Multi-lingual Information Access Tools: User Survey

Peggy Nzomo
Faculty of Information and Media Studies
Language and Information Technology Research Lab (LIT.RL)
University of Western Ontario
pnzomo@uwo.ca

Victoria L. Rubin
Faculty of Information and Media Studies
Language and Information Technology Research Lab (LIT.RL)
University of Western Ontario
Tel.: 1 (519) 661-2111 ext. 81405
vrubin@uwo.ca

Isola Ajiferu
Faculty of Information and Media Studies
University of Western Ontario
Tel.: 1 (519) 661-2111 ext. 81364
iajiferu@uwo.ca

ABSTRACT
This research presents the results of a case study on potential users of Cross Language Information Retrieval (CLIR) systems – international students at a Canadian University. The study is designed to test their awareness of Multi-Lingual Information Access (MLIA) tools on the internet and in select electronic databases. The study investigates how non-native English speakers cope with language barriers while searching for information online. We advocate for designing systems that incorporate CLIR options and other MLIA tools to support users from diverse linguistic backgrounds with varying proficiency levels.

Categories and Subject Descriptors
H.1.2 [Information Systems]: User/Machine Systems – human factors, human information processing, software psychology; H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval; H.5.2: [Information Systems]: Information Interfaces and Presentation – User Interfaces; K.4 [Computers and Society]

General Terms
Design, Human Factors, Languages

Keywords
Cross language information retrieval, multi-lingual information access tools, online information-seeking, international students, non-native English speakers, survey.

1. INTRODUCTION
Until recently CLIR research has focused on laboratory experiments entailing physical system development. Few studies exist on the users of CLIR systems. In order to bridge the gap between CLIR research and applications, an effort has to be made to identify the users of CLIR systems and to fully understand how these users can make use of such systems [e.g., 1]. The goal of this study is to gain understanding of the users’ unique needs and their information seeking behavior which provides valuable information for CLIR system designers and LIS professionals.

2. RESEARCH QUESTIONS
This study seeks to answer the following questions in the case of a Canadian (primarily English-speaking) educational setting:

1). How much are bilingual/multilingual system users aware of, and in the habit of using multilingual information access tools available in electronic databases and search engines?

2). Is English language proficiency correlated to satisfaction with results retrieved from online searches?

3). Does low English proficiency (as evidenced by attendance of ESL classes) have an effect on: a) their understanding of database features described only in English (e.g., thesaurus, Boolean operators, proximity, exact phrase); b) their language choice in query formulation; and c) their need to translate search terms from native language into English before performing a search?

4). Do bi-/multilingual non-native English speakers use other languages while searching for information on the web and in electronic databases?

3. BACKGROUND
Previous studies of international students have identified language barriers as one of the challenges they face in their information-seeking process [2,3,4]. In the past, the onus has often been placed on the user to improve their language skills. This study, however, in keeping with the current emphasis on user-centered design, advocates for designing systems that incorporate CLIR options and other MLIA tools that would support users from diverse linguistic backgrounds with varying proficiency levels.

Recent studies [5,6] have shown that due to technological advancements in machine translation, language barriers are becoming less severe. The current study provides a timely investigation on whether international students are taking full advantage of these technological advancements and/or what further steps need to be taken to reduce language barriers in online information seeking, not just for international students, but for all online information users worldwide.

4. METHODOLOGY
We surveyed potential users of multi-lingual information access tools (or CLIR systems) in a population of international students at a Canadian university – the University of Western Ontario, London, ON. This population is chosen because of the assumption that most international students are likely to speak or use at least one other language in addition to English. The survey was administered anonymously (through the Survey Monkey online service, between July 1 and August 1, 2011) preceded by an invitational e-mail to both undergraduates and graduates. Out of the 1135 invitations that were sent out, a total of 151 responses (13.3%) were collected and used as our study sample.

The survey questionnaire consisted of 3 sections. The participant demographic section included questions about gender, age, native language and level in college (i.e., undergraduate or graduate). The second and third sections elicited information about the participants’ use of language tools while searching for information on the internet and in electronic databases, respectively. Two general open-ended questions were included at the end. One was to give participants a chance to comment on what type of...
linguistic help they often needed while searching for information online, while the other sought the participants’ input on what type of library services they thought would be beneficial to them while searching for information online.

5. RESULTS

Of the 151 (n=151) responses collected, 80 (53.0%) self-identified as undergraduates, 70 (46.3%) as graduates, and 1 (0.7%) did not indicate his/her level. Most of the respondents (67.6%) were in the 18-25 age category. A total of 146 out of 151 responded to the question regarding their native language. Thirty-one languages were represented with Chinese having the most students (49).

Other languages that had at least five respondents were: Persian (17), Korean (10), French (8), Spanish (8), English (8) and German (5). Apart from their native languages, English was the language most respondents (82.4%) indicated that they used, with most claiming either an advanced or intermediate knowledge of it. Slightly over half of the respondents (54.7%) used only English to search for information online or in electronic databases. Only about half (51.5%) of the respondents indicated that they did not have difficulty understanding database features that were described only in English. Many (61.3%) indicated that they translate their search terms from their native languages into English before performing a search. When asked what type of linguistic help they needed while searching for information online, most respondents that they need help formulating search terms. However, when asked if they translated their retrieved results before using them, of the 63 students who answered this question, 47 (74.6%) indicated that they did not translate the results of their searches into their native languages.

A series of hypotheses were tested in order to find out if having low English proficiency affects various factors. The corresponding chi-square tests revealed a significant relationship between attending ESL classes and difficulty in understanding database features that were described only in English (Χ²,2=20.0, p<.001). Of the 33 students who reported that they had attended ESL classes within the last two years, 83.3% indicated that they had difficulty understanding database features. Conversely, of the 112 students who indicated that they had not attended ESL classes, only 16.7% reported that they had this difficulty. However, no significant relationship was found between having attended ESL classes and the language used in formulating the search terms (Χ²,2=1.91, p=.385). Similarly, no significant relationship was found between attending ESL classes and translation of terms from native language into English (Χ²,2=2.13, p=.344).

Respondents rated their level of English proficiency on a five-point Likert scale (from very good to very poor). Of the 81 responses, 40 (49.4%) self-rated their English proficiency as good, 22 (27.2%) very good, 17 (21.0%) average, 2 (2.5%) poor and none were very poor. The level of English proficiency had an impact on the way international students search for information online and especially in electronic databases. We tested whether the level of English proficiency, as reported by the students, was correlated to the satisfaction with the retrieved results while searching for information in electronic databases. We found a positive relationship between level of English proficiency and satisfaction with the results retrieved (rₛ(73)=.464, p<.001). In other words, the higher the level of English proficiency, the more the respondents are likely to be satisfied with their search results.

6. DISCUSSION

The findings of this study suggest that language barriers facing international students while searching for information online may be less severe now than they were previously. Most of the international students self-reported advanced to intermediate English proficiency levels, and indicated a comfort level in conducting their internet searches in English. The language choices made by the students while searching for information on the internet seemed to indicate that the students used their native languages just as much as they used English. This may be attributed to the fact that more and more information in other languages is becoming available online and English is no longer as dominant as it was some years ago[7]. However, in searching for information in electronic databases, most participants indicated that they formulated their search terms primarily in English. This may point to the lack of language choices or tools available, and/or a lack of awareness of the language tools available to them.

Corroborating results from [5, 6] we found that technological advancements have helped to ease up some of the language barriers. Translation using machine translation (e.g. Google Translate) was reported as a language tool that was used most and was therefore identified as one of the coping mechanisms employed by international students in the linguistic related challenges they faced while searching for information online.

However, our findings also revealed that language barriers do still exist especially in the process of searching for information. While the respondents in this survey self-reported a high level of English proficiency, when asked what sort of linguistic help they needed while searching for information online, most (48.7%) responded that they would need help in “finding more effective terms”. While this may not entirely be a language related issue, 32 (61.5%) out of the 52 students who responded to this question also said that they sometimes translate their search terms from their native language into English before performing a search. This may indicate that while the students did not necessarily need a translation for the documents they retrieved, they occasionally needed to translate their search terms from their native language into English. MLIA search options in library OPACs and electronic databases were selected by most as a service they would like to have. This was followed closely by library instruction. One respondent asked for an “ability to enter search parameters for a book written in other language using this language”.

7. CONCLUSIONS AND FUTURE WORK

The study has implications for bridging systems’ and users’ perspectives in developing information access tools in broader multi-lingual settings worldwide. In order to further investigate the behavior, perceptions and preferences of potential or already existing CLIR system users, the authors intend to conduct a series of follow-up studies involving observations, interviews, screen capture and eye-tracking software, and search log data analyses. As we advocate wider usage of MLIA tools in electronic databases and library OPACs, we urge LIS professionals to promote awareness of existing CLIR/MLIA systems.

8. ACKNOWLEDGEMENTS

This survey was funded by the Language and Information Technology Research Lab at the University of Western Ontario.
9. REFERENCES


