

Mutual bilingual terminology extraction

Le An Ha*, Gabriela Fernandez**, Ruslan Mitkov*, Gloria Corpas***

* University of Wolverhampton

** Universidad de Sevilla

*** Universidad de Malaga

E-mail: *{l.a.ha,r.mitkov}@wlv.ac.uk, **gfernan@us.es, ***gcorpas@ya.com

Abstract

This paper describes a novel methodology to perform bilingual terminology extraction, in which automatic alignment is used to improve the performance of terminology extraction for each language. The strengths of monolingual terminology extraction for each language are exploited to improve the performance of terminology extraction in the other language, thanks to the availability of a sentence-level aligned bilingual corpus, and an automatic noun phrase alignment mechanism. The experiment indicates that weaknesses in monolingual terminology extraction due to the limitation of resources in certain languages can be overcome by using another language which has no such limitation.

1. Introduction

The identification of terms in scientific and technical documents is a crucial issue for any application dealing with analysis, understanding, generation, and translation of such documents. Throughout the last decade, computational linguists, translators, lexicographers, and computer engineers among other specialists have been interested in automatically identifying terminology in texts. Software tools to accomplish terminological related tasks have been designed and implemented. There is also a lot of interest in bilingual terminology extraction (BLTE) (detailed in Section 2). The usual approach to BLTE is monolingual terminology extraction followed by automatic alignment. In other words, automatic alignment is the final, independent step used only to align two monolingual term candidate lists.

This paper proposes a novel approach to BLTE, where alignment plays an active role in automatic terminology extraction. Instead of performing monolingual terminology extraction for each language, we first perform monolingual terminology extraction for one language, and then align these term candidates to the other language's candidates. A term candidate in the second language is promoted if it is aligned with a term candidate in the first language. In this approach, weaknesses due to the lack of resources in the second language can be overcome by the alignment process. For example, the amount of texts available in English is much greater than the number available for Spanish, and we can exploit this fact by first extracting terms in English using sophisticated statistical measures, and then aligning them with the Spanish term candidates. By doing this, terms in Spanish can be extracted without relying on the availability of large corpora which may not be available for that language.

The proposed method relies on the automatic alignment of

term candidates. The availability of parallel texts on the Internet makes this possible. In the experiment, we rely on a parallel corpus collected from the MedlinePlus website (the medical corpus). With the help of Trados WinAlign, it is straightforward to align the parallel texts at sentence level¹.

2. Mono and multilingual terminology extraction

2.1 Monolingual terminology extraction

The main stages in terminology work can be summarised as: extraction of term candidates from a corpus, validation of the term candidates found, and organisation of validated terms by domain and sub-domain (Sauron, 2002). In this respect, a number of projects have been able to create automatic extraction tools, which identify term candidates from a corpus in electronic form. Some projects go one step further: on the basis of parallel corpora of texts and their translations they propose not only candidate terms but also possible equivalents in a target language.

Approaches to term extraction (TE) are usually classified as linguistic, statistical, or hybrid. Linguistic and statistical approaches can be further subdivided into term-based (intrinsic) and context-based (extrinsic) methods (cf. Bourigault et al., 2001; Streiter et al., 2003). Terminology Extraction tools (TETs) following a linguistic approach try to identify terms by their linguistic (morphological and syntactic) structure. For this purpose, texts are annotated with linguistic information with the help of morphological analysers, part-of-speech taggers and parsers. Then, term candidates (TCs) following certain syntactic structures are filtered from the annotated text by using pattern matching techniques. Intrinsic methods try to filter TCs according to their internal (i.e. morphological) structures (Ananiadou 1994). Extrinsic

¹ It should be noted that other tools such as ParaConc and Déjà Vu can also be used for this task.

methods, on the other hand, try to identify TCs by analysing the syntactic structure of a word or phrase, such as looking for part-of-speech sequences like NP= noun + noun (e.g. computer science). An example of this technique is represented by the program LEXTER (Bourigault, 1992). Another commonly used technique is to filter TCs by looking for commonly used text structures such as definitions and explanatory contexts like “X is defined as ...” or “X is composed of ...” (cf. Pearson, 1998).

The general assumption underlying the statistical approaches to TE is that specialised documents are characterised by the repeated use of certain lexical units or morpho-syntactic constructions. TETs based on statistics try to filter out words and phrases having a certain frequency-based statistic higher than a given threshold (Manning & Schütze 1999). Another common method is to compare the frequency of words and phrases in a specialised text to their frequency in general language texts assuming that terms tend to appear more often in specialised texts than in general language texts.

Different evaluation criteria exist for TETs. Although the most important criterion is accuracy, other criteria such as supported file formats and languages are also used. The most frequently used measures for accuracy are noise and silence, as well as recall and precision. While noise refers to the ratio between discarded TCs and the accepted ones, silence refers to the number of terms not detected by a TET. Recall and precision are two measures frequently used in IR, the former being defined as the ratio between the number of correctly retrieved terms and the number of existing terms, the latter being defined as the ratio between correctly extracted terms and the number of proposed TCs (cf. Zielinski, 2002).

TETs following a purely linguistic approach tend to produce too many irrelevant TCs (noise), whereas those following a purely statistical approach tend to miss TCs that appear with a low frequency value (silence, cf. Clematide, 2003). Linguistically-based TETs often provide better delimited TCs than statistically-based ones. However, the disadvantage of linguistically based TETs is that they are language-dependent and thus only available for major languages. Statistical TETs, on the other hand, can also be used for lesser-used languages that lack computational resources such as minority languages (cf. Streiter et al., 2003).

More recently, approaches to automatic TE and TR have moved towards using both statistical and linguistic information (Daille et al., 1994; Justeson & Katz, 1996; Frantzi, 1998). Generally the main part of the algorithm is the statistical part, but shallow linguistic information is incorporated in the form of a syntactic filter which only permits phrases having certain syntactic structures to be considered as candidate terms.

2.2 Bilingual terminology extraction (BLTE)

Most of what has been discussed so far applies to monolingual TE and TR. Lately, research has evolved towards the automatic extraction of bilingual terms. This process generally involves three steps: 1) automatically capturing bilingual terminology from existing technical texts and their translations (parallel corpora), 2) validating the candidate term pairs generated, and 3) generating terminological records in an automatic or semi-automatic manner. Several works have focused on the extraction of knowledge from bilingual corpora. All of them address the problem of aligning units across languages. Although very successful methods have been designed to align paragraphs and sentences written in two different languages, aligning units smaller than a sentence still poses a real challenge.

Thus, Gaussier (1998) relies on corpora aligned at the sentence level. Association probabilities between single words are calculated on the basis of bilingual co-occurrences of words in aligned sentences. Then these probabilities are used to find the French equivalents of English terms through a flow network model. Hull (1998) differs from Gaussier (1998) in that single-word alignment, term extraction and term alignment are three independent modules. Terms and words are aligned through an algorithm that scores the candidate bilingual pairs according to probabilistic data, chooses the highest scored pair, removes it from the pool, and repeatedly recomputes the scores and removes pairs until all the pairs have been chosen. Further improvements on Gaussier’s first model can be found in Gaussier et al. (2000) and Dejean et al. (2003).

Chambers (2000) describes a project launched in 1999 whose main aims include the automatic extraction of bilingual terminology from parallel corpora, manual validation of bilingual term pairs, and automatic generation of terminological records. The process has three major operations: monolingual extraction from the source text, monolingual extraction from the target text and bilingual matching to produce candidate term pairs.

Many methods have been proposed for extracting translation pairs from bilingual corpora, but most are based on word frequency and are, therefore, not effective in extracting low-frequency pairs. Word-frequency-based methods are language-pair-independent. Examples of these methods include Melamed (2000) and Hiemstra (1997). While popular and well-known translation pairs may already be included in existing bilingual dictionaries, newly coined and minor translation pairs are not very well-covered in available resources. In order to tackle this problem, Tsuji & Kageura (2004) present a method for extracting low-frequency translation pairs from Japanese-English bilingual corpora. Their method uses transliteration patterns that are observed in actual loan-word pairs, thus incorporating language-pair-dependent knowledge.

3. Mutual Bilingual Term Extraction

Most of the BLTE methods described above comprise two main steps: the extraction of monolingual term candidates and the alignment of those candidates together. In this way, term extraction methods from different languages do not benefit from each other. To exploit strengths and limit weaknesses in terminology extraction for each language, we propose the use of automatic alignment to help propagate the strengths of terminology extraction from one language into the other. The availability of parallel corpora aligned at sentence level makes the alignment process more accurate, and thus makes this possible. The overall process of the mutual bilingual terminology extraction methodology can be described as follows: first, a list of term candidates is extracted for the first language; then term candidates from the second language are aligned to this list. If a term candidate in the second language is aligned to a term candidate in the first language, its term score is increased: the candidate is promoted. This process can also be repeated many times. The following sections detail the proposed approach.

3.1 Monolingual terminology extraction

In general, we use a combined lexico-syntactic and statistical approach to extract terms for both of the languages. In the experiment, part-of-speech sequences ([AN]*NP?[AN]*N for English, and N[NA]*(PN)?[NA]* for Spanish) are used to select term candidates. Statistical measures such as term frequency, TF.IDF etc. are used to assign scores to these term candidates. Although other scores have been experimented with, none has been shown to be as good as frequency. As a result, in this paper we use term frequency as the statistical score.

3.2 Term candidate alignment

To align term candidates, we use a contingency table, and log-likelihood (Manning & Schütze 1999) to measure how likely a pair of English and Spanish term candidates is to be a correct pair. The contingency table is built using a parallel corpus manually aligned at sentence level (see Section 4.1 for the description of the building of the corpus). Thanks to the sentence alignment effort, we can collect statistics for the contingency table and for log-likelihood calculation. Take the pair of “lymph node”, and “ganglio linfático” as an example: in a subset of 1894 pairs of aligned sentences, “lymph node” appears in 22 English sentences, and “ganglio linfático” in 25 Spanish sentences. They appear in 18 pairs of sentences which are manually aligned. From these statistics, the contingency table can be constructed (in which 1894 is the total of pairs of aligned sentences in the corpus):

	<i>lymph node</i>	<i>lymph node</i>	total
<i>ganglio linfático</i>	18(c_{12})	7	25(c_2)
<i>ganglio linfático</i>	4	1865	1869
total	22(c_1)	1872	1894(N)

Table 1: Contingency table for “lymph node” and “ganglio linfático”

Log-likelihood value is calculated using the following formula

$$\log L\left(c_{12}, c_1, \frac{c_2}{N}\right) + \log L\left(c_2 - c_{12}, N - c_1, \frac{c_2}{N}\right) - \log L\left(c_{12}, c_1, \frac{c_{12}}{c_1}\right) - \log L\left(c_2 - c_{12}, N - c_1, \frac{c_2 - c_{12}}{N - c_1}\right)$$

in which, c_{12} : total number of pairs of aligned sentences in which “lymph node” appear in the English sentence, and “ganglio linfático” appear in the Spanish one; c_1 is the total number of English sentences in which “lymph node” appears; c_2 : the total number of Spanish sentences in which “ganglio linfático” appears; and N: the total number of pairs of aligned sentences.

From these statistics, log-likelihood value of the pair of “lymph node” and “ganglio linfático” is 76.48. Among log-likelihood values of alignment candidates for “lymph node”, this is the highest value, suggesting that “ganglio linfático” is likely to be the translation of “lymph node”. As the evaluation section (Section 4) will show, this alignment process produces an accuracy of around 0.8 (i.e. out of 10 pairs suggested by the process, 8 of them are confirmed as correct translations).

3.3 Mutual bilingual terminology extraction

We hypothesise that the term score of a term candidate in one language can be used to improve the term score of its aligned candidate in the other language, and vice versa. Hence, three algorithms in which the term scores of the candidates in one language are used to boost those in the other language have been proposed. In algorithm 1, the boosting process is performed on the target language only, whereas in algorithm 2, it is performed on both source and target languages. In algorithm 3, the boosting process is repeated for both languages until the term candidate lists are stabilised.

Definition

Source language: the language used to help the term extraction process

Target language: the language in which terms are extracted

AL(T1,T2): alignment score of the two term candidates, T1, and T2.

TCs[T]: term score of the candidate T in the source language

TCt[T]: term score of the candidate T in the target language

The initial values of those term scores are assigned using functions discussed in Section 3.1.

BT(TC1,TC2): boosting function, i.e. how the term score of the aligned term affects the target term score; one example is the simple addition: $BT(TC1,TC2)=TC1+TC2$;

For all three algorithms, the initial term scores for both source and target languages are calculated and put into two hash tables: TCs and TCt.

```

1  Foreach term candidate Tt in Target Language
2      | Ts=argmax(AL(Tt,Ti));
3      | TCt[Tt]=BT(TCs[Ts],TCt[Tt]);
4  Sort the new TCt, extract the first N candidates.

```

Algorithm 1: single boosting

```

1  Foreach term candidate Ts in Source language
2      | Tt=argmax(AL(Ts,Ti));
3      | TCs[Ts]=BT(TCs[Ts],TCt[Tt]);
4  Foreach term candidate Tt in Target Language
5      | Ts=argmax((AL(Tt,Ti));
6      | TCt[Tt]=BT(TCs[Ts],TCt[Tt]);
7  Sort the new TCt, and extract the first N candidates

```

Algorithm 2: double boosting

```

1  Repeat
2      | Foreach term candidate Ts in Source language
3      | Begin
4          | Tt=argmax((AL(Tt,Ti)); // Get the most
5          | likely aligned term candidate in Target
6          | Language
7          | TCs[Ts]=BT(Ts,Tt); // Changing the term
8          | score of Ts using the boosting function
9      | End
10     | Foreach term candidate Tt in Target Language
11     | Begin
12         | Ts=argmax((AL(Ts,Ti)); // Get the most
13         | likely aligned term candidate in Source
14         | language
15         | TCt[Tt]=BT(Ts,Tt); // Changing the term
16         | score of Tt using the boosting function and
17         | the new term score (calculated previously);
18     | End
19     | Reorder the term candidates in Target
20     | Language according to the new term score, and
21     | extract the first N candidates.
22 Until the list of the first N candidates remains
23 unchanged.

```

Algorithm 3: converge boosting

Factors that affect the outcome of the proposed algorithms are: the alignment function AL, the mechanism to calculate the initial term scores TCs and TCt, and the boosting function BT. Different combinations of these functions have been experimented with and the result indicated that the best term score function is frequency, and the best boosting function is simple addition (i.e. $BT(tc1,tc2)=tc1+tc2$)². As a result, in this paper we only present results produced by this combination.

4. Evaluation

4.1 Data, gold standard and evaluation metrics

Bilingual term extraction tools analyse aligned bilingual corpora in an attempt to identify potential terms and their translation equivalents. Therefore, the first step of the

² The following boosting functions have been experimented with, none is shown to be as good as simple addition: multiplication ($BT(tc1,tc2) = tc1*tc2$); log multiplication ($=tc1*\log(tc2)$).

evaluation process is to create a corpus. A corpus is a large collection of texts designed to meet some specific needs. It needs to contain enough samples of concepts, terms and linguistic patterns relevant to a specific domain. A bilingual parallel corpus contains texts and their translations into another language. When deciding to create a parallel corpus, one needs to have access to pairs of texts (source texts and their translations) dealing with the topic or subject field one wishes to study. In our case, we chose the medical domain as documents in this specialised subject field abound and are quite representative of a language being used for specific purposes. Furthermore, we needed to find a domain where sufficient bilingual documents are generated, and can be freely used for research purposes. Having taken these into account, we decided that MedlinePlus was the best place to start with for our project.

Medline represents the world's largest medical library, the National Library of Medicine. MedlinePlus has extensive information from the National Institutes of Health and other trusted sources on over 750 diseases and conditions. This resource can be easily browsed online for free. Furthermore, most of the information contained in the site is bilingual (English-Spanish): files in English have been translated and stored with their corresponding Spanish translations. From all the health topics covered in this site, we found that Cancer was the topic that has more bilingual documents (English-Spanish), so we decided to focus on this particular disease and on this particular language pair. As bilingual and multilingual parallel texts are less easy to find than monolingual texts, the size of a parallel corpus is bound to be very much smaller than a monolingual corpus created using the same criteria. Specifically, our corpus consisted of 9,250 sentences for each language. The English corpus was made up of 31,498 words, whereas the Spanish one contained 30,344 words.

The whole process of corpus compilation can be summarised as follows: searching for texts on the web (Medline), selecting and downloading appropriate texts from the site, storing the texts, and finally preparing the texts for the alignment process, which will be described later.

Once we have chosen the texts and converted all of them to machine-readable form, some pre-processing was needed in order to prepare texts for alignment. The pre-processing tasks mainly involved deleting superfluous line breaks as alignment programs tend to interpret them as paragraph breaks.

The process of creating links between texts is generally referred to as alignment. A number of alignment techniques have been developed and a small number of programs are available. The first stage of alignment usually involves creating links between matching paragraphs and headings in the source and target texts by matching them sequentially. Most programs will then attempt to align texts at sentence level. Trados WinAlign software was used for our project, as we were already familiar with this program. WinAlign has an inbuilt editing tool which allows users to confirm and/or rectify the alignments that have been made.

Alignment programs make a number of assumptions about the texts and their translations, some of which can be incorrect. For example, these programs assume that there is generally a sequential one-to-one correspondence between source and target text, at the very least at paragraph level and ideally at sentence level. This means that sentence 1 in the source text is assumed to correspond to sentence 1 in the target text, sentence 2 to sentence 2 and so on throughout each paragraph or text. Unfortunately, as translations do not always follow the sequential progression of the source texts, we encountered abnormal sentence pairs when we were reviewing some of the aligned texts and we had to correct them manually.

Another assumption that alignment programs make is that each sentence in the source text is translated in the target text, i.e. that all the information contained in the source text should be transferred to the target text. However, we sometimes found a number of sentences and even some entire paragraphs in the source texts that were not translated at all, creating many mismatches. When cases like these appeared, most of the file needed to be realigned manually.

Overall, it is observed that the benefits of having access to parallel texts, even when there are mismatches, outweighed the disadvantages and the whole process of reviewing the results from WinAlign did not take long. Our medical corpus consists of 9,250 sentences for each language collected from MedlinePlus. The English corpus contained 31,498 words, whereas the Spanish one contained 30,344 words.

Pairs of candidates are extracted from the medical parallel corpus and subjected to validation by a professional. In the initial experiment, 389 English terms, 442 Spanish terms, and 357 term pairs have been validated and used as a gold standard.

We use standard f-measure as our main evaluation metric to measure the impact of mutual bilingual term extraction. F-measure is calculated as $F=2/(1/P+1/R)$, where P: precision, the total number of correct terms divided by the total number of term candidates, and R: recall, the total number of correct terms divided by the total number of confirmed terms.

4.2 Results

4.2.1 Alignment accuracy

In total, the algorithm suggests 472 translation pairs, of which 374 are confirmed as correct translation³. This suggests that the accuracy of the alignment is 0.8. The accuracy of the alignment process, although an important indicator, is not the main concern of this paper.

4.2.2 Term extraction f-measures

³ Of 374 translation pairs, only 357 are term pairs, 17 are correct translations, but are not considered terms, according to the translation expert.

Table 2 and Figure 1 show the f-measure values for the five lists of English and Spanish term candidates containing 400, 500, 600, 700, and 800 term candidates identified respectively by frequency, single alignment boost, and converge alignment boost⁴. The results show that the performance of term extraction can be improved by around 10% (from 0.56 to 0.62) by using only the single boosting algorithm, and up to 25% (from 0.56 to 0.70) when converge boosting is used. The improvements are shown to be consistent for both English and Spanish.

There are several interesting observations which can be drawn from the results. The first one is that the use of the single boosting algorithm tends to produce similar results for English and Spanish regardless of the original values of f-measure, whereas converge boosting produces different results for English and Spanish (f-measure values for Spanish are slightly higher than for English). This may be due to the fact that in our parallel corpus, English is the source language, and Spanish is the target. Another experiment (in which Spanish is the source language, and English is the target) will be needed in order to confirm (or deny) this hypothesis.

It seems that the improvements peak when the number of term candidates is around 500 and 600. This suggests that the use of mutual bilingual term extraction has limitations, especially when the number of candidates becomes large. We hypothesise that as the number of candidates increases, the accuracy of the alignment would decrease as less data are available, which makes the alignment scores less accurate. This, in turn, affects the performance of mutual bilingual terms extraction. If we look at the values of f-measure when the numbers of term candidates are 400 and 700 respectively, we can see the effect of the number of candidates on mutual bilingual term extraction more clearly. At 400 term candidates, original f-measures for English and Spanish are 0.56 and 0.55 respectively; after MBTE has been applied, the values improve to 0.71 and 0.73, providing improvement ratios of 1.27 and 1.33. At 700 term candidates, these respective numbers are 0.57, 0.58, 0.63, 0.69, 1.11, and 1.19. These numbers show that not only the relative improvement ratios decrease (from roughly 1.3 to roughly 1.15) when the number of term candidates increases from 400 to 700, the absolute values of the f-measure also decrease (from over 0.7 to under 0.7). Future experiments will be designed to evaluate the impact of alignment accuracy on bilingual mutual term extraction.

5. Discussion and future directions

In this paper, it is shown that mutual terminology extraction is a promising approach to bilingual terminology extraction. Terminology extraction from a language whose resources are limited can benefit from terminology extraction from another language whose resources are more widely available. As the results of this initial experiment are very encouraging, in the future, we

⁴ Although three algorithms have been experimented with, we only report the results for the first and the third algorithm. This is due to the fact that the results from the double boosting algorithm are very similar to those of single boosting one, and thus are overlooked in this paper.

intend to explore the following research directions.

We will experiment with different termhood functions for different languages, and exploit different ways to align term candidates, including the use of dictionaries and other alignment scores. We will investigate the effect of alignment accuracy on the performance of mutual bilingual term extraction, as this has been overlooked in this paper. These additional experiments will provide better insight into the usefulness of the proposed methodology.

Additionally, we will experiment with different settings of the boosting algorithm, including the use of a voting function. We will also run the experiments on another domain (EU legislations), in order to gain insights into the domain-independent nature of the approach. We also intend to run the experiment on other pairs of languages, such as English-Hindi in order to gain insights into how the proposed methodology performs against different pairs of languages. We would especially like to study the performance difference between English-Spanish and Spanish-English pairs, to understand the effect of the source languages on the approach.

6. References

- Ananiadou, S. 1994. A methodology for Automatic Term Recognition. In Proceedings of the *15th International Conference on Computational Linguistics (COLING94)*, pp. 1034-1038. Kyoto, Japan.
- Bourigault, D., C. Jacquemin, and M. C L'Homme (ed.) 2001. *Recent Advances in Computational Terminology*. Amsterdam: John Benjamins Publishing Company.
- Chambers, D. 2000. Automatic Bilingual Terminology Extraction: A Practical Approach. In Proceedings of *Translating and the Computer 22, Aslib/IMI*.
- Daille, B., E. Gaussier, J.-M. Lange. 1994. Towards Automatic Extraction of Monolingual and Bilingual Terminology. In Proceedings of *COLING 1994*.
- Dejean, H., E. Gaussier, C. Goutte, and K. Yamada. 2003. Reducing parameter space for word alignment. In Proceedings of *HLT-NAACL 2003 Workshop on Building and Using Parallel Texts: Data Driven Machine Translation and Beyond*, Edmonton, Alberta.
- Frantzi, K. T. 1998. Automatic Recognition of Multi-Word Terms. PhD Thesis. Manchester Metropolitan University, UK.
- Gaussier, E. 1998. Flow Network Models for Word Alignment and Terminology Extraction from Bilingual Corpora. In Proceedings of *Thirty-Sixth Annual Meeting of the Association for Computational Linguistics and Seventeenth International Conference on Computational Linguistics*, pp. 444-450. San Francisco, California.
- Gaussier, E., D. Hull, and S. At-Mokthar. 2000. Term alignment in use: Machine-aided human translation. In J. Veronis (ed.). *Parallel text processing: Alignment and use of translation corpora*, pp. 253-274. Dordrecht: Kluwer Academic Publishers.
- Hiemstra, D. 1997. Deriving a bilingual lexicon for cross language information retrieval. In Proceedings of *Gronics 1997*, pp. 21-26.
- Hull, D. 1998. A practical approach to terminology alignment. In Proceedings of *CompuTerm 1998*, pp. 1-7.
- Justeson, J. S., and S. L. Katz. 1996. Technical Terminology: some linguistic properties and an algorithm for identification in text. *Natural Language Engineering* 3(2): 259-289.
- Manning, C. D., and H. Schütze. 1999. *Foundations of Statistical Natural Language Processing*. MIT Press.
- Melamed, I. D. 2000. Models of translational equivalence among words. *Computational Linguistics* 26(2): 221-249.
- Pearson, J. 1999. *Terms in context*. Amsterdam: John Benjamins.
- Sauron, V. A. 2002. Tearing out the terms: evaluating terms extractors. In Proceedings of *Translating and the Computer 24*, London, Britain.
- Streiter, O., D. Zielinski, I. Ties, and L. Voltmer. 2003. Term extraction for Ladin: An example-based approach. In Proceedings of *TANL 2003 Workshop on Natural Language Processing of Minority Languages with few computational linguistic resources*, Batz-sur la Mer.
- Tsuji, K., and K. Kageura. 2004. Extracting low-frequency translation pairs from Japanese-English bilingual corpora. In Proceedings of *CompuTerm 2004*, pp. 23-30.
- Zielinski, D., and Y. R. Safar. 2005. t-survey 2005: An Online Survey on Terminology Extraction and Terminology Management. In Proceedings of *Translating and the Computer 27*, London, Britain.

Number of term candidates	Termhood function					
	English TF	Spanish TF	English TF (single Boosted)	Spanish TF (single Boosted)	English converge boosted	Spanish converge boosted
400	0.546185	0.517327	0.594378	0.591584	0.671123	0.669963
500	0.564345	0.555066	0.623377	0.61674	0.707547	0.730473
600	0.578669	0.565476	0.616684	0.619048	0.670886	0.733399
700	0.571156	0.575812	0.590258	0.602888	0.627863	0.687106
800	0.544028	0.544702	0.564952	0.571192	0.587108	0.640199

Table 2: F-measure values

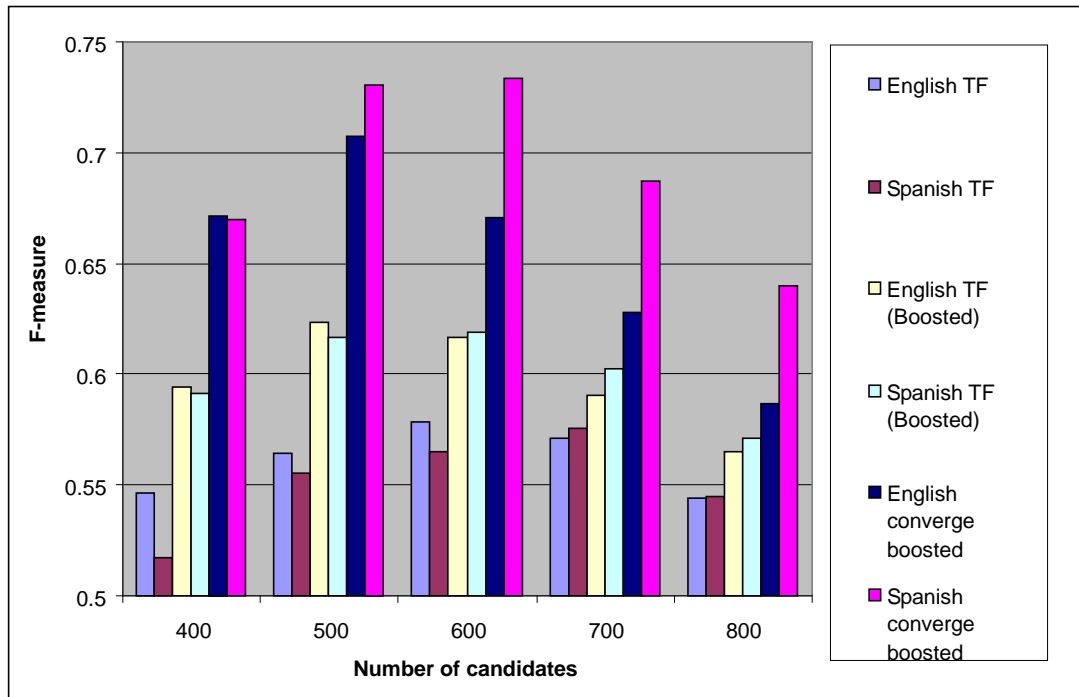


Figure 1: F-measure values graph