

# A Study of Users' Image Seeking Behaviour in Flickling

Evgenia Vassilakaki, Frances Johnson, R.J. Hartley, David Randall  
Dept. Information & Communications  
Manchester Metropolitan University  
evgenia.vassilakaki@student.mmu.ac.uk,  
f.johnson/r.j.hartley/d.randall@mmu.ac.uk

## Abstract

This study aims to explore users' image seeking behaviour when searching for a known, non-annotated image in Flickling provided by iCLEF2008 track. The task assigned to users was to search for the three first images given after first login. Users did not know in advance in which of the six languages (English, German, Dutch, Spanish, French, Italian) the images were described, forcing them to search across languages. The main focus of our study was threefold: a) to identify the reasons that determined users' choice over a specific interface, b) to examine whether users were thinking about languages when searching for images and to what extent and c) to examine if used, how helpful the translations proved to be for finding the images.

This study used four different, both quantitative and qualitative methods (questionnaires, retrospective thinking aloud, observation and interviews) to meet its research questions. Results show that two out of ten users were using only the monolingual interface because they did not feel confident with languages and the rest were switching between interfaces for a variety of reasons in which languages played a small part. Only four out of ten users were actually thinking about languages when searching for the images, while the rest were more preoccupied with finding the images and completing the task successfully. As a consequence, only four users paid attention to translations and only judged the translations in languages known to them. Overall, the translations were not considered to be helpful due to their inconsistency in coverage and their tendency to lead to irrelevant results.

## Categories and Subject Descriptors

H.3 [Information Storage and Retrieval]: H.3.3 Information Search and Retrieval; H.2.3 [Database Management]: Languages—*Query Languages*

## General Terms

Languages, Human Factors, Experimentation

## Keywords

Multilingual Information Retrieval, User Behaviour, User Image Seeking Behaviour, Flickr, Flickling, iCLEF

# 1 Introduction

Cross Language Evaluation Forum (CLEF) is an annual evaluation campaign that aims to promote the development of monolingual and multilingual information retrieval systems for European Languages. The main research interest of CLEF has gradually moved over the years, from only textual document retrieval to question answering (QA) and Geographic information retrieval [2]. In this context, a pilot interactive track, known as interactive CLEF (iCLEF), was introduced in 2001 focusing initially on document selection questions. Since then, iCLEF has been included as a regular event in CLEF.

The iCLEF tracks in CLEF2002 and 2003 focused on examining support mechanisms for query formulation and refinement, as well as user-assisted translations experiments. In the 2004 track, the participants, all using a common evaluation design, tried to assess the ability of their own interactive systems to find specific answers to specific questions in a language other than the one of the initial query. In addition, the 2005 track studied the problem of QA from a user-inclusive perspective [8]. Finally in 2006, the iCLEF track moved to Flickr, a photo-sharing multilingual database in order to promote interactive experimentation on multilingual search tasks and study users' behaviour [6]. In particular, three studies were submitted: a) UNED [1] examined the attitude of users towards cross-language searching when the search system allows three search modes (no translation, automatic translation, assisted translation), b) U. Sheffield (and IBM) [3] used bilingual Arabic-English students to test the Arabic interface of Flickr that they have developed and finally c) Swedish Institute of Computer Science (SICS) [7] focused on evaluating information access based on user satisfaction and user confidence.

In this context, the 2008 iCLEF track focuses both on acquiring a large set of search session logs for the participants to mine and on allowing participants to perform their own interactive experiments with the Flickling interface provided and adopting the task predefined by the Organizers [4]. The aim of this study is to explore users' image seeking behaviour when searching and retrieving known, non-annotated images across languages in Flickling. In particular, the research questions that will be addressed in this paper are:

- Identify the reasons that determined our users' choice of a specific interface (monolingual/multilingual).
- Examine if and/or to what extent users were thinking about languages when searching and retrieving images.
- Examine if and/or to what extent users were paying attention to translations when searching and retrieving images.

The remainder of this paper is structured as follows: a description of the Flickling interface, of our user sample, of the task given, of the four different methods that we used in assembling the data, of the way that the study was carried out and the data processed are illustrated in section 2. We provide an analysis of our findings and an extended discussion of them in sections 3 and 4 respectively. Finally, we conclude summarizing the different image seeking behaviours that our users developed while using Flickling in section 5.

## 2 Method

In this section, further details about the test object of the study, the users, the task given, the specific methods used such as retrospective thinking aloud, observation and interviews, and the data processing are presented.

### 2.1 Test Object

The test object of this study was the Flickling interface, a basic cross-language search front-end to the well-known web application Flickr. Flickr was adopted as the target collection by the iCLEF

organizers mainly for two reasons [4]: a) it is a multilingual database enabling the users to tag and comment in different languages the uploaded images and b) it can provide the baseline for a series of both realistic and challenging multilingual search tasks.

The Flickling interface is a multilingual information retrieval interface that encompasses the following functionalities [4]: a) multilingual search across six languages (English, Spanish, German, French, Dutch, Italian), b) term-to-term translations between six languages using freely available dictionaries, c) select of best target translations, d) pick/remove translations and adding of new ones by the users, f) provision of search suggestions and g) control over the game-like features of the task.

The Flickling interface was intended for two user groups: a) CLEF participants, researchers who expressed interest in conducting experiments based on the provided multilingual interface and b) Flickr/Web users, ordinary users of Flickr application that would like to participate in this game offered by CLEF organizers.

Our main reason of interest in participating in the iCLEF2008 Flickr challenge was to investigate the behaviour of users when asked to search and retrieve a known, non-annotated image across languages.

## 2.2 Users

The study was carried out with a sample of 10 users, three male and seven female, ranging in age from 20 to 40. They were all related in one way or another to Manchester Metropolitan University (MMU). In particular, from the sample of ten users, seven were research postgraduate students, one taught postgraduate student, one lecturer and one MMU staff member. In addition, four of the users were English native speakers, two Greek, one German, one Spanish, one Arabic and one Luganda (see Table 1). Moreover, one of the users was monolingual, four stated knowledge of a language other than their native and five were multilingual.

Native Language	No. Users
Arabic	1
English	4
German	1
Greek	2
Luganda	1
Spanish	1

Table 1: Users' Native Language

In addition, the users were asked to state their level of comprehension for the languages used in Flickling but also any other additional language. In particular, from the six non English native speakers, two stated an Excellent knowledge of English, three Very Good knowledge and one Basic knowledge. In regards of German, three of the nine non-German native speakers stated a Basic knowledge of German. Four out of ten users stated knowledge of French, three of whom Basic and one Good. Concerning Italian two out of ten users stated knowledge of Italian language, Basic and Good respectively. Two out of ten stated a Basic knowledge of Dutch and finally, three out of nine non Spanish native speakers stated a Basic knowledge of Spanish (see Table 2).

All ten users have searched in the past for an image on the web. In particular, four stated that they “rarely” have, three “sometimes”, two “very often” and one “often”. In addition, nine out of ten stated that they have searched for an image on the web in a language other than their native and only one had not. In particular, the nine users identified the following reasons for having done so: “university research, searching for holiday info”, “to increase the numbers of results because I could not find any relevant image by using keywords in my native language”, “because there are only few web resources that I am interested in the Greek language”, “for my course arguments, assignments”, “looking for shoes and clothing in Portuguese and French”, “because the images I wanted were provided in English [Luganda native speaker]”, “I was looking for an image of a region

Language	Basic	Good	Very Good	Excellent
English	1	0	3	2
German	3	0	0	0
French	3	1	0	0
Italian	1	1	0	0
Dutch	2	0	0	0
Spanish	3	0	0	0

Table 2: Users’ Knowledge of Languages

of Poland” and last “because language was not an issue”. The user who had not searched for images on the web in other languages justified it as “not necessary”.

In addition to the users’ previous knowledge and experience with Flickr, only nine out of ten users answered this question. Three of whom stated that “Yes” they have used Flickr in the past. When asked the reason why they have used it, one stated for searching images, one “just to see what it is” and the third user stated for uploading, sharing and searching images. The other six users gave the following justification for not having used Flickr before: “because I was not aware of it”, “not interested, security issues”, “never needed to”, “I don’t Know what Flickr is”. The participants were evenly assigned to the conditions in the experiment with no difference in gender, age, and prior knowledge of Flickling interface.

### 2.3 Task

Our users were asked to find a given image which it was not annotated from Flickr using the Flicking interface. The users did not know in advance in which of the six languages (English, German, Dutch, Spanish, French, Italian) the image was described enforcing them to use both monolingual and multilingual features to find the given image. Each of our 10 users was asked to search and retrieve the first three given images after login by using all the features and the help instructions of the Flickling interface. The images presented to users were not controlled but given randomly from a set of 100 stored in the Flickling database.

### 2.4 Retrospective Thinking Aloud

Retrospective thinking aloud is a widely used method for usability testing of software and interfaces. Its basic principle is to ask from potential users to complete a certain task with the testing object in question and to describe their thoughts and actions afterwards on the basis of a video recording their task performance [5]. This method focuses on peoples’ cognitive processes after having completed a specific task. It is a method that enables the users and not the experts to point out the problems concerning the test object in the usability test.

The use of retrospective thinking aloud method to carry out our study, like any other method, entails both drawbacks and benefits. In particular, it allows users to complete the task in their own way and pace, spending as much time as they wish and are therefore not likely to perform better or worse than usual. In addition, it enables the exact recording of the time spent for each part of the task, as it reflects the real time that a user has spent on completing the task. Moreover, it provides the possibility to users to reflect on their own actions while using the test object and highlight particular causes or milestones that they have personally encountered. In addition, retrospective thinking aloud is considered to be an appealing method for conducting tests across languages as it is less difficult for users to disclose their thoughts in a foreign language after completing the task.

Apart from benefits, the use of retrospective thinking aloud has also some drawbacks the most important of which are: a) the duration of every user session varies according to the time that user will spend on completing the task plus the time that user will need to describe the video in retrospect and b) there is also risk that users may forget what they were thinking during specific phases of the task or withhold others for reasons of social desirability.

In this context, we have used the Camtasia Studio (v.5.1), a premiere screen recorder, in order to capture the users' search sessions in individual videos and a digital recorder for the retrospective thinking aloud and for the individual interviews that followed.

## 2.5 Observation

In addition to retrospective thinking aloud, observation was also adopted. The observation method was used to form specific questions regarding preselected research areas of the test object (translations, layout of the interface, etc) in an attempt to shed light on specific behaviours of the users on specific occasions. A form was created to assist the work of the facilitator at focusing on specific areas of interest and at the same time reflecting on users' behaviour. This form was categorized according to the areas that they were to be tested. Every category had a set of predefined questions/ remarks that the facilitator had to fill in according to user's behaviour each time and write additional comments for the questions to be asked to users during individual interviews. The facilitator, one of the organizers of our study, had to fill in a set of three forms, one for each image of the task, for each user. These forms were coded according to the number assigned to each user and the order of the images (eg. 01/01, 01/02, 01/03).

## 2.6 Interviews

The last part of the study consisted of small scale individual interviews with every user after the completion of the retrospective thinking aloud. The interviews lasted no more than 10 minutes for every user. The questions asked varied according to user's answers to the questionnaire, search session, retrospective thinking aloud and the notes gathered throughout the experiment. The main goal of these questions was to clarify specific actions of the user's image seeking behaviour during the search session and expressions that the user used to describe what he/she was doing.

## 2.7 Experimental Procedure

The study was carried out in 10 individual sessions, which they were all held in the same lab and each lasted from one to two hours approximately. During each session, users were given general instructions about the way that the study will be carried out and about the task that they had to complete. These instructions were read to each user explaining that there were no more instructions to be given throughout the session and the facilitator was there to observe only. After that, users were asked to fill in the questionnaire on personal details and prior experience. Users were then instructed to register, login and start completing the task while screen recording software was taping the computer screen. Having done that, users were asked to watch the recorded session that was played back to them and describe what they were doing and what they were thinking in retrospect. Finally, a no more than 10 minutes semi-structured interview was carried out with each user.

## 2.8 Processing of the data

Once the 10 sessions were completed, transcripts were created based on users' retrospective thinking aloud and interviews, as well as analysis of the questionnaires and comments on video recordings of the users' search sessions. The analysis of the data gathered focused on the way that users were interacting and using the Flickling interface and its features to complete the given task.

The transcripts of the retrospective thinking aloud and interviews, as well as the video sessions and observation notes were examined specifically to identify the users decision to choose a specific interface (monolingual/ multilingual) and identify the extent of the role played by languages and translations in forming their image seeking strategy. These parts were then grouped, when possible, to enable better presentation and discussion of the findings.

In addition, users also occasionally experienced technology problems, such as trouble with the function of the interface (problem messages coming up, interrupting the search and thought

process of the users), the flickering of the cursor due to the software used to record users' search sessions. These problems were excluded from the study.

### 3 Findings

The analysis of data gathered by retrospective thinking aloud, video recordings and interviews focused on three distinct areas, such as: a) to identify the reasons that determined users' choice over a specific interface, b) if and/or to what extent users were thinking about languages when searching for images and c) if and/or to what extent they were paying attention to translations and how helpful they proved to be for finding the images. The findings will be presented according to the three research questions in the subsections 3.1, 3.2 and 3.3 consecutively.

#### 3.1 Reasons

The Flickling was providing to users with two different interfaces, monolingual and multilingual in order to cope with the problem of searching across languages for the target three images. This study's first research question aims at identifying the reasons that determined each time users' choice of a specific interface (monolingual/ multilingual).

Out of the ten users, two used only the monolingual interface and the rest switched between interfaces. The reasons our ten users gave for their behaviour in the thinking aloud process and interviews are stated below.

##### 1. Only Monolingual Interface

Two out of ten users did not use at all the multilingual interface, even though they were given images to search in a language unknown to them. The first user, an English native speaker with basic knowledge in French when informed by the system that the image was annotated in French, stated while still on the monolingual interface: "I kindly instantly gave up, because I am not good in French...I realized that I am never going to find it...So, I decided to give up". When asked why the subject didn't use the multilingual interface, the subject answered: "Because I did not trust my abilities with other languages, to be able to put the decent search words in...Because I did not know the keywords to search in other languages". As a final remark, the subject added: "I was not confident with the languages".

The second user, a Luganda native speaker with no knowledge of French, stated: "I went for the hint and it said that the image is described in French. So, I felt there is no need to...Well, I thought, I do not speak French, I can't understand that". When asked why the subject did not use the multilingual interface, answered: "If I knew how to use another language, then I could use the multilingual and access the same image in another language. But because my first language is not accessible in there [Flickling], then I thought I should keep to monolingual to where I know what I am looking for". When asked how the subject was planning to cope with the problem of searching a French annotated image on monolingual interface by using English keywords, the subject stated: "I thought that the image was not available and all images should be described in English [as well]. So, I thought that it was inaccessible...that I could not get it".

##### 2. Switching between Monolingual & Multilingual Interface

The remaining eight users switched between monolingual and multilingual interfaces in order to complete the given task. A variety of reasons to justify these actions were reported by the users during the retrospective thinking aloud process and interviews. In particular, users identified the following reasons why: "In order to increase or decrease the number of results, depending on the results that I had on the beginning of my search", "I have chosen to use the multilingual interface because I assumed that it would give me the highest possible number of relevant results in relation to my query", "I am trying to find the right combination of keywords", "Because of the setting of the image...I believe that this system, if you know

where the picture was from, or if you know the place then you can like recognize the language in which you can type in”, “I was looking to isolate words and translate them”, “Simple because I wasn’t getting any of the results that I wanted”, “I tried to increase my chances of getting the image...I am widening my possibilities”, “I am just trying out the system”, “So, it was not there [monolingual English], I guess it was in other language” and lastly: “For me the problem was more kind of how to find where the image was from”.

Also, hints played a significant part in users’ choice over an interface. As stated by the users: “I switched to monolingual because the hint told me that the image was described in English”, “ok, I have learned about the hints, so I gave up and asked for a hint...Off I went to multilingual to ask to translate...and then I went back to monolingual and searched for it” and “I went to ask for a hint on language just in case because that seemed to save me lots of time”.

Two users, both English native speakers, stayed on multilingual interface though after taking the hint, they both knew that their image was described in English. When asked why they haven’t switched to monolingual, they gave the following explanations for their choice consecutively: “Because I did not think that would make any difference, because I was assuming that it is in English as well” and “Well, because I was there. I did not realize that...I thought, to be honest, I thought, it’s not going to make that much difference really. It is set to do a search in English, so if it does search in other languages that does not make any difference...is not going to increase my chances in monolingual English...maybe, it would but I don’t know that”.

There were also some cases that although users were seemingly using a specific interface (monolingual or multilingual) they stated during retrospective thinking aloud process and confirmed afterward with the interviews that: “I did not really, even think about it. I was just...I was at that point...I was thinking about getting this title”, “I was not paying attention to the fact that it was multilingual. Maybe, I forgot about that and left it as it was” and “I was so focused on trying to see how to describe the image that I was not paying attention to the interface”.

### 3.2 Role of Languages

The second research question that our study was set out to explore was if and/or to what extent languages were forming the image seeking behaviour of our users. As already stated, our users did not know in advance in which of the six languages (English, German, Dutch, Spanish, French, Italian) supported by the system the given images were described. The task was set in that way so users had to include the element of different languages. As a consequence, a set of different behaviours were identified which can be grouped in the following, again through the analysis of the data gathered from retrospective thinking aloud and interviews:

1. Two users out of ten used only the monolingual interface searching in English, although they knew that the images may not be described in English (see subsection 3.1.1). In particular, the English native speaker stated: “My French are not good, so I decided to give up because I could not find the appropriate translations for my keywords, so I was never to find it”. The subject also added: “I was not confident with the languages”. The Luganda native speaker admitted that: “I would not search an image in any other language; I would only search images in English. If I would search images from my home country, from my background, then I would use my first language. But any other image, I would search in English”. When asked if the subject was thinking about languages while searching, the user answered: “No. It did not...because when I am searching for images on the Internet, I normally get them in English because I imagine that...I guess it’s a little bit of arrogance, I speak English and I imagine that images...That if you put them in Internet, they should have English tags”.
2. The other eight users who were switching between monolingual and multilingual interfaces, can be divided in two groups: a) those who were thinking about languages and b) those for

whom languages were not a variable when performing the given task. In particular, four of them stated clearly during retrospective thinking aloud that: “Now, I made the relationship of country, Florida...I write them [keywords] in English”, “I had the feeling that the building which I recognized, was described in German”, “It was not within my results, so, I guessed that it is in other language”, “Because by looking at the tortoise had written on it...it was written in English. So, I assumed that it would be in English...And I was also thinking at this time, I wonder if it is English or not...because the child got a little blue and red hat and I was thinking, maybe the child is French...Yes, I changed into multilingual because I think that maybe it is French, with the outside possibilities that it might be Italian” and lastly “Well, that’s probable a bit Anglo-centrism. You know, well, it is a picture in England”.

On the other hand, the remaining four users when asked if they were thinking about languages during the task, they said that: “To be honest, I was not thinking about languages...I did not consider it a variable that influences my results”, “I did not bother about languages...I did not really think about them...I did not focus on languages while performing my searches. Maybe, because I am not used to, is not widely used or maybe I am not using languages when retrieving information on the web”, “I was not taking languages under consideration when searching for the images” and “For me it was not a question of language...In my mind language was a very small factor in there [Flickling]. It did not really play any important role”.

### 3.3 Role of Translations

All users were given a minimum one image out of three which was described in a language unknown to them. The multilingual interface was provided to cope with this problem and help retrieve the image. The third and last research question of this study was to examine the use of translations and the influence of translations on the users’ information seeking behaviour.

We are obliged at this point to exclude the two users who used only the monolingual interface and the four users who used the multilingual interface but with no thought to the translations. The remaining four users tried both the monolingual and multilingual interfaces driven by the need to identify the language of the image and the appropriate keywords to retrieve the given images. In particular the four users, when asked if they were paying attention to the translations, stated: “Yes, but it did not translate anything. I thought like, it did not give me anything, because it did not translate anything”, “Yes, I did use them”, “Yes, at this point I am trying to figure out how this translation thing works” and “Yes, I was paying attention to the translations”.

In addition, when users asked if they could judge the translations that were given to them, users answered: “Overall, I had the feeling that the translations of the system were not that good...I switched to monolingual because the translations were not doing anything”, “I would trust the system to give me the right translations...I would have to for languages unknown to me”, “Because I went for the languages that I had a vague idea about and it did not tell me something that I did not really know” and last “I was not satisfied with the German translations because I can understand German...it’s not the right word in German for a man. So, it should have been something else...in Dutch I don’t know what the translation is, so, I had to accept it, whatever it is...Yes, I was satisfied [Dutch translations] because the computer knows the Dutch language better than I do... maybe that’s not the best translation, so, I just had to accept it. There was anything that I could do about it really”.

Finally, when the users were asked if the translations were helpful in terms of actually contributing to the retrieval of the image, users stated: “Ok, I have got the translations but they are not doing anything to me...at the end, I totally disregard the translations”, “I think that I stop searching for translations, when I stop having much confidence that it was bringing me the right translations...So, I got used to asking for a hint...I thought I am gonna ask for another hint now and then it would tell me in what it was translated it, not necessarily what translation [the system] has given me”, “The words that I was trying to isolate like particular words like London, the different translations there were not coming up...and what it was saying, like gigante in Italian

for giant, it told something that I already knew. So it was not isolating the words in the way that I wanted it to. It was just telling me the adjectives were, which I did not really need” and finally “At the end I was not paying attention to the translations, I was purely interested in finding the image as quickly as possible because once more I did not think the translations would necessarily help me”.

## 4 Discussion

The evaluation of CLIR effectiveness often does not involve the end user. On the one hand, it may be assumed that since the translation is automated the user has no role to play or possibly that the user has no interest in the translation, providing the system is effective. On the other hand, the non trivial challenges posed in the effort in designing realistic task scenarios, recruiting participants, analyzing large amounts of data to obtain user assessments or to observe search behaviour can be prohibitive. However, we take the view of Petrelli [9] that effective system design must be in accordance with the end users’ needs and to best assist users involved in cross-language information retrieval we need to understand their behaviours and the search problems they face. Petrelli’s study of users involved in CLIR presented a number of interesting findings. In particular, the users preferred the interface which hid the translation and that language knowledge and sight of the translation affected search behaviour.

The present study shows that users form image seeking behaviours according to a series of reasons from which thinking about languages is the least important one. In addition, users were so focused on completing the task that were not paying attention to interfaces, languages and translations though these were factors that affected and predefined the results. In particular, regarding the first research question, the reasons why two users used only the monolingual, we conclude they were not feeling confident of their language skills and they were not used to searching for images in languages other than English. These users would use the multilingual interface only if they could speak the language in which they were searching and that one would be other than English.

The other eight users described a variety of reasons from which languages played a small part. Users’ need to minimize or maximize the number of results was the driving forces of switching between interfaces. Moreover, they were more concerned with identifying the images in terms of the keywords and in the right combination. In the whole process, languages were playing an insignificant role. Only four users were trying to identify the language of the images from its context and use it to their benefit. Some were treating the multilingual interface as a translator, trying to isolate specific words, translate them on the multilingual interface and use the translations to retrieve the images on the monolingual interface. Another user stated that: “I think I just saw it as a translation tool and not as an integrated translation thing that already was retrieving images. I did not really use it in this way because in my mind, it was only translating my keywords”. The Hint feature was also a factor forming users’ image seeking behaviour to a large extent. Users became accustomed to using the hints after a few minutes of unsuccessfully searching for the image. On the whole, it would appear that users were so focused on completing the task, “obsessed” (as a user stated) of finding the images that even from the beginning of the task, they were not thinking really which interface they are going to use and for what reason. Even users who were concerned about languages, at the end of the task, also admitted that they were not paying much attention to the interfaces because they thought that it was not making any difference.

Going back to Petrelli’s findings, our users did not express any preference in hiding or not the translations, though they were not satisfied with the position of the translations. They stated that the translations were not obvious because they were on the right side and not on the left, close to the search box. They commented even the way that the translations were presented. It was not clear to them how the translations were working or how they could write their query and get translations. They also stated that it was taking some time to figure out how the multilingual interface was built and how they could use the translations to their benefit. Since it was not clear that the system was retrieving both their search terms and the translations, users were

typing both their search terms and the translations provided in the search box and running the search again. In short, they used it as a translator tool and not as an integrated feature of the multilingual interface. However, at no point, did the users state that the presence of translations was distracting or that they would prefer them to be hidden. On the contrary, they said that they used them in order to retrieve relevant results and disregarded them when they were not getting the results that they were expecting or hoping for.

Both Petrelli's and our study aimed to explore user information seeking behaviour in cross-language information retrieval but from a different perspective and for different reasons. As a consequence, the methodologies adopted in both studies differ. Petrelli used mainly observation and interviews whereas we used retrospective thinking aloud and interviews as the main methods to record user behaviour, supplemented with questionnaires and observation to further verify our findings. Our aim was to derive findings entirely on users' thoughts, comments and search behaviour rather than depending on the facilitator's observations, interpretations and questions asked. Although, in retrospective thinking aloud there was some risk of users not remembering what exactly they were thinking, our users were always able to recall what caused them search for an image in such a way. Not only did they justify their actions, they also revealed details about their thought process and the reasons that motivated them or made them feel distracted, uncomfortable or even bored while using Flickling. Our study based on retrospective thinking aloud revealed a complex picture of the influence (or not) of language skills and confidence therein and of perceptions of the role of the multilingual interface, language and translations in image retrieval. Most revealing and of potential interest to future study of users of CLIR is the finding that less than half of our users appeared to consider identification of the language to be essential in retrieving the image. The majority either lacked confidence in using different languages or were so focused on finding the given images and completing the task that were not thinking at all about languages. Indicative of this was the comment *"...completing the task successfully. What was success for me? That you find the image. In any way I possible could. I was not focusing on translations...I thought my task is to find that image and I will do whatever I could to find it"*. Of those for which languages played a significant role in the process of identifying keywords to search for the images, the translations were judged to be poor as either the translations were not coming up, were not corresponding to the users' keywords or were judged to be resulting in the retrieval of irrelevant results. As a consequence, users were losing interest and trust in translations resulting in no usage of them or not paying attention to them.

One of the initial aims of this study was to try and look in greater detail at how working with the translations affected search behaviour with regards to the actual search terms entered by users. Unfortunately, this study could not reach a conclusion because only four out of ten users used the translations and this was in a way not anticipated. The reasons why this happened are:

- the Experiment Design: the fact that users were given a specific task to complete, in a context of a game, was so overwhelming that they were not thinking about anything else other than carrying out the task. The task to find clues for unambiguous query terms from the images was sufficiently challenging in the Flickling interface that perhaps our users chose to ignore the language and translations. Our users were even feeling that they were failing the task or disappointing the organizers in some way when they had to give up. As a result, completing the task was regarded as such a challenge that they were not paying attention to anything else.
- the Flickling Interface: the way that the Flickling interface was developed, providing the two different interfaces, monolingual and multilingual, to search across languages created some issues in studying users' information seeking behaviour. The option of the two interfaces created a confusion to most of the users because they are accustomed to using only one search box on one interface. In addition, users had to interpret what monolingual and multilingual meant, something that again users are not accustomed in doing when using a search engine. Others driven by this fact were not paying attention to interfaces, they were just using what was in front of them without making any informed choice. The users who

used the multilingual interface did so for various reasons but, as stated, languages played an insignificant role. The four users who used the translations to enhance their chances in finding the target image rather than viewing them as an integrated feature of the multilingual interface could not be regarded as forming their search strategy based on the translations.

Although, it would have been interesting to look at how translations influenced or not users' information seeking behaviour because of the reasons stated above we were not able to collect data as we might have hoped.

## 5 Conclusion

This study aimed at investigating the users' image seeking behaviour when retrieving a known, non-annotated image in Flickling. The task assigned to ten users was to retrieve the first three images given after first login. The images could have been described in any of the six languages (English, German, Dutch, French, Spanish, Italian) supported by the Flickling. As a consequence, users had to use monolingual and multilingual interface to search across languages and retrieve the images.

This study used a combination of four different methods: a) questionnaires, b) retrospective thinking aloud, c) observation and d) interviews. These both quantitative and qualitative methods contributed to meeting the research questions of our study. In particular, we identified the reasons why two of our users were choosing to search only on the monolingual interface and the eight switching between interfaces. We demonstrated that only four users were thinking about languages when trying to retrieve the given images while the rest of our users were more preoccupied with finding the images and completing "successfully" the task. Consequently, we showed that only these four users were paying attention to translations provided by the system. These stated that translations were not helpful or they were not making much difference in finding the given images since the results were irrelevant to what they were looking for.

This small study has also shown that if we are to ask whether a CLIR system should display query translations or not, then the answer is no. Our users were either not interested in the translations or found them to be poor. However taking the findings to such conclusion would be foolhardy given the complexity of the activity highlighted in the users' comments that they were so engaged in finding the image that language or translations played little or no part. Rather than reaching firm conclusions, this small study has suggested the need for more research into users' search behaviour with translations (and in image retrieval) if we are to design CLIR systems which will not place additional or unnecessary cognitive demands on the user and will support effective search behaviour and performance.

## References

- [1] J. Artiles, J. Gonzalo, F. López-Ostenero, and V. Peinado. Are users willing to search cross-language? an experiment with the flickr image sharing repository. In *CLEF*, pages 195–204, 2006. [http://www.clef-campaign.org/2006/working\\_notes/workingnotes2006/artilesCLEF2006.pdf](http://www.clef-campaign.org/2006/working_notes/workingnotes2006/artilesCLEF2006.pdf).
- [2] M. Braschler, G. Di Nunzio, N. Ferro, J. Gonzalo, C. Peters, and M. Sanderson. From clef to trebleclef: promoting technology transfer for multilingual information retrieval. In *Second DELOS Conference on Digital Libraries, 5-7 December 2007, Tirrenia, Pisa (Italy)*, pages 1–7, 2007. <http://www.trebleclef.eu/getfile.php?id=38>.
- [3] P. Clough, A. Al-Maskari, and K. Darwish. Providing multilingual access to flickr for arabic users. In *CLEF*, pages 205–216, 2006. [http://www.clef-campaign.org/2006/working\\_notes/workingnotes2006/cloughCLEF2006.pdf](http://www.clef-campaign.org/2006/working_notes/workingnotes2006/cloughCLEF2006.pdf).

- [4] P. Clough, J. Gonzalo, J. Karlgren, E. Barker, J. Artilles, and V. Peinado. Large-scale interactive evaluation of multilingual access systems: the iclef flickr challenge. In *Workshop on Novel Methodologies for Evaluation in Information Retrieval, 30 March 2008, Glasgow, Scotland*, 2008. <http://nlp.uned.es/iCLEF/ECIR-evaluation-workshop.pdf>.
- [5] M. Haak, M. de Jong, and P. J. Schellens. Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue. *Behaviour & Information Technology*, 22(5):339–351, 2003. <http://www.students.cs.uu.nl/people/jpwester/W01/Artikelen/p339.pdf>.
- [6] J. Karlgren, J. Gonzalo, and P. Clough. iclef 2006 overview: Searching the flickr www photo-sharing repository. In *CLEF*, pages 186–194, 2006. <http://eprints.sics.se/321/01/iclef-2006-overview-v4.pdf>.
- [7] Fredrik Olsson and Jussi Karlgren. Trusting the results in crosslingual keyword-based image retrieval. In *Proceedings of Evaluation of Multilingual and Multi-modal Information Retrieval, 7th Workshop of the Cross-Language Evaluation Forum, CLEF 2006, Alicante, Spain, September 20-22, 2006, Revised Selected Papers*, page 3, Alicante, Spain, 2007. <http://eprints.sics.se/319/01/iclef-2006-SICS.pdf>.
- [8] C. Peters. Comparative evaluation of cross-language information retrieval systems. In *From Integrated Publication and Information Systems to Virtual Information and Knowledge Environments*, pages 152–161, 2005. <http://dienst.isti.cnr.it/Dienst/Repository/2.0/Body/ercim.cnr.isti/2004-TR-43/pdf?tiposearch=cnr&langver=>.
- [9] Daniela Petrelli, Micheline Beaulieu, Mark Sanderson, George Demetriou, Patrick Herring, and Preben Hansen. Observing users, designing clarity: A case study on the user-centered design of a cross-language information retrieval system. *JASIST*, 55(10):923–934, 2004. <http://dblp.uni-trier.de/db/journals/jasis/jasis55.html#PetrelliBSDHH04>.