Lexical ambiguity in cross-language image retrieval: a preliminary analysis.

Borja Navarro-Colorado, Marcel Puchol-Blasco, Rafael M. Terol,
Sonia Vázquez and Elena Lloret.
Natural Language Processing Research Group (GPLSI)
Department of Software and Computing Systems
University of Alicante.
{borja,marcel,rafamt,svazquez,elloret}@dlsi.ua.es

Abstract

In this paper we calculate and analyse the lexical ambiguity of queries in a cross-lingual Image Retrieval (Flickling) and compare it with the results obtained by users. We want to know to what extent the lexical ambiguity of a query influences the correct localization of an image in a multilingual framework. With this, our final objective is to determine the necessity of Word Sense Disambiguation systems in Information Retrieval tasks, according to user behaviour. The results show that users really deal with lexical ambiguity: in general terms, users find the correct image with low ambiguity queries.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: .3.1 - Content Analysis and Indexing; H.3.3 Information Search and Retrieval; I.2 [**Artificial Intelligence**]: I.2.7 - Natural Language Processing

General Terms

Human Factors, Measurement, Performance, Experimentation

Keywords

Lexical ambiguity, Interactive Image Retrieval, Word Sense Disambiguation

1 Introduction

In this paper[1] we present a preliminary study about how users deal with lexical ambiguity during the interaction with an Information Retrieval system. Taking advantage of Flickling system[2], we have analysed user behaviour with lexical ambiguity in a multilingual image retrieval task.

The Natural Language Processing community has not reached an agreement as to whether Word Sense Disambiguation (WSD) systems are useful or not in Information Retrieval (IR) tasks. In other words, there is not an agreement about IR systems improving the retrieval process with a WSD system.

---

[2]http://cabrillo.lsi.uned.es/flickling

In general terms, on the one hand, papers like [14, 15] said that a WSD system really does not improve the text retrieval process. On the other hand, papers like [5, 8] said that indexing the text collection with senses, the retrieval process improves. During the CLEF 2008 some papers argued about this topic. For example, [9] said that only in most specific cases, would a WSD system improve the success of IR system; while [6, 7, 10, 1] said that there is no improvement with WSD systems. However, papers like [11, 12, 3] prove that there is a real improvement on Information Retrieval with WSD systems. At present, it seems that only for ambiguous words whose senses do not have semantic relations among them (homographs), a WSD system is useful in IR tasks.

There are different approaches to WSD, with different disambiguation algorithms and lexical resources, and different problems [2]. However, all of them introduce errors and are time-consuming in an IR system, therefore nowadays it is difficult to know to what extent they could be useful in IR tasks.

In our opinion, the key question is to know if lexical ambiguity is really a decisive factor in IR tasks for a user: if it is, WSD systems are necessary in IR. If not, they are unnecessary.

Following Robins [13], in order to design an effective IR system, it is necessary to know how users interact with these systems. Taking advantage of iCLEF 2009 task, we have analysed the lexical ambiguity of the queries used in Flickling system by different users, and we have compared it with the results obtained by them. With this, we have extracted interesting data of user behaviour about lexical ambiguity on an IR task that could be useful to throw light on this question.

In the next section, we introduce the mathematical formula created to calculate the lexical ambiguity of a query. Then we show some data extracted from the iclef log: the lexical ambiguity and the image retrieved by the system, the lexical ambiguity and the images correctly found by the users, and the lexical ambiguity of a specific user. Finally we extract some preliminary results and future work.

2 The lexical ambiguity of the queries

In order to achieve our objective, we need to know the general ambiguity of each query, and then compare it with the success of the query. That is, according to the general ambiguity of a query, we want to know if the image has been located or not, and also if both aspects are related.

In this sense, we must represent each query with a value that shows the general ambiguity of this query.

On the one hand, the ambiguity of each word is the number of senses that it has in WordNet [4] or EuroWordNet [16]. On the other hand, a query is made up by more than one word. This set of words has a relation of co-occurrence (for the user, the image will probably be annotated with these words in the title, description or tags) and the senses of each word have influence on one another. With this frame, we have developed a formula to calculate the general ambiguity of a query with two variables: the number of words in the query and the number of senses of each word.

$$\alpha = \left(1 - \frac{1}{\prod s}\right) * \frac{1}{n}$$

where α represents the general ambiguity of the query, s the number of senses for each word in WordNet or EuroWordNet, and n the number of words in the query.

Next, we are going to split the formula into different parts, in order to explain each in more detail. The ambiguity of one word is uniquely related to the number of associated senses to it in the WordNet or EuroWordNet lexical resource, but the ambiguity of a query is not the sum of the word senses. The user introduces a word in a query with a specific sense. Therefore, we take into account the probability of each sense for each word. The formula $\frac{1}{s}$ calculates the ambiguity of the word: the probability that the user is using only one sense for a word. The scores obtained by this formula are in the range between 0 (upper ambiguity) and 1 (lower ambiguity).

Moreover, considering that a query is composed of a set of n words, the ambiguity of a query will be associated to the ambiguity of its n related words. The formula that computes this ambiguity

is defined such as $\frac{1}{s_1} * \frac{1}{s_2} * ... * \frac{1}{s_n}$ that simplified is defined as such $\frac{1}{\prod_{i=1}^{n} s_i}$. Finally, with the aim of considering the opposite situation (values near zero indicate lower ambiguity while values near one indicate upper ambiguity), we introduce the opposite element '1' in the formula being $1 - \frac{1}{\prod_{i=1}^{n} s_i}$ the final formula that computes the ambiguity of each query.

The second part of the formula ($\frac{1}{n}$) is the reciprocal of the number of words in the query. The number of words in a query has a specific weight during the retrieval process and in the general ambiguity of the query too. The words that make up a query are the context in which each ambiguous word is disambiguated. That is, for a user, each word has a specific sense in this specific context (the rest of words in the query). Therefore, the number of words in the context influences the ambiguity of the query.

From a semantic point of view, the fact that different languages are used in the same query is a method of lexical disambiguation. In this sense, users employ translation to deal with the semantic ambiguity. The same word in different languages could have a different amount of senses (for example, the Spanish word "jugar" -5 senses- and the English translation "to play" -29 senses-). In these cases, we consider this set of words (that is, a word and its translation to another language) as the same word. Considering that users use translation to reduce the ambiguity, we can conclude that the ambiguity of these complex words are the intersection of their senses, that is, we consider the language with less senses for the same word. In the previous example, we considered that the ambiguity of "jugar - to play" is the amount of senses of only "jugar": 5 senses.

There is a general problem with words that do not appear in WordNet or EuroWordNet. In these cases, we assume that that word is either a proper noun or a technical term. In both cases, we consider that these words do not have ambiguity, and assume only one sense for each[3].

In a nutshell, from our point of view, a query with only one word and five senses has more ambiguity that a query with five words each with one sense. In each of these queries, the number of senses is the same, but not the number of words. This formula tries to represent this fact.

3 Log analysis and results

For this paper we take into account only three languages: English, Spanish and Italian[4]. We have calculated the general lexical ambiguity of each query with the previous formula, and we have extracted some data related to the success of these queries. Specifically, we have extracted data about these three aspects:

- lexical ambiguity of the queries and images retrieved by the system,

- lexical ambiguity of the queries and images found by the users,

- lexical ambiguity of the queries and images found by a specific user.

Let's see the data extracted for each of these aspects.

3.1 Lexical ambiguity of the queries and images retrieved by the system

The average number of images returned for a specific ambiguity in a question has been extracted. Graph 3.1 shows the average of images retrieved by the system according to the level of lexical ambiguity within the queries.

---

[3]Another possibility is that a word does not appear in the lexicon due to a typographic error and spelling mistake by the user. This case is not taken into account. However, as we will show later, this assumption has introduced some errors in the results.

[4]Portuguese and Dutch have not been taken into account. The ambiguity associated with words of these languages will be the minimum number of senses where we have translation. Queries with no results are not relevant for our experiments because the words contained in the queries usually are words with spelling mistakes by the users.

These data show that the amount of image retrieved by the system increases according to the level of the queries' lexical ambiguity. It is meaningful the increment of images retrieved from an ambiguity level of 0.5 to 0.9. With these data, we can say that the system has more precision with low ambiguity queries.

However, it is curious the amount of retrieved images with non-ambiguous queries (0). It is higher than the images retrieved with ambiguity 0.1 or 0.2. This is an exception to the general behaviour of data.

The reason for this exception is the words that do not appear in the lexicon. As we said before, the words that do not appear in the lexicon are considered non-ambiguous words (proper nouns or technical terms). However, it has been a hard decision due to, on one hand, proper nouns and technical terms possibly being ambiguous; and, on the other hand, there are ambiguous words that simply do not appear in the lexicon and words written in other languages. In all these cases, we have considered these words as non-ambiguous words, and, in many times, it is not true.

3.1. Ambiguity and average of image retrieved
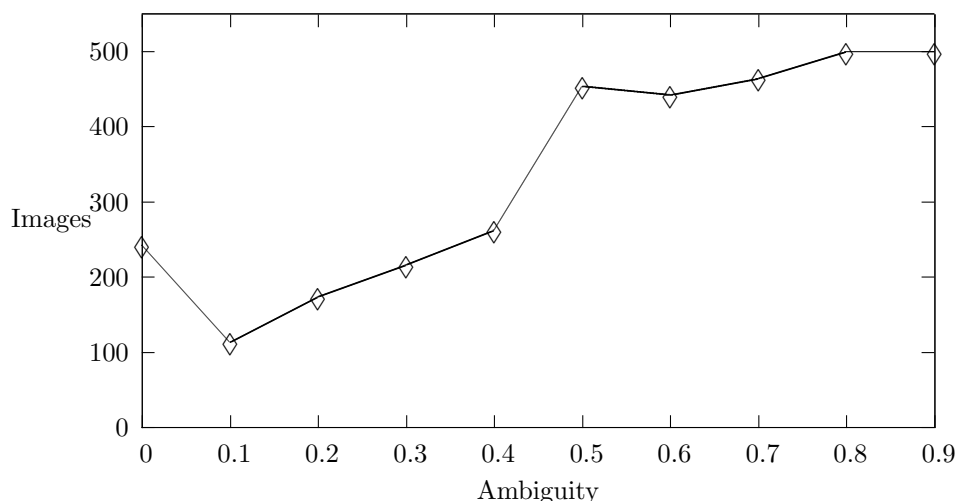
This is the reason why the amount of queries with no ambiguity is high. However, not all these queries really have no ambiguity. It must be also taken into account that, in a multilingual framework, we depend totally on the completeness of each lexicon. This aspect of the formula must be reviewed.

Independently of this fact, the data show a clear increase in the image retrieved according to the increment of the queries ambiguity. With the maximum ambiguity of the queries (0.8 and 0.9), the system retrieves the maximum amount of images allowed (500). With this we can make a preliminary conclusion: with more lexical ambiguity in the query, a higher number of images are retrieved and the system has less precision. In this sense, a WSD system could be useful for IR tasks.

## 3.2 Ambiguity of the queries and images found by the users

Table 1 and graphs 3.2 and 3.2b show the average of images located correctly by all users according to the queries' lexical ambiguity. Images not located by the users are included too.

To clarify the data, we have divided them into two graphs. The first one shows the data with queries with 0 ambiguity up to 0.5 ambiguity. The second one, from 0.5 ambiguity to 0.9.

Ambiguity	Images not found	Images found	Total
0	323 (23.59%)	1046 (74.4%)	1369
0.1	268 (21.77%)	963 (78.22%)	1231
0.2	307 (20.1%)	1220 (79.89%)	1527
0.3	155 (24.56%)	476 (75.43%)	631
0.4	51 (16.94%)	250 (83%)	301
0.5	45 (54.21%)	38 (45.78%)	83
0.6	18 (64.28%)	10 (35.71%)	28
0.7	9 (100%)	0 (0%)	9
0.8	7 (50%)	7 (50%)	14
0.9	1 (50%)	1 (50%)	2
Total	1184 (22.8%)	4011 (77.2%)	5195

Table 1: Ambiguity and image found and not found

3.2. Ambiguity and images found
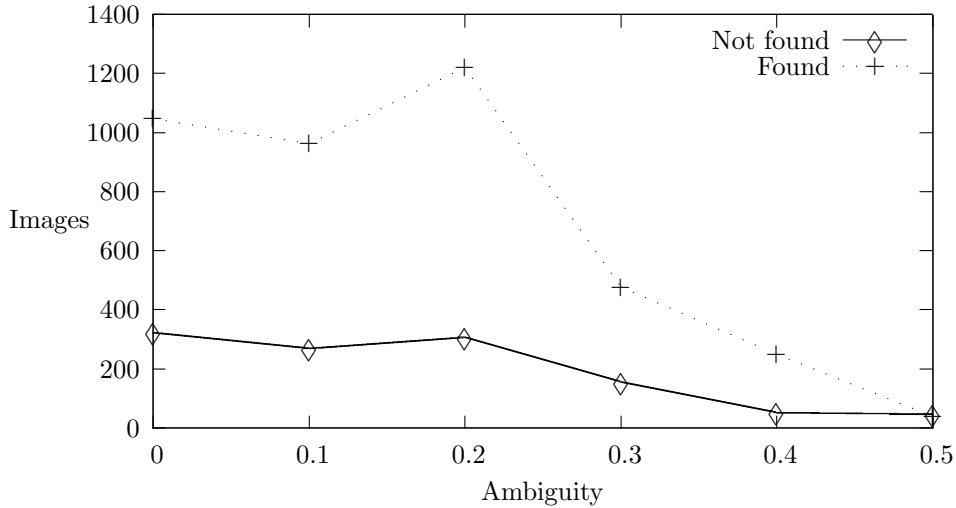
These data clearly show that with low ambiguous queries, the users correctly find more much images than with queries with high ambiguity: 71.3% of the images have been correctly located with queries with low ambiguity (between 0 to 0.3). However, an unambiguous query doesn't always mean that the user correctly finds an image, or an ambiguous query signify that the user doesn't find the image. For example, as shown the table, with queries with ambiguity 0, a 74.4% of the images are correctly located and a 23.5% are not located.

For queries with high ambiguity (second graph) the situation is different. Actually, these data do not show relevant behaviour: the amount of images located and not located by the user are more or less the same (approximately 50%), but the total amount of images located and not located by the users with higher ambiguity queries is low.

With these data we can achieve interesting conclusions. The great majority of images correctly located has queries with low ambiguity: 1046 images for queries with ambiguity 0; 963 images for queries with ambiguity 0.1, or 1220 images for queries with ambiguity 0.2, from a total of 4011. 92.37% of the images correctly located have queries with low ambiguity (between 0 and 0.3). However, the 88.93% of the images not located have queries with the same level of ambiguity.

The reasons why an image is not located could be different: for example, cases where the user does not know the language of the image or the correct translation, or cases where the user does not find the correct words.

However, we notice that, in the majority of cases, when a user correctly find an image, the ambiguity of the query is low. From the total amount of images, 71.3% have been correctly located with queries with low ambiguity (between 0 to 0.3), against the 20.26% of images not located with queries with the same level of ambiguity.

3.2b. Ambiguity and images found
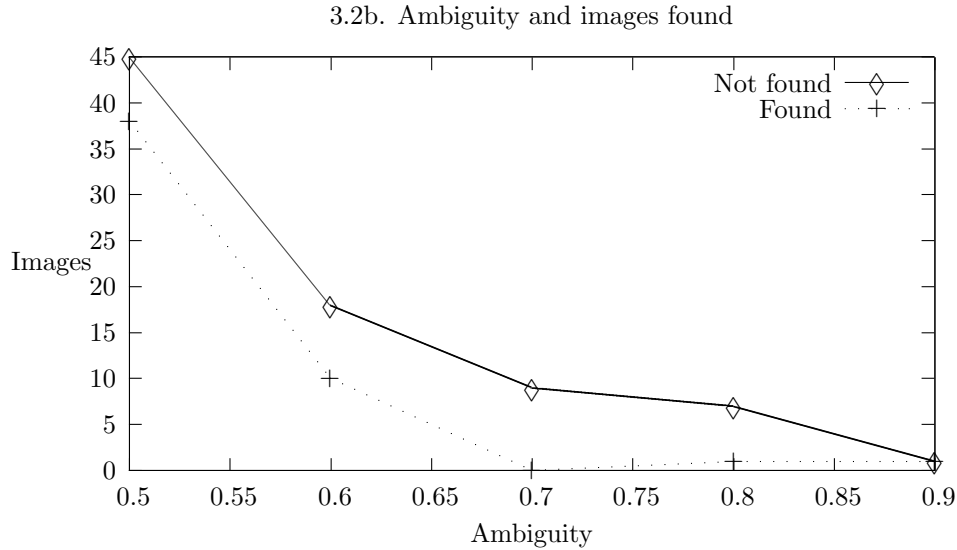
According to these data we can conclude that, in the majority of cases, the query's lexical ambiguity influences the precision of the user finding images with a Multilingual Image Retrieval system. However, at the same time, due to the users correctly finding some images with queries with high ambiguity and they can not find correctly some images with queries with low ambiguity, our conclusion can not be categorical: the ambiguity of the query is an important (and maybe decisive) factor, but it is not the unique factor for a correct image retrieval by a user.

In this sense, the use of a WSD system in order to improve the IR tasks is important, and maybe decisive. In any case, it is interesting that, according to these data, it is not necessary to disambiguate with one sense per word. Up a 0.3 level of ambiguity is acceptable for a user to find images correctly. Therefore, a coarse-grained WSD system[5] could be useful for IR tasks.

## 3.3 Ambiguity of the queries and images found by a specific user

Another interesting fact is how a specific user deals with the lexical ambiguity of the queries. In order to show this, we have extracted the lexical ambiguity of the queries from the best user (the user that has located more images correctly). They are shown in graph 3.3 and table 2. We have taken into account only the last query: the query in which the user has found the image.

---

[5]That is, WSD systems that could disambiguate with a low level of granularity, with more than one related sense per word, or with general semantic categories.

Ambiguity of the query	Amount of images found
0	190
0.14	1
0.16	22
0.175	1
0.178	2
0.183	1
0.1875	1
0.194	1
0.19	1
0.2	1
0.22	4
0.229	1
0.237	2
0.245	1
0.248	1
0.25	42
0.267	3
0.278	3
0.3	5
0.305	1
0.32	1
0.33	1
0.33	12
0.37	1
0.4	5
0.416	9
0.428	1
0.43	3
0.44	6
0.45	1
0.464	1
0.5	36
0.67	23
0.75	9
0.8	3
Total	396

Table 2: Ambiguity for the best user

3.3. Ambiguity for the best user
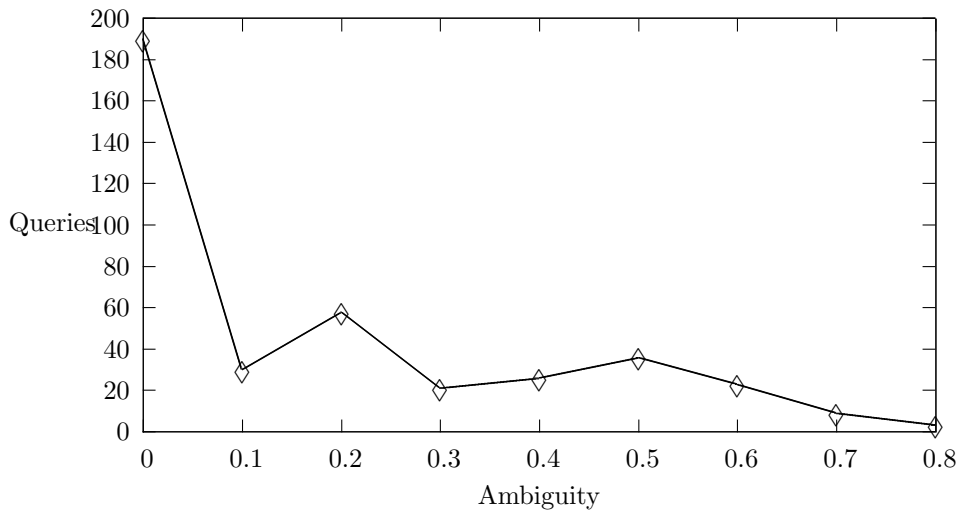
In this case, the data are clear: the majority of images has been correctly found with non-ambiguous queries (47.9% - 190 images). With low ambiguity (between 0.1 and 0.3) the 27.5% (109 images) has been correctly found. However, 17.9% (71 images) of the images have been found with high ambiguity queries (between 0.5 and 0.9).

The behaviour of data with a specific user is similar to the average of users: the majority of images has been found correctly with low ambiguity queries (75.4% for queries with ambiguity between 0 and 0.3). However, the user can find images correctly with high ambiguous queries (17.9%). Therefore, the conclusion is the same. For this user, according to these data, the lexical ambiguity of the query is an important factor: with less ambiguous queries, the user can find more images correctly. However, it is not the unique factor, because there are some cases in which the user can find images correctly with high ambiguous queries.

4   Conclusions and future work

According to all these data, we assume the following preliminary conclusions:

- In general terms, the lexical ambiguity of the queries influences the precision of a user finding images: with less ambiguous queries, the user can find more images correctly.

- The user can find images correctly with low ambiguous queries: ambiguity between 0 and 0.3. For a human, it is not necessary to have a fine grain disambiguation: with an ambiguity of 0 to 0.3, he can find the majority of images correctly.

- The user tries to make up the least ambiguous query to improve the retrieval task.

- Although the lexical ambiguity of the query is an important factor to find images correctly, it is not the unique nor absolute factor: they can find some images correctly with ambiguous queries (on average, less than with low ambiguous queries).

Extrapolating these results to the general case about the usefulness of WSD systems in IR tasks, we think that it is necessary to apply these kind of systems to improve the precision of the IR systems. However, it is not necessary to have a fine-grain disambiguation. Otherwise, coarse-grained disambiguation could be useful and enough. For example, to disambiguate with more than one related sense (in WordNet), to use lexical resources with low ambiguity, to disambiguate with semantic classes, generic concepts or domains, etc.

As a future work, there are some aspects that must be extensively developed:

- The formula must be improved in order to correctly represent words that do not appear in the lexical resource. Some options are: to use more than one lexicon (general and technical) or to apply Named Entities Recognition systems to the query.

- In order to prove that coarse-grained disambiguation is enough for an IR task, it is necessary to repeat the experiment with other lexical resources with less granularity than WordNet, or to reduce the granularity of WordNet with some well-known technique (domains, clustering of related senses, semantic classes, etc.).

- Finally, all these data have been extracted from an Image Retrieval system. It is necessary to complete the analysis with a more generic Text Retrieval system.

References

[1] Otavio Costa Acosta, André Pinto Geraldo, Viviane Moreira Orengo, and Aline Villavicencio. UFRGS@CLEF2008: Indexing Multiword Expressions for Information Retrieval. In *Working Notes for the CLEF 2008 Workshop*, 2008.

[2] Eneko Agirre and Philip Glenny Edmonds, editors. *Word Sense Disambiguation: algorithms and applications.* Springer, 2006.

[3] Pierpaolo Basile, Annalina Caputo, and Giovanni Semeraro. UNIBA-SENSE at CLEF 2008: SEmantic N-levels Search Engine. In *Working Notes for the CLEF 2008 Workshop*, 2008.

[4] Christiane Fellbaum, editor. *WordNet. An Electronic Lexical Database.* MIT Press, 1998.

[5] Julio Gonzalo, Felisa Verdejo, Irina Chugur, and Juan M. Cigarrán. Indexing with WordNet synsets can improve Text Retrieval. In *Usage of WordNet in Natural Language Processing Systems. Coling-ACL Workshop.*, 1998.

[6] Jacques Guyot, Gilles Falquet, Saïd Radhouani, and Karim Benzineb. UNIGE Experiments on Robust Word Sense Disambiguation. In *Working Notes for the CLEF 2008 Workshop*, 2008.

[7] Andreas Juffinger, Roman Kern, and Michael Granitzer. Exploiting Co-occurrence on Corpus and Document Level for Fair Cross-language Retrieval. In *Working Notes for the CLEF 2008 Workshop*, 2008.

[8] Robert Krovetz. On the Importance of Word Sense Disambiguation for Information Retrieval. In *Creating and Using Semantics for Information Retrieval and Filtering. State of the Art and Future Research. Third International Conference on Language Resources and Evaluation (LREC) workshop*, 2002.

[9] Fernando Martínez-Santiago, José M. Perea-Ortega, and Miguel A. García-Cumbreras. SINAI at Robust WSD Task @ CLEF 2008: When WSD is a Good Idea for Information Retrieval tasks? In *Working Notes for the CLEF 2008 Workshop*, 2008.

[10] Sergio Navarro, Fernando Llopis, and Rafael Muñoz. IRn in the CLEF Robust WSD Task 2008. In *Working Notes for the CLEF 2008 Workshop*, 2008.

[11] Arantxa Otegi, Eneko Agirre, and German Rigau. IXA at CLEF 2008 Robust-WSD Task: using Word Sense Disambiguation for (Cross Lingual) Information Retrieval. In *Working Notes for the CLEF 2008 Workshop*, 2008.

[12] José R. Pérez-Agüera and Hugo Zaragoza. UCM-Y!R at CLEF 2008 Robust and WSD tasks. In *Working Notes for the CLEF 2008 Workshop*, 2008.

[13] D. Robins. Interactive Information Retrieval: contexts and basic notions. *Informing Science*, 3(2):51–61, 2000.

[14] Mark Sanderson. Word Sense Disambiguation and Information Retrieval. In *Proceedings of the 17th ACM SIGIR Conference*, pages 142–151, 1994.

[15] Ellen M. Voorhees. Using WordNet to Disambiguate Word Senses for Text Retrieval. In *Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 171–180, 1993.

[16] Piek Vossen, editor. *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. Kluwer Academic Publishers, 1998.