Rim Symposium, pp. 83-89.
- Kaji, H. & Morimoto, Y. (2002). Unsupervised word sense disambiguation using bilingual comparable corpora. In Proceedings of the 19th International Conference on Computational Linguistics (COLING), August 24 – September 1, Taipei, Taiwan, pp. 1-7.
- Kilgarriff, A. (1997). I don't believe in word senses. *Computers and the Humanities* 31(2), pp. 91-113.
- Koehn, P. (2005). Europarl: A Parallel Corpus for Statistical Machine Translation. In Proceedings of the MT Summit X, Phuket, Thailand, pp. 79-86.
- Lesk, M. (1986). Automated sense disambiguation using machine-readable dictionaries: How to tell a pine cone from an ice cream cone. In Proceedings of the 1986 SIGDOC Conference, Toronto, Canada, June, pp. 24-26.
- Mihalcea, R. & Moldovan, D. I. (2001). Automatic generation of a coarse grained WordNet. In Proceedings of the 14th International Florida Artificial Intelligence Research Society Conference (FLAIRS), May 21-23, pp. 454-458.
- Miháltz M. (2005). Towards a Hybrid Approach to Word-Sense Disambiguation in Machine Translation. Workshop on Modern Approaches in Translation Technologies (RANLP-2005), Borovets, Bulgaria.
- Miller, G. A. & Charles, W. G. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes*, 6(1), pp. 1-28.
- Navigli, R. (2006). Meaningful Clustering of Senses Helps Boost Word Sense Disambiguation Performance. In Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics (COLING-ACL), Sydney, Australia, pp. 105-112.
- Ng, H.T., Wang B. & Chan, Y. S. (2003). Exploiting Parallel Texts for Word Sense Disambiguation: An Empirical Study. In Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL), Sapporo, Japan, pp. 455-462.

- Nießen, S. & Ney, H. (2004). Statistical Machine Translation with Scarce Resources Using Morpho-syntactic Information. *Computational Linguistics*, 30(2), pp. 181-204.
- Pantel, P. & Lin., D. (2002). Discovering word senses from text. In *Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining 2002*, Edmonton, Canada, pp. 613-619.
- Peters, W., Peters, I. & Vossen, P. (1998). Automatic sense clustering in EuroWordNet. In *Proceedings of the 1st International Conference on Language Resources and Evaluation (LREC)*, Granada, 28-30 May, pp. 409-416.
- Purandare, A. & Pedersen, T. (2004). Word Sense Discrimination by Clustering Contexts in Vector and Similarity Spaces. In *Proceedings of the Conference on Computational Natural Language Learning (CONLL)*, 6-7 May, Boston, MA, pp. 41-48.
- Resnik, P. (2004). Exploiting Hidden Meanings: Using Bilingual Text for Monolingual Annotation. In Gelbukh, A. (Ed.), *Lecture Notes in Computer Science 2945: Computational Linguistics and Intelligent Text Processing: Proceedings of the Fifth International Conference CICLing*, pp. 283-299.
- Schmid, H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. In *Proceedings of the International Conference on New Methods in Language Processing*, Manchester, pp. 44-49.
- Schütze, H. (1998). Automatic Word Sense Discrimination. *Computational Linguistics*, Vol. 24, Number 1, pp. 97-123.
- Simard, M. & Langlais, P. (2003). Statistical Translation Alignment with Compositionality Constraints. In *Proceedings of HLT-NAACL Workshop on Building and Using Parallel Texts: Data Driven Machine Translation and Beyond*, Edmonton, Canada, May 31, pp. 19-22.
- Snyder, B. & Palmer, M. (2004). The English All-Words Task. In *Proceedings of the Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text (Senseval-3)*, Barcelona, Spain, July 25-26, pp. 41-43.
- Specia, L., Das Graças Volpe Nunes, M., Castelo Branco R. G. & Stevenson, M. (2006) Multilingual versus Monolingual WSD. In *Proceedings of the EACL Workshop on Making Sense of Sense: Bringing Psycholinguistics and Computational Linguistics Together*, April 3-7, Trento, pp. 33-40.
- Tufiş, D., Ion, R. & Ide, N. (2004) 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 (COLING)*, Geneva, pp. 1312-1318.
- Van der Plas, L. & Tiedemann, J. (2006) Finding Synonyms Using Automatic Word Alignment and Measures of Distributional Similarity. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics (COLING-ACL)*, July, Sydney, Australia, pp. 866-873.
- Véronis, J. (2004) Hyperlex: lexical cartography for information retrieval. *Computer, Speech and Language, Special Issue on Word Sense Disambiguation*, 18(3), pp. 223-252.
- Vickrey, D., Biewald L., Teysier, M. & Koller, D. (2005) Word-Sense Disambiguation for Machine Translation. In *Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP)*, October 6-8, Vancouver, Canada, pp. 771-778.
- Wilks, Y. & Stevenson, M. (1996) The Grammar of Sense: Is word-sense tagging much more than part-of-speech tagging? University of Sheffield, Department of Computer Science, Research Memoranda, CS-96-05.