

Are Users Willing to Search Cross-Language? An Experiment with the Flickr Image Sharing Repository

Javier Artiles, Julio Gonzalo, Fernando López-Ostenero and Víctor Peinado*
NLP Group, ETSI Informática, UNED
c/ Juan del Rosal, 16, E-28040 Madrid, Spain
javart@bec.uned.es, {julio, flopez, victor}@lsi.uned.es

Abstract

This paper summarizes the participation of UNED in the CLEF 2006 interactive task. Our goal was to measure the attitude of users towards cross-language searching when the search system provides the possibility (as an option) of searching cross-language, and when the search tasks can clearly benefit from searching in multiple languages.

Our results indicate that, even in the most favorable setting (the results are images that can be often interpreted as relevant without reading their descriptions, and the system can make translations in a transparent way to the user), users often avoid translating their query into unknown languages.

Categories and Subject Descriptors

H.3 [Information Storage and Retrieval]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval; H.4 [Information Systems Applications]: H.4.m Miscellaneous

General Terms

interactive information retrieval, cross-language information retrieval

Keywords

CLEF, iCLEF, Flickr, online photo sharing, multilingual image search, user studies

1 Introduction

CLEF,¹ NTCIR² and TREC³ evaluation campaigns have contributed, along the years, to create an extensive corpus of knowledge on Cross-Language Information Retrieval from an algorithmic perspective. Little is yet known, however, on how users will benefit from Cross-Language retrieval facilities.

iCLEF⁴ (the CLEF interactive track) has been devoted, since 2001, to study Cross-Language Retrieval from a user-centered perspective. Many things have been learned in the iCLEF framework about how a system can best assist users when searching cross-language. But all iCLEF

*Authors are listed in alphabetical order

¹See <http://www.clef-campaign.org/>.

²See <http://research.nii.ac.jp/ntcir/>.

³See <http://trec.nist.gov/>.

⁴See <http://nlp.uned.es/iCLEF>.

experiments so far were slightly artificial, because users were forced to search in a foreign language. Previous UNED experiments [2, 3, 4], for instance, consisted of native Spanish speakers which were asked to search a news collection written entirely in English. The task was artificial because the contents they were searching were also available in other news collections in Spanish, their native language.

iCLEF 2006 proposes a radically new task, which consists of searching images in a naturally multilingual database, Flickr,⁵ which has millions of photographs shared by people all over the planet, tagged and described in a mixture of most languages spoken on earth.

We have used the iCLEF 2006 task design to find out how users react to a system that provides cross-language search facilities on a naturally multilingual image collection. If searching cross-language is an option of the system (rather than a requisite of an experiment), will users take advantage of this possibility? How will their language skills influence their use of the system? How will the nature of the search task influence the degree of multilinguality they achieve when searching? Will there be an inertial effect from the dominant search mode (exact match, all words conjunctively, no expansion/translation) used by all major search engines? Do they perceive the system as useful for their own search needs?

Rather than focusing on the outcome of the search process, we have therefore focused on the search behavior of users. We have designed a multilingual search interface (a front-end for the Flickr database) where users can search in three modes: *no translation*, *automatic translation* in the languages selected by the user, and *assisted translation*, where users can change the translations initially picked up by the system. Our users have conducted the three search tasks prescribed by the iCLEF design [1], and we have studied (through observations and log analysis) their search behavior and their usage of cross-language search facilities. Finally, we have also contrasted this information with the subjective opinions of the users, stated in a post-experiment questionnaire.

2 Experiment Design

Our experiment follows iCLEF guidelines [1]. In summary:

Test collection The collection to be searched is all public photos in Flickr uploaded before 21 June 2006. This date is fixed so that everyone is searching exactly the same collection. This is a collection of more than 30 million photographs annotated with title, description and tags. Tags are keywords freely chosen by users; community usage of tags creates a so-called “folksonomy”.

Access to the collection The collection can be accessed via Flickr’s search API,⁶ which allows only two search modes: search tags (either all tags in the query or any tag in the query) and full search (search all query terms in title, description and tags). No exact statistics on the collection (images, size, vocabulary, term frequencies) were available for the experiment.

Search tasks iCLEF guidelines prescribed at least three tasks to be performed by users, which could employ a maximum of twenty minutes per task:

- Ad-hoc task: *Find as many European parliament buildings as possible, pictures from the assembly hall as well as from the outside.*
- Creative task: *Find five illustrations to the article “The story of saffron”, a one-page text about cultivation of saffron in Abruzzo, Italy.*⁷
- Visually oriented task: *What is the name of the beach where this crab is resting?, along with a picture of a crab lying in the sand.*⁸

⁵See <http://www.flickr.com>.

⁶For further details about the Flickr’s API and its documentation see <http://www.flickr.com/services/api/>.

⁷The English version of the text is available at <http://nlp.uned.es/iCLEF/saffron.txt>.

⁸The picture is available at <http://nlp.uned.es/iCLEF/topic3.jpg>.

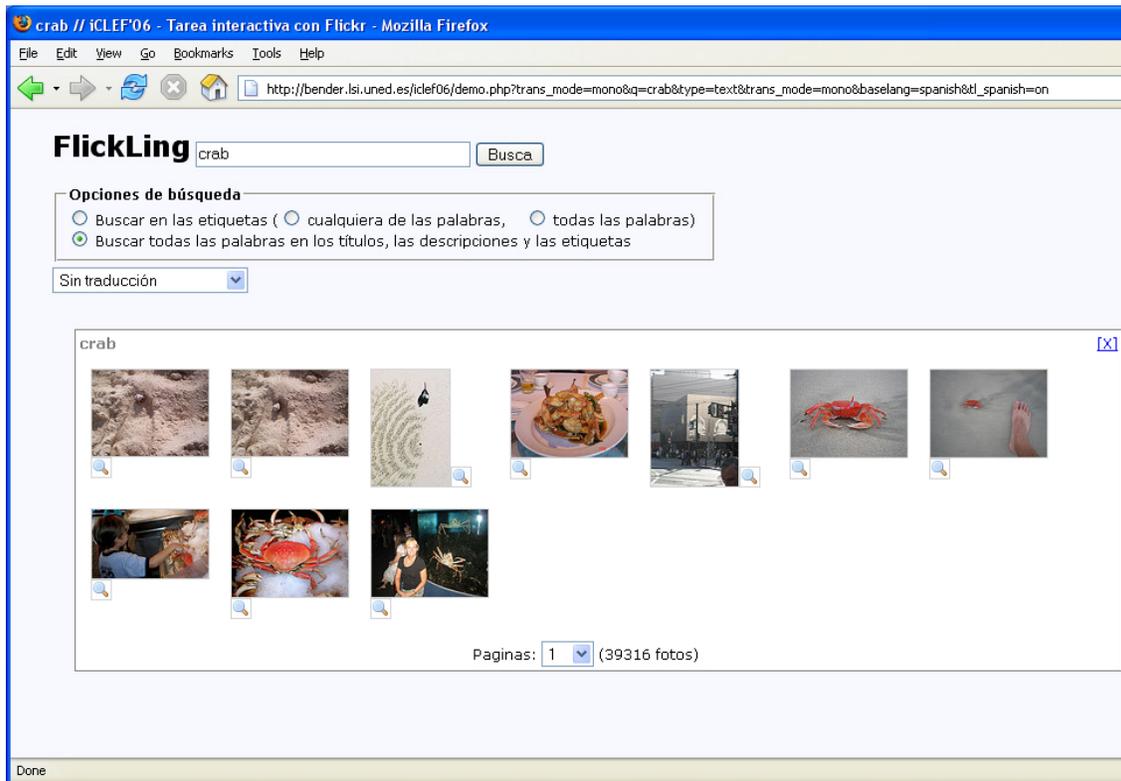


Figure 1: Search interface, *no translation* mode.

The name of the beach is included in the Flickr description of the photograph, so the task is basically finding the photograph, which is annotated in German (a fact that the users ignore) and identifying the name of the beach.

All tasks can benefit from a multilingual search: Flickr has photographs of European parliament buildings described in many languages, photographs about the Abruzzo area and saffron are only annotated in certain languages, and the crab photograph can only be found with German terms.

With these constraints, we have designed an experiment that involves:

- 22 users, all of them native Spanish speakers and with a range of skills in other languages.
- A search front-end to Flickr (see Figure 1).
- A pre-search questionnaire, asking users about their experience searching images and using Flickr, and about their language skills in the six languages proposed by our interface.
- A post-search questionnaire, asking users about the perceived usefulness of cross-language search facilities and the degree of satisfaction with the results.

The key of the experiment is the design of the search front-end with Flickr. It has three search modes: *no translation*, *automatic translation*, and *assisted translation*. Users can switch between these search modes at will. In both translation modes, users can select Spanish, English, French, Italian, Dutch and German as source language, and any combination of them as target languages.

In the *no translation* mode, the interface simply launches the queries against the Flickr database using its API. Figure 1 shows how the interface displays results for the query “crab”. Users may

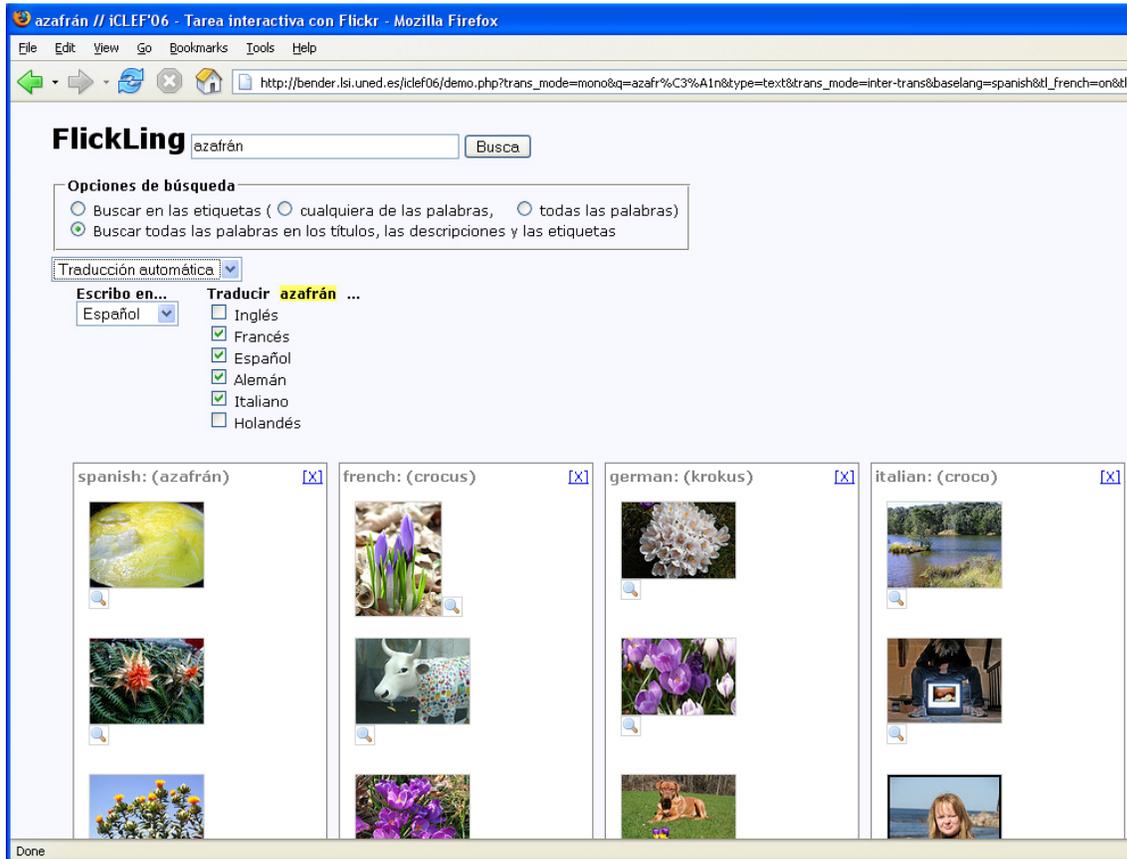


Figure 2: Search interface, *automatic translation* mode.

choose between searching only the tags, or searching titles, tags and descriptions (full text). In the tags mode, they can select a conjunctive or a disjunctive search mode. In the full text search mode, the search is always conjunctive. These are the original search options with Flickr’s API.

In the *automatic translation* mode, the user can choose the source language and the target languages. In Figure 2, the query “azafrán” (saffron) has been launched with translations into French, Spanish, German and Italian. The system shows four result boxes, one per language, together with the query as it is formed in every language. The user cannot manipulate the translation process directly. As the full text search is always conjunctive in the Flickr API, we chose to translate each original term in the query with only one term in every target language, to avoid over-restrictive queries.

In the *assisted translation*, the user has the additional possibility of changing the default translations chosen by the system. A code color is used to signal alternative translations coming from the same source term. The user can change the preferred translation for every term as desired; the query is automatically re-launched (only in the language where it was refined) with the new term. Figure 3 shows the results for the query “parlamento Oslo” translated into all languages except Italian. In German, for instance, “parlament” is chosen as translation for “parlamento”, but the user might change this term to “unterredung” or “verhandlung”. As a help to select the most appropriate translation, the system displayed inverse translation when moving the mouse over a translated term.

For the experiment, we have implemented three versions of this search interface which facilitate performing each task and providing results for the user. For instance, Figure 4 shows a snapshot of the dedicated interface for the “find the crab” visually-oriented task. In the left-hand side of the interface, the original crab image is always shown, together with the task description. Immediately

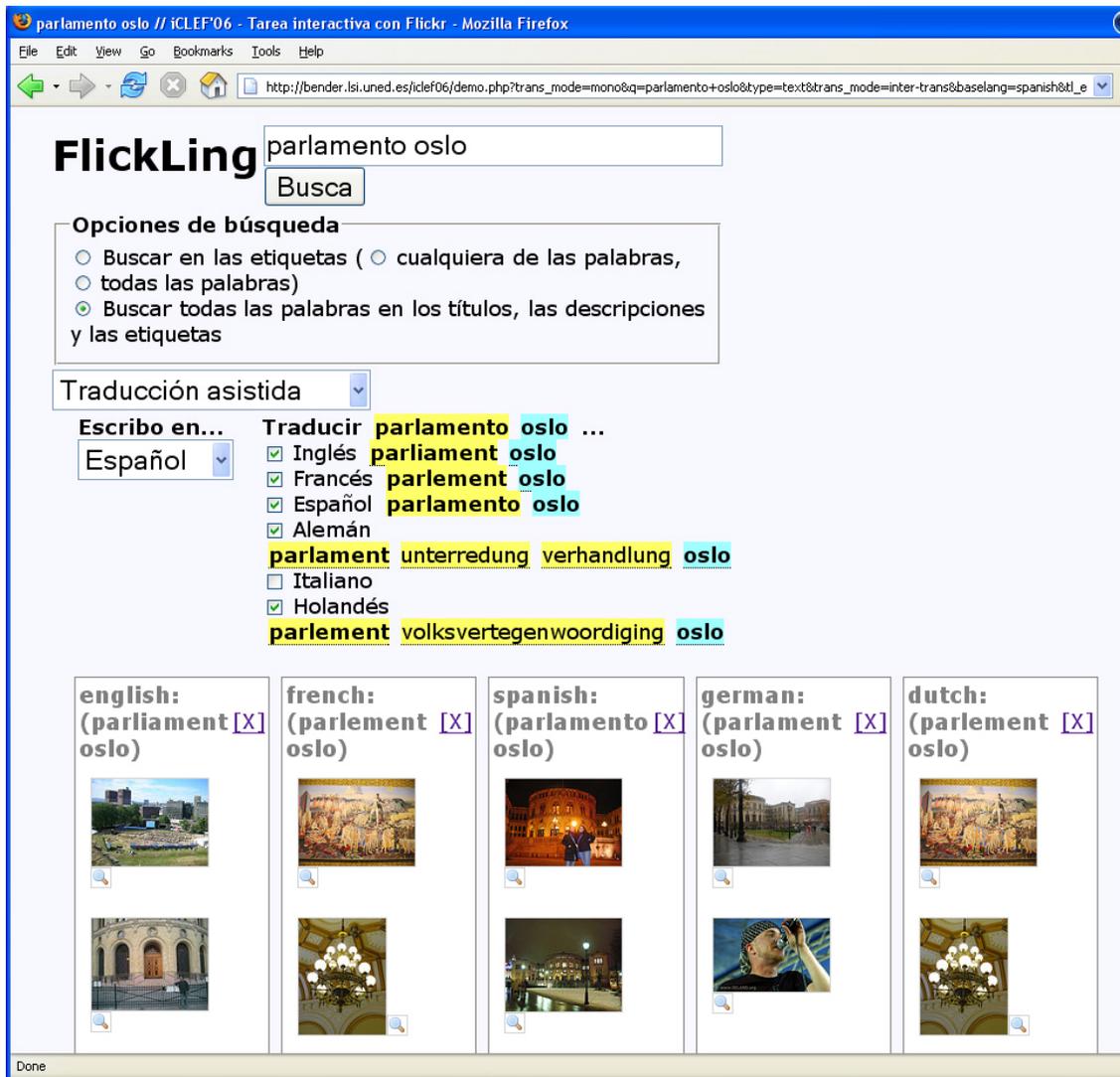


Figure 3: Search interface, *assisted translation* mode.

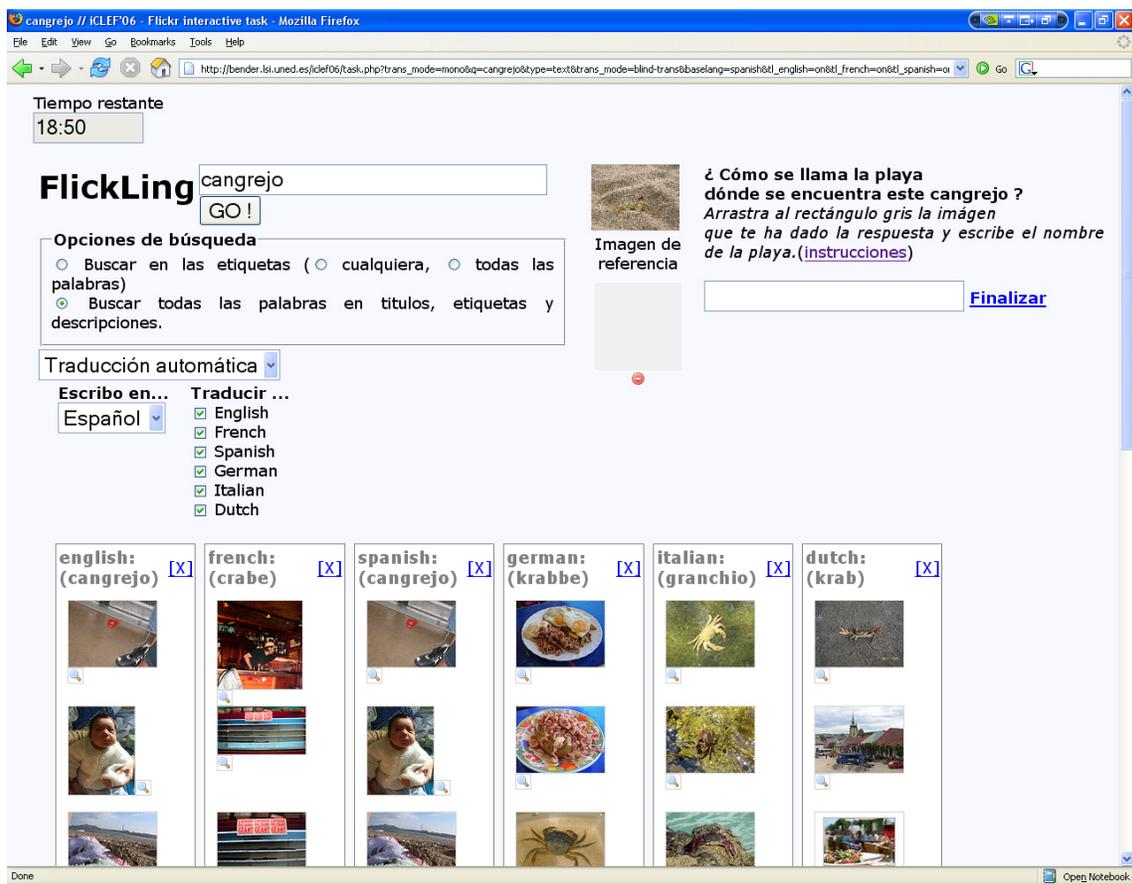


Figure 4: Dedicated search interface for the “find the crab” task.

below there is an empty box where the user must drag-and-drop the image that provides the answer (name of the beach), a text box where s/he can write the answer, and a “Done” button to end the task. In the upper left corner, a clock indicates how much time is left to complete the task.

For translation we have only used freely available resources. We have built databases for all language pairs from bilingual word lists (with approx. 10,000 entries per language) provided by the Universal Dictionary Project⁹. All entries were stemmed using a specific Snowball stemmer¹⁰ for each language. The *automatic translation* mode simply picks up the first translation in the dictionary for every word in the query. If a given word is not found in the database, it remains untranslated.

3 Results and Discussion

3.1 Comparison among Monolingual, Automatic Translation and Assisted Translation Search Modes

As explained above, our interface implements three different search modes: *no translation*, which launches the query as it is, *automatic translation* where users can select target languages and the systems picks up the first translation for each query term, and *assisted translations*, where users can also select target translations.

Figure 5 shows the proportion of total subjects using every search mode for each task. It seems that the *assisted translation* was the most popular search mode across tasks. It allowed both to select the right translation for each query term when the *automatic translation* was not accurate enough and to learn or remember new equivalent translations.

It is noticeable that most users begin the activities with the *no translation* mode (the system default) but it’s discarded relatively soon (mostly, after the first minute) as soon as subjects realize that a translation facility is needed to complete a task. This fact is confirmed in Figure 5a. Specifically, for the “find this crab” task, we can also see how the figures tend to lower as the time goes by: the number of subjects still on search decreased because some users finished the task before the 20-minute limit.

For the “European parliaments” task (Figure 5b), we can see that the number of subjects on search remains stable: most of the users employed the available 20 minutes in order to grab as many photographs as possible.

In the “saffron” task, in Figure 5c, it is interesting to note that the *no translation* mode in higher than in other tasks. This seems to point out that, for a creative activity, some users do not consider essential to retrieve pictures annotated in other languages.

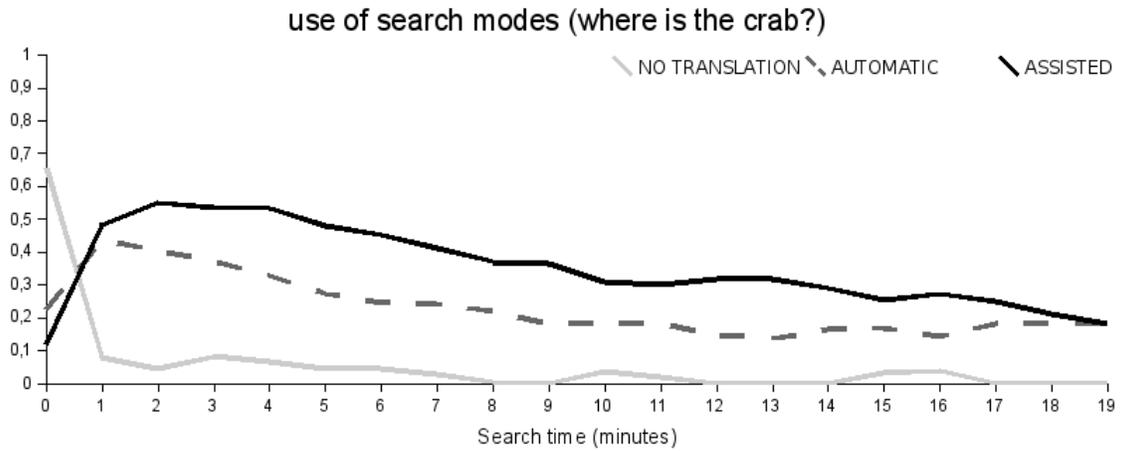
3.2 Comparison of Searches in Native, Familiar and Unknown Languages

In Figure 6 we show the ratio of subjects using a *native*, *passive* or *unknown* language as target during each task. For comparison purposes, we have classified subjects according to the information compiled in the pre-session questionnaires as follows: the users who declared to be native or fluent speakers are represented as “native”. The ones who declared to be able to read but not to write fluently are represented as “passive” speakers. Finally, the ones who claimed not to understand a single word are shown as “unknown”.

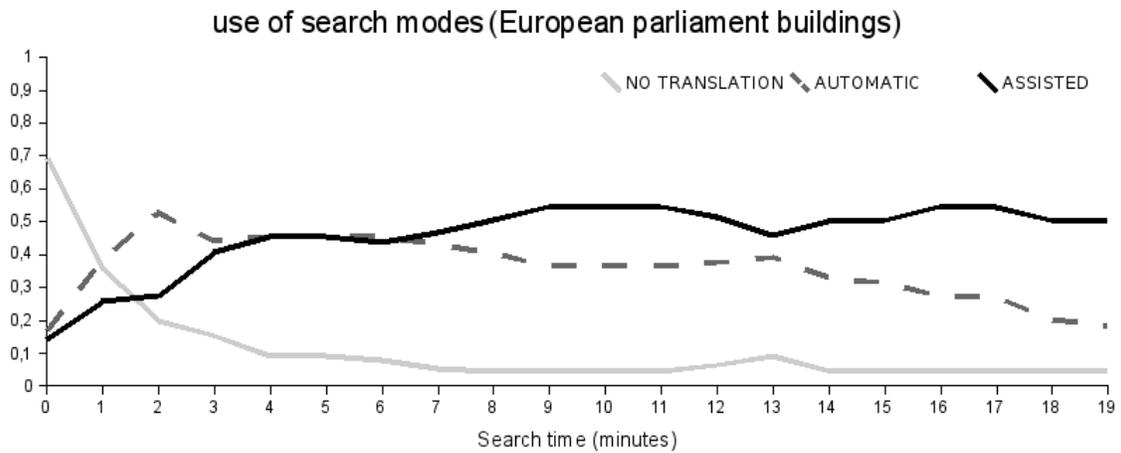
For the “find this crab” task, as shown in Figure 6a, we can see that users preferred to perform their searches using languages in which they may feel confident rather than using unknown languages. Notice that there is only one picture containing the right answer and, therefore, this is clearly a precision-oriented task. The photograph was annotated in German, which is an unknown language for most of our users. Since they were not willing to search in German, that made the task hard to complete. Indeed, only 9 out of 22 subjects were able to find the crab. As subjects

⁹Dictionaries can be downloaded from <http://dicts.info/udd1.php>.

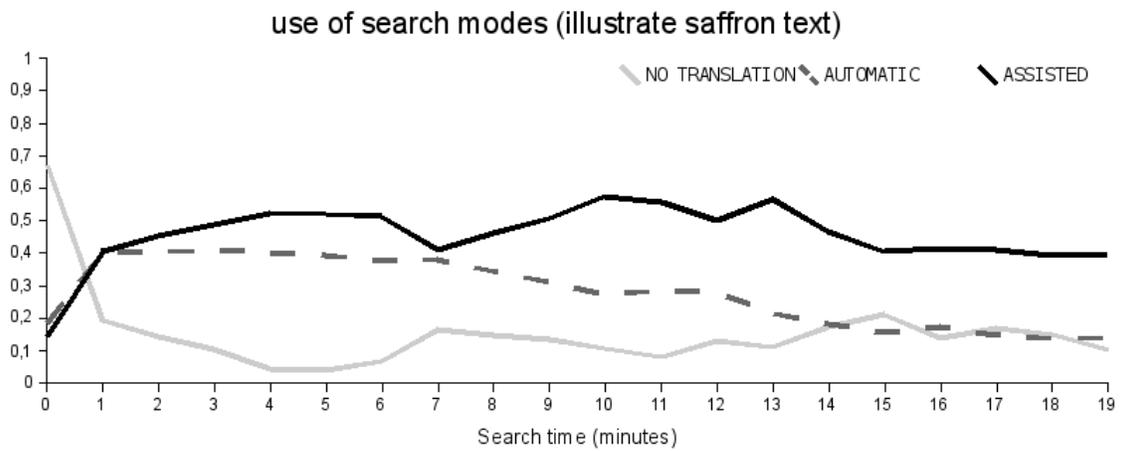
¹⁰See <http://www.snowball.tartarus.org/> for further information about Snowball.



(a)

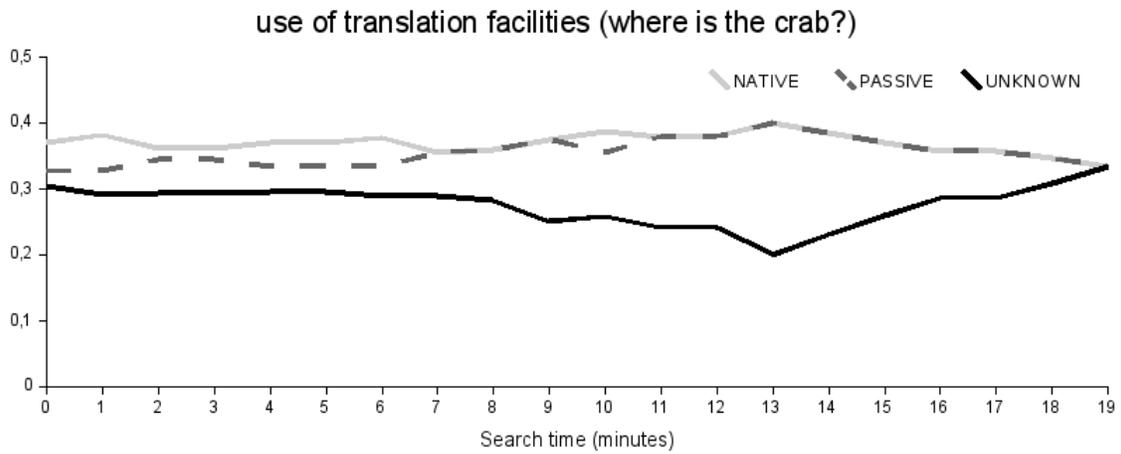


(b)

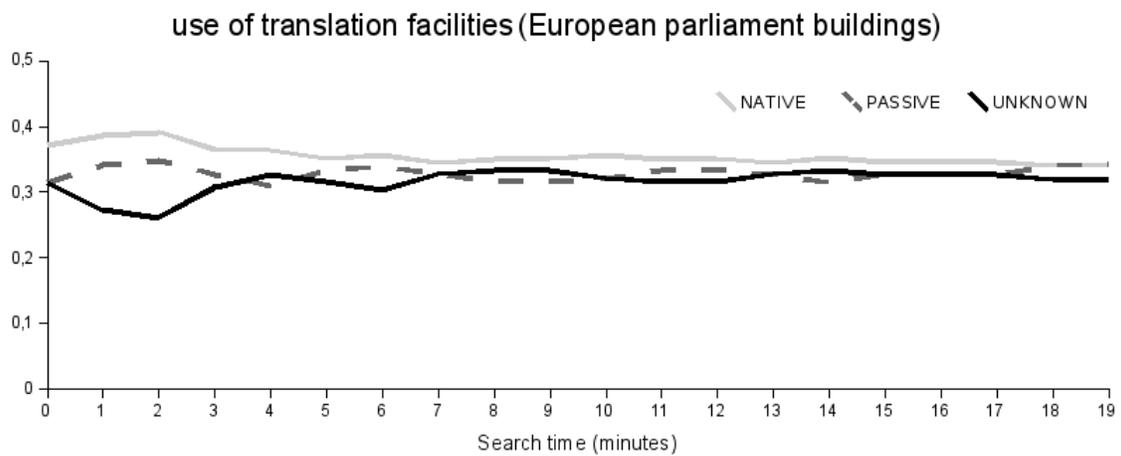


(c)

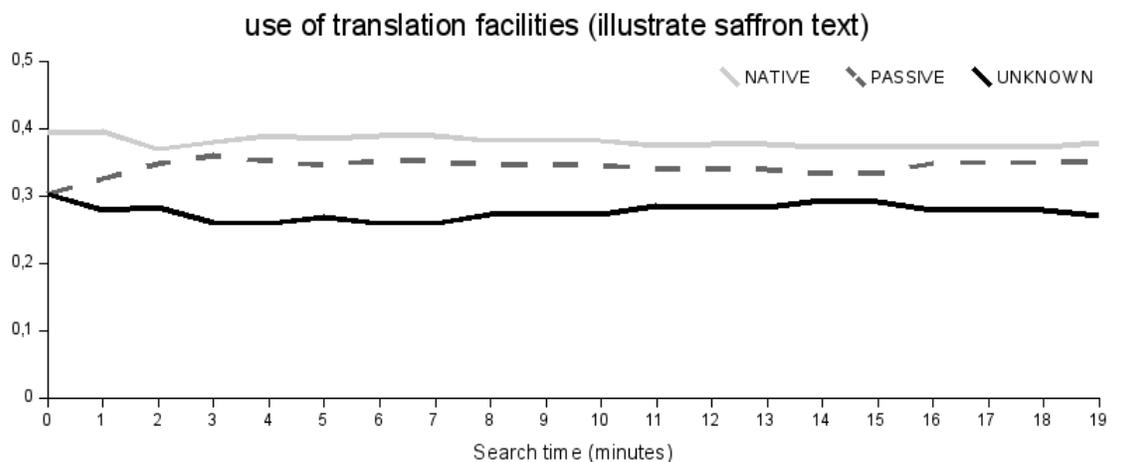
Figure 5: Use of the search modes implemented in the system: *no translation*, *automatic translation* (users select target languages) and *assisted translations* (users can also select target translations).



(a)



(b)



(c)

Figure 6: Use of target languages classified according to self-declared user language skills.

approach the time limit, the use of unknown languages rises up. This can be interpreted as a final attempt to come up with the crab.

On the contrary, the “European parliaments” task can be considered as a recall-oriented activity. Figure 6b shows the same proportion of subjects performing searches over known and unknown target languages, because they tried to grab pictures from any possible source.

Lastly, the “saffron” task is the most creative one. It is hard to evaluate whether our subjects performed well or not. They had to choose photographs related to the content in order to enhance the text. As said in section 3.1, most of the users considered enough to search in known languages (see Figure 6c).

3.3 Observational Study

Twelve out of twenty-two subjects performed the experiments remotely; Therefore only ten people were observed during the experiment and gave input for the observational analysis.

These are the most remarkable facts applicable to all tasks:

- The task is strongly visual: in general, users decide on relevance with just a quick look at the image thumbnail.
- Users feel more confident when searching in languages they know. More specifically, our Spanish users were reluctant to search in Dutch or German in spite of the relevance judgments being mostly visual.
- Remarkably, our users seemed to assume they could find everything in English. Only after some minutes searching they realized that this was not the case in our experiment.
- As the user acquires experience with the front-end, s/he uses the *assisted translation* search mode more often.
- The more experience the user has, the more s/he notices the mistranslations and search options. In general, however, the possibility of changing translations was largely unexploited.
- Most of the queries were submitted in Spanish. English (the most popular foreign language for our users) and Italian (which is very close to Spanish) were also used.
- We have identified three broad types of search strategies: i) a “depth first” strategy in which the user poses a general, broad query and then exhaustively inspects all the results; ii) a “breadth first” strategy, where the user makes many query variations and refinements, and makes only a quick inspection of the first retrieved results; iii) a “random” behavior, where the actions taken by the user can only be explained by “impulse”.

Now let’s examine some task-dependent remarks:

Find the crab

- Most of the users tagged German as an unknown language and some of them did not select it as a target language, in spite of the fact that it was necessary to search over German pictures to find the right one.
- Some of the users learned the right location of the picture (Imbassai) by browsing across Flickr tags, using our front-end or even searching in Google. Regarding the nine people who found the correct answer, it took them, on average, around three minutes to confirm that Imbassai was indeed the name of the beach.
- The picture could be located in the top ten results by using the keywords “cangrejo arena” (crab sand) and selecting German as target language. This was the most common way of finding the image, and corresponds to the “breadth first” strategy. There was, however, some users that made a more general query and spend some time inspecting the first one hundred results.

European parliaments

- The best strategy was to combine the term “parliament” together with country or city names. Most of our users tried successfully this approach; some of them even made use of maps of Europe found within Flickr or using Google.
- In order to retrieve as many pictures as possible, some of the users did not pay attention to the description of the pictures, and they decided to grab all pictures looking like an official building.

The saffron text

- Most of the queries were submitted in Spanish, English and Italian as target languages. This was the only query in which users could make some presumptions on which languages would find more results.
- Some users tried to find pictures by searching the place names or proper nouns present in the text (such as towns, cities, regions and family names), sometimes combining them with “saffron” and other food names.
- The selected pictures usually were flowers, crop fields, Abruzzo landscapes and risotto dishes.
- Just a few users noticed that the correct Italian translation of “saffron” was “zafferano” (which was not offered by the translation system), which can be learned after reading the description of some of the pictures.

3.4 User Questionnaires

Users were asked to fill in two different questionnaires. In the pre-search questionnaire, we asked users about their age and education, their previous experience searching images in the Web and using Flickr, and their language skills in the six languages proposed by our interface.

Two out of twenty-two users were 20-25 years old, twelve were 25-35 and eight were 35-45. Among them, there were four people with PhD studies, fourteen with different university degrees and four with high school diplomas. All of them were native Spanish speakers. Most people had some English skills and passive knowledge of other languages (mainly French and Italian), and most of them declared Dutch and German to be cryptic to them.

Lastly, after the experiment, users filled in an additional questionnaire about the perceived usefulness of the cross-language search facilities for every task and the degree of satisfaction with their performance.

Figure 7 sums up the data compiled in this last questionnaire. As shown in the first set of bars, most users were rather pleased with their overall performance, with slight differences among tasks. Then, we can also see that the more precision-oriented the task was, the more the users needed cross-lingual mechanisms. Lastly, for the “crab” and the “saffron” tasks, most subjects declared translation facilities did not help improve their results: in the first case, some users felt frustrated and they might have underestimated the usefulness of the multilingual search; in the second one, being an open-ended task it was not strictly necessary to search in other languages to complete a reasonable job.

4 Conclusions

In this paper, we presented the participation of UNED in the CLEF 2006 interactive task. Our goal was to measure the attitude of users towards cross-language searching when the search system provides the possibility (as an option) of searching cross-language, and when the search tasks can clearly benefit from searching in multiple languages.

Our results indicate that, even in the most favorable setting (the results are images that can be often interpreted as relevant without reading their descriptions, and the system can make

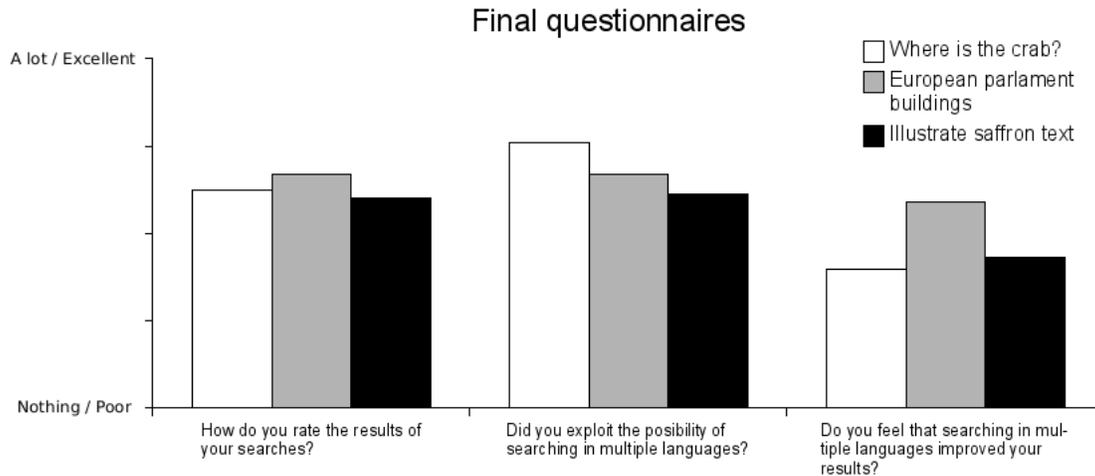


Figure 7: Results of post-experiment questionnaire.

translations in a transparent way to the user), users often avoid translating their query into unknown languages. On the other hand, the learning curve to use cross-language facilities was fast. At the end of the experience, most users were using multilingual translations of their queries, although they rarely interacted with the system to fix an incorrect translation.

Acknowledgments

This work has been partially supported by the Spanish Government under project R2D2-Syembra (TIC2003-07158-C04-02) and by the European Commission, project MultiMatch (FP6-2005-IST-5-033104). Javier Artiles and Víctor Peinado hold PhD grants by UNED (*Universidad Nacional de Educación a Distancia*).

References

- [1] Clough, P., Gonzalo, J., Karlgren, J.: iCLEF 2006 overview: searching the Flickr WWW photo-sharing repository. Working Notes of CLEF 2006 (this volume). 2006.
- [2] López-Ostenero, F., Gonzalo, J., Verdejo, F.: UNED at iCLEF 2003: Searching Cross-Language Summaries. In: Comparative Evaluation of Multilingual Information Access Systems. Results of the CLEF 2003 Evaluation Campaign. Springer Verlag. Lecture Notes in Computer Science, vol. 3237. 450-461. 2004.
- [3] López-Ostenero, F., Gonzalo, J., Peinado, V., Verdejo, F.: Interactive Cross-Language Question Answering: Searching Passages versus Searching Documents. In: Results of the CLEF 2004 Evaluation Campaign. Springer Verlag. Lecture Notes in Computer Science, vol. 3491. 323-333. 2005.
- [4] Peinado, V., López-Ostenero, F., Gonzalo, J., Verdejo, F.: UNED at iCLEF 2005: Automatic highlighting of potential answers. In: 6th Workshop of the Cross-Language Evaluation Forum (CLEF 2005). Revised Selected Papers. Springer Verlag. Lecture Notes in Computer Science, vol. 4022. 2006.