A new proposal of Word Sense Disambiguation for nouns on a Question Answering System*

S. Ferrández[1], S. Roger[1,2], A. Ferrández[1], A. Aguilar[1], and P. López-Moreno[1]

[1] Natural Language Processing and Information Systems Group
Department of Software and Computing Systems
University of Alicante, Spain
[2] Natural Language Processing Group
Department of Computing Sciences
University of Comahue, Argentina
{sferrandez,sroger,antonio}@dlsi.ua.es,
{P.Lopez}@ua.es

Abstract. This paper describes the impact of the application of a Word Sense Disambiguation (WSD) algorithm for nouns on AliQAn [16], the Question Answering system with which we have participated in the CLEF-2005. Applying the traditional WSD decreases the performance in 4.7% on the Mean Reciprocal Rank (MRR). To solve this problem, we propose two different uses of WSD: (1) to choose a set of synsets instead of the traditional use of WSD, in which only one synset is chosen; (2) to disambiguate the words not present in EuroWordNet (EWN). Using our proposal of WSD the MRR increases a 6.3% with regard to the baseline without WSD. Furthermore, our proposal of WSD increases the MRR with regard to the traditional use of WSD in an 11%. Finally, the implementation of our approach of WSD is computationally efficient by means of a preprocessing of EWN.

1 Introduction

In this paper we analyze the benefits of a Word Sense Disambiguation (WSD) algorithm for nouns in AliQAn [16], a Spanish Question Answering (QA) system, with which we have participated in the CLEF-2005[1] competition.

QA objective consists of identifying the answer of questions in large collections of documents. QA is not a simple task of Information Retrieval (IR), QA tries to go beyond and returns a concrete answer in response to an arbitrary query. For the users, it is very interesting to find accurate information, thanks to the increment of available information. The QA systems are capable to answer questions formulated by the users in natural language.

* This research has been partially funded by the Spanish Government under project CICyT number TIC2003-07158-C04-01 and by the Valencia Government under project number GV04B-268.

[1] http://www.clef-campaign.org/

© A. Gelbukh (Editor)
Advances in Natural Language Processing
Research in Computing Science 18, 2006, pp. 83-92

Current approaches to QA are mainly based on NLP tools or machine learning. There are different and possible implementations for QA systems. Generally, most of them are based on NLP tools [1, 2, 9, 13, 15], like Part of Speech (PoS) taggers, syntactic parsers, WSD, knowledge bases consisting of dictionaries, lexical-semantic data bases, ontologies and many others. Nevertheless, other systems use machine learning techniques with statistical models [5], such as Hidden Markov Models or Maximum Entropy. This is, in outline, the present situation.

The AliQAn system uses the NLP techniques. Our system has been developed during the last two years in the Department of Language Processing and Information Systems at the University of Alicante. It is based on complex pattern matching using NLP tools. Beside, WSD is applied to improve the system.

WSD algorithm is used in the phases of indexation and the search. In the first case, this algorithm allows disambiguation of the corpora words, and in the second one, it resolves ambiguities in the question words.

WSD has several critical problems. The running time of WSD algorithms makes difficult its use on huge corpora, as QA systems require. On the other hand, the low precision of WSD algorithms makes that this technique is not appropriated to be applied in QA systems. These two reasons do not allow to obtain interesting results applying WSD in real time QA systems. In order to solve these problems, we propose a concrete WSD algorithm that reduces its running time in 98.9%, due to a preprocess of EWN and improves the QA precision by means of: (1) selecting a set of synsets per word (instead of only one); (2) disambiguating words that are not presented in EWN.

The rest of the paper is organized as follows; section two describes the backgrounds of QA with regard to WSD; section three details AliQAn system with a brief description; section four explains our proposal of WSD algorithm on the AliQAn system; section five shows the evaluation results and finally, section six exposes our conclusions and discusses future works.

2 Backgrounds of QA with regard to WSD

Most of current monolingual QA systems [7, 10, 14, 18] do not apply any WSD algorithm. Nowadays, the use of WSD algorithms on QA and IR usually produce a decrease on the overall accuracy and an increase in time running.

Only in IR systems, the WSD techniques have been applied [8, 19, 17]. In the first analyzed system [8], the indexation with wordnet synsets improves the results of the IR system to 29% (from 30% up to 60%) but has the disadvantage that is carried out manually.

The project MEANING [19], has developed tools for the automatic acquisition of lexical knowledge that will help WSD. The obtained lexical knowledge is stored in the Multilingual Central Repository [4], which is based on the design of the EWN database. This implementation is based on the use of WSD with domains [12]. The problem using this technique is that the domains have to be

created in the previous phase. The observed precision in this system is 65.9%, which means an increase of 2% (Table 1, System A) with WSD.

Finally, the last system [17], uses a combination of high precision techniques and sense frequency statistics in an attempt to reduce the impact of erroneous disambiguation on retrieval performance. The incremented precision in the final result is 1.73% over 62.1% (Table 1, System B).

Table 1. Different IR systems results with the obtained improvements once WSD is applied

	System A	System B
Without WSD	65.9%	62.1%
WSD Improvement	2%	1.73%

3 The AliQAn system

AliQAn (Figure 1), is a monolingual open-domain QA system based on the intensive use of NLP tools. AliQAn has participated in the Spanish QA CLEF-2005 competition [16], in which it was ranked third.

The AliQAn architecture is divided in two main phases: Indexation phase and Search phase. In both phases, both in the document corpora and the questions, the same NLP process is applied: PoS tagging, partial parsing and WSD.

In order to make the syntactic analysis, SUPAR [6] system is used, which works in the output of MACO [1] PoS tagger. Beside, WSD is applied using EWN.

AliQAn identifies the different grammatical structures of the sentence, named syntactic blocks (SB). This is realized using the output of SUPAR, which performs partial syntactic analysis. These blocks are verb phrases (VP), nominal phrases (NP) or prepositional phrases (PP).

For example in the sentence: *"Kim Il Sung died at the age of 80"*, the obtained list of SB is: [NP,kim∗il∗sung][VP,to die][PP, at: age [PP, of: 80]].

Indexation phase, the first phase of AliQAn, consists of arranging the data where the system tries to find the answer of the questions. Two different indexation are carried out: IR-n [11] and QA indexation.

In the second phase, the search, the following three tasks are sequentially performed: question analysis, selection of relevant passages and extraction of the answer.

In the first task the system detects the type and the case of the question. The WordNet Based-Types and EWN Top-Concepts have been considered for the question type, the question case determines the set of syntactic patterns to use in the extraction of the answer.

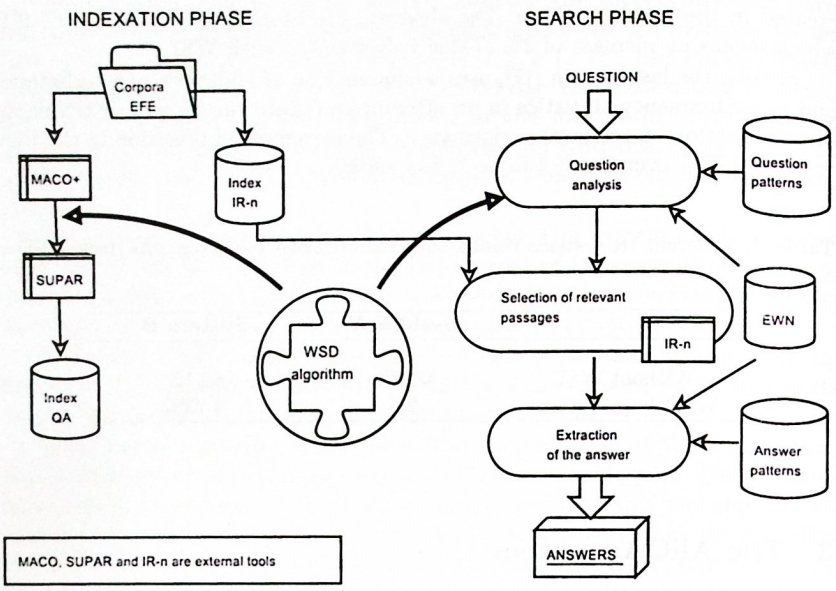

Fig. 1. AliQAn system architecture

Besides, the selection of the question terms (keywords) is carried out. These keywords are used in the next task, the selection of relevant passages, which is developed by the IR-n [11] system. IR-n returns a list of passages where the system applies the extraction of the answer, the last task of AliQAn, where it tries to extract the correct answer to the question using the syntactic patterns.

Next, an example (question 114, *In Workshop of Cross-Language Evaluation Forum (CLEF 2003)*) of resolution of one question, where the system chooses the correct solution.

- **Question:** A qué primer ministro abrió la Fiscalía de Milán un sumario por corrupción? (To whom prime minister the Office of the public prosecutor of Milan opened a summary for corruption?)
- **Type:** person
- **Case:** 3
- **List of SB:**
  - **NP1:** ([NP, primer*ministro])
  - **VP:** ([VP, abrir])
  - **NP2:**([NP, fiscalia [PP, de: milan]])([NP, sumario [PP, por: corrupcion]])
- **Text where the correct solution is:** "[...] la **Fiscalía de Milán abrió**, hoy martes, un **sumario** al **primer ministro**, *Silvio Berslusconi*, por un supuesto delito de **corrupción** [...]"
- **Text where the incorrect solution is:** "[...] **primer ministro** y líder socialista, *Bettino Craxi*, al que el pasado 21 de septiembre Paraggio **abrió** un **sumario** relacionado con el proyecto Limen por supuestos delitos de **corrupción** [...]"
- **Answer:** Silvio Berlusconi

4 The WSD algorithm

The WSD algorithm analyzed in this paper is the one proposed by Agirre and Rigau [3]. This WSD algorithm is based on the use of conceptual distance in order to try to provide a basis for determining the closeness in meaning among words, taking as reference the hierarchical structure of EWN. The conceptual distance is captured by a Conceptual Density (CD) formula. Given a concept c, the CD is calculated using the next equations:

$$CD(c,m) = \frac{\sum_{i=0}^{m-1} nhyp^i}{descendants_c} \quad (1)$$

$$descendants_c = \sum_{i=0}^{h-1} nhyp^i \quad (2)$$

where $nhyp$ is number of hyponyms, m is the number of marks of words senses and h is the height of the subhierarchy.

Unlike the algorithm proposed by Agirre and Rigau [3], our implementation selects not just one synset, but a set of the most probable synsets. This is because the algorithm attempts to discard only completely wrong synsets (keeping related synsets) in order to improve the precision of the system. Moreover, the proper nouns that are not in EWN are disambiguated too with respect to the following synsets: 05369359 (person) 07451540 (object) 08229827 01218276 (place).

The behavior of our implementation is shown in the next example:

- In the sentence
  - *El presidente de Guinea, Obiang, sugirió hoy, viernes, que su Gobierno podría rechazar la ayuda internacional* (The president of Guinea, Obiang, suggested today, Friday, that his Government could refuse the international help)
- Synsets for the word *"presidente"* sorted by [3]
  - 09400170 006140480 11176111 01090427 08956043
- Correct synset of the word *"presidente"* in this sentence
  - 00614048
- Synset returned by Agirre tool [3]
  - 09400170
- Set of synsets selected of the word *"presidente"* using our implementation
  - 09400170 00614048

In this example, the word *"Obiang"* that is not in EWN is also disambiguated. In this case, Agirre tool [3] returns the right synset: 05369359 (person).

Our proposals allow to improve the precision of our QA system as the evaluation section will show, because the WSD algorithms usually have a low precision (about 66%). It makes quite probable to discard the right synset, which is worst than keeping all the synsets of the word. In this way, our proposal only discards completely wrong synsets, which are not used when the question terms are compared with the document terms. For example, in the question "¿Cuántos muertos al año causan las minas antipersona en el mundo? (How many deaths

per year are caused by the anti-personnel mines in the world?)", the document sentence that speaks about a "coal mine" will not be processed. Moreover, our proposal overcomes the disambiguation that uses EWN domains [19] because it does not require any grouping of EWN, which could also introduce errors in the system.

The next example shows the improvement obtained when our WSD implementation is used.

- For the question:
  - *Quién es el director de la CIA?* (Who is the director of CIA?)
- Without using disambiguation, AliQAn returns:
  - *Servicio Central de Información* (Information Central Service)
  - *Servicio Central de Inteligencia* (Intelligent Central Service)
  - *Servicio Central de Información de EEUU* (Information Central Service of USA)
- Using disambiguation, AliQAn returns:
  - William Colby
  - Robert Gates
  - James Woolsey

The Agirre and Rigau [3] WSD algorithm presents a precision of 66%. Our WSD implementation of this algorithm has been evaluated in the EFE corpora (detailed in the evaluation section), where its precision stays at 60% when we select only one synset. When a 40% of the synsets of a word is selected, its precision stays at 76%. Regarding to the precision of our proposal for nouns that are not in EWN, the precision stays at 65%. Finally, the running time of the algorithm has been reduced by means of storing all the required information shown in equations (1) and (2), such as the number of descendants, for each synset in EWN. In this way, the running time has been reduced in a 98.9% (from 1400 seconds to 15 seconds in disambiguating 9 files).

5 Evaluation

5.1 Dataset

The experiments described in this section have been carried out using the AliQAn system with CLEF 2003 questions and corpora, i.e. Spanish corpora EFE 1994. The collection, which was indexed by the IR-n system, contains approximately 215.738 documents for a total of 509 Megabytes.

The our test set has been extracted from the CLEF 2003 competition task, and it includes a total of 200 questions. The type of questions capable to contain noun as answer are: definition, abbreviation, event and person. This corresponds to the 65% of all questions and the 35% remaining corresponds to the following type of questions: date, month, percentage, object, quantity, economic, age, measure, period, year. Finally, the average ambiguity factor for each term in EuroWordNet 1.6 is ZZZ

5.2 Evaluation Measures

In order to be able to evaluate the system we need a measure that values the general results of the system. Mean Reciprocal Rank (MRR) is the measure used in the CLEF 2003 campaign for evaluating the systems.

$$MRR = (\sum_{i=1}^{Q} \frac{1}{far(i)})/Q$$

where Q is the number of questions (200 in our case) and $far(i)$ indicates the position of the first correct answer. The value of $1/far(i)$ will be 0 if the system has not found the answer.

5.3 Results analysis

The architecture and behavior of our system have been described in previous sections. Now we are going to present the obtained results and the study about the performance of the system when WSD is used in the corpora and questions.

The comparison will be performed using the AliQAn system with three algorithms, which have different WSD levels. These algorithms are shown in Table 2.

1. **Baseline without WSD:** (first column of the Table 2) The baseline does not have WSD algorithm, i.e. all the synsets per word are used. This base system has got a MRR of 44.5%.
2. **1-Sense WSD algorithm:** (second column of the Table 2). In this case, the WSD algorithm used is the propose by Agirre et al. [3]. This algorithm chooses one synset for disambiguating a word and the MRR obtained is 42.4%.
3. **Our proposal of WSD algorithm:** (third column of the Table 2). In this case, the set of most probable synsets is chosen. The selected synsets are the 40% of total word synsets. The system has got a MRR of 47.3%.

Table 2. Results using different applications of WSD on 200 questions of CLEF 2003

	1 Baseline without WSD	2 1-Sense WSD	3 Our proposal of WSD
MRR	44.5	42.4 (-4.75%)	47.3 (+6.3%)
% of First Correct Answer (FCA)	39	37.5 (-3.85%)	42.5 (+8.97%)
FCA Improvement (200 questions)		3 questions	7 questions
		+11% of improvement in the MRR with regard to the 1-Sense WSD	

Therefore, our WSD proposed has increased the MRR compared to the 1-sense WSD algorithm up to 11% and a 6.3% regarding the baseline (see Figure 2), it increases the number of queries answered in first place in 7.

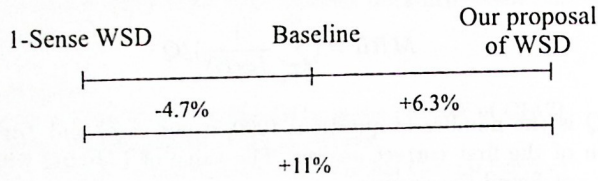

Fig. 2. Relationship with regard to percentages of the WSD system presented in Table 2

In comparison with the 42.4 of the 1-sense WSD algorithm, the percentage of increment in the MRR is 8.3% when a 10% of the synsets is selected. It is 9.2% when a 20% is selected. It is 10.4% when a 30% is selected. Finally, it is 11% when a 40% is selected.

We have carried out an analysis of the number of the disambiguated nouns in the 200 questions (159,778), where 46,194 nouns have only one synset in WordNet, and the remaining nouns (113,584) have 3.9 synsets in WordNet on average. Our proposal of WSD selects 1.9 synsets on average (in case of draw in the score between 2 synsets, both are selected). Moreover, 44,247 nouns that are not in WordNet have been disambiguated.

On the other hand, normally the time of carrying out the corpus processing WSD is too long or at worst NP complete. Many researchers have been confronted with this problem when they apply their approaches of WSD, for this reason they only apply WSD to questions. With our WSD algorithm, we have achieved to reduce considerably the running time required. Initially, we needed 1400 seconds for the processing of 9 files, now we do it within 15 seconds, which supposes a decrease of 98.9%.

6 Conclusions and future works

In this paper we propose an algorithm that aims at enhancing WSD for nouns on a QA system. This algorithm is based on the algorithm proposed by Agirre et al. [3]. The difference is that while it considers one synset for disambiguating a word, our proposal selects the most relevant synsets and adds the disambiguation of the proper nouns that are not included in EWN 1.6. In order to evaluate our algorithm, a number of comparisons have been carried out. Results confirm the viability of our algorithm, showing an improvement up to 11% over the traditional WSD algorithm and the 6.3% over the baseline. Furthermore, we have greatly reduced the computational cost for WSD process by 98.9% by means of a

preprocessing of EWN. It is important to emphasize that the algorithm proposed does not require any previous grouping (such as domains) for the disambiguation process, which may involve some error.

The contribution of our paper to the WSD research area is that the traditional 1-sense WSD algorithms do not improve QA, as it is stated in Table 2 and Figure 2 (-4.7% in the MRR). That is because of their low precision. Our proposal allows selecting a percentage of synsets instead of only one, as the traditional 1-sense WSD algorithms do. In this way, the MRR of a 1-sense WSD algorithm is improved in 11%, and it improves the MRR of a QA system without WSD in 6.3%. Moreover, our proposal overcomes the traditional drawback of WSD that is its high computational cost, which makes too difficult its application to huge corpora. Our approach reduces the running time of the WSD algorithm in a 98.9%.

The results are promising. Therefore we expect to analyze the results on CLEF 2004 and CLEF 2005. In the future, we are going to develop other WSD algorithm to prove that our proposal is validated independently of the WSD algorithm itself.

References

1. S. Acebo, A. Ageno, S. Climent, J. Farreres, L. Padró, R. Placer, H. Rodriguez, M. Taulé, and J. Turno. MACO: Morphological Analyzer Corpus-Oriented. *ESPRIT BRA-7315 Aquilex II, Working Paper 31*, 1994.
2. A. Ageno, D. Ferrés, E. González, S. Kanaan, H. Rodríguez, M. Surdeanu, and J. Turmo. TALP-QA System for Spanish at CLEF-2004. *In Workshop of Cross-Language Evaluation Forum (CLEF)*, pages 425 – 434, 2004.
3. E. Agirre and G. Rigau. A proposal for word sense disambiguation using conceptual distance. *1st Intl. Conf. on recent Advances in NLP. Bulgaria*, 1995.
4. J. Atserias, L. Villarejo, G. Rigau, E. Agirre, J. Carrol, and B. Magnini P. Vossen. The meaning multilingual central repository. *In Proceeding of the Second International WordNet Conference-GWC.*, pages 23–30, 2004.
5. C. de Pablo, J.L. Martínez-Fernández, P. Martínez, J. Villena, A.M. García-Serrano, J.M. Goñi, and J.C. González. miraQA: Initial experiments in Question Answering. *In Workshop of Cross-Language Evaluation Forum (CLEF)*, pages 371 – 376, 2004.
6. A. Ferrández, M. Palomar, and L. Moreno. An Empirical Approach to Spanish Anaphora Resolution. *Machine Translation. Special Issue on Anaphora Resolution In Machine Translation*, 14(3/4):191–216, December 1999.
7. J. M. Gómez, E. Bisbal, D. Buscaldi, and P. Rosso E. Sanchis. Monolingul and Cross-Language QA using a QA-oriented Passage Retrieval System. *In Workshop of Cross-Language Evaluation Forum (CLEF)*, September 2005.
8. J. Gonzalo, F. Verdejo, I. Chugur, and J. Cigarrán. Indexing with wordnet synsets can improve text retrieval. *In the Proceedings of the ACL/COLING Workshop on Usage of WordNet for Natural Language Processing*, pages 38–44, 1998.
9. J. Herrera, A. Peñas, and F. Verdejo. Question Answering Pilot Task at CLEF 2004. *In Workshop of Cross-Language Evaluation Forum (CLEF)*, pages 445 – 452, 2004.

10. D. Laurent, P. Séguéla, and S. Negre. Cross Lingual Question Answering using QRISTAL for CLEF 2005. *In Workshop of Cross-Language Evaluation Forum (CLEF)*, September 2005.
11. F. Llopis and J.L. Vicedo. Ir-n, a passage retrieval system. *In Workshop of Cross-Language Evaluation Forum (CLEF)*, 2001.
12. Bernardo Magnini, Carlo Strapparava, Giovanni Pezzulo, and Alfio Gliozzo. The role of domain information in word sense disambiguation. *Nat. Lang. Eng.*, 8(4):359–373, 2002.
13. E. Méndez-Díaz, J. Vilares-Ferro, and D. Cabrero-Souto. COLE at CLEF 2004: Rapid prototyping of a QA system for Spanish. *In Workshop of Cross-Language Evaluation Forum (CLEF)*, pages 413 – 418, 2004.
14. G. Neumann and B. Sacaleanu. DFKI's LT-lab at the CLEF 2005 Multiple Language Question Answering Track. *In Workshop of Cross-Language Evaluation Forum (CLEF)*, September 2005.
15. M. Pérez-Coutiño, T. Solorio, M. Montes y Gómez, A. López-López, and L. Villaseñor-Pineda. The Use of Lexical Context in Question Answering for Spanish. *In Workshop of Cross-Language Evaluation Forum (CLEF)*, pages 377 – 384, 2004.
16. S. Roger, S. Ferrández, A. Ferrández, J. Peral, F. Llopis, A. Aguilar, and D. Tomás. AliQAn, Spanish QA System at CLEF-2005. *In Workshop of Cross-Language Evaluation Forum (CLEF)*, 2005.
17. Christopher Stokoe, Michael P. Oakes, and John Tait. Word sense disambiguation in information retrieval revisited. In *SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 159–166, New York, NY, USA, 2003. ACM Press.
18. R. F. E. Sutcliffe, M. Mulcahy, and I. Gabbay. Cross-Language French-English Question Answering Using the DLT system at CLEF 2005. *In Workshop of Cross-Language Evaluation Forum (CLEF)*, September 2005.
19. P. Vossen, G. Rigau, I. Alegria, E. Agirre, D. Farwell, and M. Fuentes. Meaningful results for information retrieval in the meaning project. *Third International WordNet Conference*, 2006.