An Investigation on Text-Based Cross-Language Picture Retrieval Effectiveness through the Analysis of User Queries

PETRELLI, Daniela and CLOUGH, Paul

Published version

PETRELLI, Daniela and CLOUGH, Paul (2012). An Investigation on Text-Based Cross-Language Picture Retrieval Effectiveness through the Analysis of User Queries. The Electronic Library incorporating Library Computing, 30 (2), 197-219.

Analysing User's Queries for Cross-Language Image Retrieval from Digital Library Collections

Purpose:

This paper describes a study of the queries generated from a user experiment for cross-language information retrieval (CLIR) from a historic image archive. Italian speaking users generated 618 queries for a set of known-item search tasks. The queries generated by user's interaction with the system have been analysed and the results used to suggest recommendations for the future development of cross-language retrieval systems for digital image libraries.

Methodology:

A controlled lab-based user study was carried out using a prototype Italian-English image retrieval system. Participants were asked to carry out searches for 16 images provided to them, a known-item search task. User's interactions with the system were recorded and queries were analysed manually quantitatively and qualitatively.

Findings:

Results highlight the diversity in requests for similar visual content and the weaknesses of Machine Translation for query translation. Through the manual translation of queries we show the benefits of using high-quality translation resources. The results show the individual characteristics of user's whilst performing known-item searches and the overlap obtained between query terms and structured image captions, highlighting the use of user's search terms for objects within the foreground of an image.

Limitations and Implications:

This research looks in-depth into one case of interaction and one image repository. Despite this limitation, the discussed results are likely to be valid across other languages and image repository.

Value:

The growing quantity of digital visual material in digital libraries offers the potential to apply techniques from CLIR to provide cross-language information access services. However, to develop effective systems requires studying user's search behaviours, particularly in digital image libraries. The value of this paper is in the provision of empirical evidence to support recommendations for effective cross-language image retrieval system design.

Keywords: Cross-language information retrieval, image retrieval, historical photo archives, automatic machine translation.

Classification: Research Paper

1 Introduction

Digital libraries are amassing an increasing amount of digital content; indeed users and curators of digital libraries are being confronted with large quantities of digital material that are increasingly diverse in nature: multi-media, multi-cultural and multi-language (Borgman, 1997; Crane, 2006). For example, Europeana, the European digital library, museum and archive, aims to provide users access to around 2 million digital objects, including photos, paintings, sounds, maps, manuscripts, books, newspapers and archival papers.

An important aspect in providing effective information access is making the content of a digital library widely accessible. Supporting Multi-Lingual Information Access (MLIA) and Cross-Language Information Retrieval (CLIR) in digital libraries has long been recognised as important in providing universal access to digital content (Oard, 1997; Borgman, 1997; Bian & Chen, 2000; Peters & Sheridan, 2001). The mechanisms involved in MLIA and CLIR are shown graphically in Figure 1, where the user's query is translated into the language(s) of the document collection (destination language) and retrieved documents are translated into the user's source language. The material in Europeana, for example, exists in multiple languages and there are plans to provide end users with cross-language access to reduce the effects of language as a barrier to accessing content.

TAKE IN FIGURE 1

Figure 1. The process of cross-language information retrieval between a user's query language (source language) and the language of the document collection (destination language).

An area of research that is receiving increased attention from the academic community is cross-language image retrieval (Clough & Sanderson, 2006; Müller et al., 2010) where cross-language text retrieval is applied to image captions. As a form of visual media, images can be regarded as "language-independent" and therefore offer a clear user case for adopting CLIR in practice. This is because the relevance of an image, with respect to a given query, can often be judged regardless of the user's linguistic abilities and foreign language skills. As Oard (1997) comments: "an image search engine based on cross-language free text retrieval would be an excellent early application for cross-language retrieval."

Providing cross-language access to collections of visual media could open up image and photographic digital libraries to wider audiences. There have been significant advances in dealing with the many technical issues of providing effective image retrieval, such as developing computer vision algorithms to identify objects within an image from low-level features, e.g. colour and shape; mapping low-level features (e.g. shapes) to linguistic concepts to offer automated generation of linguistic image metadata; designing novel input mechanisms for image retrieval, e.g. sketching the input query. However, key to providing effective information access is also understanding the users and their search behaviours and this has received less attention from the research community than more technical issues. For example, as Robins (2000:57) states "most IR systems are used by people and we cannot design effective IR systems without some knowledge of how users interact with them". There has been little previous work on analysing user's cross-language image searching behaviours, particularly in digital image libraries, and investigating the effects of varying quality of translation on user queries harvested from an interaction log.

In this paper we investigate user's search behaviour when interacting with a digitised image library of historical photographs from St Andrews University, Scotland, that provides further insights into data gathered in a previous study (Petrelli & Clough, 2006). A prototype cross-language image retrieval system was developed to search for images described using British English metadata, such as title and description, and queried using Italian. An interactive user evaluation was instrumental in gathering user's queries which have been analysed, both quantitatively and qualitatively, to gain a better understanding of user's image searching behaviours and their response to the problems of translation errors. Based on the results of this study, a set of

guidelines is proposed that will enable the development of more effective interactive cross-language image retrieval systems.

Section 2 summarises related work in image retrieval and cross-language image retrieval and analysing user's interaction with such systems. Section 3 provides a summary of the experimental setup including the data collection used in the study, the prototype cross language image retrieval system and the user study itself. Section 4 analyses the results of the user study with respect to the queries issue by participants in the user study and problems caused by query translation errors. Recommendations for the design of future cross-language image retrieval systems based on the findings of the user study are described in Section 5 and the paper concludes in Section 6 with a summary.

2 Related Work

2.1 Image Retrieval

There are a number of approaches for retrieving visual objects (e.g. images and videos) depending on the information associated with an object (Enser, 1995; Goodrum, 2000). Concerning images, one approach, 'content-based', uses techniques from the field of computer vision to extract low-level features, such as colour, shapes and texture, for indexing and retrieval. However, despite the many advances in content-based approaches, a significant challenge remains because of the (semantic) gap that exists between the automatically extracted low-level image features and the visual (and semantic) content of the image itself (Sandom and Enser, 2002).

In practice the images stored in a digital library are often accompanied by textual metadata (e.g. assigned free-text or terms from a vocabulary, tags or captions), which describe properties of the image, such as its contents or creation history. This associated text can be used to provide simple text-based access to images and is often users' preferred method for image retrieval interaction (Eakins et al., 2004). Such approaches are often referred to as 'description-based' (or text-based) and are commonly the most popular form of image retrieval, particularly on the web. However, images in many digital library collections are associated with limited amounts of text, if any at all, which may limit the success of retrieving images due to a mismatch between the terms user's issue to find images and the (limited) text used to describe the image. In fact, in many cases retrieval can be reduced to a "hit or miss" situation, i.e. retrieval is only successful if the metadata has been indexed and matches at least one term in the user's query. Research by Zhang et al. (2004) showed that techniques which can often improve document retrieval, such as helping users

identify more relevant terms to reformulate their queries, were not successful at improving image retrieval. The effects of vocabulary mismatch are further compounded in the case of cross-language image retrieval where user's linguistic skills and the effects of translation resources may further increase the problem of vocabulary mismatch (e.g. an MT system could translate a user's query 'road' as 'highway' and not match on the term 'road' as used in the document collection).

2.2 Cross-Language Image Retrieval

The implementation of cross-language image retrieval systems is typically a combination of standard cross-language information retrieval and image retrieval techniques. Leveraging the use of both content- and concept-based techniques has shown to offer the most effective form of cross-language image retrieval across multiple search requests, as demonstrated within the academic community. For example, events such as ImageCLEF, an initiative for evaluating cross-language image retrieval systems in a standardised manner thereby allowing comparison between the various approaches, have helped to systematically investigate the effectiveness of different algorithms (Müller et al., 2010).

An example of early academic cross-language image retrieval was the Eurovision system (Sanderson & Clough, 2002; Clough & Sanderson, 2006). This system operated on a historic image archive, a collection of historic images from St Andrews University library (described in Section 3.1) and used machine translation for the translation of user's queries, the user interface and search results. More recently, the PanImages[1] cross-language image search engine provides access to images on the web and utilises dictionary-based query translation rather than using MT (Colowick, 2008). The system also includes features, such as auto-completion, and also allows users to add their own (preferred) translations.

Much of the academic work in cross-language image retrieval has focused on developing systems that can return as many relevant images as possible to a given search topic rather than on users and their interactions with a systems. This is also the case for other forms of cross-language information retrieval, with exceptions such as work on the Clarity project (Petrelli et al., 2004) and at events such as the interactive track of the Cross Language Evaluation Forum (CLEF), known as iCLEF[2] (see, e.g. (Oard & Gonzalo, 2002)). The 2008-2009 iCLEF events focused on cross-language image retrieval from the user's

[1] http://www.panimages.org
[2] http://nlp.uned.es/iCLEF/

perspective and the task was organised based on users participating in an interactive image search experiment. Organisers provided a default multilingual search system which accessed images from Flickr.com, with the whole iCLEF experiment run as an online game. Interaction by users with the system was recorded in log files which were shared with participants for further analyses and provide a future resource for studying various aspects on user-orientated cross-language image search (Clough et al., 2008; Clough et al., 2010).

In the spirit of user-oriented cross-language image retrieval research, this paper investigates user's interaction in more depth and focuses on analysing on the queries issued during interaction and the effect of query translation on the results. This work aims to complement the aforementioned studies and provide more insights into users and their cross-language image search behaviours.

3 Experimental Setup

3.1 The St Andrews Image Archive

The data collection used in this experiment consists of 28,133 historic images (postcards and photographs) from the library at St Andrews University in Scotland (Reid, 1999). This collection was also adapted for use in the system-oriented evaluation of cross-language image retrieval systems (Clough et al., 2006; Müller et al., 2010). All images are accompanied by a textual caption consisting of 8 distinct fields, which can be used individually or collectively to facilitate image retrieval (see, Figure 2). All captions are written by domain experts and in *British* English, thereby often containing colloquial expressions and historical terms. Approximately 81% of captions contain textual data in all fields; the rest generally lack text in the 'description' field. On average, text in the description field of the image is a grammatical sentence or paragraph of around 15 words. The majority of images (82%) are black and white or sepia; of the colour images, some are hand painted postcards (the Plymouth lighthouse, no. 6 in Appendix A); others are recent photos (nos. 2 and 5 in Appendix A).

Figure 2 shows an example image and accompanying caption text (descriptive metadata), which is organised into separate fields that includes archival information, such as the name of the photographer, date and location; and a textual description of the visual contents of the image. Ancillary text is typically provided in a 'notes' field which contains information commonly not directly related to the image itself, but of use for the collection owners. The 'categories' provide a set of terms drawn from a controlled vocabulary that assist with organising the photographs, e.g. 'horse drawn vehicles' and

'processions – state'. The example shown in Figure 2 is similar to how photographs in many photographic libraries are stored. The associated text, however, is not consistent across the collection; some fields are left blank, and the quantity of text is highly variable across images. This can result in varying retrieval effectiveness for different queries and cause unexpected results for the user of a retrieval system indexing this collection.

TAKE IN FIGURE 2

Figure 2. An example photograph with the associated text from the St Andrews photographic library.

3.2 System Prototype

The corpus of user-generated queries used in this study was collected during the evaluation of a text-based cross-language interactive image retrieval system, which is described in (Petrelli & Clough 2006). In interactive cross-language retrieval there are two points where users can interact with the search process (see Figure 1): in input, during query formulation and reformulation (Oard et al., 2008) and in output, when exploring the search results and examining individual documents (Oard et al., 2004). The focus of this paper is not the design of the visual aspects of an interactive information retrieval system, but rather the analysis of the queries resulting from the user's interaction with such a system[3]. We use the data collected from a user experiment to provide empirical evidence supporting technical aspects that should be considered in creating more effective interactive cross-language image retrieval systems. In essence, in this paper we are more concerned with the back-end of the system than with the front-end.

The user interface for the system used to gather data is shown in Figure 3 and provides a simple cross-language image retrieval interface. The query entered by user (in this case Italian) was automatically translated into English using a publicly-accessible online translation tool (Yahoo! BabelFish[4]). In this setup the user has little opportunity to interact with query translation, however this is often commonplace when MT systems are used for translating user's queries.

The translated query (in this case English) was used to search the St

[3] In this paper we are not concerned with the layout of the interface and how it could be improved. Readers specifically interested in the design and evaluation of a user interface for cross-language retrieval could refer to (Petrelli et al. 2002; Petrelli et al., 2004).

[4] http://babelfish.yahoo.com/

Andrews image collection via a standard version of the Okapi search engine (a probabilistic retrieval model based on the BM25 weighting function (Robertson et al. 1995)). The inverted index used by Okapi had been composed using selected fields from the metadata: the long title, description and category fields[5] (Figure 2). The results (in English) were composed in a graphical user interface that was translated into Italian using the web page translation service offered by Babelfish, before being displayed on the user's screen (Figure 3). An additional feature of the prototype system is the use of 'concept hierarchies' to automatically extract terms (or concepts) from the captions of the results to provide a way of exploring the results set (the left-hand hierarchy of images shown in Figure 3). The extracted terms are organised into a hierarchy of terms (e.g. vehicle > car) where travelling lower down the hierarchy represents a narrowing or subset of the results set.

TAKE IN FIGURE 3

Figure 3. The prototype cross-language image retrieval system user interface.

3.3 Experimental Setup

A user[6] evaluation was conducted as a means of investigating user's behaviour with the prototype system and their ability to carry out a series of pre-defined search tasks. During the user evaluation participants were shown an image from the St Andrews image collection and asked to find it again, a form of known-item search task. Although the setting is not completely natural as users are searching for information needs that are not their own, the set of queries generated in this way were effective at capturing linguistic variations for a given topic. Our goal here was in collecting comparable data rather than simulating a realistic setting for image retrieval. Indeed we recognise that a controlled, lab-based situation does not provide insights on how the system could be used in practice. However, at early phases in the design process it is essential to fully understand the implications of selecting different technical approaches before moving toward more naturalistic use (Petrelli, 2008). Moreover, asking users to perform a known-item search task has a number of advantages: the success of carrying out the task is easy to measure – the user either finds the required image or not; the data collected are directly comparable; for participants in the experiment the task is easy to understand

[5] The other metadata fields were omitted because the index was also used to generate a hierarchical summary of results which we found to be less effective when text not related to the visual content of the image was used.

[6] Note we use the term user and participant interchangeably throughout the paper.

and does not rely on digesting and interpreting a written description of an information need; presenting people with an image enables the study of what kinds of queries people generate for a given visual object and provides a wide diversity of queries to analyse.

A set of 18 randomly selected images from the St Andrews image archive was used in the experiment: 16 tasks for testing (Appendix A) and 2 for training purposes. (We refer to the search tasks as *topics*.) The decision on the number of images to use in the experiment was a balance between providing a good variety of test images and practical constraints. In particular, we aimed to keep the length of entire user evaluation under 3 hours in order not to excessively tire participants. A within-subjects experimental design was used where participants carried out all search tasks. A Latin-Square arrangement was used to counterbalance the experiment and ensure the data collected was unbiased due to possible task-order learning effects. Participants were given one image at the time (no text or metadata was shown) and requested to find it again. They were nominally given a 5 minutes limit to complete the task; as the system did not automatically block the interaction after the allotted 5 minutes, more time was granted if participants wished to continue searching.

On arrival, participants were introduced to the evaluation protocol and received an explanation of the purpose of the study. They were also asked to fill out a questionnaire providing information regarding their demographics and prior searching experiences. Participants performed the tasks individually and were observed by an experimenter. Whilst participants were constrained to find a specific search task they were free to choose their interaction strategy, such as the number and subject of queries to issue; the number of results pages to view; the number of images to inspect. Two tasks were used to familiarise users with the system, and in this phase participants were allowed to ask any questions about the system. User's interactions (e.g. queries issued, items clicked and navigational aids used) were automatically recorded in a custom log file and a user satisfaction questionnaire was submitted to collect user's opinions about each search task. At the end of the experiment, participants were invited to comment on their overall search experience. The entire evaluation exercise lasted 3 hours and a total of 618 queries were input by 8 participants. All were native Italian speakers, 5 male and 3 female, recruited through the University of Sheffield mailing list for volunteers. All participants were bilingual students or researchers, although their knowledge of the English language and UK culture awareness varied[7]. All were computer literate and searched the web daily;

[7] A wide variation was registered with respect to culture awareness depending on the time spent in the UK.

searching the University library or using commercial search software was less popular. All participants stated they were aware of machine translation and had previously used image search tools on the web.

4 Results

This section reports the analysis carried out on the data gathered in the user experiment with respect to a number of dimensions: the users and success at individual tasks; the success of query translation; and query-caption term overlap. The analysis is both quantitative and qualitative making use of observations, manual inspection and analysis of the different phases of the cross-language information retrieval process. The combination of these two analytical methods is instrumental in obtaining a fuller understanding of the phenomena under observation (Petrelli 2008). Indeed, whilst a quantitative analysis provides an indication of where problems might lie, a qualitative analysis can then be used to explain why problems happen. Therefore, the many findings derived from the different analyses are then combined to provide a more well-rounded explanation.

TAKE IN TABLE I

Table 1. A summary of query variation by topic (success/failure rate is also reported).

4.1 Analysis of Users, Topics and Retrieval Success

The success with which participants were able to complete the task (i.e. re-find a given image) varied widely across the topics. Table 1 shows a summary of average success for each topic (the proportion of participants who found the topic), the average number of queries issued per topic, and the average number of terms used in the queries for each topic. From Table 1, topics 4, 9, 12 and 14 were retrieved by all participants (100% success); however, images 5[8], 7 and 10[9] were only retrieved by a single participant (12.5% success). Even among the 4 images with 100% success, the effort required was unbalanced: topics 4 and 14 were retrieved by issuing a single query; topic 9 was harder (13 queries in total, mean by participant = 1.6) and topic 12 harder still (36 queries in total, mean by participant = 4).

In general, the number of queries issued per topic varied greatly (min=8, max=68); whilst the number of terms issued per query was limited (min=4, max=10.3). Minimum values collected for "simpler" images (e.g. 4 and 14) show users do generate different terms even when a single query is entered. As

[8] This task was the retrieval of the image in Fig. 3.
[9] This task was the retrieval of the image in Fig. 2.

expected, a higher number of queries corresponded to the hardest topics. However, the number of terms does not seem to follow the same pattern: topic 5 was found by just one participant using an average of 7.9 terms; topic 12 had 100% success rate with 8.4 terms entered. This means that both simple and difficult searches could stimulate a high number of terms. A possible explanation for this is the ease with which users can describe an image: topic 12 has many visual details that can be included; topic 5 offers less detail to describe in the query.

A multiple regression analysis showed no correlation between success rate and number of queries or number of terms, suggesting that retrieving an image does not depend on the effort the user put in the search task. Rather, it seems to depend on the amount of text used to describe an image and its overlap with the user's query (see discussion below in 4.3).

TAKE IN TABLE II

Table 2. Query variation by users with success rate. None of the 8 participants had 100% retrieval success.

An interesting observation of the queries issued by participants is that many queries were not well-formed phrases, a common condition for system-oriented evaluation of CLIR systems. In this study the behaviour of users formulating full phrases seems to be an individual trait rather than the norm: only participant U1 consistently typed phrases; another 3 users did it occasionally (U2, U5, U6), and 4 users never formulated a phrase but used separated keywords (U3, U4, U7, U8).

The number of queries issued (min=3, max=7) and their length (min=1.5, max=10.6) differs widely between users, as well as between topics (see Table 2). The variation appears related to how difficult a certain image is to retrieve for each user and their attitude in fully exploring the results. By observing users' behaviour, it was noticed that U2, U5, U8 checked the first page and then changed the queries; U1, U3 and U6 looked at the first few pages before changing the queries, while U4 and U7 methodically checked each results before modifying the query. Data collected is not enough to determine whether extensively exploring the results caused a change in the thread of the search (i.e. all new terms in the next query). The sample is too small to derive any concrete conclusions, but the varying degree to which users explored the results suggests that the best interface design is to display as many images as possible given that some users do not view past the first page of results and rather

prefer to reformulate their query.

4.2 Analysis of Query Translation

In CLIR, a good translation is vital for effective retrieval, as demonstrated by Clough & Sanderson (2004), whose study showed a significant correlation between the quality of query translation, as assessed by human experts, and retrieval effectiveness across a number of languages. Therefore, the success of query translation depends on the quality of available resources and the approach used for translation (e.g. the use of machine translation versus a bilingual dictionary). Free machine translation (MT) services offered by Yahoo! were used to translate queries from English to Italian and resulting English image captions to Italian (as shown in Figure 1). To analyse the effectiveness of the MT system, the translated queries were compared with the original queries. Of the total 1,018 different query terms, 83% were correctly translated. However, for 114 terms (11%) the translations failed and a further 61 terms (6%) were ill translations, often due to BabelFish selecting the wrong sense, or preferring verbs to nouns for the same spelling. This is because MT systems are designed to operate on larger text segments and will therefore translate well-formed phrases more effectively than short queries. In the vast majority of cases, users' queries are not well-formed phrases: only 126 queries (12%) could be considered proper phrases (i.e. include prepositions and articles) and were nearly all generated by the same participant (U1); the other queries were lists of terms, though simple strings (e.g. noun + adjective) occurred.

TAKE IN TABLE III

Table 3. Summary of ill translated terms; all were checked in AltaVista while the correct translation is from an online Italian dictionary (Garzanti Linguistica www.garzantilinguistica.it). The reason column provides an explanation of possible reasons why the machine translation may have failed.

Besides obvious mistakes, such as the selection of the wrong sense of an ambiguous term, some ill translations were simply unexplainable. A summary is provided in Table 3. Some of the nonsensical translations resulted in unexpected and confusing results for users. For example, when the query 'signora vestito bianco' (lady white dress) was issued for image 12 (see Appendix A), results included many portraits of men. This is because the query was mis-translated as 'mrs. dressed white man' and participants could not understand why portraits of men were retrieved (the translated query was not shown to users).

Some ill translated or non-translated terms were frequent in the corpus and Tables 4 and 5 summarise those used by more than one participant. A few terms were used by all or almost all users and correspond to significant visual features. For example, all participants used 'tempio' (temple) to retrieve image number 5, 'faro' (lighthouse) for number 6, 'carrozza' (carriage) to retrieve number 10 and 'bianco' (white) for number 12. When these high-frequency terms were indexed, failing to properly translate them drastically decreased retrieval effectiveness. However, in some cases a correct translation would not have helped as the description did not contain the translation: this is the case of 'macchine' (cars) in Table 3 or 'cornamuse' (bagpipes) in Table 4. In another set of cases (like for 'reali' (royals) in Table 3 and 'gotico' (gothic) in Table 4) the term is in the image metadata but that field was not indexed; therefore retrieval was impossible.

TAKE IN TABLE IV

Table 4. Summary of the ill translations by users and tasks; only terms used by more than one person have been listed.

TAKE IN TABLE V

Table 5. Summary of the missed translations by users and tasks; only terms used by more than one person have been listed.

Upon inspection of the two sets of translations (queries and results) and comparing between them, we found that the MT service used lacked translation symmetry, i.e. terms translated from source to destination languages (e.g. from Italian into English) were not always translated back the same (from English to Italian). This is a known problem in Machine Translation (Rapp 2009) and can hamper human comprehension (Yamashita & Ishida 2006). In our case, problems arose when terms were back-translated in the output (results to user) but were not recognised in the input (user to retrieval system) and created confusion for the users. For example, in our experiments 'portrait' in the caption was properly translated into 'ritratto', but the translation failed when the Italian term was input. As users made use of terms from the displayed results in order to make the search more effective, translation asymmetry negatively affected the interaction, as well as the perception and user satisfaction with the system. For example, the terms 'croce Celtica' (Celtic cross) were picked up by participants and used in follow-up queries to focus the search.

How much users could be influenced by distorted translations is suggested by 8 cases (0.8%) when participants used terms that were ill translations with alternate results. For example, 'golf club' was translated from English to Italian as 'randello' (the correct term is 'mazza [da golf]') then translated back into English as 'club' thus maintaining retrieval effectiveness; at the opposite end, 'bridge' in the caption was translated as 'ponticello' ('ponte' would be correct) that was not translated from Italian into English. This shows that users are willing to accommodate the system's behaviour in an attempt to improve retrieval effectiveness. Indeed, in a further 11 queries (1%) of participants inputted English words to overcome (real or perceived) system limitations, which is consistent with the findings in (Petrelli et al 2004). Examples include 'bagpipes' and 'lighthouse' after MT failed to translate the corresponding Italian words, and also 'cottage' or 'clubhouse' which were inputted directly by participants. In another 1.5% (15 cases), queries contained proper names (e.g. 'Plymouth', 'Robert Burns', 'Wallace') or nouns (e.g. 'ballgown', 'golfers') which were found in the results. By inputting English terms in the Italian query the users were actually bypassing the translation mechanism and those terms were then used unchanged in the retrieval, therefore increasing effectiveness.

TAKE IN TABLE VI

Table 6: Summary of the overlap between image captions and queries; *nQ* is the total number of queries issued by all users; *found* the number of users who successfully found the image; *overlap* the proportion of matching terms between users; *GT* gold translation; *MT* machine translation.

4.3 Analysis of the Overlap Between Query and Caption Terms

Whilst the previous section analysed mis-translations, this section compares the overlap of the translated query with all terms in all caption fields vs. the fields indexed for this study (long title, description, and categories) which we refer to as *full text* or *caption* respectively. A gold translation[10] (GT) for every query was generated by an Italian native speaker. To assure a correct comparison with an automatic system, each term in the GT was checked against the Garzanti dictionary[11] and was included only if the translation was listed in the dictionary. This approach roughly simulates the result of dictionary-based and user-checked query translation step and provides an upper bound for

[10] By "gold translation" we mean a manual translation of the query performed by a native speaker using a dictionary, which we trust to be the correct translated query.
[11] http://www.garzantilinguistica.it

retrieval effectiveness.

To compute term overlap between translated queries and captions, English stop words were removed first followed by term suffixes removed using the Snowball implementation of the English Porter stemming algorithm[12]. The results are shown in Table 6 where: *nQ* represents the total number of queries submitted by all users for a given topic, *found* the percentage of users who found the image, and *overlap* the proportion of matching query terms between users.

On average the similarity between the terms in user's queries is 78.4% indicating that, in general, the same terms are used by many users (min=62.5% for topic 14; max=86.3% for topic 5). As an example of the variety of individual's queries, Table 7 shows 8 successful queries: 1 for each participant (U1 to U8) submitted for topic 14. It is interesting to see the differences between queries: the varying length, the common use of the term "family" but the variation in additional terms; the description of visual properties and the use of the background as well as the foreground.

TAKE IN TABLE VII

Table 7: Variation between the 8 queries submitted by each user for topic 14. All queries successfully retrieved the target image.

Table 6 also shows that, as expected, translation does impact the degree of overlap between query-caption terms with the MT having higher mismatch compared to a "perfect" translation (i.e. high-quality resource). Comparing GT full text with MT full text, the gold translation achieves 22% higher term overlap than the MT version indicating limitations in the translation approach used. Higher overlap is achieved using all metadata (full text) compared to the fields indexed in this study (caption). Overlap with caption was lower for both GT and MT showing that important details which match user's query terms are often described in ancillary fields and indexing those could compensate poor MT results and thereby improve recall.

To further analyse the way users describe images, the description field of each caption (in which an indexer has described the visual content of an image) was split into *focus* (the most prominent object), *foreground* (the focus + prominent details), and *background* (caption - foreground). The query-description overlap was calculated for each of the three. For example, for topic

[12] http://snowball.tartarus.org/

1 (see Appendix A) the description "Commemorative cross in Celtic style, on grass sward beside sea loch, with sandy bay and houses on low promontory", was split into focus as "Commemorative cross in Celtic style", foreground as "Commemorative cross in Celtic style, on grass sward beside sea loch" and background as "with sandy bay and houses on low promontory". As the task required participants to retrieve a given image, we would assume that users would typically use words describing the focus or the foreground. Results shown in Table 8 support this statement: on average, 16.6% of terms match the foreground, 10.7% of terms match the focus and only 1.7% the background[13]. For example, 16 queries in topic 1 used the term "cross" which matches clearly the main subject of the image: a Celtic cross.

The low overlap with the background can be explained by some images having only foreground descriptions and no apparent background visual content. It would perhaps suggest that indexing on terms which are more descriptive of prominent foreground objects would better match the kinds of queries generated by users.

TAKE IN TABLE VIII

Table 8: Percentage of user's query terms which overlap terms in the focus, foreground and background of the image

5 Recommendations for Design

The analysis of user's queries and searching behaviour carried out in this study provides a detailed picture of the linguistic interaction between a user and a cross-language image retrieval system. These empirical results lead us to suggest guidelines (or principles) that designers and developers of text-based cross-language image retrieval should follow to create more effective systems. Further helpful information can be found in the best practice guidelines for developing multilingual information access systems from the system- and user-oriented perspectives by Braschler & Gonzalo (2009)[14].

5.1 Query Translation

It is common in CLIR systems, particularly in research, to use machine translation (MT), especially freely available online services[15]. However, our

[13] Note that the remaining terms did not match text in the image caption at all.

[14] An online version of the document is available here: http://www.trebleclef.eu/jsbestpractices.php

[15] See, for example, the many submissions to the Cross Language Evaluation Forum (CLEF) campaign for evaluating multilingual information access systems (http://www.clef-campaign.org).

results have shown that users do not generally input well-formed phrases, but instead prefer to use isolated terms or simple expressions (e.g. adjective + noun). This limited grammatical structure is critical as MT systems use syntax and context to correctly translate text and non-grammatical input may cause poor translation, which directly impinges on retrieval performance (see, e.g. (Clough & Sanderson, 2004)). The effects of retrieval on the results for image retrieval may be less critical than document retrieval as the user may be able to recognise relevant items without understanding the associated text. In this study the task was a known-item search and the focus query/result translation, we are not concerned with retrieval effectiveness.

With ambiguous words, the lack of context forces a "default" selection of one of the possible translations that, in our experiment, often resulted in choosing the wrong sense (see Table 3). The output of an MT system is one translation: if the term picked is wrong there is no opportunity to correct the error before performing the search. It is then essential that the user is involved in checking the translation; indeed previous research has shown that allowing the user monitoring the query translation increases the performance of CLIR systems (He et al. 2002, Petrelli et al. 2002).

As even the best translation does not ensure retrieval success (see Table 6), it is essential to use a good quality bilingual dictionary with appropriate linguistic resources in an attempt to ensure the best quality translation. Essential features include: identifying bi-grams (e.g. adjective + noun) to assist the correct translation of phrases (e.g. translation of 'famiglia reale' (royal family) as an atomic unit rather than the individual words), to identify proper names that should not be translated, and finally dealing with out-of-vocabulary terms.

5.2 System Design

Table 6 shows that even if the full text associated with an image is indexed and the best translation used, the maximum success rate in retrieving the correct image is around 50% (max. 53.8% for topic 14). Moreover, for some images the text in fields considered ancillary (e.g. location) can be essential for matching user's queries (e.g. 'Egypt' in the retrieval of topic 5). It seems important to index all caption terms, but this can come at the cost of reducing precision.

Based on the overlap between query and terms identified as being prominent in an image, we contend that weighting terms which more likely describe visual content of the image foreground could improve retrieval and the

re-ranking of search results. This could be achieved by considering the position of the term in the caption, as the description of the foreground always precedes the description of the background. Whilst this solution does not require any change in the database, a more radical solution would be to modify the metadata and explicitly distinguish between foreground and background. This would have the advantage of stimulating the archivist in creating richer descriptions and therefore potentially increasing the searchability of the collection.

Using retrieval based on indexing the different fields would also enable field normalisation (term weights based on their occurrence in the field and normalised to the length of each field rather than entire caption). A further option would be to offer users the possibility to select which fields to match their query against, very much like advanced search allows users to distinguish between the title, the abstract or the full text.

5.3 Interface Design

Guidelines for the interface and interaction design of CLIR are related to both inputs and outputs. Evidence from this study (and previous work (He et al. 2002, Petrelli et al. 2004)), suggests to design the interface in such a way that the translated query is presented to the user who can then better understand why a certain image was retrieved and correct the translated query if needed.

Another important finding of this study is the fact that users pick up terms displayed by the system in follow-up queries. This encompasses correct translations and proper names, but also mistranslations. Asymmetry in the dictionary can occur, but mitigations are possible. Some form of local dictionary, kept by the interface that records asymmetrical terms can be used to foster the translation by substituting non-translated terms before the search is done.

Users tend to formulate their query on what is prominent in the image but describing part of the background could improve the retrieval for a subset of images (Table 8). To stimulate the production of more complex queries the interface could use two search boxes: one labelled foreground and the other background. How effective this strategy could be needs to be measured in a further evaluation, however if the metadata contained this distinction, as suggested above, then it is likely this type of interface would result in higher retrieval effectiveness.

The observation of the user's behaviour has highlighted several interaction styles: from carefully looking at every screen of images to skimming the first page before immediately changing the query. For the latter approach

to be successful the image must have been retrieved and ranked at the very top, quite an unlikely event given the result of the overlap analysis. Previous research has shown users can scan images much more quickly than text. The strategy here is to display as many pictures as possible as this would increase the chance for the user to pick up the desired one and faster.

6 Conclusions

Many digital libraries contain content that may benefit from cross-language information access. A common question of such systems is why users would want to retrieve items they presumably cannot read? One such argument in favour of CLIR is that some documents, such as visual media, are "language independent" and therefore offer a practical use case for the application of CLIR. However, although there is a clear trend that the linguistic diversity of users and online content is increasing, many digital libraries still provide limited multi-language support. Despite the large amount of research carried out on CLIR systems there is still need to further understand the users of such systems and their environments. Studying the user while performing realistic tasks with a real system is fundamental to understand the interaction; this would then support the design of CLIR systems on the basis of empirical evidence.

Although this paper has addressed a specific scenario of cross-language information retrieval – locating images from a historic collection of images from St Andrews University library, the problems faced by users and the findings from the user study carried out in the work are applicable to a wider range of digital image libraries. Through a prototype CLIR system we have generated 618 queries from a user study involving 8 Italian-speaking participants. The participants queried the system (in Italian) for 16 known-item search tasks; the system automatically translated the queries and the results using a freely-available online MT system. The queries generated by user's interaction with the system were analysed, particularly with respect to the success of query translation. Results highlighted the problems users face when queries are mistranslated and the direct that this has on retrieval effectiveness. The wide variety in user's queries, particularly noticeable when users were attempting to describe the contents of an image, make query translation difficult and often error-prone.

Further challenges for cross-language image retrieval systems, particularly in digital libraries and archives, includes the consistency, quality and length of the image descriptors written by archivists: often image descriptions are short and written in a way that includes colloquial terms. The shorter the

text associated with the indexed images the higher the chance of vocabulary mismatch. A further challenge is whether all metadata fields are indexed for retrieval. Although this can improve recall, without some form of re-ranking it can drastically reduce precision.

Observations from this user study have led to suggestions for improving the development of cross-language image retrieval system in the future. This paper is a first step in this direction, though a full understanding of cross-language image retrieval needs to consider aspects of perception and cognition. This is left for future research.

Acknowledgements

The work was carried out when the first author was at the Information School at the University of Sheffield. The authors wish to thank the participants for their time and thoughts.

References

Bian, G. and Chen, H. (2000), "Cross Language Information Access to Multilingual Collections on the Internet", *Journal of American Society for Information Science & Technology*, Special Issue on Digital Libraries, Vol. 5 No.3, pp. 281-296.

Borgman, C.L. (1997), "Multi-Media, Multi-Cultural and Multi-Lingual Digital Libraries: or How do we exchange data in 400 languages?" *D-Lib Magazine*. Available online: http://dlib.ukoln.ac.uk/dlib/june97/06borgman.html (accessed March 2011).

Braschler, M. and Gonzalo, M. (2009), "Best Practices in System-oriented and User-oriented Multilingual Information Access", TrebleCLEF Deliverable 3.3. Available online: http://www.trebleclef.eu/getfile.php?id=254 (accessed March 2011).

Clough, P. and Sanderson, M. (2004), "Assessing Translation Quality for Cross Language Image Retrieval", In Comparative Evaluation of Multilingual Information Access Systems, Eds (Peters, C., Gonzalo, Braschler, M. and Kluck, M.), Lecture Notes in Computer Science (LNCS), Springer, Heidelberg, Germany, Volume 3237/2004, pp. 594-610.

Clough, P., Sanderson, M. and Reid, N. (2006), "The Eurovision St Andrews Collection of Photographs", *ACM SIGIR Forum*, Vol. 40 No.1, pp. 21-30.

Clough, P. and Sanderson, M. (2006), "User Experiments with the Eurovision Cross-Language Image Retrieval System", *Journal of the American Society for Information Science and Technology*, Vol. 57 No. 5, pp. 697-708.

Clough, P., Gonzalo, J., Kargren, J., Barker, E., Artiles, J. and Peinado, V. (2008), "Large-Scale Interactive Evaluation of Multilingual Information Access Systems - the iCLEF Flickr Challenge", *Proceedings of Workshop on novel methodologies for evaluation in information retrieval,* 30th European Conference on Information Retrieval, Glasgow, 30th March-3rd April 2008.

Clough, P., Gonzalo, J. and Karlgren, J. (2010), "Creating Re-useable Log Files for Interactive CLIR", *In SIGIR 2010 Workshop on the Simulation of Interaction (SimInt)*, Geneva 23 July 2010.

Colowick, S. (2008), "Multilingual Search with PanImages", *MultiLingual*, Vol. 19 No. 2, pp. 61–63. Available online: http://turing.cs.washington.edu/PanImMultilingual.pdf (accessed March 2011).

Crane, G. (2006), "What do you do with a million books?" *D-Lib Magazine*, Vol. 12 No. 3. Available online: http://www.dlib.org/dlib/march06/crane/03crane.html (accessed March 2011).

Eakins, J., Briggs, P. and Burford, B. (2004), "Image Retrieval Interfaces: A User Perspective", *In Proceedings of Conference on Image and Video Retrieval (CIVR'04)*, pp. 628-637.

Enser, P.G.B. (1995), "Pictorial information retrieval", *Journal of Documentation*, Vol. 51 No. 2, pp. 126-170.

Goodrum, A. (2000), "Image information retrieval: An overview of current research", *Informing Science*, Vol. 3 No. 2, pp. 63-66.

He, D., Wang, J., Oard, D. and Nossal, M. (2002), "Comparing User-assisted and Automatic Query Translation", *In Working notes of Cross Language Evaluation*

Forum (CLEF 2002), pp. 267-278.

Müller, H., Clough, P., Deselaers, T. and Caputo, B. (2010), *ImageCLEF - Experimental Evaluation of Visual Information Retrieval*, Springer: Heidelberg, Germany. 495 pages. http://dx.doi.org/10.1007/978-3-642-15181-1.

Oard, D. (1997), "Serving Users in Many Languages – Cross-Language Information Retrieval for Digital Libraries", *D-Lib Magazine*, December 1997. Available online: http://www.dlib.org/dlib/december97/oard/12oard.html (accessed March 2011).

Oard, D. and Gonzalo, J. (2002), "The CLEF 2001 Interactive Track", Evaluation of Cross-Language Information Retrieval Systems, Lecture Notes in Computer Science (LNCS), Springer, Heidelberg, Germany, Volume 2406, pp. 308-319.

Oard, D., Gonzalo, J., Sanderson, M., López-Ostenero, F. and Wang, J. (2004), "Interactive Cross-Language Document Selection", *Information Retrieval*, Vol. 7 Issue 1-2, pp. 205-228.

Oard, D., He, D. and Wang, J. (2008), "User-assisted query translation for interactive cross-language information retrieval", *Information Processing & Management*, Vol. 44 No. 1, January 2008, pp. 181-211.

Peters, C. and Sheridan, P. (2001), "Multilingual information access", In M. Agosti, F. Crestani, and G. Pasi (Eds.), Lectures on information Retrieval, New York: Springer LNCS 1980, pp. 51-80.

Petrelli, D., Hansen, P., Beaulieu, M. and Sanderson, M. (2002), "User requirement elicitation for Cross-Language Information Retrieval", *New Review of Information Behaviour Research*, Vol. 3, pp. 17-35.

Petrelli, D., Hansen, P., Beaulieu, M., Sanderson, M., Demetriou, G. and Herring, P. (2004), "Observing Users - Designing Clarity: A Case study on the user-centred design of a cross-language retrieval system." In *Journal of the American Society for Information Science and Technology*, Vol. 55 No. 10, pp. 923-934.

Petrelli, D. and Clough, P. (2006) "Using Concept Hierarchies in Text-Based Image Retrieval: A User Evaluation", In Paul Clough, Julio Gonzalo, Thomas Mandl, Thomas Deselaers, Henning Müller, Horacio Rodríguez, Sven Hartrumpf, Felisa Verdejo, Alicia Ageno, Víctor Peinado Eds. "Accessing Multilingual Information Repositories", Lecture Notes in Computer Science (LNCS), Springer, Heidelberg, Germany, Volume 4022/2006, pp. 297-306.

Petrelli, D. (2008), "On the role of User-Centred Evaluation in the Advancement of Interactive Information Retrieval", *Information Processing and Management*, Vol. 44 No. 1, pp. 22-38.

Rapp, R. (2009), "The Back-translation Score: Automatic MT Evaluation at the Sentence Level without Reference Translations", *In Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pp. 133–136.

Reid, N. H. (1999), "The photographic collections in St Andrews University Library", *Scottish Archives*, Vol. 5, pp. 83-90.

Robertson, S. E., Walker, S., Jones, S., Hancock-Beaulieu, M. M. and Gatford, M. (1995), "Okapi at TREC-3", In Harman, D. K., editor, The Third Text Retrieval Conference (TREC-3), pp. 109-126.

Robertson, S.E., Walker, S., Beaulieu, M.M., Gatford, M. and Payne, A. (1995), "Okapi at TREC-4", In: Harman, D.K. (ed.), NIST Special Publication 500-236: The Fourth Text REtrieval Conference (TREC-4), Gaithersburg, MD. pp. 73-97.

Robins, D. (2000), "Interactive Information Retrieval: Context and basic Notions", *Informing Science*, Vol. 3 No. 2, pp. 57-61.

Sanderson, M. and Clough, P. (2002), "Eurovision - an image-based CLIR system", Workshop "Cross-Language Information Retrieval: A Research Roadmap" held at the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, University of Tampere, Finland, August 15th 2002, pp. 56-59.

Sandom, C. and Enser, P. (2002), "VIRAMI - Visual Information Retrieval for Archival Moving Imagery", *Library and information report 129*. London: The

council for museums, archives and libraries.

Salton, G. (1973), "Experiments in multi-lingual information retrieval", *Information Processing Letters*, Vol. 2 No. 1, pp. 6–11.

Yamashida, N. and Ishida, T. (2006) "Effects of machine translation on collaborative work", *In Proceedings of 2006 ACM Conference on Software and Computer Work (CSCW '06)*, pp. 515-524.

Zhang, C., Chai, J. and Jin, R. (2005), "User term feedback in interactive text-based image retrieval", *In Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '05)*. ACM, New York, NY, USA, pp. 51-58.

Appendix A

The images used in the evaluation. For each image its title and description are given.

TAKE IN TABLE IX

Daniela Petrelli
Art & Design Research Centre
Cultural, Communication and Computing Research Institute
Furnival Building, 153 Arundel Street
S1 2NU - Sheffield, UK
d.petrelli@shu.ac.uk

Bio: An expert in Human-Computer Interaction and User Centred Design, Dr. Petrelli conducted this research while at the University of Sheffield, at the Information School. Her research focuses on methods for and practice of designing interactive information systems to ease their use by people at work and at home. She has explored different domains, e.g. multimedia and multilingual, in different contexts, e.g. information professionals, digital archives for scholars and the public, organizational memory and its use in large corporations. Other research interests include the design of innovative devices for accessing personal multimedia memories.

Paul Clough
Information School
University of Sheffield
Regent Court, 211 Portobello Street
S1 4DP – Sheffield, UK
p.d.clough@shef.ac.uk

Bio: A Senior Lecturer in Information Retrieval within the Information School, University of Sheffield. Dr. Clough has conducted research in a range of topics related to information storage and retrieval. These include multilingual information access, geographical information retrieval, system-oriented and user-oriented evaluation of searching systems, and the creation of standardised resources for evaluation. Clough has been on the steering committee of the Cross Language Evaluation Forum (CLEF) since 2003 and co-organised tasks for evaluating cross-language image retrieval systems and interactive cross-language information access systems. This work has since been written up as a book entitled "ImageCLEF - Experimental Evaluation of Visual Information Retrieval". He is currently co-authoring another book entitled "Multilingual Information Access: From Research to Practice" due for publication in 2011.

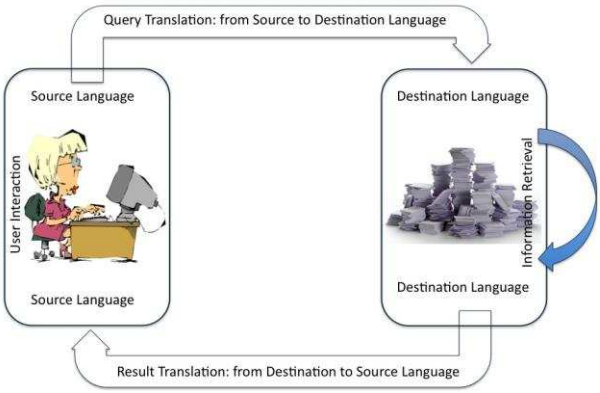
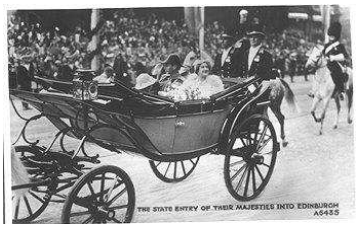


Figure 1



Record ID:	JV-A.005435
Short title:	The State Entry into Edinburgh.
Long title:	The State Entry of Their Majesties into Edinburgh.
Location:	Midlothian, Scotland
Description:	Open landau carriage with three people sitting inside and two footmen on back; mounted soldiers and crowds in background.
Date:	Registered 2 July 1937
Photographer:	J Valentine & Co
Categories:	[horse drawn vehicles][soldiers][army][royal visits][processions - state][regalia][dress - uniforms - military][M Loth all views][Collection - J Valentine & Co]
Notes:	JV-A5435 R jf/ mb DETAIL: King George VI and Queen Elizabeth with Princess Margaret (and possibly Princess Elizabeth, unseen) in landau. Guardsman following carriage. King in naval (Admiral of the Fleet ?) uniform with several medals; footmen and guardsman also wearing medals.

Figure 2


CiQuest
cattedrale st andrews "timpano orientale"
GO
Browser Di Frase
st (930)
uomini (40)
timpano (37)
chapel (33)
protezione (30)
timpano orientale (12)
corvo (8)
costruzione dell'università (6)
headstone (5)
- 12 documenti per i andrews della st del cattedrale e "il orientale di timpano"
cattedrale st andrews "orientale di timpano"
[controlli tutti | Uncheck tutto]
St Andrews. Regola del regulus della st o della st. Chapel e Towe
t Andrews. Regola del regulus della st o della st. Chapel e
orretta, con il timpano orientale della cattedrale.
Segno: 15.874
ola del regulus della st o della st. Chapel e Towe
t Andrews. Regola del regulus della st o della st. Chapel e
orretta, con il timpano orientale della cattedrale.
Segno: 15.871
Rovine della cattedrale. st Andrews. Vista del timpano orientale

Figure 3

Topic	Success (as % and users)	Av. No. Queries	Av. No. Terms
1	75% (6)	3.9	6.5
2	62.5% (5)	3.1	7.1
3	62.5% (5)	5.1	7.4
4	100% (8)	1.1	4.2
5	12.5% (1)	7.5	7.9
6	87.5% (7)	2.6	4.7
7	12.5% (1)	7.1	9.2
8	75% (6)	4	6.2
9	100% (8)	1.6	4.9
10	12.5% (1)	7	10.2
11	62.5% (5)	5.6	8.9
12	100% (8)	4.5	8.4
13	37.5% (3)	5.1	8
14	100% (8)	1	4
15	62.5% (5)	3.2	5.5
16	62.5% (5)	4.2	8.2

Table I

User ID	1	2	3	4	5	6	7	8
Success	13 72%	10 56%	12 67%	13 72 %	11 61%	11 61%	13 72 %	15 83 %
Average No. of Queries	3	4	3	3	7	5	4	5
Average No. of stop words	12	2	0	0	4	1	0	0
Average No. of Terms	20	14	14	9	17	8	6	19
Average No. of Unique Terms	10	7	8	5	7	5	5	8
Average Length of Queries (words)	10.6	4	4.6	3	3	1.8	1.5	3.8

Table II

Query	Meaning(s)	MT translation	Reason (assumed)
bianco	white (adj.)	white man	very colloquial sense
signora	madam, lady, ms., mrs., woman	mrs.	multiple senses, mrs. is for formal written communications
vestito	dress, dressed	dressed	multiple senses, noun an verb (past tens)
reale [famiglia reale]	real, royal [royal family]	real family	multiple senses
lanterna	lantern	spider	
prato	lawn	Prato	Prato is a city in Italian
riva	seaside, (river) bank	river	
sala	hall, sitting room	it knows it	
cappelli	hats	nails head	
coppia	couple	brace	
primo piano	foreground	association of Bologna	inexplicable (Bologna is an Italian city)
bianco e nero	white and black	R-bianco.e.nero	
ingresso	entry, entrance	income	
macchine	machines, cars	it blots some	

Table III

Task	Term	Ill translation	Correct translation	Participants who used it
2	macchine	it bolts some	cars	1, 3
6	faro	beacon	lighthouse	1, 2, 3, 4, 5, 6, 7, 8
6	bianco	white man	white	1, 2, 3
12	bianco	white man	white	1, 2, 3, 4, 5, 6, 7, 8
9	letti	read	beds	3, 4, 8
10	reale/reali	real	royal	1, 2, 5, 6, 7, 8
	ingresso	income	Entrance (hall)	task 3: participants 5, 8 task 11: participants 3, 8
	ingresso	income	entrance (hall)	task 3: participants 5, 8 task 11: participants 3, 8
12	vestito	dressed	dress	5, 6

Table IV

Term	Translation	Tasks and participants who used it
celtica	Celtic	task 1: participants 1, 2, 4, 5, 7
citta'	city, town	task 2: participants 1, 3
tempio	Temple	task 5: participants 1, 2, 3, 4, 5, 6, 7, 8
vagone/vagoni	Carriage	task 7: participants 1, 2, 3, 4, 5, 6
carrozza	Carriage,coach	task 7: participants 5, 7 task 10: participants 1, 2, 3, 4, 5, 6, 7, 8
vetrata/vetrate	glass door	task 11: participants 1, 6
lampadario	chandelier	task 11: participants 1, 3, 6, 7, 8
candelabro	candelabrum	task 11: participants 2, 6
gotico/gotica	Gothic	task 11: participants 2, 3, 7
ritratto	portrait	task 12: participants 1, 4, 8 task 14: participants 1, 5, 7, 8 task 16: participants 1, 3, 4
cornamuse	bagpipes	task 13: participants 1, 2, 3, 5, 6, 7
tamburi	drums	task 13: participants 1, 2, 3, 4, 6
lungomare	seashore	task 15: participants 4, 5

Table V

Topic	nQ	Found (% & users)	Overlap (%)	GT full text (%)	GT caption (%)	MT full text (%)	MT caption (%)
1	21	75% (6)	79.4	33.3	25	28.6	20.7
2	24	62.5% (5)	76.4	14.8	0	9.4	0
3	31	62.5% (5)	74.7	29.7	13.5	18.4	7.9
4	9	100% (8)	68.2	41.7	41.7	45.5	45.5
5	58	12.5% (1)	86.3	18.5	0	9.7	0
6	21	87.5% (7)	75.3	41.2	23.5	25	12.5
7	56	12.5% (1)	81.7	22.9	11.4	17.1	7.3
8	33	75% (6)	81.6	40	16.7	39.4	18.2
9	13	100% (8)	75.4	26.7	26.7	22.2	22.2
10	64	12.5% (1)	85.1	29.4	8.8	16.2	5.4
11	34	62.5% (5)	82.8	46.4	25	23.7	13.2
12	36	100% (8)	82	16.7	13.3	16.1	12.9
13	41	37.5% (3)	83.2	27.6	10.3	19.4	6.5
14	8	100% (8)	62.5	53.8	46.2	50	42.9
15	26	62.5% (5)	80.6	16.7	16.7	16.7	16.7
16	34	62.5% (5)	79.2	29	12.9	23.5	11.8
Avg	**31.8**	**64.1**	**78.4**	**30.5**	**18.2**	**23.8**	**15.2**

Table VI

U1	family portrait of a sitting couple with two children
U2	family two children dark background
U3	photo family man woman white children dark background
U4	family child sitting
U5	family portrait
U6	family photo
U7	family portrait
U8	family portrait children formal

Table VII

Topic	Focus (%)	Foreground (%)	Background (%)
1	8.3	16.7	8.3
2	0	0	0
3	10.8	10.8	2.7
4	33.3	41.7	0
5	0	0	0
6	23.5	23.5	0
7	8.6	11.4	0
8	10	13.3	3.3
9	13.3	26.7	0
10	5.9	5.9	2.9
11	7.1	21.4	3.6
12	13.3	13.3	0
13	3.4	10.3	0
14	15.4	46.2	0
15	11.1	11.1	5.6
16	6.5	12.9	0
Average	**10.7%**	**16.6%**	**1.7%**

Table VIII

 1. War Memorial, Isle of Iona. Commemorative cross in Celtic style, on grass sward beside sea loch, with sandy bay and houses on low promontory.	 2. North Street, St Andrews. No description.	 3. East Gable, St Andrews. Remains of turreted cathedral gable framed within recessed entrance arch; many gravestones in grounds; tower in background.	 4. Setting Drag Nets, Lough Neagh. Clinker built rowing boat, with two fishermen, one casting nets and the other rowing; low harbour wall with slipway.
 5. Montus Temple, Nag-el-Medamoud. No description.	 6. The Smeaton Tower, Plymouth. Red and white striped lighthouse on coastal cliff with harbour and town beyond, and substantial building on cliff terrace below.	 7. Tay Bridge Disaster, girder. Beached section of fallen railway bridge showing mangled rails and remains of passenger carriages resting on side girders.	 8. Joe Conrad (USA) in practice. Golfer with raised club following course of ball watched by other players and spectators in front of R& A Clubhouse.
 9. Roman Camp Hotel, Callander. Interior of bedroom with beds, chairs, fireplace,	 10. The State Entry into Edinburgh. Open landau carriage	 11. The Lobby, House of Commons. Ornate entrance hall, with bas-relief wall	 12. Netta [Gilmour]. Young woman in silk or

carved overmantel, fire screen, footstool, screen, bell pull.	with three people sitting inside and two footmen on back; mounted soldiers and crowds in background.	frieze, statue; chandelier, clock on double doors, decorated floor tiles.	satin dress with a buttoned panel on bodice and trimmed with flounces, in conservatory, holding fan.
 13. Home Guard Parade, St Andrews. Boys' Brigade and Boy Scouts marching through town led by pipes and drums.	 14. Mr & Mrs Alexander Keiller. Family studio portrait, seated: man in frock coat, woman in dress with bows on bodice; two children in pin-tucked smocks.	 15. Port Bannatyne. Children paddling on the sea edge; boats in harbour, church, houses and shops round bay; hill with trees behind.	 16. Robert Burns. Robert Burns; Burns Cottage; Auld Brig o' Doon; Burns Monument, Ayr; Auld Brig o' Ayr.

Table IX