



HAL
open science

Term Translation Validation by Retrieving Bi-terms

Brigitte Grau, Anne-Laure Ligozat, Isabelle Robba, Anne Vilnat

► **To cite this version:**

Brigitte Grau, Anne-Laure Ligozat, Isabelle Robba, Anne Vilnat. Term Translation Validation by Retrieving Bi-terms. Peters, Carol. Accessing Multilingual Information Repositories: 6th Workshop of the Cross-Language Evaluation Forum, CLEF 2005, Vienna, Austria, 21-23 September, 2005, Revised Selected Papers, Springer Science & Business Media, 2006. hal-02307365

HAL Id: hal-02307365

<https://hal.science/hal-02307365>

Submitted on 7 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Term Translation Validation by Retrieving Bi-Terms

Brigitte Grau, Anne-Laure Ligozat, Isabelle Robba,
and Anne Vilnat

LIR group, LIMSI-CNRS, BP 133 91403 Orsay Cedex, France
`firstName.lastName@limsi.fr`

Abstract. For our second participation to the Question Answering task of CLEF, we kept last year's system named MUSCLEF, which uses two different translation strategies implemented in two modules. The multilingual module MUSQAT analyzes the French questions, translates "interesting parts", and then uses these translated terms to search the reference collection. The second strategy consists in translating the question into English and applying QALC our existing English module. Our purpose in this paper is to analyze term translations and propose a mechanism for selecting correct ones. The manual evaluation of bi-terms translations leads us to the conclusion that the bi-term translations found in the corpus can confirm the mono-term translations.

1 Introduction

This paper presents our second participation to the Question Answering task of the CLEF evaluation campaign. This year we have participated in two tasks: a monolingual task (in French) for which we submitted one run, and a bilingual task (questions in French, answers in English) for which we submitted two runs. Concerning the bilingual task, we used the same two strategies as last year:

- translation of selected terms issued of the question analysis module, then search in the collection; this first system is called MUSQAT
- question translation thanks to a machine translation system, then application of QALC our monolingual English system

Most systems make use of only one of these strategies [6], but our system, MUSCLEF[1], follows both approaches, by combining MUSQAT and QALC. In this article, we focus on the evaluation of the different translation techniques used in MUSCLEF. This study leads us to propose a mechanism for selecting correct term translations.

We will first present an overview of our system (section 2), then we will focus on our recognition of terms in documents, realized by Fastr (3), and their translation (4). We will then present an evaluation of these translations (5) followed by results concerning term validation (6) and our global results at the QA task (7).

2 Overview of MUSCLEF

MUSCLEF architecture is illustrated in Figure 1. First, its question analysis module aims at deducing characteristics which may help to find answers in selected passages. These characteristics are: the expected answer type, the question focus, the main verb and some syntactic characteristics. They are deduced from the morpho-syntactic tagging and syntactic analysis of the question. For this campaign, we developed a grammar of question and used the Cass robust parser¹ to analyze the English questions that were translated using Reverso².

We conducted a quick evaluation of our French question analysis, which revealed that 77% of the French questions were attributed the correct expected answer types. We corrected some of the errors, and a more up-to-date question analysis reached 97% of correct type attribution. Though these results are quite satisfactory, the question analysis is much deteriorated on the translated questions, and this problem will have to be taken into account for next year's evaluation.

A new type of questions, temporally restricted questions, was introduced in this year's campaign. We have adjusted the question analysis to the category of the question. When a temporal restriction was to be found, we tried to detect it, and to classify it according to the three following types: date, period, and event. The answering strategy was then adapted to the type of temporal constraint.

On the 29 temporally restricted French questions, 18 of them contained dates or periods, and 12 contained event-related restrictions (one question contained both). Our system was able to detect and classify the dates and periods for all 18 questions, the classification consisting in separating dates and periods, and for the periods, in detecting if the period concerned days, months or years.

For querying the CLEF collection and retrieving passages we used MG³. Retrieved documents are then processed: they are re-indexed by the question terms and their linguistic variants, reordered according to the number and the kind of terms found in them, so as to select a subset of them.

Named entity recognition processes are then applied. The answer extraction process relies on a weighting scheme of the sentences, followed by the answer extraction itself. We apply different processes according to the kind of expected answer, each of them leading to propose weighted answers.

The first run we submitted corresponds to the strategy implemented in MUSQAT: translation of selected terms. For the second run, we added a final step consisting in comparing the results issued from both strategies: the translated questions and the translated terms. This module named fusion, computes a final score for each potential answer. Its principle is to boost an answer if both chains ranked it in the top 5 propositions, even with relatively low scores.

¹ <http://www.vinartus.net/spa/>

² <http://www.reverso.net>

³ MG for Managing Gigabytes <http://www.cs.mu.oz.au/mg/>

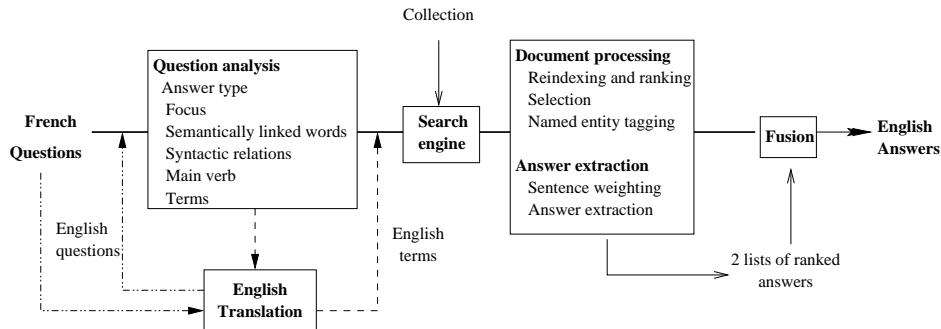


Fig. 1. MUSCLEF architecture

3 Searching Terms and Variants

Term recognition in retrieved documents is performed by FASTR, a transformational shallow parser for the recognition of term occurrences and variants ([3]). Terms are transformed into grammar rules and the single words building these terms are extracted and linked to their morphological and semantic families.

The morphological family of a single word w is the set $M(w)$ of terms in the CELEX database ([2]) which have the same root morpheme as w . For instance, the morphological family of the noun *maker* is made of the nouns *maker*, *make* and *remake*, and the verbs *to make* and *to remake*. The semantic family of a single word w is the union $S(w)$ of the synsets of WordNet1.6 ([4]) to which w belongs. A synset is a set of words that are synonymous for at least one of their meanings. Thus, the semantic family of a word w is the set of the words w' such that w' is considered as a synonym of one of the meanings of w . The semantic family of *maker*, obtained from WordNet1.6, is composed of three nouns: *maker*, *manufacturer*, *shaper* and the semantic family of *car* is *car*, *auto*, *automobile*, *machine*, *motorcar*. Variant patterns that rely on morphological and semantic families are generated through metarules. They are used to extract terms and variants from the document sentences in the selected documents.

For instance, the following pattern, named *NtoSemArg*, extracts the occurrence *making many automobiles* as a variant of the term *car maker*:

$$NN(car)NN(maker)- > VM('maker')RP?PREP?ART?(JJ|NN|NP|VBD|VBG)0 - 3NS('car')$$

In this pattern, *NN* are nouns and *NP* proper nouns, *RP* are particles, *PREP* prepositions, *ART* articles, and *VBD*, *VBG* verbs. *VM('maker')* is any verb in the morphological family of the noun *maker* and *NS('car')* is any noun in the semantic family of *car*.

Relying on the above morphological and semantic families, *auto maker*, *auto parts maker*, *car manufacturer*, *make autos*, and *making many automobiles* are

extracted as correct variants of the original term *car maker* through the set of metarules used for the QA-track experiment. Unfortunately, some incorrect variants are extracted as well, such as *make those cuts in auto* produced by the preceding metarule.

4 Term Translation in MUSQAT

In order to achieve term translation we considered the easiest method, which consists in using a bilingual dictionary to translate the terms from the source language to the target language. Last year, we chose Magic-dic⁴, a GPL dictionary, because of its increasing capacity: terms can be added by any user, but they are verified before being integrated, and to prevent its incompleteness, we used this year another dictionary FreeDict⁵ and merged their translations. FreeDict had added 424 different translations of the 690 words. However, these new translations are mainly other synonyms rather than new translations of unknown words. For example the query for the French word *mener* to Magic-Dic gives the following results: *conduct, lead, guide*; while *accord* is only translated by *agreement*. FreeDict added five translations for *accord*: *accord, accordance, concurrence, chord, concord*, and gave translations for *frontière* that was unknown by MagicDic. However, *occidental* (western) remained not translated.

We illustrate the strategy defined in our multilingual module MUSQAT on the following example: “*Quel est le nom de la principale compagnie aérienne allemande?*”, which is translated in English “*What is the name of the main German airline company?*”.

The first step is the parsing of the French question that provides a list of the mono-terms and all the bi-terms (such as *adjective/common noun*) which are in the question, and eliminates the stop words. The bi-terms are useful, because they allow a disambiguation by giving a (small) context to a word. In our example, the bi-terms (in their lemmatized form) are: *principal compagnie, compagnie aérien, aérien allemand*; and the mono-terms: *nom, principal, compagnie, aérien, allemand*.

With the help of the dictionaries, MUSQAT attempts to translate the bi-terms (when they exist), and the mono-terms. All the proposed translations are taken into account. All the terms are grammatically tagged. If a bi-term cannot be directly translated, it is recomposed from the mono-terms, following the English syntax. For our example, we obtained for the bi-terms: *principal company/main company, air company, air german*; and for the mono-terms: *name/appellation, principal/main, company, german*.

When a word does not exist in the dictionaries, we keep it as it is without any diacritic, which is often relevant for proper nouns. Then, all the words are weighted relatively to their existence in a lexicon that contains the vocabulary found in the Latimes of the Trec collection, so that each word is weighted according to its specificity within this corpus. If a word is not found in this lexicon, we

⁴ <http://magic-dic.homeunix.net/>

⁵ <http://www.freedict.de/>

search with MG if documents contain it (or rather its root because MG indexing was made using stemming). If it is not the case, MUSQAT eliminates it from the list of translated terms. In this way, MUSQAT discarded 72 non-translated words (out of 439 non-translated mono-terms, the remaining ones often being proper nouns). As we form boolean requests, it was important not to keep non existing words.

English terms plus their categories (given by the Tree Tagger) were then given as input to the other modules of the system, instead of the original words. The translation module did not try to solve the ambiguity between the different translations. We account on the document retrieval module to discard irrelevant translations. This module has been improved this year: it always selects passages (the collection was preliminary split), but in a very smaller number. It first generates boolean requests, based on proper nouns, numbers and specificity of the words. It aims at retrieving 200 passages maximum, and makes the smaller request with the more specific terms so as to obtain a minimum number of passages, set to 50. Each term of the request is made of the disjunction of the different translations. If the boolean query retrieves too few or too much documents, passage retrieval is made thanks to a ranked research with a query that hold all the terms. If there are synonyms for certain terms, relevant documents are then retrieved with these synonyms. If a word is incoherent within the context, we suppose its influence is not sufficient to generate noise. This hypothesis can only be verified if the question is made of several words.

5 Magic-dic Term Evaluation

We manually evaluated the bi-term translations for the 200 questions of CLEF04 given by this module. Table 1 presents the results of this evaluation.

The system found 375 bi-terms. Among them, 135 are correct translated bi-terms (OK) such as *CERN member*. 24 are bi-terms contextually false i.e. for which one word is not a good translation in the context of this bi-term, such as *accretion hormone* instead of *growth hormone* to translate *hormone de croissance*. 74 bi-terms are due to an erroneous bi-term constitution (False Bi-Terms), such as *able animal* in question asking to *Give an animal able to....* Finally, 142 bi-terms (a) are completely erroneous translations (False Translation), such as *overground escort* instead of *main company* (110) or (b) have a translation which was absent from the dictionary (Absent Translations), such as *olympique*, where the French word has been kept instead of the English term *olympic* (32).

It is obvious on this table that a lot of terms are wrong for different reasons. We decided to confirm those that must be kept by considering their presence or absence in the selected documents. To do so, we used FASTR results to evaluate the bi-terms or their variants which are retrieved in the documents. Table 2 shows the results of this evaluation. The second column gives the results obtained by FASTR without considering the semantic variations. The third column includes these semantic variations. The last column indicates the percentage of bi-terms

Table 1. MagicDic terms evaluation

Bi-Terms	#	%
OK	135	36
Contextually False	24	6.4
False	74	19.7
False Transl	110	29.3
Absent Transl	32	8.5
Total False	240	64
Total	375	

FASTR confirms, taking into account the semantic variations.

Table 2. MagicDic terms validated by Fastr

Bi-terms	#	#retrieved without sem.var.	#retrieved including sem.var.	%
OK	135	61	83	61.5
Context. False	24	4	7	29.2
False	74	11	15	20.3
False Transl	110	7	19	17.3
Absent Transl	32	0	0	0
Total	375	82	120	32

The correct bi-terms are mostly confirmed by FASTR. The contextually false bi-terms obtain a rather high percentage of confirmation due to the semantic variations which lead to recognize correct synonyms of non accurate translated terms. The false bi-terms can be considered as co-occurrences rather than bi-terms. As co-occurrences, they are retrieved by FASTR in the documents and just a few false translations are retrieved.

6 Evaluation of Terms Extracted from Question Translations

We also proceeded to a similar evaluation of the terms extracted from the questions translated last year by Systran.

As a first step we proceeded to an evaluation of the question translations themselves. We evaluated the syntactic quality of the translations, and classified them in correct, false, or quite correct. Table 3 recapitulates these results.

Table 3. Questions translations evaluation

Questions	Correct	Quite Correct	False	Total
#	73	12	115	200
%	36.5	6.0	57.5	100

We also evaluated the terms extracted from these translated questions by our monolingual system QALC. We use the same notations as in table 1. Results are given in Table 4.

Table 4. Evaluation of terms from translated questions

Bi-Terms	#	%
OK	126	75.4
Contextually False	0	0
False	41	24.6
False Transl	0	0
Absent Transl	0	0
Total False	41	24.6
Total	167	

These results are quite interesting: despite the moderate quality of the translations, QALC is able to identify good terms from these questions. We can also notice that we obtain a smaller number of terms following this procedure because there is only one translation by word.

7 Results

Table 5 gives the results that our system obtained at the CLEF04 and CLEF05 campaigns, with the different strategies: (a) with the translation of the terms (MUSQAT), (b) with QALC applied on the translated questions and searching the collection. The evaluation was made by an automatic process that looks for the answer patterns in the system answers, applying regular expressions. These results were computed with 178 answer patterns that we built for the 200 questions of CLEF04 and 188 for the CLEF05 questions.

The first line indicates the number of correct answers found in the 5 first sentences given by MUSQAT (using term translation) and QALC. The second line, “NE answers”, gives the number of correct answers on questions the system categorized as waiting for a Named Entity (the total is 107 in CLEF04 for MUSQAT and 97 for QALC and 91 in CLEF05 for MUSQAT and 66 for QALC). Our total number of questions for this category is far beyond the real number in CLEF05. The third line, “non NE answers”, concerns the other questions (the complement to 178 in CLEF04 and to 188 in CLEF05).

Results are presented when the system just gives one answer and when it gives 5 answers. The last line indicates the best official result of our system on the 200 questions. The official score of MUSQAT was 22 (11%) in CLEF04 and 28 (14%) in CLEF05, thus we can observe that merging answers obtained by different strategies enabled a significant gain. We also can notice that if our CLEF05 system better selects sentences, it is less efficient on extracting the named entity answers.

Table 5. Results at CLEF04 and CLEF05

		MUSQAT04	QALC04	MUSQAT05	QALC05
Sentences	5 first ranks	56 (31 %)	65 (37 %)	78 (41 %)	87(46 %)
NE answers	Rank 1	17	26	16	9
	5 first ranks	32	37	24	11
Non NE answers	Rank 1	7	3	16	16
	5 first ranks	12	8	22	24
Total	Rank 1	24	29	32	25
	%	12%	14.5%	17%	13%
	5 first ranks	44	45	46	35
Fusion (official results)		38 (19 %)		38 (19 %)	

According to the manual evaluation results of bi-terms translations, we have tested an automatic process for filtering Magic-dic translations on CLEF04 questions. So, if a bi-term or a variant form was found in the selected documents, we kept it as a valid translation and we kept its lemmas as valid mono-term translations. When a validated translation existed for a term, the non-validated translations were taken out. When no translation of a bi-term was found in corpus, we assumed that mono-term translations were wrong and we kept Systran translations. In order to improve the coverage of our translation, we added Systran translation for terms absent from the dictionary.

In this way, we selected 253 bi-terms in 112 questions, and added 37 translations, with 12 bi-terms, which concerns 35 questions. The last improvement consisted in adding Systran translations that were different from Magic-dic translations (138 terms in 96 questions) to the filtered terms. This last set of terms was composed of 1311 translations for 836 terms in 200 questions (522 terms with 1 translation, 199 with 2 translations, 81 with 3 translations, 25 with 4 translations, 6 with 5 translations and 3 with 6 translations).

We tested MUSQAT with this new selection. Results are shown Table 6. We see that MUSQAT finds relevant documents for 7 additional questions (increase of 4%).

MUSQAT extracts 7 additional correct answers in the top 5 short answers, with 29 answers in rank 1. MUSQAT obtains here slightly better results than QALC with Systran translations, both for short and long answers. We also measured the number of questions for which the selection process based on

Table 6. MUSQAT new results

		MUSQAT
Sentences	5 first ranks	67
NE answers	Rank 1	25
	5 first ranks	41
Non NE answers	Rank 1	4
	5 first ranks	10
Total	Rank 1	29 (14,5%)
	5 first ranks	51

FASTR indexing provides documents containing the answer pattern. In the original MUSQUAT, it was possible to find the answer for 80% of questions. Term selection allows to improve this value to 85%.

These improvements are not significant enough so we had not incorporated them in this year’s version, even if we think that this kind of translation validation is worth being tried. So we plan to realize bi-term validation on a larger corpus. Concerning the absence of translations, we began to increase manually our dictionary from lexicons and gazetteers we use for named entities recognition, specially for acronyms and location names, and we plan to use a bilingual aligned corpus.

8 Conclusion

In [7], we can find the results of all participants to CLEF 2005, in all question-answering tasks. For the monolingual task the average of best scores is 42.6%, while for the cross-language task it is 24.5%. The best score for monolingual task is obtained in Portuguese task with 64.5% of right answers, and the best score for cross-language task is obtained in an English-French run with 39.5% of right answers. The best score for task with English language as a target is 25.5% (with German as source language). As always, cross-language task results are far lower than monolingual task results.

Our system MUSCLEF was ranked third among the 12 systems participating to the task with English as target language. Thanks to the use of a second dictionary, we improved MUSQAT results. Moreover, both systems MUSQAT and QALC got better results concerning sentence selection (+10%), because we modified the elaboration of the query sent to the search engine, allowing us to retrieve more relevant documents.

But in spite of these enhancements, our final results remained exactly the same as last year’s: 19%. This can be explained by two reasons: firstly, due to a technical problem, we could not use this year the web resource; secondly the question analysis module was much deteriorated on the translated questions.

Nevertheless, this second participation to CLEF evaluation was encouraging, and we plan to continue to improve our system; doing for example a better

selection of good translations thanks to the multi-terms, as explained in this paper and also in [5].

References

1. G. Bourdil, F. Elkateb-Gara, O. Ferret, B. Grau, G. Illouz, B. Mathieu, L. Monceaux, I. Robba and A. Vilnat. 2005. Answering French questions in English by exploiting results from several sources of information. *LNCS, Vol 3491, Revised selected papers from Workshop CLEF 2004*, p.470-481.
2. CELEX. 1998. http://www ldc.upenn.edu/readme_files/celex.readme.html, UPenns, Eds., *Proceedings of Consortium for Lexical Resources*.
3. C. Jacquemin. 1999. Syntactic and paradigmatic representations of term variation. *Proceedings of ACL 1999*, p.341-348.
4. C. Fellbaum. 1998. WordNet: An Electronic Lexical Database. Cambridge, MA: MIT Press.
5. M. Negri, H. Tanev and B. Magnini. 2003. Bridging languages for question-answering: Diogene at CLEF 2003. *Working Notes of CLEF 2003 workshop, 21-22 August, Trondheim, Norway*.
6. C. Peters, M. Braschler, G. Di Nunzio and N. Ferro, CLEF 2004: Ad Hoc Track Overview and Results Analysis, *Fifth Workshop of the Cross-Language Evaluation Forum (CLEF 2004), Lecture Notes in Computer Science (LNCS), Springer, Heidelberg, Germany*.
7. A. Vallin, D. Giampiccolo, L. Aunimo, C. Ayache, P. Osenova, A. Peas, M. de Rijke, B. Sacaleanu, D. Santos, R. Sutcliffe. 2005 Overview of the CLEF 2005 Multilingual Question Answering Track. *Working Notes of CLEF 2005 workshop, 21-23 September, Vienna, Austria*.