

Using lexico-semantic information for query expansion in passage retrieval for question answering

Lonneke van der Plas

LATL

University of Geneva

Switzerland

lonneke.vanderplas@lettres.unige.ch

Jörg Tiedemann

Alfa-Informatica

University of Groningen

The Netherlands

j.tiedemann@rug.nl

Abstract

In this paper we investigate the use of several types of lexico-semantic information for query expansion in the passage retrieval component of our QA system. We have used four corpus-based methods to acquire semantically related words, and we have used one hand-built resource. We evaluate our techniques on the Dutch CLEF QA track.¹ In our experiments expansions that try to bridge the terminological gap between question and document collection do not result in any improvements. However, expansions bridging the knowledge gap show modest improvements.

1 Introduction

Information retrieval (IR) is used in most QA systems to filter out relevant passages from large document collections to narrow down the search for answer extraction modules in a QA system. Accurate IR is crucial for the success of this approach. Answers in paragraphs that have been missed by IR are lost for the entire QA system. Hence, high performance of IR especially in terms of recall is essential. Furthermore, high precision is desirable as IR scores are used for answer extraction heuristics and also to reduce the chance of subsequent extraction errors.

Because the user's formulation of the question is only one of the many possible ways to state the information need that the user might have, there is

often a discrepancy between the terminology used by the user and the terminology used in the document collection to describe the same concept. A document might hold the answer to the user's question, but it will not be found due to the TERMINOLOGICAL GAP. Moldovan et al. (2002) show that their system fails to answer many questions (25.7%), because of the terminological gap, i.e. keyword expansion would be desirable but is missing. Query expansion techniques have been developed to bridge this gap.

However, we believe that there is more than just a terminological gap. There is also a KNOWLEDGE GAP. Documents are missed or do not end up high in the ranks, because additional world knowledge is missing. We are not speaking of synonyms here, but words belonging to the same subject field. For example, when a user is looking for information about the explosion of the first atomic bomb, in his/her head a subject field is active that could include: war, disaster, World War II.

We have used three corpus-based methods to acquire semantically related words: the SYNTAX-BASED METHOD, the ALIGNMENT-BASED METHOD, and the PROXIMITY-BASED METHOD. The nature of the relations between words found by the three methods is very different. Ranging from free associations to synonyms. Apart from these resources we have used categorised named entities, such as *Van Gogh* IS-A *painter* and synsets from EWN as candidate expansion terms.

In this paper we have applied several types of lexico-semantic information to the task of query expansion for QA. We hope that the synonyms retrieved automatically, and in particular the synonyms retrieved by the alignment-based method, as these are most precise, will help to overcome the

©2008. Licensed under the *Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported* license (<http://creativecommons.org/licenses/by-nc-sa/3.0/>). Some rights reserved.

¹The Cross-Language Evaluation Forum (<http://clef-qa.itc.it/>)

terminological gap. With respect to the knowledge gap, we expect that the proximity-based method would be most helpful as well as the list of categorised named entities. For example, knowing that Monica Seles is a tennis player helps to find relevant passages regarding this tennis star.

2 Related work

There are many ways to expand queries and expansions can be acquired from several sources. For example, one can make use of collection-independent resources, such as EWN. In contrast, collection-dependent knowledge structures are often constructed automatically based on data from the collection.

The results from using collection-independent, hand-built sources are varied. Moldovan et al. (2003) show that using a lexico-semantic feedback loop that feeds lexico-semantic alternations from WordNet as keyword expansions to the retrieval component of their QA system increments the scores by 15%. Also, Paşca and Harabagiu (2001) show substantial improvements when using lexico-semantic information from WordNet for keyword alternation on the morphological, lexical and semantic level. They evaluated their system on question sets of TREC-8 and TREC-9. For TREC-8 they reach a precision score of 55.3% without including any alternations for question keywords, 67.6% if lexical alternations are allowed and 73.7% if both lexical and semantic alternations are allowed.

However, Yang and Chua (2003) report that adding additional terms from WordNet's synsets and glosses adds more noise than information to the query. Also, Voorhees (1993) concludes that expanding by automatically generated synonym sets from EWN can degrade results.

In Yang et al. (2003) the authors use external knowledge extracted from WordNet and the Web to expand queries for QA. Minor improvements are attained when the Web is used to retrieve a list of nearby (one sentence or snippet) non-trivial terms. When WordNet is used to rank the retrieved terms, the improvement is reduced. The best results are reached when structure analysis is added to knowledge from the Web and WordNet. Structure analysis determines the relations that hold between the candidate expansion terms to identify semantic groups. Semantic groups are then connected by conjunction in the Boolean query.

Monz (2003) ran experiments using pseudo relevance feedback for IR in a QA system. The author reports dramatic decreases in performance. He argues that this might be due to the fact that there are usually only a small number of relevant documents. Another reason he gives is the fact that he used the full document to fetch expansion terms and the information that allows one to answer the question is expressed very locally.

A global technique that is most similar to ours uses syntactic context to find suitable terms for query expansion (Grefenstette, 1992; Grefenstette, 1994). The author reports that the gain is modest: 2% when expanded with nearest neighbours found by his system and 5 to 6%, when applying stemming and a second loop of expansions of words that are in the family of the augmented query terms.² Although the gain is greater than when using document co-occurrence as context, the results are mixed, with expansions improving some query results and degrading others.

Also, the approach by Qiu and Frei (1993) is a global technique. They automatically construct a similarity thesaurus, based on what documents terms appear in. They use word-by-document matrices, where the features are document IDs, to determine the similarity between words. Expansions are selected based on the similarity to the query concept, i.e. all words in the query together, and not based on the single words in the query independently. The results they get are promising.

Pantel and Ravichandran (2004) have used a method that is not related to query expansion, but yet very related to our work. They have semantically indexed the TREC-2002 IR collection with the ISA-relations found by their system for 179 questions that had an explicit semantic answer type, such as *What band was Jerry Garcia with?* They show small gains in performance of the IR output using the semantically indexed collection.

Recent work (Shen and Lapata, 2007; Kaiser and Webber, 2007) that falls outside the scope of this paper, but that is worth mentioning successfully applies semantic roles to question answering.

3 Lexico-semantic information

We have used several types of lexico-semantic information as sources for candidate expansion terms. The first three are automatically acquired

²i.e. words that appear in the same documents and that share the first three, four or five letters.

from corpora by means of distributional methods.

- Nearest neighbours from proximity-based distributional similarity
- Nearest neighbours from syntax-based distributional similarity
- Nearest neighbours from alignment-based distributional similarity

The idea behind distributional methods is rooted in the DISTRIBUTIONAL HYPOTHESIS (Harris, 1968). Similar words appear in similar context. The way words are distributed over contexts tells us something about their meaning. Context can be defined in several ways. The way the context is defined determines the type of lexico-semantic knowledge we will retrieve.

For example, one can define the context of a word as the n words surrounding it. In that case proximity to the head word is the determining factor. We refer to these methods that use unstructured context as PROXIMITY-BASED METHODS. The nearest neighbours resulting from such methods are rather unstructured as well. They are merely associations between words, such as *baby* and *cry*. We have used the 80 million-word corpus of Dutch newspaper text (the CLEF corpus) that is also part of the document collection in the QA task to retrieve co-occurrences within sentences.

Another approach is one in which the context of a word is determined by syntactic relations. In this case, the head word is in a syntactic relation to a second word and this second word accompanied by the syntactic relation form the context of the head word. We refer to these methods as SYNTAX-BASED METHODS. We have used several syntactic relations to acquire syntax-based context for our headwords. This method results in nearest neighbours that at least belong to the same semantic and syntactic class, for example *baby* and *son*. We have used 500 million words of newspaper text (the TwNC corpus parsed by Alpino (van Noord, 2006)) of which the CLEF corpus is a subset.

A third method we have used is the ALIGNMENT-BASED METHOD. Here, translations of word, retrieved from the automatic word alignment of parallel corpora are used to determine the similarity between words. This method results in even more tightly related data, as it mainly finds synonyms, such as *infant* and

baby. We have used the Europarl corpus (Koehn, 2003) to extract word alignments from.³

By calculating the similarity between the contexts words are found in, we can retrieve a ranked list of nearest neighbours for any headword. We gathered nearest neighbours for a frequency-controlled list of words, that was still manageable to retrieve. We included all words (nouns, verbs, adjectives and proper names) with a frequency of 150 and higher in the CLEF corpus. This resulted in a ranked list of nearest neighbours for the 2,387 most frequent adjectives, the 5,437 most frequent nouns, the 1,898 most frequent verbs, and the 1,399 most frequent proper names. For all words we retrieved a ranked list of its 100 nearest neighbours with accompanying similarity score.

In addition to the lexico-semantic information resulting from the three distributional methods we used:

- Dutch EuroWordNet (Vossen, 1998)
- Categorised named entities

With respect to the first resource we can be short. We selected the synsets of this hand-built lexico-semantic resource for nouns, verbs, adjectives and proper names.

The categorised named entities are a by-product of the syntax-based distributional method. From the example in (1) we extract the apposition relation between *Van Gogh* and *schilder* ‘painter’ to determine that the named entity *Van Gogh* belongs to the category of painters.

- (1) Van Gogh, de beroemde schilder huurde een atelier, Het Gele huis, in Arles.
‘Van Gogh, the famous painter, rented a studio, The Yellow House, in Arles.’

We used the data of the TwNC corpus (500M words) and Dutch Wikipedia (50M words) to extract apposition relations. The data is skewed. The Netherlands appears with 1,251 different labels. To filter out incorrect and highly unlikely labels (often the result of parsing errors) we determined the relative frequency of the combination of the named entity and a category with regard to the frequency of the named entity overall. All categorised named entities with relative frequencies under 0.05

³In van der Plas and Tiedemann (2006) there is more information on the syntax-based and alignment-based distributional methods.

Lex. info	Nouns	Adj	Verbs	Proper
Proximity	5.3K	2.4K	1.9K	1.2K
Syntax	5.4K	2.3K	1.9K	1.4K
Align	4.0K	1.2K	1.6K	
Cat. NEs				218K
EWN	44.9K	1.5K	9.0K	1.4K

Table 1: Number of words for which lexico-semantic information is available

were discarded. This cutoff made the number of unwanted labels considerably lower.

In Table 1 we see the amount of information that is contained in individual lexico-semantic resources. It is clear from the numbers that the alignment-based method does not provide nearest neighbours for all head words selected. Only 4.0K nouns from the 5.4K retrieve nearest neighbours. The data is sparse. Also, the alignment-based method does not have any nearest neighbours for proper names, due to decisions we made earlier regarding preprocessing: All words were transformed to lowercase.

The proximity-based method also misses a number of words, but the number is far less important. The amount of information the lists of categorised named entities provide is much larger than the amount of information comprised in the list provided by distributional methods. EWN also provides more information than the distributional methods, except for adjectives.⁴

4 Methodology

In order to test the performance of the various lexico-semantic resources we ran several tests. The baseline is running a standard full-text retrieval engine using Apache Lucene (Jakarta, 2004). Documents have been lemmatised and stop words have been removed.

We applied the nearest neighbours resulting from the three distributional methods as described in section 3. For all methods we selected the top-5 nearest neighbours that had a similarity score of more than 0.2 as expansions.

For EWN all words in the same synset (for all senses) were added as expansions. Since all synonyms are equally similar, we do not have similarity scores for them to be used in a threshold.

The categorised named entities were not only used to expand named entities with the corre-

⁴Note that the number of nouns from EWN is the result of subtracting the proper names.

sponding label, but also to expand nouns with named entities. In the first case all labels were selected. The maximum is not more than 18 labels. In the second case some nouns get many expansions. For example, a noun, such as *vrouw* ‘woman’, gets 1,751 named entities as expansions. We discarded nouns with more than 50 expansions, as these were deemed too general and hence not very useful.

The last two settings are the same for the expansions resulting from distributional methods and the last two types of lexico-semantic information.

- Expansions were added as root forms
- Expansions were given a weight such that all expansions for one original keyword add up to 0.5.

5 Evaluation

For evaluation we used data collected from the CLEF Dutch QA tracks. The CLEF text collection contains 4 years of newspaper text, approximately 80 million words and Dutch Wikipedia, approximately 50 million words. We used the question sets from the competitions of the Dutch QA track in 2003, 2004, and 2005 (774 in total). Questions in these sets are annotated with valid answers found by the participating teams including IDs of supporting documents in the given text collection. We expanded these list of valid answers where necessary.

We calculated for each run the Mean Reciprocal Rank (MRR).⁵ The MRR measures the percentage of passages for which a correct answer was found in the top- k passages returned by the system. The MRR score is the average of $1/R$ where R is the rank of the first relevant passage computed over the 20 highest ranked passages. Passages retrieved were considered relevant when one of the possible answer strings was found in that passage.

6 Results

In Table 2 the MRR (Mean Reciprocal Rank) is given for the various expansion techniques. Scores are given for expanding the several syntactic categories, where possible. The baseline does not

⁵We used MRR instead of other common evaluation measures because of its stronger correlation with the overall performance of our QA system than, for example, coverage and redundancy (see Tiedemann and Mur (2008)).

SynCat	MRR				
	EWN	Syntax	Align	Proxi	Cat.NEs
Nouns	51.52	51.15	51.21	51.38	51.75
Adj	52.33	52.27	52.38	51.71	
Verbs	52.40	52.33	52.21	52.62	
Proper	52.59	50.16		53.94	55.68
All	51.65	51.21	51.02	53.36	55.29

Table 2: MRR scores for the IR component with query expansion from several sources

SynCat	#questions (+/-)				
	EWN	Syntax	Align	Proxi	Cat.NEs
Nouns	27/50	28/61	17/58	64/87	17/37
Adj	3/6	1/2	1/2	31/47	
Verbs	31/51	5/10	8/32	51/56	
Proper	3/2	30/80		76/48	157/106
All	56/94	56/131	25/89	161/147	168/130

Table 3: Number of questions that receive a higher (+) or lower (-) RR when using expansions from several sources

make use of any expansion for any syntactic category and amounts to 52.36.

In Table 3 the number of questions that get a higher and lower reciprocal rank (RR) after applying the individual lexico-semantic resources are given. Apart from expansions on adjectives and proper names from EWN, the impact of the expansion is substantial. The fact that adjectives have so little impact is due to the fact that there are not many adjectives among the query terms.⁶

The negligible impact of the proper names from EWN is surprising since EWN provides more entries for proper names than the proximity-based method (1.2K vs 1.4K, as can be seen in 1). The proximity-based method clearly provides information about proper names that are more relevant for the corpus used for QA, as it is built from a subset of that same corpus. This shows the advantage of using corpus-based methods. The impact of the expansions resulting from the syntax-based method lies in between the two previously mentioned expansions. It uses a corpus of which the corpus used for QA is a subset.

The type of expansions that result from the proximity-based method have a larger effect on the performance of the system than those resulting from the syntax-based method. In Chapter 5 of van der Plas (2008) we explain in greater detail that the proximity-based method uses frequency cut-

⁶Moreover, the adjectives related to countries, such as *German* and *French* and their expansion *Germany*, *France* are handled by a separate list.

offs to keep the co-occurrence matrix manageable. The larger impact of the proximity-based nearest neighbours is probably partly due to this decision. The cutoffs for the alignment-based and syntax-based method have been determined after evaluations on EuroWordNet (Vossen, 1998) (see also van der Plas (2008)).

The largest impact results from expanding proper names with categorised named entities. We know from Table 1 in section 3, that this resource has 70 times more data than the proximity-based resource.

For most of the resources the number of questions that show a rise in RR is smaller than the number of questions that receive a lower RR, except for the expansion of proper names by the categorised named entities and the proximity-based method.

The expansions resulting from the syntax-based method do not result in any improvements. As expected, the expansion of proper names from the syntax-based method hurts the performance most. Remember that the nearest neighbours of the syntax-based method often include co-hyponyms. For example, *Germany* would get *The Netherlands* and *France* as nearest neighbours. It does not seem to be a good idea to expand the word *Germany* with other country names when a user, for example, asks the name of the Minister of Foreign Affairs of Germany. However, also the synonyms from EWN and the alignment-based method do not result in improvements.

The categorised named entities provide the most successful lexico-semantic information, when used to expand named entities with their category label. The MRR is augmented by almost 3,5%. It is clear that using the same information in the other direction, i.e. to expand nouns with named entities of the corresponding category hurts the scores. The proximity-based nearest neighbours of proper names raises the MRR scores with 1,5%.

Remember from the introduction that we made a distinction between the terminological gap and the knowledge gap. The lexico-semantic resources that are suited to bridge the terminological gap, such as synonyms from the alignment-based method and EWN, do not result in improvements in the experiments under discussion. However, the lexico-semantic resources that may be used to bridge the knowledge gap, i.e. associations from the proximity-based method and categorised

CLEF score					
EWN	Syntax	Align	Proxi	Cat.NEs	Baseline
46.3	47.0	46.6	47.6	47.9	46.8

Table 4: CLEF scores of the QA system with query expansion from several sources

named entities, do result in improvements of the IR component.

To determine the effect of query expansion on the QA system as a whole we determined the average CLEF score when using the various types of lexico-semantic information for the IR component. The CLEF score gives the precision of the first (highest ranked) answer only. For EWN, the syntax-based, and the alignment-based nearest neighbours we have used all expansions for all syntactic categories together. For the proximity-based nearest neighbours and the categorised named entities we have limited the expansions to the proper names as these performed rather well.

The positive effect of using categorised named entities and proximity-based nearest neighbours for query expansion is visible in the CLEF scores as well, although less apparent than in the MRR scores from the IR component in Table 2.

6.1 Error analysis

Let us first take a look at the disappointing results regarding the terminological gap, before we move to the more promising results related to the knowledge gap. We expected that the expansions of verbs would be particularly helpful to overcome the terminological gap that is large for verbs, since there is much variation. We will give some examples of expansion from EWN and the alignment-based method.

- (2) Wanneer werd het Verdrag van Rome getekend?
‘When was the Treaty of Rome signed?’

Align: *teken* ‘sign’ → *typeer* ‘typify’, *onderteken* ‘sign’
EWN: *teken* ‘sign’ → *typeer* ‘typify’, *kentekenen* ‘characterise’, *kenmerk* ‘characterise’, *schilder* ‘paint’, *kenschets* ‘characterise’, *signeer* ‘sign’, *onderteken* ‘sign’, *schets* ‘sketch’, *karacteriseer* ‘characterise’.

For the example in (2) both the alignment-based expansions and the expansion from EWN result in a decrease in RR of 0.5. The verb *teken* ‘sign’ is ambiguous. We see three senses of the verb represented in the EWN list, i.e. drawing, characterising, and signing as in signing an official document. One out of the two expansions for the alignment-based method and 2 out of 9 for EWN are in princi-

ple synonyms of *teken* ‘sign’ in the right sense for this question. However, the documents that hold the answer to this question do not use synonyms for the word *teken*. The expansions only introduce noise.

We found a positive example in (3). The RR score is improved by 0.3 for both the alignment-based expansions and the expansions from EWN, when expanding *explodeer* ‘explode’ with *ontplof* ‘blow up’.

- (3) Waar explodeerde de eerste atoombom?
‘Where did the first atomic bomb explode?’

Align: *explodeer* ‘explode’ → *ontplof* ‘blow up’.
EWN: *explodeer* ‘explode’ → *barst los* ‘burst’, *ontplof* ‘blow up’, *barst uit* ‘crack’, *plof* ‘boom’.

To get an idea of the amount of terminological variation between the questions and the documents, we determined the optimal expansion words for each query, by looking at the words that appear in the relevant documents. When inspecting these, we learned that there is in fact little to be gained by terminological variation. In the 25 questions we inspected we found 1 near-synonym only that improved the scores: *gekkekoeienziekte* ‘mad cow disease’ → *Creutzfeldt-Jacob-ziekte* ‘Creutzfeldt-Jacob disease’.

The fact that we find only few synonyms might be related to a point noted by Mur (2006): Some of the questions in the CLEF track that we use for evaluation look like back formulations.

After inspecting the optimal expansions, we were under the impression that most of the expansions that improved the scores were related to the knowledge gap, rather than the terminological gap. We will now give some examples of good and bad expansions related to the knowledge gap.

The categorised named entities result in the best expansions, followed by the proximity-based expansions. In (4) an example is given for which categorised named entities proved very useful:

- (4) Wie is Keith Richard?
‘Who is Keith Richard?’

Cat. NEs: *Keith Richard* → *gitarist* ‘guitar player’, *lid* ‘member’, *collega* ‘colleague’, *Rolling Stones-gitarist* ‘Rolling Stones guitar player’, *Stones-gitarist* ‘Stones guitar player’.

It is clear that this type of information helps a lot in answering the question in (4). It contains the answer to the question. The RR for this question goes from 0 to 1. We see the same effect for the

question *Wat is NASA?* ‘What is NASA?’.

It is a known fact that named entities are an important category for QA. Many questions ask for named entities or facts related to named entities. From these results we can see that adding the appropriate categories to the named entities is useful for IR in QA.

The categorised named entities were not always successful. In (5) we show that the proximity-based expansion proved more helpful in some cases.

- (5) Welke bevolkingsgroepen voerden oorlog in Rwanda?
‘What populations waged war in Rwanda?’

Proximity: *Rwanda* → Zaire, Hutu, Tutsi, Ruanda, Rwandees ‘Rwandese’.

Cat. NEs: *Rwanda* → bondgenoot ‘ally’, land ‘country’, staat ‘state’, buurland ‘neighbouring country’.

In this case the expansions from the proximity-based method are very useful (except for Zaire), since they include the answer to the question. That is not always the case, as can be seen in (6). However, the expansions from the categorised named entities are not very helpful in this case either.

- (6) Wanneer werd het Verdrag van Rome getekend?
‘When was the treaty of Rome signed?’

Proximity: *Rome* → paus ‘pope’, Italië, bisschop ‘bishop’, Italiaans ‘Italian’, Milaan ‘Milan’.

Cat. NEs: *Rome* → provincie ‘province’, stad ‘city’, hoofdstad ‘capital’, gemeente ‘municipality’.

IR does identify *Verdrag van Rome* ‘Treaty of Rome’ as a multi-word term, however it adds the individual parts of multi-word terms as keywords as a form of compound analysis. It might be better to expand the multi-word term only and not its individual parts to decrease ambiguity. *Verdrag van Rome* ‘Treaty of Rome’ is not found in the proximity-based nearest neighbours, because it does not include multi-word terms.

Still, it is not very helpful to expand the word *Rome* with *pope* for this question that has nothing to do with religious affairs. We can see this as a problem of word sense disambiguation. The association *pope* belongs to Rome in the religious sense, the place where the Catholic Church is seated. Rome is often referred to as the Catholic Church itself, as in *Henry VIII broke from Rome*. Gonzalo et al. (1998) showed in an experiment, where words were manually disambiguated, that a substantial increase in performance is obtained when query words are disambiguated, before they

are expanded.

We tried to take care of these ambiguities by using an overlap method. The overlap method selects expansions that were found in the nearest neighbours of more than two query words. Unfortunately, as Navigli and Velardi (2003), who implement a similar technique, using lexico-semantic information from WordNet, note, the COMMON NODES EXPANSION TECHNIQUE works very badly. Also, Voorhees (1993) who uses a similar method to select expansions concludes that the method has the tendency to select very general terms that have more than one sense themselves. In future work we would like to implement the method by Qiu and Frei (1993), as discussed in section 2, that uses a more sophisticated technique to combine the expansions of several words in the query.

7 Conclusion

We can conclude from these experiments on query expansion for passage retrieval that query expansion with synonyms to overcome the terminological gap is not very fruitful. We believe that the noise introduced by ambiguity of the query terms is stronger than the positive effect of adding lexical variants. This is in line with findings by Yang and Chua (2003). On the contrary, Paşca and Harabagiu (2001) were able to improve their QA system by using lexical and semantic alternations from WordNet using feedback loops.

The disappointing results might also be due to the small amount of terminological variation between questions and document collection.

However, adding extra information with regard to the subject field of the query, query expansions that bridge the knowledge gap, proved slightly beneficial. The proximity-based expansions augment the MRR scores with 1.5%. Most successful are the categorised named entities. These expansions were able to augment the MRR scores with nearly 3.5%.

The positive effect of using categorised named entities and proximity-based nearest neighbours for query expansion is visible in the CLEF scores for the QA system overall as well. However, the improvements are less apparent than in the MRR scores from the IR component.

Acknowledgements

This research was carried out in the project *Question Answering using Dependency Relations*, which is part of the research program for *Interactive Multimedia Information eXtraction*, IMIX, financed by NWO, the Dutch Organisation for Scientific Research and partly by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n 216594 (CLASSIC project: www.classic-project.org).

References

- Gonzalo, J., F. Verdejo, I. Chugur, and J. Cigarran. 1998. Indexing with WordNet synsets can improve text retrieval. In *Proceedings of the COLING/ACL Workshop on Usage of WordNet for NLP*.
- Grefenstette, G. 1992. Use of syntactic context to produce term association lists for text retrieval. In *Proceedings of the Annual International Conference on Research and Development in Information Retrieval (SIGIR)*.
- Grefenstette, G. 1994. *Explorations in automatic thesaurus discovery*. Kluwer Academic Publishers.
- Harris, Z. S. 1968. *Mathematical structures of language*. Wiley.
- Jakarta, Apache. 2004. Apache Lucene - a high-performance, full-featured text search engine library. <http://lucene.apache.org/java/docs/index.html>.
- Kaisser, M. and B. Webber. 2007. Question answering based on semantic roles. In *Proceedings of the ACL workshop on deep linguistic processing*.
- Koehn, P. 2003. Europarl: A multilingual corpus for evaluation of machine translation.
- Moldovan, D., M. Passça, S. Harabagiu, and M. Surdeanu. 2002. Performance issues and error analysis in an open-domain question answering system. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*.
- Moldovan, D., M. Pasça, S. Harabagiu, and M. Surdeanu. 2003. Performance issues and error analysis in an open-domain question answering system. *ACM Transactions on Information Systems.*, 21(2):133–154.
- Monz, C. 2003. *From Document Retrieval to Question Answering*. Ph.D. thesis, University of Amsterdam.
- Mur, J. 2006. Increasing the coverage of answer extraction by applying anaphora resolution. In *Fifth Slovenian and First International Language Technologies Conference (IS-LTC)*.
- Navigli, R. and P. Velardi. 2003. An analysis of ontology-based query expansion strategies. In *Proceedings of the Workshop on Adaptive Text Extraction and Mining (ATEM), in the 14th European Conference on Machine Learning (ECML 2003)*.
- Pantel, P. and D. Ravichandran. 2004. Automatically labeling semantic classes. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT/EMNLP)*.
- Pasça, M. and S Harabagiu. 2001. The informative role of wordnet in open-domain question answering. In *Proceedings of the NAACL 2001 Workshop on WordNet and Other Lexical Resources*.
- Qiu, Y. and H.P. Frei. 1993. Concept-based query expansion. In *Proceedings of the Annual International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 160–169.
- Shen, D. and M. Lapata. 2007. Using semantic roles to improve question answering. In *Proceedings of EMNLP*.
- Tiedemann, J. and J. Mur. 2008. Simple is best: Experiments with different document segmentation strategies for passage retrieval. In *Proceedings of the Coling workshop Information Retrieval for Question Answering*. To appear.
- van der Plas, L. and J. Tiedemann. 2006. Finding synonyms using automatic word alignment and measures of distributional similarity. In *Proceedings of COLING/ACL*.
- van der Plas, Lonneke. 2008. *Automatic lexico-semantic acquisition for question answering*. Ph.D. thesis, University of Groningen. To appear.
- van Noord, G. 2006. At last parsing is now operational. In *Actes de la 13eme Conference sur le Traitement Automatique des Langues Naturelles*.
- Voorhees, E.M. 1993. Query expansion using lexical-semantic relations. In *Proceedings of the Annual International Conference on Research and Development in Information Retrieval (SIGIR)*.
- Vossen, P. 1998. EuroWordNet a multilingual database with lexical semantic networks.
- Yang, H. and T-S. Chua. 2003. Qualifier: question answering by lexical fabric and external resources. In *Proceedings of the Conference on European Chapter of the Association for Computational Linguistics (EACL)*.
- Yang, H., T-S. Chua, Sh. Wang, and Ch-K. Koh. 2003. Structured use of external knowledge for event-based open domain question answering. In *Proceedings of the Annual International Conference on Research and Development in Information Retrieval (SIGIR)*.