Using WordNet into UKB in a Question Answering System for Basque

Olatz Perez de Viñaspre, Maite Oronoz and Olatz Ansa
University of the Basque Country
Donostia
operezdevina001@ikasle.ehu.es, {maite.oronoz, olatz.ansa}@ehu.es

Abstract

This paper presents the use of semantic information at chunk level in a Question Answering system called *Ihardetsi*. The semantic information has been added through a tool called UKB. For this experiment, UKB uses the Basque WordNet to compute the similarity between the chunks. We use this added information to help *Ihardetsi* to choose the correct answer among all the extracted candidates. Along with the description of the system, we outline its performance presenting an experiment and the obtained results.

1 Introduction

Question answering systems deal with the task of finding a precise and concrete answer for a natural language question on a document collection. These systems use Information Retrieval (IR) and Natural Language Processing (NLP) techniques to understand the question and to extract the answer.

Ihardetsi (Ansa et al., 2009), a Basque question answering system, takes questions written in Basque as input and obtains the results from a corpus written in Basque too. The stable version of *Ihardetsi* incorporates tools and resources developed in the IXA group, such as the morphosyntactic analyzer (Aduriz et al., 1998) and the named entity recognizer and classifier (Fernandez et al., 2011). Nevertheless, we can assume that the use of more syntactic and semantic information in *Ihardetsi*, will probably improve the quality of the obtained answers. Let us see, for instance, the next question in Basque:

"Nor izendatu zuten EEBBtako lehendakari 1944. urtean?" ("Who was appointed president of the US in the year 1944?")

This question belongs to the *Gold Standard* question bank defined for Basque for the CLEF2008 conference (Forner et al., 2008). In this bank, the answer given as correct for this question is the following:

"Harry Trumanek Franklin Roosevelt ordezkatu zuen EEBBetako lehendakaritzan 1944. urtean." ("Harry Truman replaced Franklin Roosevelt in the presidency of the US in 1944.")

The search of the named entity "EEBB" (*"US"*), the common noun *"lehendakari"* ("president") and the date "1944" separately, does not guarantee that the president to be found by the system will be from the US. For example, searching in Google these three elements, we obtain among others the sentence "The decision of **President** Edwin Barclay (1930-**1944**) to adopt the **US** dollar as the sole legal tender in Liberia..." in which the president is from "Liberia". The use of chunks, that is noun and verbal phrases, in the question answering system, i.e. *"EEBBtako lehendakari"* ("president of the US"), would reduce the searching space of the system.

On the other hand, and as we can see in the previous example, sometimes the terms used in the question and in the possible answers, although are not the same ("president of the US" in the question, and "presidency of the US" in the best answer) the terms refer to the same concept. That is one of the reasons why we decided to use the semantic similarity of the chunks to try to improve *Ihardetsi*. The similarity algorithm we use has the Basque WordNet (Pociello et al., 2010) as its base-ontology. As this ontology lacks of named entities we have included some of them with their corresponding synsets to the dictionary used by the algorithm.

As seen in the previous example, the use of shallow syntactic information and semantic information seems to be helpful, so we have integrated more linguistic knowledge in Ihardetsi. We have integrated the IXATI chunker (Aduriz et al., 2004) in the analysis chain and we have used a similarity

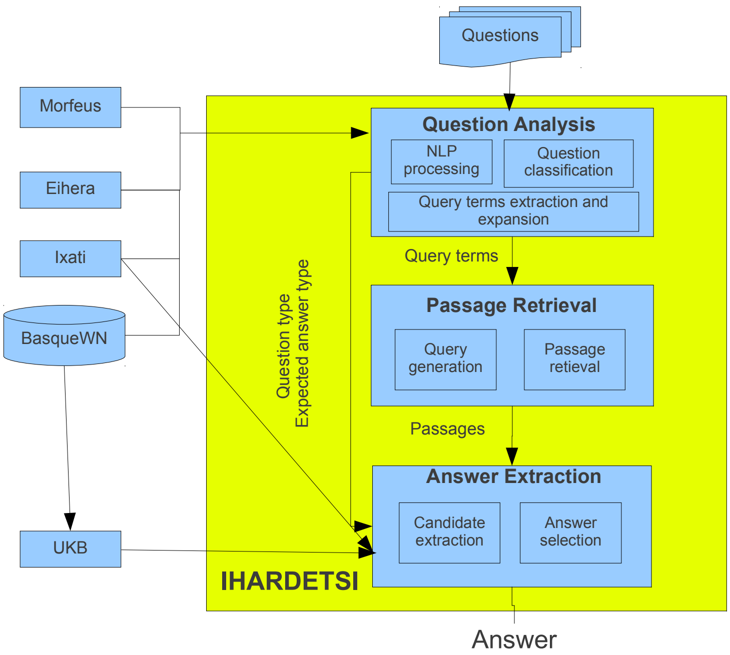

Figure 1: General architecture of the system.

algorithm that is implemented into a tool called UKB (Agirre et al., 2009). The chunks are obtained both in the question and in all the candidate answer-passages.

The remainder of the paper is organized as follows. Section two is devoted to introduce the general architecture of the system. In section three we describe the work done when comparing semantically the chunks from the questions and from the candidate answers. In section four evaluation issues are discussed. Finally, section five contains the conclusions and suggestions for future research.

2 Ihardetsi - A QA System for Basque Language

The principles of versatility and adaptability have guided the development of *Ihardetsi*. It is based on web services and integrated by the SOAP (Simple Object Access Protocol) communication protocol. The linguistic tools previously developed in the IXA group are reused as autonomous web services, and the QA system becomes a client that calls these services when needed. This distributed model allows to parameterize the linguistic tools, and to adjust the behavior of the system.

As it is common in question answering systems, *Ihardetsi* is based on three main modules: the question analysis module, the passage retrieval module and the answer extraction module. Those modules can be seen in the figure 1.

Question Analysis: the main goal of this module is to analyze the question and to generate the information needed for the next tasks. Concretely, a set of search terms is extracted for the passage retrieval module, and the expected answer type along with some lexical and syntactic information is passed to the answer extraction module. Before our contributions, this module used to analyze the questions at morphological level with an analyzer called *Morfeus* (Aduriz et al., 1998), and a named entity recognizer called *Eihera* (Fernandez et al., 2011). After the changes described in this paper, the chunker called *Ixati* is added to this module, enriching this way, the question analysis linguistic chain.

Passage Retrieval: basically an information retrieval task is performed, but in this case the retrieved units are passages and not entire documents. This module receives as input the selected query terms and produces a set of queries that are passed to a search engine.

Answer Extraction: in this module two tasks are performed in sequence: the candidate extraction and the answer selection. Basically, the candidate extraction consists of extracting all the candidate answers from the retrieved passages, and the answer selection consists of choosing the best answers among the considered as candidates. The chunker is applied to the candidate answer passages extracted by the stable version of Ihardetsi that uses a kind of “bag of words” technique. For the work presented in this paper, a re-ranking of the candidate answers is performed using the semantic similarity algorithm from UKB. The number of candidates to be shown could be parameterized but usually five answers are presented to the user.

3 Comparison at chunk level using WordNet

Having applied shallow syntax to the text involved in the QA process, it is possible to compare syntactically the chunks from the question with the according chunks from the candidate answer passages; but also the semantic similarity of the chunks could be measured. Although we have used both syntactic and semantic information to re-rank the answers, we will focus on the semantic area in this paper. The next section describes deeply this work.

3.1 Semantic similarity - UKB similarity

UKB is a collection of programs to perform graph-based Word Sense Disambiguation and lexical similarity/relatedness using a pre-existing knowledge base. It applies the so-called Personalized PageRank on a Lexical Knowledge Base (LKB) to rank the vertices of the LKB and thus it performs disambiguation. The algorithm can also be used to calculate lexical similarity/relatedness of words/sentences (Agirre et al., 2010a) (Agirre and Soroa, 2009).

We took the decision of using UKB according to different reasons: i) it is developed by our same research group, the IXA group; ii) it is language independent as it only needs the semantic knowledge of a language in order to be used, so having a Basque WordNet we can use it for our language; iii) it is free and it is free software as well; iv) it is robust and that is the reason why some analyzers have already integrated it, for example Freeling (Padró et al., 2010).

UKB needs two sources of knowledge that could be extracted from a WordNet to work: on the one hand, a graph containing relations and glosses between concepts, and on the other hand, a dictionary in which word-forms are linked to their corresponding concept. For specific domains as the medical one, other sources as the Unified Medical Language System (UMLS) has been successfully used instead of WordNet in UKB (Agirre et al., 2010b).

Similarity algorithms measure the semantic similarity and relatedness between terms or texts. This concrete algorithm in UKB is able to estimate the similarity measure between two texts, based on the relations of the LKB senses. The method has basically two steps: first, it computes the Personalized PageRank over WordNet separately for each text, producing a probability distribution over WordNet synsets. Then, it compares how similar these two discrete probability distributions are by encoding them as vectors and computing the cosine among the vectors.

When using UKB and WordNet applied to the question answering area, we have found some problems related to the semantic ambiguity of the chunks and to the lack of information in WordNet. These problems will be extensively explained in section 4.

3.2 Procedure to get a weight for each candidate answer

The re-ranking of the candidate answers in *Ihardetsi* is performed by normalizing the weights obtained after the analysis of several syntactic and semantic characteristics. Before explaining this process, in this section we will explain some linguistic phenomena used for the definition of the weights, and then, we will show by means of an example, which features are taken into account for the weight assignment.

We have defined some syntactic patterns at shallow syntax level in order to describe the behavior of some interrogative pronouns in Basque, such as *"Nor"* ("Who"), *"Non"* ("Where"), *"Noiz"* ("When") and *"Zein"* ("Which"). They all belong to the *factoid* question type, i.e. the question is asking about a simple fact or relationship, and the answer is easily expressed, usually by means of a named entity. We found some interesting phenomena when defining the patterns:

- *"Zein". A noun following the pronoun.*

  The interrogative pronoun "Zein" ("Which") is more complicated than the others. When a noun (or noun phrase) is following the interrogative pronoun ("Which nation is...") it behaves different, because this noun specifies which the answer should be (a nation in the example). In other words, we know which semantic concept (or synset) are we looking for, so we can find different word forms or instances related to the same concept.

- *"Zein"-"Non"/"Noiz": overlap in the meaning.*

  Continuing with the "Zein" (*"Which"*) interrogative pronoun, we have noticed that depending on the meaning of the noun following the pronoun, the pattern is similar to the patterns in other interrogative pronouns. For example, *"Zein lekutan dago gailurrik altuena?"* ("In which place is the highest peak?") is equivalent to "Non" (*"Where"*). In these cases we have added the "Non" patterns to the "Zein" patterns as their meaning is the same. The interrogative pronoun "Noiz" is similar to "Non", so we added the "Noiz" patterns to the "Zein" patterns as well.

As we have shown, an analysis of the question patterns gives tips to find related concepts in the answers (specific locations and dates) and tips to share patterns among interrogative pronouns.

After having all the chunks tagged in the question and candidate answer passages and having the syntactic patterns properly defined, we perform pattern matching. For the evaluation of the system, we obtain a weight for each element to be compared. These are the elements that are compared: the noun phrase, the verb phrase and the noun (this element is used to evaluate the noun following the interrogative pronoun "Which").

The following example identifies each of these elements in a question and in one of its candidate answers. In all the cases the lemma of the words is used in order to make easier the deal with the high inflection of the Basque language.

- Question: *"Nor izendatu zuten EEBBtako lehendakari 1944. urtean?"* ("Who was appointed president of the US in the year 1944?")
  - Noun phrase: *"EEBB lehendakari 1944 urte"* ("president of the US in the year 1944")
  - Verb phrase: *"izendatu izan"* ("be appointed")
- Candidate answer: *"Harry Trumanek Franklin Roosevelt ordezkatu zuen EEBBetako lehendakaritzan 1944. urtean."* ("Harry Truman replaced Franklin Roosevelt in the presidency of the US in 1944.")
  - Noun phrases: *"Harry Truman"*, *"Franklin Roosevelt"* and *"EEBB lehendakaritza 1944 urte"* ("in the presidency of the US in 1944")
  - Verb phrase: *"ordezkatu izan"* ("be replaced")

Following the example, we compare the noun phrase in the question (*"EEBB lehendakari 1944 urte"*) with all the noun phrases of the candidate answer (*"Harry Truman"*, *"Franklin Roosevelt"* and *"EEBB lehendakaritza 1944 urte"*). Thus, we get the marks given by the similarity script for the three comparisons. We will choose the highest mark as the noun phrase weight to evaluate this candidate answer. In this case, we expect that the chunk *"EEBB lehendakaritza 1944 urte"* is the semantically closest chunk, and thus, this will be the representative chunk of the candidate answer.

Let us show the marks obtained in this example. In case of the noun phrases *"Harry Truman"* and *"Franklin Roosevelt"* the marks obtained from UKB are the same. This is caused by the fact that both are named entities, and they do not appear in our dictionary. As we will explain in the following section, in those cases we get the synset of the entity type, that is, the synset of person. It is surprising that we obtain worse marks comparing the noun phrase *"EEBB lehendakaritza 1944 urte"* with the question's chunk. As we have mentioned before, we were expecting this chunk to be best one, but we get the mark 0.0025 over the 0.0040 from the other two. We should explain that the mark 0.0025 seems to be low, but it is a really good mark, considering that the similarity obtained when comparing the chunk with itself (*"EEBB lehendakari 1944 urte"*) is 0.0037. Anyway, we will take the highest mark, 0.0040, to compare with the other candidate answers. This process is repeated for the verb phrases, and the whole comparison is again applied to the other candidate answers. Thus, we will be able to normalize the weights obtained and according to this normalization to get the best candidate answer.

We can not forget too the cases in which we find a noun following the "Which" interrogative pronoun. In these cases, we have an additional weight: we compare the noun from the question with each one of the concrete answers (not with the passages). For example, for the question *"Zein herritan jaio zen Mikel Laboa?"* ("In which town was Mikel Laboa born?") the noun following the interrogative pronoun is “town”. In this case, we will get the semantic similarity between this noun and the candidate answers. This way we are looking for instances of the concept (“town”), such as “Donostia” or “London”, and we are excluding other kind of entities or nouns, such as “EHU” (an organization) or “research”.

3.3 Problems with WordNet in QA

Clark et al. (2008) expose very clearly the important limitations that WordNet has for supporting textual QA. One of the biggest challenge and in the same proportion important task is the recognition of textual entailment. For this task, we found different semantic knowledge requirements to take into account, such as derivational links, synonyms or world knowledge. Some of those are included in the current WordNet such as synonyms, hypernyms or relations. Other requirements are well oriented for the English WordNet such as the morphosemantic links. In our case, we should follow the path shown on Clark et al. to extend the current Basque WordNet.

Otherwise, named entities have a weighty relevance in QA systems. To treat this lack in WordNet Toral et al. (2008) present a new version of the WordNet: Named Entity WordNet (NEWN). This extension includes named entities extracted from the Wikipedia. Suchanek et al. (2007) developed the YAGO ontology, that is based on WordNet and Wikipedia. Both choices are very interesting for the work presented in this paper but both are offering a solution for the English language. NEWN and YAGO could be translated from English to Basque but as it is a very arduous task, we have found a middle way solution: we have extended the dictionary given to UKB with named entities (person names and location names). The entries added to the dictionary have been extracted from several lists of the Academy of the Basque Language (*Euskaltzaindia*[1]). More concretely, we have used toponyms, exonyms and person names in Basque. On the other hand, some named entities identified by Eihera are added to the dictionary following these steps: i) UKB always looks for the target word in the dictionary; if this word is a named entity and it is found in the dictionary, its synset number is used by UKB; ii) if the named entity does not appear in the dictionary, the synset number of the general category obtained by *Eihera* (person, organization or location) is passed to UKB; iii) if a word does not appear neither in the dictionary nor among the entities, a synset number that does not exist is given to the entity, to avoid a crash from UKB. We should continue expanding more the dictionary, for example, with named entities found in the Basque Wikipedia.

Furthermore, we think that by adding acronyms to this dictionary the accuracy of the system will improve. In addition, we will overcome one of the lacks of the Basque WordNet.

4 Experiment and Results

This section describes the results we obtained in our first evaluation of the new version of *Ihardetsi*.

4.1 Experiment

Our corpus for evaluating the system is the “Gold Standard” question bank from CLEF. Those questions were created to evaluate Basque-Basque QA systems. The question bank is composed of 500 questions and their corresponding answers, and we have filtered them just to get the question types we have defined syntactic patterns for. Thus, we only get 63 questions from 500. This subset of the original corpus has been divided into two groups, one for training, and the other one for testing: we get 39 questions for training and 24 for testing.

During the evaluation of the training question bank (39 questions), we noticed that as the question group is small, it is very difficult to draw any conclusion. In addition, some of the questions of the bank were not useful for our system for two main reasons: i) there were problems in the analysis chain, and ii) the stable version of *Ihardetsi* working with a “bag of words” technique returned no results, so the chunking could not be performed. When we start evaluating the training corpus, we realized that, before continuing with the evaluation, a deeper analysis of the re-ranking and the errors in the analysis chain was necessary. Thus, we decided not to evaluate the test corpus. The results in table 1 correspond to the training question bank.

As far as the corpus is concerned, it is composed of all the documents of the *Euskaldunon Egunkaria* (a newspaper wholly written in Basque language) from the years 2000 to 2002, with the amount of 24 million up words (in total). Additionally, the corpus has a version of the Basque Wikipedia from the year 2006 with 1.5 million up

[1]http://www.euskaltzaindia.net

words.

4.2 Results

As mentioned in section 2, *Ihardetsi* usually shows the best five candidate answers. In the evaluation we decided that it is very interesting to measure if the correct answer is close to the first position and in which position it is.

Table 1 shows the results obtained when testing the following systems: *Ihardetsi* in the original/stable mode (Ihard), and *Ihardetsi* using semantic information (Ihard+S).

	Ihard	**Ihard+S**
Correct answer (first candidate)	7	8
Answer in the first five candidates	8	5
Not in the first five candidates	8	10
Total	23	23

Table 1: Results.

As shown in the results, taking into account the few data we deal with, no general conclusions could be reached, but the overall impression is good. The correct answers improve when Ihardetsi uses semantic information.

If we go into details, it could be surprising to see that the semantic contribution obtains more correct answers out from the first five candidates. We think that this is caused by the behaviour of UKB. We think that UKB gives higher marks to the less ambiguous chunks and lower marks to those in which the words inside the chunk have a high ambiguity. Usually, in question answering systems the answers have among others named entities. In our system it is very important to analyze if they must be tagged in a general way, giving them an unique general synset (e.g. “location” for “Barcelona”) or concretely assigning them all the possible synsets they have (e.g. “Autonomous city” synset 08524735 and “Province” synset 08654360n for “Barcelona”). This decision could change the obtained results. On the other hand, the lack of named entities affects the results too. As UKB works with the dictionary extracted from the Basque Wordnet an it has very few named entities (among others the ones added using lists as described in section 3.3), we lose accuracy.

We find interesting to go step by step analyzing each phase of the linguistic analysis chain to try to understand the results. In the same way, the reader will be able to understand the problems of the system and how these barriers could be broken down in order to obtain better results. We will explain those steps with the paper-example: *“Nor izendatu zuten EEBBetako lehendakari 1944. urtean?”* (“Who was appointed president of the United States in the year 1944?”).

Our module takes two main input files: the analysis of the question and the analysis of all the candidate answer-passages. In the first step we generate a file that contains all the information we need from the question:

```
Question structure: nor VP NP ?
Answer structure: [ENTI_PER] {VP} NP
Noun phrase: eebb lehendakari 1944. urte
      (president United States year 1945)
Verb phrase : izendatu edun (appoint)

UKB format
  Noun phrase: EEBB#n#1#0
             lehendakari#n#2#0 urte#n#3#0
  Verb phrase: izendatu#v#1#0 edun#v#2#0
```

In the second step we extract the information from the candidate answers analysis and we generate a similar file with all the information about all the candidate answers. The similarity algorithm assigns weights to the chunks in the way explained in section 3.2. The semantic ambiguity of the chunks and the numbers and specifity of the synset assigned to the named entities change the results.

Otherwise, we lose some information when we translate the chunks into the UKB format. This is the case of the numbers (i.e. the year “1944”). UKB just works with nouns, verbs, adjectives and adverbs, and as there is not place for numbers, they are tagged as determiners. Although usually numbers are not important for lexical semantics, they are significant in question answering systems.

As a consequence of these problems, the results of our module are not better than the results obtained using the original *Ihardetsi*. Anyway, as we have been able to identify the problems, we will be able to correct them.

5 Conclusions and Future work

In this paper we have presented important contributions to *Ihardetsi*, a Question Answering system for Basque, adding semantic information to *Ihardetsi* by means of a chunker and an algorithm that performs semantic similarity. We have exposed the use of UKB and WordNet for QA, and we have listed some problems related to WordNet in QA, finding a middle way solution to the lack of named entities.

In order to extract more concrete conclusions about the behavior of UKB in *Ihardetsi*, we need a bigger evaluation question bank. The number of valid questions will increase improving the linguistic analysis chain. A newer version of the linguistic tools is ready to be integrated in *Ihardetsi* giving us the chance to improve the results of the system.

The similarity script output has been used as given, but we think that it is necessary to study the impact of the terms ambiguity in the chunk and perhaps these results need to be tuned. Due to the relevance of the named entities in QA systems, we need to expand the dictionary for example with named entities found in the Basque Wikipedia or using other additional ontologies.

The results obtained in the experiment give us a promising way to research. In our opinion, the use of semantic information is very interesting not only for the re-ranking of the candidate answers, but also to help in the candidate extraction task.

Acknowledgments

This research is supported by the Ministry of Science and Innovation of the Spanish Government (KNOW2 (TIN2009-14715-C04-01)) and the Basque Government (IT344-10 and Berbatek (IE09-262)).

References

Itziar Aduriz, Eneko Agirre, Izaskun Aldezabal, Iñaki Alegria, Olatz Ansa, Xabier Arregi, Mari Arriola Jose, Xabier Artola, Arantza Díaz de Ilarraza, Nerea Ezeiza, Koldo Gojenola, Montserrat Maritxalar, Maite Oronoz, Kepa Sarasola, Aitor Soroa, Ruben Urizar, and Miriam Urkia. 1998. A Framework for the Automatic Processing of Basque. In *Proceedings of Workshop on Lexical Resources for Minority Languages. First LREC Conference. Granada. 1998.*

Itziar Aduriz, Maxux Aranzabe, Jose Mari Arriola, Arantza Díaz de Ilarraza, Koldo Gojenola, Maite Oronoz, and Larraitz Uria. 2004. A Cascaded Syntactic Analyser for Basque. In *Computational Linguistics and Intelligent Text Processing. 2945 pg. 124-135. ISBN 3-540-21006-7 LNCS Series. Springer Verlag. Berlin. 2004.*

Eneko Agirre and Aitor Soroa. 2009. Personalizing PageRank for Word Sense Disambiguation. In *Proceedings of the 12th conference of the European chapter of the Association for Computational Linguistics (EACL-2009). pp. 33-41. ISBN: ISBN 978-1-932432-16-9.*

Eneko Agirre, Aitor Soroa, Enrique Alfonseca, Keith Hall, Jana Kravalova, and Marius Pasca. 2009. A Study on Similarity and Relatedness Using Distributional and WordNet-based Approaches. In *Proceedings of NAACL-HLT. Pages 19-27. Boulder, Colorado. ISBN: 978-1-932432-41-1.*

Eneko Agirre, German Rigau, Aitor Soroa, and Montse Cuadros. 2010a. Exploring Knowledge Bases for Similarity. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC10). European Language Resources Association (ELRA). ISBN: 2-9517408-6-7. Pages 373–377.*

Eneko Agirre, Aitor Soroa, and Mark Stevenson. 2010b. Graph-based Word Sense Disambiguation of Biomedical Documents. *Bioinformatics*, 26:2889–2896.

Olatz Ansa, Xabier Arregi, Arantxa Otegi, and Ander Soraluze. 2009. Ihardetsi: A Basque Question Answering System at QA@CLEF 2008. In *Evaluating Systems for Multilingual and Multimodal Information Access 9th Workshop of the Cross-Language Evaluation Forum, CLEF 2008, Aarhus, Denmark, September 17-19, 2008. Lecture Notes in Computer Science, Vol. 5706/2009, pp. 369-376. Springer Berlin / Heidelberg. ISSN 0302-9743 ISBN 978-3-642-04446-5.*

Peter Clark, Christiane Fellbaum, and Jerry Hobbs. 2008. Using and Extending WordNet to Support Question-Answering. In *Proceedings of the 4th Global WordNet Conference*, Szeged, Hungary.

Izaskun Fernandez, Iñaki Alegria, and Nerea Ezeiza. 2011. Semantic Relatedness for Named Entity Disambiguation using a small Wikipedia. In *TSD 2011: 14th Conference on Text, Speech and Dialogue*.

Pamela Forner, Anselmo Penas, Eneko Agirre, Iñaki Alegria, Corina Forascu, Nicolas Moreau, Petya Osenova, Prokopis Prokopidis, Paulo Rocha, Bogdan Sacaleanu, Richard Sutcliffe, and Erik Tjong Kim Sang. 2008. Overview of the CLEF 2008 Multilingual Question Answering Track. In *Working Notes of the Cross-Lingual Evaluation Forum, Aarhus, Denmark. ISBN 2-912335-43-4, ISSN 1818-8044*.

Llus Padró, Samuel Reese, Eneko Agirre, and Aitor Soroa. 2010. Semantic Services In Freeling 2.1: WordNet and UKB. In Piek Vossen eds. P. Bhattacharyya, Christiane Fellbaum, editor, *Principles, Construction and Application of Multilingual Wordnets: Proceedings of the 5th Global Wordnet Conference. ISBN: 978-81-8487-083-1*. Narosa publishing house.

Elisabete Pociello, Eneko Agirre, and Izaskun Aldezabal. 2010. Methodology and Construction of the Basque WordNet. In *Language Resources and Evaluation. Springer. ISSN 1574-020X.*

Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2007. YAGO: a core of semantic knowledge. In *Proceedings of the 16th international conference on World Wide Web*, pages 697–706. ACM.

Antonio Toral, Rafael, Munoz, and Monica Monanchini. 2008. Named Entity Wordnet. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC08)*, Marrakech, Morocco.