



Lola García-Santiago is a Doctor of Documentation, University of Granada, where she works as professor. At present she teaches in both the University School of Documentation and the School of Translation and Interpretation. She is also part of the associated unit of the SCImago research group, under Spain's Higher Council for Scientific Research (CSIC), where her lines of work and published articles focus mainly on Web-mining, Internet information retrieval, QA systems and grey literature

Dr. García can be reached at mdolo@ugr.es.

María-Dolores Olvera-Lobo holds a Ph.D. in Documentation and is professor in the Department of Communication and Documentation of the University of Granada, and teaches as well in the School of Translation and Interpretation. As member of the associated unit of the SCImago team, she participates in a number of research projects now underway, and has authored or co-authored several books and numerous specialized journal articles.

Dr. Olvera-Lobo can be reached at molvera@ugr.es.

Front Page

Select one of the previous 50 issues.

Select an issue:

- [Index 1997-2009](#)
- [TJ Interactive: Translation Journal Blog](#)
- **Translator Profiles**
- [Twenty Years of Steady Workload](#)
by Andrei Gerasimov
- **The Profession**
- [The Bottom Line](#)
by Fire Ant & Worker Bee
- **Translators Around the World**
- [Where Can I Find a Chinese Sworn Translator in Rio de Janeiro?](#)
by Danilo Nogueira and Kelli Semolini

- **Cultural Aspects of Translation**
- [Culture-Specific Items in Literary Translations](#)
by Sepideh Firoozkoobi

Automatic Web Translators as Part of a Multilingual Question-Answering (QA) System:

Translation of Questions

by *Lola García-Santiago and María-Dolores Olvera-Lobo*
CSIC, Unidad Asociada Grupo SCImago, Madrid, España
University of Granada, Department of Library and Information Science,
Granada, Spain

Abstract

The traditional model of information retrieval entails some implicit restrictions, including: a) the assumption that users search for documents, not answers; and that the documents per se will respond to and satisfy the query, and b) the assumption that the queries and the document that will satisfy the particular informational need are written in the same language. However, many times users will need specific data in response to the queries put forth. Cross-language question-answering systems (QA) can be the solution, as they pursue the search for a minimal fragment of text—not a complete document—that applies to the query, regardless of the language in which the question is formulated or the language in which the answer is found. Cross-language QA calls for some sort of underlying translating process. At present there are many types of software for natural language translation, several of them available online for free. In this paper we describe the main features of the multilingual Question-Answering (QA) systems, and then analyze the effectiveness of the translations obtained through three of the most popular online translating tools (Google Translator, Prompt and Worldlingo). The methodology used for evaluation, on the basis of automatic and subjective measures, is specifically oriented here to obtain a translation that will serve as input in a QA system. The results obtained contribute to the realm of innovative search systems by enhancing our understanding of online translators and their potential in the context of multilingual information retrieval.

Keywords: *information retrieval, question-answering systems, machine translation, machine translation evaluation.*

1. Introduction

Information Retrieval (IR) is the collection of tasks implemented by the user to locate and access the information sources that are appropriate for the resolution of the information problem proposed. In these tasks, documental languages, abstracting techniques, and the description of the documental object play key roles, largely determining how fast and efficient retrieval is (Belkin & Croft, 1987). Normally there is a balance between the precision and recall of information retrieved. This aspect is increasingly important, as the World Wide Web diffuses vast quantities of new contents every day, with a great variety of formats and languages.

Cross-language question-answering systems can be the solution, as they pursue the search for a minimal fragment of text, not a complete document.

When a need for information arises, a process called the "search strategy" is set in motion, which leads to the supply of documents by the system. This process entails seven basic stages: a) definition of the informational need; b) selection of the information sources to be used; c) translation of the user query expressed in natural language (as we communicate normally) into the indexing language of the information source; d) translation of the expression from the indexing language to the query language of each information system; e) implementation of expressions obtained from the query language; f) results assessment and redefinition of the query expressions if the obtained results are not relevant; and g) selecting and obtaining the documents that respond to the user's needs. In sum, this traditional model of information retrieval entails some implicit restrictions, including: a) the assumption that users search for documents, not answers; and that the documents per se will respond to and satisfy the query, and b) the assumption that the queries and the document that will satisfy the particular informational need are written in the same language.

One step in the evolution toward improved IR resides in the use of question-answering (QA) systems, which pursue the supply of specific data instead of documents, and respond to the questions formulated by users in natural language. If this response derives from documents that are found in other languages, we are talking about a cross-language or multilingual question-answering system. This type of system is particularly complex, as it incorporates the capacities of a cross-language information retrieval (CLIR) system, while also working as a QA system.

Usually QA systems that deal with multiple languages rely on a translation module, as shown in Figure 1. The user enters his specific query, generally including some interrogative adverb (How? When? Where?...) in a given natural source language. This question is translated by an automatic translator. In the stage of query analysis, the QA system examines the user's question and determines what type of information is being demanded. The classification of the questions is key for

Medical Translation

- [How Many Varieties of Medical Practice Are There?](#)
by Rafael A. Rivera, M.D.,
FACP

Science & Technology

- [Translating a Patent: Translator's Templates](#)
by Kriemhild (Karen) Zerling

Translators and the Computer

- [Automatic Web Translators as Part of a Multilingual Question-Answering \(QA\) System: Translation of Questions](#)
by Lola Garcia-Santiago and
María-Dolores Olvera-Lobo
- [The Efficacy of Round-trip Translation for MT Evaluation](#)
by Milam Aiken and Mina
Park

Arts & Entertainment

- [Empirical Study of Subtitled Movies](#)
by Maria Bernschütz, Ph.D.

Literary Translation

- [La influencia de Voltaire en el primer Hamlet español](#)
Laura Campillo Arnaiz

Translators' Education

- [English Language Teaching Through the Translation Method \(A Practical Approach to Teaching Mongolian CPAs\)](#)
by Dr. Naveen K. Mehta

Translators' Tools

- [Pondering and Wondering](#)
by Jost Zetzsche
- [Translators' Emporium](#)

Caught in the Web

- [Web Surfing for Fun and Profit](#)
by Cathy Flick, Ph.D.
- [Translators' On-Line Resources](#)
by Gabe Bokor
- [Translators' Best Websites](#)
by Gabe Bokor
- [Call for Papers and Editorial Policies](#)

the system, as this information will be utilized in the search stage, and in the selection and extraction of the potential responses (García Cumberas et al, 2005). The resulting search expression will be, then, the *input*, or the formulation of the query to be used by the search engine of the system for comparing and matching it with the documents in the database. Once the documents that are relevant to the query are located, the system breaks them up into sections, selects the excerpts that include the candidate responses, and selects a final response. This response, along with its location in the corresponding document, is finally delivered to the user.

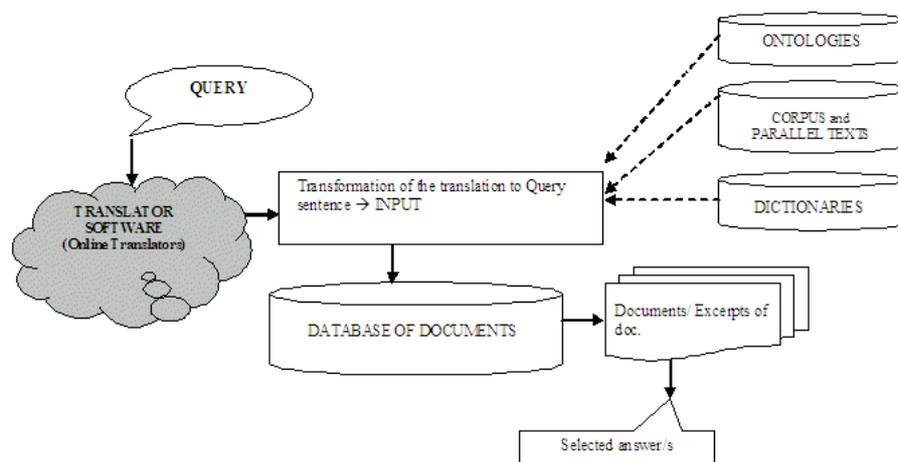


Figure 1: Basic elements of a cross-language QA system

To obtain good results, the questions asked by the users are to be specific, short and good well-structured. Usually the query has an interrogative adverb that seeks an answer to questions about people and institutions, places, dates, quantities,...

Besides the user's interface there is translation software in the cross-language QA systems architecture. Nowadays, automatic translation (AT) shows a number of different facets and views. All of them have in common that this translation must be carried out by software in a more or less automatic way. The rate and quality of this translation can vary. But even the most sophisticated AT systems cannot yet produce translations on a large scale that do not need absolutely any revision by a person. The AT systems also have restrictions about the nature of the texts that they can translate better.

Our study focuses on the first module of the cross-language QA systems, designed to translate the original user query. In the following sections we present a comparative study of the quality of the different automatic translation tools that may be used online for no charge, applying three that translate from German and French into the Spanish language. Our perspective is a documental one; that is, we analyze the functionality of the translator as a mediating instrument in the search for answers. To this end we apply well-known (both objective and subjective) assessment measures of machine translation. Finally, we analyze the results obtained and arrive at some succinct conclusions.

2. Evaluation of online translators as search tools when dealing with translingual questions/responses

One of the objectives of this study is that to identify *which* would be the most adequate online translator for a given QA system entailing a collection of documents in Spanish. In this particular case, the questions would be formulated in French or in German, and as part of the process, they would have to be translated to Spanish in order to constitute system *input* before proceeding. We used a collection of questions with 200 queries in German and another 200 in French. The questions already translated to Spanish by each one of the online translators were both manually and automatically analyzed, applying objective and subjective criteria for the evaluation of automatic translation with the aid of EvalTrans software.

2.1. Online translators evaluated

The *Google Translator*, *Prompt* and *Worldingo* were selected for this study because they allow us to translate and compare results using the language pairs German to Spanish and French to Spanish. Moreover, they are services with a wide diffusion, they are quick in translating, and show reasonable quality at first glance, making them appropriate for a study of this nature. There are limitations regarding the maximum amount of text (from 150 to 300 characters) with which the free online translators can work—except *Google Translator*, which admits much more extensive texts—that do not interfere with the purposes of our study, since a QA system deals with specific questions whose formulation is not that long.

2.2. Sample and types of questions

Our study involved a collection of 200 CLEF (Cross-Language Evaluation Forum) questions formulated in German and in French, supplied for experimentation in the translingual retrieval of information. These questions were formulated in natural language, and attempt to gather precise data regarding a given subject. They may be classified into three types (CLEF, 2008):

- **Factual questions:** These solicit data on the person, place, time, organization, measurement,

recount, or object, among others (Who? When? Where?, etc.)

- **Definitional questions:** The information requested is a bit more open in terms of synonymy and formulation with respect to a person, object, organization or concept (What/who is it?)
- **Closed list:** Questions that call for a response enumerating various data (What/is...? How...? or Name...)

2.3. Measures for evaluating the automatic translation of questions

Evaluation of machine translation is an unresolved research problem that has been addressed by numerous studies in recent years. The most extensively used assessment tools are classified into two major groups: automatic objective methods, and subjective methods (Tomás, Mas & Casacuberta, 2003). The objective evaluation methods compare a set of correct translations of reference against the set of translations produced by the translation software under evaluation. The units of measurement most often used work at the lexical level, comparing strings of text.

Our study evaluated the online translators in light of the following parameters based on the comparison of the Levenstein distance or the edit distance between the two strings of characters: WER (*Word Error Rate*); aWER (*all references WER*); SER (*Sentence Error Rate*); and aSER (*all references SER*) (Tillman et al., 1997; Tomás, Más and Casacuberta, 2003; Vidal, 1997

All the means of measurement mentioned are applied automatically. Therefore, the translations and the reference phrases are compared without any specific determination of the type of error or discrepancy occurring between the two strings under consideration. For this purpose, there exist other types of metrics that require human intervention for the evaluation. In the context of cross-language QA systems that include machine translation, the aim of translation is more practical; so other evaluation measures of a subjective nature, such as sSER (*Subjective Sentence Error Rate*), were applied. Again, our aim was not to find a "perfect" translation but rather a translation capable of maintaining the characteristics of the questions, so that the QA system could locate appropriate responses.

The evaluation process was carried out using EvalTrans software (following Nießen et al, 2000) in its graphic version designed for use with Windows (Tomás, Mas and Casacuberta, 2003). This tool can be used online for free for evaluating automatic translation.

3. Analysis of results

The results of analysis of the online translators includes the values obtained applying the measurements described above, and the values that resulted from the human assessment of each question translated.

Tables 1 and 2 show the values in terms of WER and SER for the *Google Translator*, *Promt*, and *Wordlingo* in automatic evaluations, from German and French, respectively, into Spanish.

GERMAN	GOOGLE	WORDLINGO	PROMT
WER	41.9%	57.6%	54.4%
SER	95%	98.5%	99%

Table 1: Automatic evaluation of the translations from German to Spanish

FRENCH	GOOGLE	WORDLINGO	PROMT
WER	43.2%	40.8%	39.6%
SER	95.5%	93%	90%

Table 2: Automatic evaluation of the translations from French to Spanish

The high values obtained for the rate of error from the SER phrases can be attributed to the need for the translator to find an identical string (with the same words, and in the same order) as in the reference phrase. Any variation, even a minor one, is interpreted as an erroneous phrase and is left out. In tables 3 and 4 below we see that the sSER measure aspires to amending the deficiencies of the SER measurement, since it is based on the evaluation proceeding from human supervision and the corresponding acceptance or not of the phrase supplied by the online translator that is judged as correct or incorrect. In our case, however, the aim is to identify the best translating software among the three tested, understood as the one that generates an input found to be functional for a cross-language QA system.

In contrast, the coefficients corresponding to aWER, sSER and aSER do indeed vary in conjunction with human intervention (see Tables 3 and 4). For instance, the sSER measurement takes the scores for each one of the phrases already translated and evaluated. The aWER measurement, meanwhile, gathers all the reference phrases that have been considered subsequently as such by a human translator. These tend to be proposed by the human evaluator as *new reference* after the reduction of the edit distance; or else, a candidate phrase is scored with a maximum mark. The evaluating

program adopts the reference phrase that is most similar to the group of reference phrases already existing, not only in regard to the first sentence of reference sentence included a priori. And as explained previously, the aSER rate determines the percentage of sentences that do not coincide precisely with those of the reference sentences in view of all the existing references, and not just the first reference.

The ranking of the online translation programs analyzed with regard to their effectiveness in the translation of specific questions is established on the basis of the means of measurement commented upon here. Therefore, the best translator of the three would be the one showing the lowest rate (indicating lower occurrence of errors), especially evident with sSER and aWER, taking human assessments into account.

Due to the fact that the applied measures for automatic evaluation do not carry out a thorough or comprehensive syntactic analysis (noting the position of the words in the phrase), the error rates are seen to be greater in German. As explained earlier, the edit distance works in terms not only of the existence of words in the sentence, but also their position, leading to higher error rates when the German language is involved: any alteration in the order of the elements in the phrase is identified as an error (Tillman et al., 1997). Grammatical similarities between the French and Spanish languages lead to a more meager harvest of errors (see Tables 1 and 2). Noteworthy is the fact that only in the case of *Google Translator* were the error rates in conjunction with words (WER) higher for French than for German.

Having carried out the subjective assessments of the translations, the results were as follows (obviously, the WER, and SER rates remained invariable):

GERMAN	GOOGLE	WORDLINGO	PROMT
aWER	57.6%	54.6%	50.4%
sSER	90.2%	91.3%	77%
aSER	88%	94%	91%

Table 3: Indicators calculated with human assessment of the translations from German to Spanish

FRENCH	GOOGLE	WORDLINGO	PROMT
aWER	36.1%	29.7%	27.5%
sSER	70.4%	53.7%	55.5%
aSER	87.5%	78.5%	75%

Table 4: Indicators calculated with human assessment of the translations from French to Spanish

Practically all the values are seen to decrease with human assessment, meaning the error rates are reduced. The consideration of various alternatives as acceptable leads to a greater yield of reference questions for calculating aWER and aSER.

Likewise, the percentages derived from the errors per phrase, whether subjective (sSER) or automatic (SER and aSER), are smaller for the translation from French than from German. One of the reasons would be that these means of measurement do not register any coincidence of words when the automatic translator has not maintained the exact word order as presented within the reference phrase.

For the manual evaluation of the translations generated by automatic online translators, we applied the Likert scale, using six levels. Taking into account the finality of the translation, the assessment implied that errors such as the position of the elements in the string would not have to be penalized to the same degree as ambiguity, or the loss of some characteristic of the question (interrogative adverb, or the entity to which the question refers, among others). These values were then used to calculate the rates given below.

As we saw in the section above, according to the sSER rate, in the case of German, *Prompt* (77%) is the best translator, followed by the *Google Translator* (with 90.2%). In contrast, when dealing with French, *Wordlingo* (53.7%) is the best translator, although *Prompt* (55.5%) has a very similar sSER rate.

4. Conclusions

We carried out a study of automatic online translators *Google Translator*, *Prompt* and *Wordlingo*, applying different means of evaluation. It is shown how strictly automatic evaluation (in the absence of subjective assessment) produces high rates of error that are not highly representative. As commented in the text, the error rates obtained through automatic evaluation are higher when the translation is from the German language into Spanish, because the most frequently used

measurements for evaluating translations utilize indicators that compare word-by-word, looking for the very same order of elements in the translation produced online as in the initial reference phrase. Therefore, the "detected" syntactic errors are more numerous because of basic grammatical differences between the source language (German) and the target language (Spanish). Also greater are the error rates resulting from subjective assessment as obtained here, owing to the capacities of the tools themselves in translating from German into Spanish. The grammatical similarities between French and Spanish tend to produce a lower rate of error.

It would moreover prove beneficial if the different tools now being used or developed for the evaluation of translations—such as EvalTrans—and the various research studies undertaken were to use the same scale of human assessment. This would make it easier to introduce data and to quantify the measures that human assessments (like sSER) apply.

The results reported here show that the tools and linguistic resources used by automatic translators for German-to-Spanish translations are more limited and less efficient than the French-to-Spanish online translators.

In future studies our research team will follow this line deeper into the design of efficient and effective multilingual QA systems.

References

Belkin, N.J. & Croft, W.B. 1987. Retrieval techniques. *Annual Review of Information Science and Technology*. 22, 109-146

CLEF (2008). GUIDELINES for the PARTICIPANTS in QA@CLEF 2008. Available online: http://nlp.uned.es/clef-qa/index.php?option=com_content&view=article&id=16&Itemid=39
[Last access: 22nd December 2009].

García Cumbereras, M. Á. et al. 2005. Búsqueda de respuestas multilingüe. Clasificación de preguntas en español basadas en aprendizaje. *Procesamiento del lenguaje natural*, ISSN 1135-5948, N^o. 34, 2005 pp. 31-40. Available online: <http://dialnet.unirioja.es/servlet/articulo?codigo=1173352>. [Last access: 22nd December 2009].

Nießen, S. et al. 2000. An Evaluation Tool for Machine Translation: Fast Evaluation for MT Research. In: *Proceedings of the 2nd International Conference on Language Resources and Evaluation*. Available online: http://www-i6.informatik.rwth-aachen.de/PostScript/InterneArbeiten/Paper_2000/Niessen_Evaluation_LREC_2000_afterSubmissionCorrection.ps
[Last access: 23rd November 2009]

Tillman, C. et al. 1997. A DP based Search Using Monotone Alignments in Statistical Translation. In: *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, Madrid, Spain.

Tomás, J., Mas, J.A. & Casacuberta, F. 2003. A Quantitative Method for Machine Translation Evaluation. *Proceedings of the of EACL 2003 workshop on Evaluation Initiatives in Natural Language Processing*, 2003, Budapest, Hungary. Available online: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.12.8410&rep=rep1&type=pdf>
[Last access: 23rd November 2009]

Vidal, E. 1997. Finite-State Speech-to-Speech Translation. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, Munich, Germany.

Websites

Cross-Language Evaluation Forum (CLEF). Available online: <http://www.clef-campaign.org/>

EvalTrans. Available online: <http://www-i6.informatik.rwth-aachen.de/web/Software/EvalTrans/index.html>

Google Translator. Available online: http://www.google.es/translate_t?hl=es [Consulted: 4th May 2008].

Prompt Translator. Available online: <http://www.online-translator.com/>

Text REtrieval Conference (TREC). Available online: <http://trec.nist.gov>

Worldlingo. Available online: http://www.worldlingo.com/en/products_services/worldlingo_translator.html>