An Exploration of Users' Needs for Multilingual Information Retrieval and Access

Evgenia Vassilakaki[1(✉)], Emmanouel Garoufallou[2], Frances Johnson[3], and R.J. Hartley[3]

[1] Department of Library Science and Information Systems,
Technological Educational Institute of Athens, Egaleo, Athens, Greece
`evevasilak@gmail.com`
[2] Department of Library Science and Information Systems, Alexander Technological
Educational Institute of Thessaloniki, Sindos, Thessaloniki, Greece
`mgarou@lib.teithe.gr`
[3] Department Languages, Information and Communications,
Manchester Metropolitan University, Manchester, UK
`{f.johnson,r.j.hartley}@mmu.ac.uk`

Abstract. The need for promoting Multilingual Information Retrieval (MLIR) and Access (MLIA) has become evident, now more than ever, given the increase of the online information produced daily in languages other than English. This study aims to explore users' information needs when searching for information across languages. Specifically, the method of questionnaire was employed to shed light on the Library and Information Science (LIS) undergraduate students' use of search engines, databases, digital libraries when searching as well as their needs for multilingual access. This study contributes in informing the design of MLIR systems by focusing on the reasons and situations under which users would search and use information in multiple languages.

1 Introduction

The available information on the Web is increasing exponentially. Information is produced in different types (i.e. images, videos, text) but most importantly in different languages. Although English is still the predominant language used on the Internet, there is a steady increase of the amount of content available in other languages namely Russian, German, Chinese, Japanese, Spanish, French [1]. Therefore, it becomes now, more than ever, apparent the need to develop systems that will support multilingual information retrieval (MLIR) and access (MLIA).

MLIS is defined as the task of searching for relevant information by using criteria in a chosen language (preferable in mother tongue) and retrieve all documents which match all the search criteria, regardless of the language of the documents or the indexed language and present them in a unified list [2,3]. There is an ongoing discussion regarding the reason why would anyone wish to find information that he/she would not be able to read. However, development and use of translation tools has offered an adequate way of accessing the content of the desirable information.

This study aims to investigate users' needs for multilingual information. Specifically, it has the following objectives:

- to identify users' preferences in databases, digital libraries and search engines while searching for information;
- to explore whether knowledge of foreign languages affects users' search behaviours while searching across languages;
- to explore the type of multilingual information users search;
- to investigate the situations under which users search across languages;

It should be noted that here only preliminary results of an overall study conducted regarding users' information seeking behaviour in multilingual digital libraries are presented. This study contributes in providing an insight into the situations under which users would search across languages, the search engines, databases and digital libraries they would employ and their needs for multilingual information. This in turn, will assist in informing the design of the multilingual information systems and interfaces developed. This research addresses the developments in Digital Libraries and the needs of multilingual users for searching searching information in different languages.

This paper is structured as follows. A review of the literature on users' information seeking behaviour when searching across languages is critically presented. Details on the methodological approach adopted in this study are illustrated. Preliminary findings of this study are presented and critically discussed with the relevant literature.

2 Literature Review

Research on MLIR and MLIA is vast [4]. However, this research mainly focuses on developing MLIR systems [5,6], evaluating translation techniques [7,8], testing approaches and models [9]. Only a few studies investigated users' information seeking behaviour in multilingual environments.

These studies explored mainly four different aspects of user behaviour when searching for information across languages [4]. Wu, Ge and He [10], Wu, Luo and He [12] and Wu, He and Luo [11] focused on exploring users' expectations and needs for multilingual information. Wu, Ge and He [10] and Wu, Luo and He [12] argued that translations are essential in order to improve MLIA for the end users. Wu, He and Luo [11] found that users have many multilingual needs, especially when searching across databases for academic purposes.

Bilal and Bachir [13], Ghorab et al. [14], Takaku et al [15], Petrelli and Clough [16] and Vassilakaki [25], Vassilakaki, Johnson and Hartley [17] and Vassilakaki et al. [18,19] explored users' information seeking behaviour in multilingual environments. Users indicated a preference to the basic interface, searching in just one language [17–19,25] and using the browsing feature [13,17–19,25]. Ghorab et al. [14] found that users from different linguistic or cultural backgrounds demonstrate different search behaviours whereas Takaku et al. [15]

argued that novice users will perform a small number of actions and would need a longer time in searching.

Users' knowledge of foreign languages as a factor affecting their multilingual searches was considered by Clough and Eleta [20] and Vassilakaki [25], Vassilakaki, Johnson and Hartley [17] and Vassilakaki et al. [18,19]. Knowledge of other languages proved to be a significant factor affecting users' search behaviour namely in terms of judging and/or relying on translations to retrieve the needed information. Vassilakaki [25], Vassilakaki, Johnson and Hartley [17] and Vassilakaki et al. [18,19] also found that knowledge of languages was also a factor affecting users' levels of trust in the system and confidence in their skills in searching.

Finally, Stafford et al. [21], Shiri et al. [22] and Shiri et al. [23] and Sastry, Manjunath and Reddy [24] evaluated different MLDL interfaces. Stafford et al. [21], Shiri et al. [22] and Shiri et al. [23] found that users would use the thesaurus-enhanced feature and language options, browsing and visualization provided. Sastry, Manjunath and Reddy [24] suggested that when developing MLDL interface different challenges should be considered namely user, search, content and network oriented challenges.

Overall, it could be argued that the number of studies exploring users' information needs in multilingual environments is still limited. However, different factors affecting users' information seeking behaviour across languages were identified namely system's interface, users' knowledge of foreign languages. The importance of these studies is evident when designing and offering multilingual information retrieval interfaces and systems to users.

3   Methodology

A questionnaire was designed based on the relevant literature [11,25]. Lime survey, an open source survey software was used to distribute the questionnaire to users. The questionnaire consisted of 15 closed type questions and relevant subquestions and was divided in three parts (see table 1). The first part inquired on users personal information (i.e. age, gender, level of comprehension with foreign languages) to define user characteristics and knowledge of languages. The second part referred to the frequency of use of specific Databases, Digital Libraries, Search Engines and online OPACs. Finally, the third part addressed the types of information users search across languages, and the means they employ. Only closed type questions were used in order to minimize the time participants had to spend in completing the questionnaire and also as a way to increase users' engagement.

The questionnaire was distributed to the first and fourth year undergraduate students of the Departments of Library Science and Information Systems in both Athens and Thessaloniki. Moodle was used to email the questionnaire to students of both departments and was available throughout May 2015. In total, 219 students participated from both Departments. Lime was also used to analyse the data and excel for creating the necessary graphs. This study collected 219,

Table 1. List of Questions

No	Question
01	Gender
02	Age
03	Degree obtained
04	Please tick the following according to your level of comprehension for each language. For languages not stated here, please fill in the table accordingly
05	How often do you search for information on the following Databases?
06	How often do you search for information on the following Digital libraries?
07	How often do you use the following Search engines to search for information (multiple choices)?
08	How often do you search for information on the following Library OPACs?
09	Besides documents in your native language, do you read any of the following?
10	Under which situations do you search for information in other languages?
11	When searching information for your research/teaching/work, do you feel that it is hindered by not having access to material in other languages?
12	What means do you employ for searching information across languages?
13	Translation tools used before
14	Overall, how do you feel about the effectiveness of the translation tools?
15	For the following sentences please indicate your degree of agreement choosing from 1= agree to 5 disagree

from which 40 (18.3%) were not completed. The data was analysed collectively to shed light on users' behaviour while searching for information across languages.

4   Findings

The majority of the users were female (64%) of age between 18 to 21 years old (64.4%). In terms of users knowledge of foreign languages, the majority of participants (70.8%) are Greek native speakers while they speak English "very good" to "excellent" and have a basic knowledge of French, German, Italian and Spanish.

In terms of the databases users most often use for searching, it was found that the majority of users use "always" Google scholar and then ERIC, Emerald, Web of Science and Elsevier Science Direct. They "sometimes" to "almost never" use PubMed, EBSCO, JSTOR, SpringerLink.

The majority of users use Google Book and Library of Congress World Digital Library to search when it comes to searching in digital libraries. They "very often" search to Europeanna whereas "almost never" to Perseus Digital Library, IEEE Xplore, IEEE Computer Science DL, ACM DL and Alexandria DL.

Google is the most used search engine followed by Yahoo, Bing, ASK and finally Excite. In terms of the OPACs used, the majority of the users search on the Union Catalog of Hellenic Academic Libraries and ATEI Thessaloniki's OPAC followed by the TEI of Athens and Library of Congress OPAC.

Users were asked to define the type of information they read in a language other than their native. The majority of users (44.3%) reads electronic articles

followed by e-books (35.2%), electronic periodicals (30.1%) and electronic newspapers (20.1%). A 16% of the users stated that they only search in their mother tongue (see Appendix, Figure 1).

The main reasons users reported for searching for information across languages were for "completing course assignments" (48%), "shopping online" (44.3%), "going abroad" (36.1%), "interested in international news/affairs" (35.6%), "getting film reviews" (34.7%) and "socializing with friends" (29.7%) among others. It worth noted that only 1.8% of the users stated that do not search across languages (see Appendix, Figure 2).

When users were asked if their search for information is hindered by not having access to material in other languages, users were divided between hindered (33%) and not hindered (30%). Finally, users rely mainly on their own languages skills, online translators and dictionaries when searching across languages (see Appendix, Figure 3).

5   Discussion and Conclusions

This study reported on preliminary results on users' multilingual information needs and behaviours. Specifically, users would continue searching in their mother tongue (Greek), although they do know that the available information in this language is limited. This finding is inline with Bilal and Bachir [13], Vassilakaki [25], Vassilakaki et al. [18,19] and Vassilakaki, Johnson and Hartley [17]. This could be attributed to users' limited knowledge of foreign languages and/or even low confidence in their languages and searching skills as suggested in Vassilakaki [25], Vassilakaki et al. [18,19] and Vassilakaki, Johnson and Hartley [17]. However, further research with more qualitative methods is needed to shed light on users' thoughts and explanations of their preference of a single language for searching.

In terms of the reasons why users would search for multilingual information, it was found that multilingual information was sought both for educational purposes as well as for pleasure and fun. It was also interesting to find that the number of users stating whether finding the needed information is hindered or not by not showing information in different languages was equal. Further analysis revealed that both groups of users have the same level of knowledge of foreign languages, use the same databases, digital libraries and search engines. Therefore, there is no apparent explanation of this finding. Further research is needed to explore the reasons why both groups feel the way they feel about language barriers in information.

This research has also some implications. MLIR systems have to consider users' knowledge and experience in searching across languages; users' familiarity and use of translation tools as well as users' tendency to use just one interface (the basic one) and just one language (their mother tongue) to search for multilingual information.

This study had also some limitations. A quantitative method was used informing our knowledge on users' tendencies and preferences but not on the

reasons why they would search for multilingual information in a specific way. This study employed students in LIS that could be considered experts in searching for information, therefore future research should explore also novice users' needs for searching across languages.

References

1. W3Techs: Usage of content languages on websites, March 2015. http://w3techs.com/technologies/overview/content_language/all (accessed May 24, 2015)
2. Jorna, K., Davies, S.: Multilingual thesauri for the modern world: no ideal solution. Journal of Documentation **57**(2), 284–295 (2001)
3. Chen, A., Gey, F.C.: Multilingual Information Retrieval using Machine Translation, relevance feedback and decompounding. Information Retrieval **7**(1–2), 149–182 (2004)
4. Vassilakaki, E., Garoufallou, E.: Multilingual Digital Libraries: A review of issues in system-centered and user-centered studies, information retrieval and user behavior. The International Information & Library Review **45**(1–2), 3–19 (2013)
5. Pavlov, P., Nisheva-Pavlova, M., Lliev, A., Rousseva, K., Apostolova, N.: Authoeing tools for an academic digital library. Pregled Nacionalnog Centra Za Digitalizaciju, **17**, 9–16 (2010). http://elib.mi.sanu.ac.rs/files/journals/ncd/17/ncd17009.pdf (accessed May 24, 2015)
6. PROMISE: Deliverable 4.1 first report on alternative evaluation methodology (2011). http://www.promise-noe.eu/documents/10156/a0d664fe-16e4-4df6-bcf9-1dc3e5e8c18e (accessed May 24, 2015)
7. Stiller, J.: Leveraging user interaction and collaboration for improving multilingual information access in digital libraries. In: Proceeding of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval e SIGIR 2010 (2010). doi:10.1145/1835449.1835689 (accessed May 24, 2015)
8. Mahajan, P., Singh, N.K.: Online guides to Indian languages with particular reference to Hindi, Punjabi, and Sanskrit. Library Philosophy and Practice1e (2012). http://digitalcommons.unl.edu/libphilprac/749 (accessed May 24, 2015)
9. Budzise-Weaver, T., Chen, J., Mitchell, M.: Collaboration and crowdsourcing: the cases of multilingual digital libraries. The Electronic Library **30**(2), 220–232 (2012). doi:10.1108/02640471211221340 (accessed May 24, 2015)
10. Wu, D., Gu, N., He, D.: The usages and expectations of multilingual information access in Chinese academic digital libraries. In: The Fifth Annual iConference, Champaign, IL, February 2010, pp. 317–322. https://www.ideals.illinois.edu/bitstream/handle/2142/14933/541_camera_ready.pdf?sequenceZ2 (accessed May 24, 2015)
11. Wu, D., He, D., Luo, B.: Multilingual needs and expectations in digital libraries: a survey of academic users with different languages. The Electronic Library **30**(2), 182–197 (2012). doi:10.1108/02640471211221322 (accessed May 24, 2015)

12. Wu, D., Luo, B., He, D.: How multilingual digital information is used: a study in Chinese academic libraries. In: 2010 International Conference on Management and Service Science (MASS), pp. 1–4. doi:10.1109/ICMSS.2010.5576827 (accessed May 24, 2015)
13. Bilal, D., Bachir, I.: Childrens interaction with cross- cultural and multilingual digital libraries. II. Information seeking, success, and affective experience. Information Processing & Management **43**(1), 65–80 (2007)
14. Ghorab, M.R., Leveling, J., Zhou, D., Jones, G.J.F., Wade, V.: Identifying common user behaviour in multilingual search logs. In: Peters, C., Di Nunzio, G.M., Kurimo, M., Mandl, T., Mostefa, D., Peñas, A., Roda, G. (eds.) CLEF 2009. LNCS, vol. 6241, pp. 518–525. Springer, Heidelberg (2010). http://doras.dcu.ie/16037/1/Identifying_Common_User_Behaviour_in.pdf
15. Takaku, M., Egusa, Y., Saito, H., Ishikawa, D., Kando, N., Terai, H., et al.: CRES at LogCLEF 2010: toward understanding user behaviors through analysis of search sessions, search units and click ranks. In: CLEF2010 Workshop, pp. 1–11. http://clef2010.org/resources/proceedings/clef2010labs_submission_120.pdf (accessed May 24, 2015)
16. Petrelli, D., Clough, P.: Analysing users queries for cross-language image retrieval from digital library collections. The Electronic Library **30**(2), 197–219 (2012)
17. Vassilakaki, E., Johnson, F., Hartley, R.J.: Image seeking in Multilingual Environments: a study of the user experience. Information Research **17**(4), 1–12 (2012)
18. Vassilakaki, E., Johnson, F., Hartley, R.J., Randall, D.: A study of users' image seeking behaviour in flickling. In: Peters, C., et al. (eds.) CLEF 2008. LNCS, vol. 5706, pp. 251–259. Springer, Heidelberg (2009)
19. Vassilakaki, E., Johnson, F., Hartley, R.J., Randall, D.: Users' perceptions of searching in flickling. In: ECDL CLEF2009 Workshop, Greece, Corfu, October 01–03, 2009
20. Clough, P., Eleta, I.: Investigating language skills and field of knowledge on multilingual information access in digital li- braries. International Journal of Digital Libraries Systems **1**(1), 89–103 (2010)
21. Stafford, A., Shiri, A., Ruecker, S., Bouchard, M., Mehta, P., Anvik, K., Rossello, X.: Searchling: user-centered evaluation of a visual thesaurus-enhanced interface for bilingual digital libraries. In: Christensen-Dalsgaard, B., Castelli, D., Ammitzbøll Jurik, B., Lippincott, J. (eds.) ECDL 2008. LNCS, vol. 5173, pp. 117–121. Springer, Heidelberg (2008)
22. Shiri, A., Ruecker, S., Doll, L., Bouchard, M., Fiorentino, C.: An evaluation of thesaurus-enhanced visual interfaces for multilingual digital libraries. In: Gradmann, S., Borri, F., Meghini, C., Schuldt, H. (eds.) TPDL 2011. LNCS, vol. 6966, pp. 236–243. Springer, Heidelberg (2011)
23. Shiri, A., Ruecker, S., Bouchard, M., Stafford, A., Mehta, P., Anvik, K., et al.: User evaluation of searchling: a visual interface for bilingual digital libraries. The Electronic Library **29**(1), 71–89 (2011)
24. Sastry, H., Manjunath, G., Reddy, L.C.: User interface design challenges for digital libraries. International Journal of Computer Applications **15**(6), 7–13 (2011)
25. Vassilakaki, E.: Users' image seeking behaviour in a multilingual environments: a grounded theoretical approach. Manchester Metropolitan University, Faculty of Humanities, Law & Social Sciences, Department of Information & Communications, Doctoral Thesis (2012)
26. Wu, D., He, D., Luo, B.: Multilingual needs and expectations in digital libraries. The Electronic Library **30**(2), 182–197 (2012)
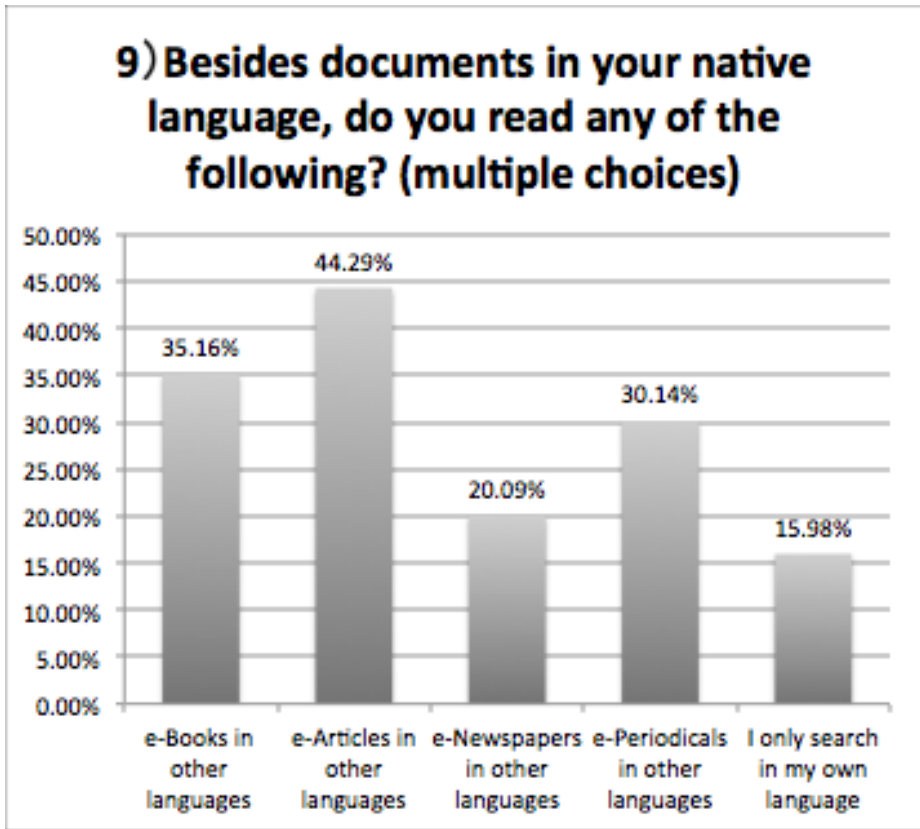
6   Appendix

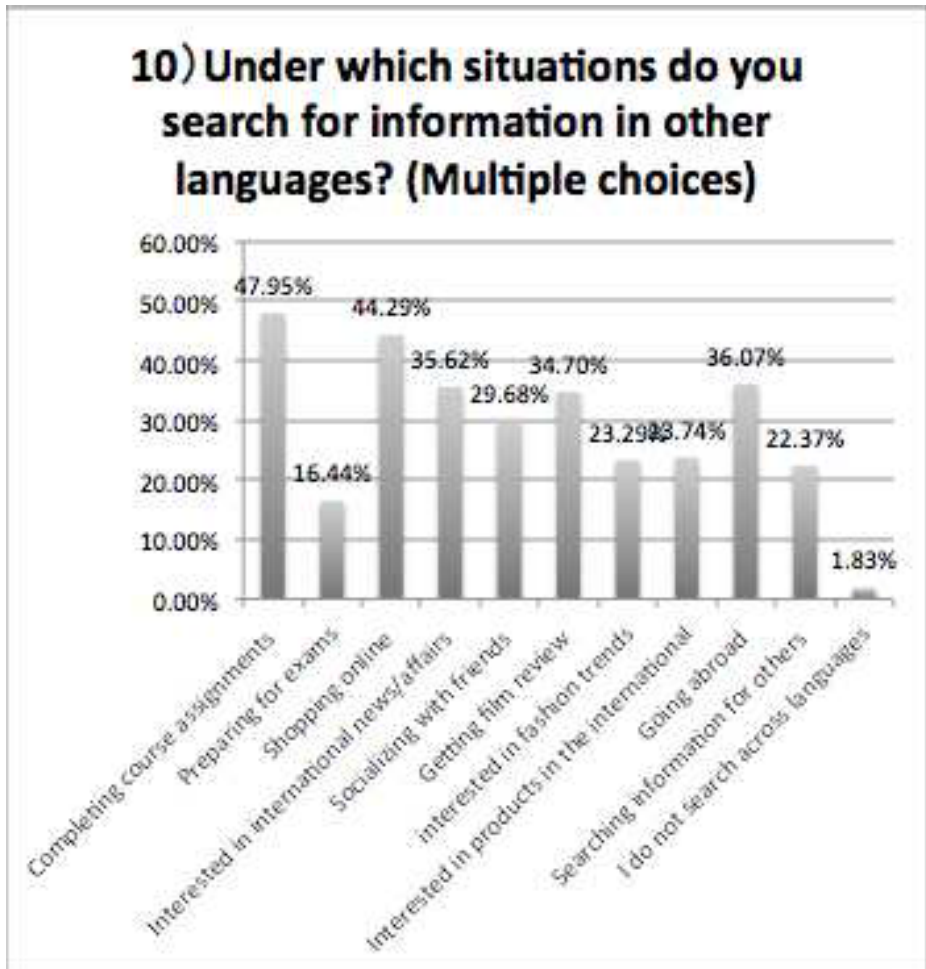**Fig. 1.** Type of documents users read in foreign languages
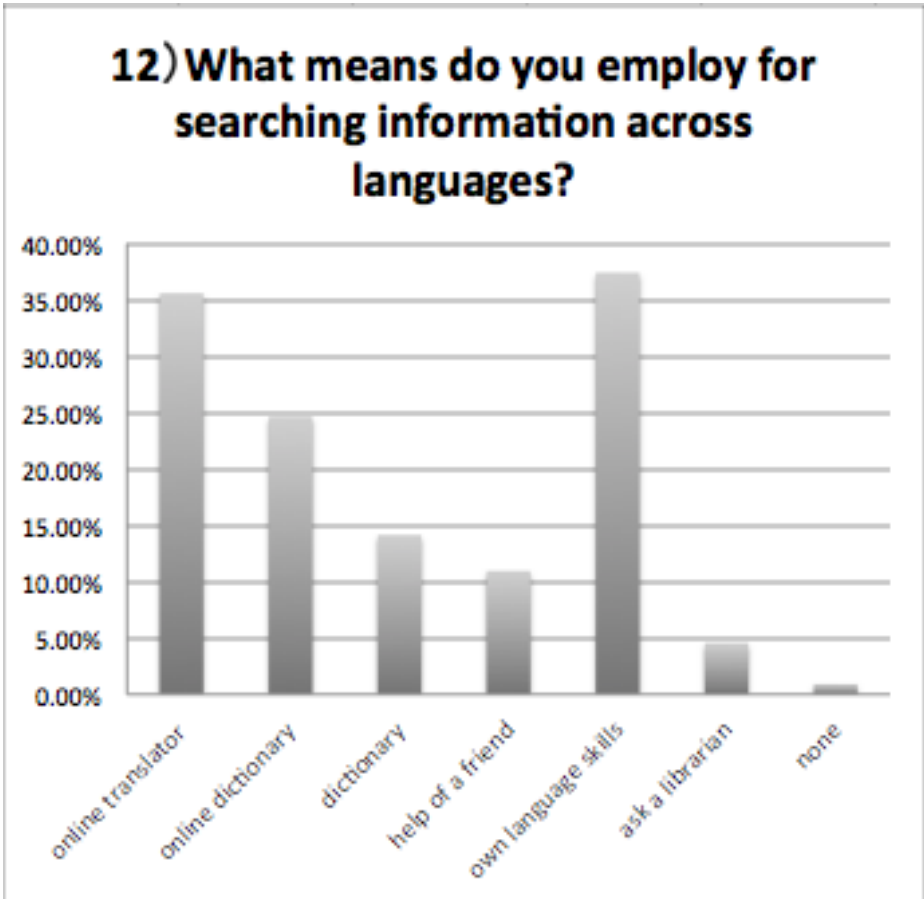
Fig. 2. Reasons for searching across languages

Fig. 3. Means of searching across languages