Cross-Language Multimedia Information Retrieval

Sharon Flank
eMotion, Inc.
2600 Park Tower Dr., Vienna, VA 22180 USA
sharon.flank@emotion.com

Abstract

Simple measures can achieve high-accuracy cross-language retrieval in carefully chosen applications. Image retrieval is one of those applications, with results ranging from 68% of human translator performance for German, to 100% for French.

1 Introduction

Information is increasingly global, and the need to access it crosses language barriers. The topic of this paper, cross-language information retrieval, concerns the automatic retrieval of text in one language via a query in a different language. A considerable body of literature has grown up around cross-language information retrieval (e.g. Grefenstette 1998, TREC-7 1999). There are two basic approaches. Either the query can be translated, or each entire document can be translated into the same language as the query. The accuracy of retrieval across languages, however, is generally not good. One of the weaknesses that plagues cross-language retrieval is that we do not have a good sense of who the users are, or how best to interact with them.

In this paper we describe a multimedia application for which cross-language information retrieval works particularly well. eMotion, Inc. has developed a natural language information retrieval application that retrieves images, such as photographs, based on short textual descriptions or captions. The captions are typically one to three sentences, although they may also contain strings of keywords. Typical queries are, as in most Web search applications, two to three words in length. At this point, all of the captions are in English. eMotion hosts a large database of images for sale and for licensing, PictureQuest. At least 10% of PictureQuest's user base is outside the United States. The tests were performed on the PictureQuest database of approximately 400,000 images.

Recent Web utilization data for PictureQuest indicate that of the 10% of users from outside the United States, a significant portion come from Spanish-speaking, French-speaking, and German-speaking countries. It is expected that adding appropriate language interfaces and listing PictureQuest in foreign-language search engines will dramatically increase non-English usage.

2 The Cross-Language Multimedia Retrieval Application

This paper offers several original contributions to the literature on cross-language information retrieval. First, the choice of application is novel, and significant because it simplifies the language problem enough to make it tractable. Because the objects retrieved are images and not text, they are instantly comprehensible to the user regardless of language issues. This fact makes it possible for users to perform a relevance assessment without the need for any kind of translation. More important, users themselves can select objects of interest, without recourse to translation. The images are, in fact,

associated with caption information, but, even in the monolingual system, few users ever even view the captions. It should be noted that most of the images in PictureQuest are utilized for advertising and publishing, rather than for news applications. Users of history and news photos do tend to check the captions, and often users in publishing will view the captions. For advertising, however, what the image itself conveys is far more important than the circumstances under which it was created.

Another significant contribution of this paper is the inclusion of a variety of machine translation systems. None of the systems tested is a high-end machine translation system: all are freely available on the Web.

Another key feature of this paper is the careful selection of an accuracy measure appropriate to the circumstances of the application. The standard measure, percent of monolingual performance achieved, is used, with a firm focus on precision. In this application, users are able to evaluate only what they see, and generally have no idea what else is present in the collection. As a result, precision is of far more interest to customers than recall. Recall is, however, of interest to image suppliers, and in any case it would not be prudent to optimize for precision without taking into account the recall tradeoff.

The PictureQuest application avoids several of the major stumbling blocks that stand in the way of high-accuracy cross-language retrieval. Ballesteros and Croft (1997) note several pitfalls common to cross-language information retrieval:

(1) The dictionary may not contain specialized vocabulary (particularly bilingual dictionaries).

(2) Dictionary translations are inherently ambiguous and add extraneous terms to the query.

(3) Failure to translate multi-term concepts as phrases reduces effectiveness.

In the PictureQuest application, these pitfalls are minimized because the queries are short, not paragraph-long descriptions as in TREC (see, e.g., Voorhees and Harman 1999). This would be a problem for a statistical approach, since the queries present little context, but, since we are not relying on context (because reducing ambiguity is not our top priority) it makes our task simpler. Assuming that the translation program keeps multi-term concepts intact, or at least that it preserves the modifier-head structure, we can successfully match phrases. The captions (i.e. the documents to be retrieved) are mostly in sentences, and their phrases are intact. The phrase recognizer identifies meaningful phrases (e.g. *fire engine*) and handles them as a unit. The pattern matcher recognizes core noun phrases and makes it more likely that they will match correctly.

Word choice can be a major issue as well for cross-language retrieval systems. Some ambiguity problems can be resolved through the use of a part-of-speech tagger on the captions. As Resnik and Yarowsky (in press) observe, part-of-speech tagging considerably reduces the word sense disambiguation problem. However, some ambiguity remains. For example, the decision to translate a word as *car*, *automobile*, or *vehicle*, may dramatically affect retrieval accuracy. The PictureQuest

system uses a semantic net based on WordNet (Fellbaum 1998) to expand terms. Thus a query for *car* or *automobile* will retrieve essentially identical results; *vehicle* will be less accurate but will still retrieve many of the same images. So while word choice may be a significant consideration for a system like that of Jang et al., 1999, its impact on PictureQuest is minimal.

The use of WordNet as an aid to information retrieval is controversial, and some studies indicate it is more hindrance than help (e.g. Voorhees 1993, 1994, Smeaton, Kelledy and O'Donnell 1995). WordNet uses extremely fine-grained distinctions, which can interfere with precision even in monolingual information retrieval. In a cross-language application, the additional senses can add confounding mistranslations. If, on the other hand, WordNet expansion is constrained, the correct translation may be missed, lowering recall. In the PictureQuest application, we have tuned WordNet expansion levels and the corresponding weights attached to them so that WordNet serves to increase recall with minimal impact on precision (Flank 2000). This tuned expansion appears to be beneficial in the cross-language application as well.

Gilarranz, Gonzalo and Verdejo (1997) point out that, for cross-language information retrieval, some precision is lost in any case, and WordNet is more likely to enhance cross-linguistic than monolingual applications.

In fact, Smeaton and Quigley (1996) conclude that WordNet is indeed helpful in image retrieval, in particular because image captions are too short for statistical analysis to be useful. This insight is what led us to develop a proprietary image retrieval engine in the first place: fine-grained linguistic analysis is more useful that a statistical approach in a caption averaging some thirty words. (Our typical captions are longer than those reported in Smeaton and Quigley 1996).

3 Translation Methodology

We performed preliminary testing using two translation methodologies. For the initial tests, we chose European languages: French, Spanish, and German. Certainly this choice simplifies the translation problem, but in our case it also reflects the most pressing business need for translation. For the French, Spanish, and German tests, we used Systran as provided by AltaVista (Babelfish); we also tested several other Web translation programs. We used native speakers to craft queries and then translated those queries either manually or automatically and submitted them to PictureQuest. The resulting image set was evaluated for precision and, in a limited fashion, for recall.

The second translation methodology employed was direct dictionary translation, tested only for Spanish. We used the same queries for this test. Using an on-line Spanish-English dictionary, we selected, for each word, the top (top-frequency) translation. We then submitted this word-by-word translation to PictureQuest. (Unlike AltaVista, this method spell-corrected letters entered without the necessary diacritics.) Evaluation proceeded in the same manner. The word-by-word method introduces a weakness in phrase recognition: any phrase recognition capabilities in the retrieval system are defeated if phrases are not retained in the input. We can assume that the non-English-speaking user will, however, recognize phrases in her or his own language, and look

them up as phrases where possible. Thus we can expect at least those multiword phrases that have a dictionary entry to be correctly understood. We still do lose the noun phrase recognition capabilities in the retrieval system, further confounded by the fact that in Spanish adjectives follow the nouns they modify. In the *hombre de negocios* example in the data below, both AltaVista and Langenscheidt correctly identify the phrase as multiword, and translate it as *businessman* rather than *man of businesses*.

The use of phrase recognition has been shown to be helpful, and, optimally, we would like to include it. Hull and Grefenstette 1996 showed the upper bound of the improvements possible by using lexicalized phrases. Every phrase that appeared was added to the dictionary, and that tactic did aid retrieval. Both statistical co-occurrence and syntactic phrases are also possible approaches. Unfortunately, the extra-system approach we take here relies heavily on the external machine translation to preserve phrases intact. If AltaVista (or, in the case of Langenscheidt, the user) recognizes a phrase and translates it as a unit, the translation is better and retrieval is likely to be better. If, however, the translation mistakenly misses a phrase, retrieval quality is likely to be worse. As for compositional noun phrases, if the translation preserves normal word order, then the PictureQuest-internal noun phrase recognition will take effect. That is, if *jeune fille* translates as *young girl*, then PictureQuest will understand that *young* is an adjective modifying *girl*. In the more difficult case, if the translation preserves the correct order in translating *la selva africana*, i.e. *the African jungle*, then noun phrase recognition will work. If, however, it comes out as *the jungle African*, then retrieval will be worse. In the architecture described here, fixing this problem requires access to the internals of the machine translation program.

4 Evaluation

Evaluating precision and recall on a large corpus is a difficult task. We used the evaluation methods detailed in Flank 1998. Precision was evaluated using a crossing measure, whereby any image ranked higher than a better match was penalized. Recall per se was measured only with respect to a defined subset of the images. Ranking incorporates some recall measures into the precision score, since images ranked too low are a recall problem, and images marked too high are a precision problem. If there are three good matches, and the third shows up as #4, the bogus #3 is a precision problem, and the too-low #4 is a recall problem.

For evaluation of the overall cross-language retrieval performance, we simply measured the ratio between the cross-language and monolingual retrieval accuracy (C/M%). This is standard; see, for example, Jang et al. 1999.

Table 1 illustrates the percentage of monolingual retrieval performance we achieved for the translation tests performed. In this instance, we take the precision performance of the human-translated queries and normalize it to 100%, and adjust the other translation modalities relative to the human baseline.

Language	**Raw Precision (%)**	**C/M (%)**
French (Human)	80	100
French (AltaVista)	86	100
French (Transparent Language)	66	83

Language	Raw Precision (%)	C/M (%)
French (Intertran)	44	55
Spanish (Human)	90	100
Spanish (AltaVista)	53	59
Spanish (Langenscheidt Bilingual Dictionary)	63	70
German (Human)	80	100
German (AltaVista)	54	68

Several other factors make the PictureQuest application a particularly good application for machine translation technology. Unlike document translation, there is no need to match every word in the description; useful images may be retrieved even if a word or two is lost. There are no discourse issues at all: searches never use anaphora, and no one cares if the translated query sounds good or not.

In addition, the fact that the objects being retrieved were images greatly simplified the endeavor. Under normal circumstances, developing a user-friendly interface is a major challenge. Users with only limited (or nonexistent) reading knowledge of the language of the documents need a way to determine, first, which ones are useful, and second, what they say. In the PictureQuest application, however, the retrieved assets are images. Users can instantly assess which images meet their needs.

In conclusion, it appears that simple on-line translation of queries can support effective cross-language information retrieval, for certain applications. We showed how an image retrieval application eliminates some of the problems of cross-language retrieval, and how carefully tuned WordNet expansion simplifies word choice issues. We used a variety of machine translation systems, none of them high-end and all of them free, and nonetheless achieved commercially viable results.

5 Appendix: Data

Source	Example	Score
OrigSp	1. hombres reparando carretera	
Human	men repairing road	100
AV	men repairing wagon	0
Lang.	man repair road	100
OrigSp	2. mujer vestida de rojo comprando en una tienda	
Human	woman wearing red shopping in store	100
AV	woman dressed red buying in one tends	90 (2 of 20 bad)
Lang.	woman clothe red buy in shop	wearing red is lost 75 (5 of 20 bad)
OrigSp	3. carros manejando por la autopista	
Human	cars driving on the highway	100
AV	cars handling by the freeway	80 (4 of 20 bad)
Lang.	cart handle for the expressway	0
OrigSp	4. leones cazando en la selva africana	
Human	lions hunting in the African forest	80 (1 of 5 bad)
AV	lions hunting in the African forest	80 (1 of 5 bad)
Lang.	lion hunt in the jungle	45 (11 of 20 bad)
OrigSp	5. malabarista con usando bolas de colores	
Human	juggler using colorful balls	67 (1 of 3 bad)
AV	juggler with using balls of colors	50 (4 of 8 bad)
Lang.	juggler by means of use ball colour	(0; 1 should be there)
OrigSp	6. niños rubios jugando con canicas	

Source	Example	Score
Human	blonde children playing with marbles	90 (#3 should be #1; remainder of top 20 ok)
AV	blond children playing with marbles	90 (2 of 20 bad)
Lang.	young fair play by means of marble	50 (1 of 2 bad)
OrigSp	7. poder adquisitivo	
Human	buying power	
AV	spending power	45 (11 of 20 bad)
Lang.	purchasing power	100
OrigSp	8. exitoso hombre de negocios en oficina	
AV	successful businessman in office	60 (8 of 20 bad)
Lang.	successful businessman in office	6 (8 of 20 bad)
OrigSp	9. madre e hija horneando pan en la cocina	
Human	mother and daughter baking bread in the kitchen	100 (but no full matches)
AV	mother and daughter [horneando-removed] bread in the kitchen	30 (14 of 20 bad)
Lang.	mother and child bake bread in the kitchen	100 (but no full matches)
OrigSp	10. vejez y soledad	
Human	old age and loneliness	100
AV	oldness and solitude	0
Lang.	old age and loneliness	100

5.1 Spanish

Human translations, tested on PictureQuest: 90% (normalize to 100%)

AltaVista: 53% (59% normalized)

Langenscheidt, word-by-word: 63% (70% normalized)

5.1.1 AltaVista

For AltaVista, we left out the words that AltaVista didn't translate.

5.1.2 Langenscheidt

Langenscheidt, word-by-word: 63% (70% normalized)

For the Langenscheidt word-by-word, we used the bilingual dictionary to translate each word separately as if we knew no English at all, and always took the first translation. We made the following adjustments:

1. Left out "una," since Langenscheidt mapped it to "unir" rather than to either *a* or *one*

2. Translated "e" as *and* instead of *e*

5.2 French

Human translations, tested on PictureQuest: 80%

AltaVista: 86% (100% normalized)

Transparent Language (freetranslation.com): 66% (83% normalized)

Intertran (www.intertran.net:2000): 44% (55% normalized)

[French examples originally drawn from http://humanities.uchicago.edu/ARTFL/projects/academie/1835.searchform.html: French-French]

Source	Example	Score
OrigFr	1. signes du zodiaque	
Human	signs of the zodiac	100
AV	signs of the zodiac	100
TrLang	sign zodiaque	0
IntrTran	[signes] any zodiac	100
OrigFr	2. poisson dans l'eau	
Human	fish in water	30 (14 of 20 bad)
AV	fish in water	30 (14 of 20 bad)
TrLang	fish in water	30 (14 of 20 bad)
IntrTran	fish at water	30 (14 of 20 bad)
OrigFr	3. Les maux d'oreille	

Source	Example	Score
	douloureux	
Human	painful earaches	100
AV	Painful earaches	100
TrLang	the painful ear evil	0
IntrTran	the [maux] [doreille]' distressing	0
OrigFr	4. Prendre un lapin par les oreilles	
Human	to take a rabbit by the ears	65 (7 of 20 bad)
AV	To take a rabbit by the ears	65 (7 of 20 bad)
TrLang	take a rabbit by the ears	65 (7 of 20 bad)
IntrTran	capture a bunny by the ears	80 (1 of 5 bad)
OrigFr	5. Chat qui vit dans les bois	
Human	cat which lives in wood	45 (11 of 20 bad)
AV	Cat which lives in wood	45 (11 of 20 bad)
TrLang	cat that lives in wood	65 (7 of 20 bad)
IntrTran	cat thanksgiving lives at the forest	70 (6 of 20 bad)
OrigFr	6. Sortir d'une maison	
Human	to leave a house	60 (8 of 20 bad)
AV	To leave a house	60 (8 of 20 bad)
TrLang	to go out of a house	95 (1 of 20 bad)
IntrTran	come out dune' dwelling house	90 (18 of 20 bad)
OrigFr	7. Instrument de charpentier	
Human	carpenter's tool	95 (1 of 20 bad)
AV	Instrument of carpenter	100
TrLang	instrument of carpenter	100
IntrTran	implement any carpenter	35 (13 of 20 bad)
OrigFr	8. jouer du violon	
Human	to play the violin	100
AV	to play of the violin	100
TrLang	to play the violin	100
IntrTran	gamble any violin	0
OrigFr	9. Les plaisirs du corps	
Human	pleasures of the body	100
AV	Pleasures of the body	100
TrLang	the pleasures of the body	100
IntrTran	the delight any body	0
OrigFr	10. une jeune fille mange du fruit	
Human	a girl eats fruit	100
AV	a girl eats fruit	100
TrLang	a girl eats fruit	100
IntrTran	a girl am eating any fruit	65 (7 of 20 bad)

5.3 German

Human translations, tested on PictureQuest: 80% (100% normalized)

AltaVista 54% (68% normalized)

Source	Example	Score
OrigGr	1. Golfplatz	
Human	boys golf course	95
AV	golf course	95
OrigGr	2. künstliche Paradiese	
Human	artificial paradise	100
AV	artificial paradiese	0
OrigGr	3. Solarenergie für Autos	
Human	solar energy for automobiles	95
AV	solar energy for auto	95
OrigGr	4. wanderungen durch den Wald	
Human	hiking through the forest	90
AV	migrations by the forest	0
OrigGr	5. Elefanten im Zoo	
Human	an elephant in a zoo	25 (#17 should be #2)
AV	elephant in the zoo	100
OrigGr	6. die Synthese der Desoxynribonukleinsäure	
Human	the synthesis of desoxyribonucleic acid	100
AV	the synthesis of the Desoxynribonukleinsaeure	0
OrigGr	7. schwarze Autos	
Human	black cars	100
AV	black auto	100
OrigGr	8. jungen zusammen spielen	
Human	playing together	60
AV	young together play	35

Source	Example	Score
OrigGr	9. Damen im blau	
Human	women in blue	65
AV	Ladies in blue	75
OrigGr	10. Damen auf Arbeit	
Human	woman at work	65
AV	Ladies on work	40

6 Acknowledgements

I am grateful to Doug Oard for comments on an earlier version of this paper.

7 References

Ballesteros, Lisa, and W. Bruce Croft, 1997. "Phrasal Translation and Query Expansion Techniques for Cross-Language Information Retrieval," in *AAAI Spring Symposium on Cross-Language Text and Speech Retrieval*, Stanford University, Palo Alto, California, March 24-26, 1997.

Fellbaum, Christiane, ed., 1998. *WordNet: An Electronic Lexical Database.* Cambridge, MA: MIT Press.

Flank, Sharon. 2000. "Does WordNet Improve Multimedia Information Retrieval?" Working paper.

Flank, Sharon. 1998. "A Layered Approach to NLP-Based Information Retrieval," in *Proceedings of COLING-ACL, 36th Annual Meeting of the Association for Computational Linguistics*, Montreal, Canada, 10-14 August 1998.

Gilarranz, Julio, Julio Gonzalo and Felisa Verdejo. 1997. "An Approach to Conceptual Text Retrieval Using the EuroWordNet Multilingual Semantic Database," in AAAI Spring Symposium on Cross-Language Text and Speech Retrieval, Stanford University, , Palo Alto, California, March 24-26, 1997. (http://www.clis.umd.edu/dlrg/filter/sss/papers)

Grefenstette, Gregory, ed., 1998. *Cross-Language Information Retrieval.* Norwell, MA: Kluwer.

Hull, David A. and Gregory Grefenstette, 1996. "Experiments in Multilingual Information Retrieval," in *Proceedings of the 19th International Conference on Research and Development in Information Retrieval (SIGIR96)* Zurich, Switzerland.

Jang, Myung-Gil, Sung Hyon Myaeng, and Se Young Park, 1999. "Using Mutual Information to Resolve Query Translation Ambiguities and Query Term Weighting," in *Proceedings of 37th Annual Meeting of the Association for Computational Linguistics*, College Park, Maryland.

McCarley, J. Scott, 1999. "Should We Translate the Documents or the Queries in Cross-Language Information Retrieval?"

Resnik, Philip and Yarowsky, David, in press. "Distinguishing Systems and Distinguishing Sense: New Evaluation Methods for Word Sense Disambiguation," *Natural Language Engineering.*

Smeaton, Alan F., F. Kelledy and R. O'Donnell, 1995. "TREC-4 Experiments at Dublin City University: Thresholding Posting Lists, Query Expansion with WordNet and POS Tagging of Spanish," in Donna K. Harman (ed.) NIST Special Publication 500-236: *The Fourth Text REtrieval Conference (TREC-4)*, Gaithersburg, MD, USA: Department of Commerce, National Institute of Standards and Technology. (http://trec.nist.gov/pubs/trec4/t4_proceedings.html)

Smeaton, Alan F. and I. Quigley, 1996. "Experiments on Using Semantic Distances Between Words in Image Caption Retrieval," in *Proceedings of the 19th International Conference on Research and Development in Information Retrieval (SIGIR96)* Zurich, Switzerland.

Voorhees, Ellen M. 1994. "Query Expansion Using Lexical-Semantic Relations," in *Proceedings of the 17th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 61-70.

Voorhees, Ellen M. 1993. "Using WordNet to Disambiguate Word Senses for Text Retrieval," in *Proceedings of the 16th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 171-180.

Voorhees, Ellen M. and Donna K. Harman, editors, 1999. *The 7th Text Retrieval Conference (TREC-7).*