

Study on the Application of Web Information Retrieval in the Teaching of Language Translation

<http://dx.doi.org/10.3991/ijet.v11i04.5550>

Lin Zhang

Luoyang Institute of Science and Technology, Luoyang, Henan, China

Abstract—The development of information technology at the end of the 20th century fundamentally transformed people's daily lives and greatly influenced education as well, actually not all the undergraduate students of translation have the skills to access, utilize, and benefit from the continually updating online information. This poses a challenge to traditional orthodox translation teaching which is still dominant. Therefore, it is necessary to introduce Web Information Retrieval (WIR) skills to the translation learners. This research focuses on the translation curriculum reform and the effects of incorporating WIR in translation teaching. From data tests, questionnaires, and curriculum experiment prove that WIR incorporating significantly improves translation quality, translators and translation instructors both request that WIR skills should be incorporated in the teaching of language Translation.

Index Terms—Web information retrieval; searching skills; cognitive translation studies; translation teaching

I. INTRODUCTION

With the development and popularization of the internet, utilizing its potential assistance has become a really necessary skill for translators. Valuable information can be accessed through the internet, including text, image, audio and video. For translators however, it can be quite a challenge to promptly find the precise information in need from a sea of cyber information. Therefore, finding ways that translators can quickly find the very information among the enormous resources available online has emerged as a vital topic of conversation and research among those in the field, as well as how they can promote the quality and efficiency of translation with the help of the internet. For more than a decade, web-based resources have become an indispensable tool for both professional translators and translation students. Web information retrieval (WIR) is integrated with artificial intelligence, data mining, natural language processing, computer networking, database technology, cognitive sciences, etc. Search engines only partially meet users' information needs. The utility and ubiquity of web search is making WIR an increasingly important and prevalent research project. Actually not all the undergraduate students of translation have the skills of WIR to access, utilize, and benefit from the continually updating online information.

This paper firstly focuses on WIR in translation, i.e. how to find information in need more quickly and precisely if Boolean operators are used in translating, and if necessary, how to use advanced searching together with basic search methods. Secondly, the question of how to critical-

ly use web information to improve translation is explored. Thirdly, this paper illustrates a multi-case study carried out to explore the web search behavior of a small group of undergraduate translation students in their first year of translation courses. Finally, the main results of the study are outlined and possible implications for translator training are discussed on the information competence of translators.

II. LITERATURE REVIEW

Internet not only transformed peoples' daily life, but deeply influenced human being's views on anything as well. Translation is a kind of highly complicated human behavior. But now more and more people explore new approaches to promote it.

At least 13 years ago, NI C., et al [1] discussed on the scientificity and feasibility of web search engine assisted translating. Later, ZHU and XIE [2] argued that taking full advantage of the search engine, translators can obtain the information they want quickly and definitely, for a better translation of Chinese words and phrases into English. Similarly, WANG Y [3] believed search engine is a good helper in the translation process. It can help translators locate relevant literature, find out equivalent terms in the target language and check whether the translated version is idiomatic or not. After that, ZHOU Jie [4] discussed the function of search engine in translation, especially in improving efficiency and accuracy of the translation. Besides these, Wang J L [5] explained how professional translators can make use of the internet resources such as online dictionaries, search engines, online machine translation websites, translation BBS, translation blogs and so on. WU Y [6] summarized several applications of search engine in translation practice, such as assisting the translators to translate special words and texts, including the application of search engine in translation studies. Based on the above research, GE Y [7] advised that search engines should be fully used in information retrieval to promote translation quality, including idiom translation. MEI Z [8] had a distinctive view on the search engine, that is to utilize cyber corpus for making clear the usages of words. In that way, translation can be bettered as well. These are the pioneer work of WIR for translation.

As engineers, LIU D., et al [9] introduced a new pattern combing search technology and translation technology well, with which an assisted translation search engine system named Yiba is fulfilled. It is a significant experiment in realizing translators' dream in computer information engineering programs. Further studies were carried

by big companies such as Google and Baidu, but there is no article found because of patent protecting.

Of all the studies, only Liang J., et al [10] proposed that translation teaching should incorporate search engine, but there is no concrete experiments and related materials. Since WIR is crucial for translation, it is strictly necessary to carry out research for incorporating WIR in translation, in this paper; we will analyze three major factors that influence students' translation, namely: motivation, beliefs and strategies.

The remainder of this paper is organized as follows. Section 3 is about the research design of this study. Section 4 concludes and analyzes the data, and discusses the enlightenments of this study. Conclusions are summarized in section 5.

III. METHODOLOGY

Translators' information competence is also referred to as "documentary competence" or "instrumental competence". In translation studies, current research literature has typically been situated within multidimensional notions of translation competence. The level of importance that the literature assigns to translation competence varies according to various multi-competence models of translation. Here, skills and knowledge are usually clustered around a number of interrelated sub-competences that make up the notion of translation competence. PACTE [11] viewed translators' instrumental sub-competence as being "made up of knowledge related to the use of documentation sources and information technologies applied to translation". Instrumental sub-competence is made up predominantly of procedural knowledge related to the use of documentation sources, information, and communication technologies applied to translation; i.e., reference books of all kinds [11-12].

On the basis of the models mentioned above, Göpferich developed a translation competence model, which forms the framework of reference for the longitudinal study TransComp [13].

As illustrated in Fig. 2, Göpferich [13] differentiated between tools and research competence. This corresponds to the PACTE group's "instrumental subcompetence" and comprises the ability to use translation-specific conventional and electronic tools, from reference works.

Translator's information competence has also been defined from the perspective of translation-oriented documentation and hence referred to as "documentary competence". The specific models for the development of information competence for translators were known as "INFOLITRANS" (information literacy for translators).

The above model draws, among others, on information literacy competency standards like the ones developed by the Association of College & Research Libraries [14], which is a division of the American Library Association. It defines information literacy as "a set of abilities requiring individuals to recognize when information is needed and have the ability to locate, evaluate, and effectively use the needed information" (ibid.). It is a term widely used in the area of higher education in connection with the instruction of information skills that may form the basis for lifelong learning. It is therefore a notion "common to all disciplines" [14]. The model illustrates the five standards listed in Table I.

Acquiring information literacy in translation courses has gradually been recognized by the instructors. WIR is often taught at the MTI (Master of Translation and Interpreting) level, which is increasingly being introduced at the undergraduate level. The purpose of this survey aims to investigate the status quo of incorporating WIR in translation teaching. The study attempts to identify any

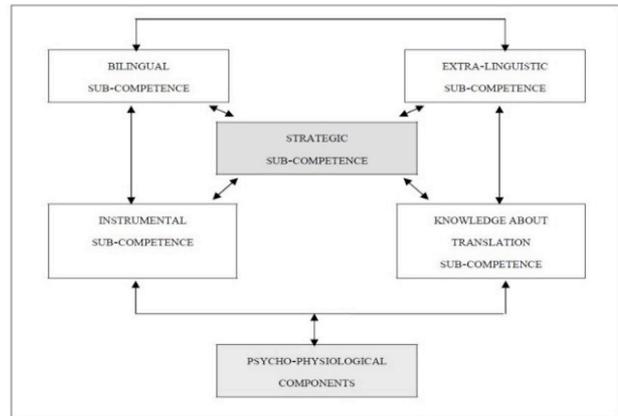


Figure 1. PACTE Translation Competence Model

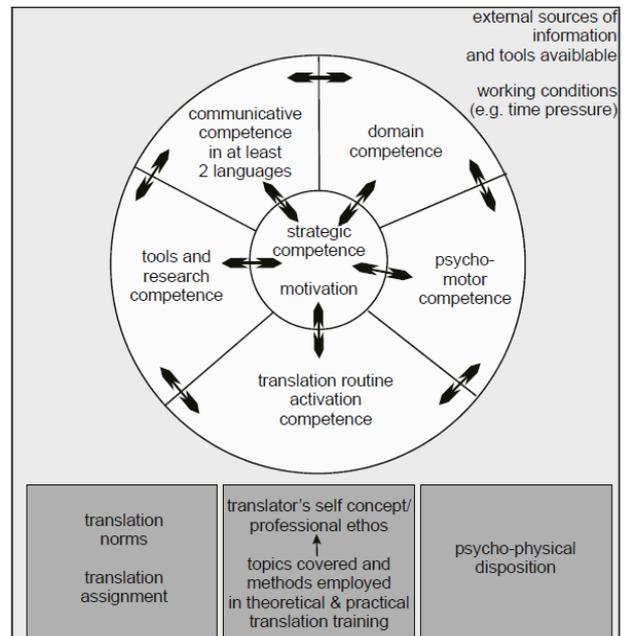


Figure 2. Göpferich Translation Competence Model



Figure 3. Information Literacy Competency Standards for Higher Education

STUDY ON THE APPLICATION OF WEB INFORMATION RETRIEVAL IN THE TEACHING OF LANGUAGE TRANSLATION

TABLE I.
FIVE STANDARDS OF INFORMATION LITERACY COMPETENCY

Standard 1	The information literate student determines the nature and extent of the information needed.
Standard 2	The information literate student accesses needed information effectively and efficiently.
Standard 3	The information literate student critically evaluates information and its sources and incorporates selected information into his or her knowledge base and value system.
Standard 4	The information literate student, individually or as a member of a group, uses information effectively to accomplish a specific purpose.
Standard 5	The information literate student understands many of the economic, legal, and social issues surrounding the use of information and accesses and uses information ethically and legally.

existing problems in translation instruction by integrating the information literacy competency model in Luoyang Institution of Science and Technology (LIT). There are altogether 80 subjects of English major students (translation oriented) involved in this investigation. In order to find the necessity and effects of incorporating WIR in translation, the author carried out a comparative research based on tests and questionnaires before and after incorporating WIR as shown in Figure 4.

IV. RESULTS OF ANALYSIS AND DISCUSSION

A. Curriculum Reform Based On Wir

According to the data from this investigation, 60.0% of students claim that teachers spend more time lecturing; i.e., the teacher is dominant in the translation course. 38.8% of students think that teachers play the major role in the instruction process. Furthermore, 68.12% of students hold the view that they have no knowledge about web-aided translation technologies, while only 2% students attempted the finite experiment of WIR in translating. Via the data, we can observe that the majority of students lack the abilities to access, evaluate, and use information effectively in the translating process. On the other hand, students show their inventiveness and initiative in achieving the goal of constructing the original knowledge of WIR, not just in translation but all aspects of the learning process. There is no doubt that the searching skills and theories come to be more and more important in the practice of language translation.

Moreover, other surveys established that additional instruction aimed at developing WIR skills may also take place within optional courses that make up the curriculum, such as "IR and Web Search" "Computer Science (or IT) Applied to Translation", "Terminology" or "Electronic Tools and Resources for Translation". The importance of WIR skills is highlighted in Chinese universities. Well ahead of the rest of the items, the translation trainers consider the foremost need of a specialized translator to be the skill of IR (68%). This is second to subject knowledge (34%), which is the product of a sound global information competence continually enriched and updated for whatever field of expertise (ibid: 61).

Nevertheless, when teachers were asked which elements of their instruction needed to be added to the curriculum, they felt "students had not yet acquired the necessary IR skills", suggesting "reinforcing knowledge of documentation techniques in general" and of IR in particular (ibid: 62). With regard to the latter, Pinto Molina and Sales Salvador state that "teachers are no doubt aware that their students tend to conduct their IRs in an intuitive fashion, low on structuring and only vaguely systematized, and suppose that a search engine like Google 'knows everything'" (ibid.).

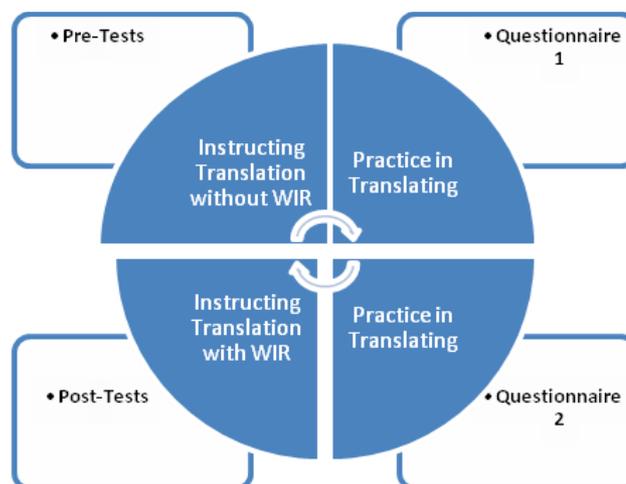


Figure 4. Research Methodology

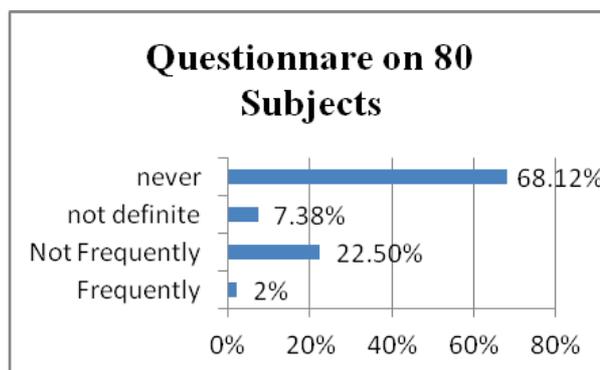


Figure 5. Student's experience on using the WIR

The results of the above survey also show that 65% of the teachers instructed that "the core component under the existing curriculum does not fully meet the students' needs" for developing their documentary competence. In most universities, there is no link established among these three areas [computing, documentary research, and terminology] or between them and practical translation activities.

The theoretical basis for this study was primarily informed by the approach of integrating WIR skills, but also by other cognitive as well as affective-oriented information-seeking models. Most of the researchers view IR as the purposive seeking of information in relation to a goal, where "the emphasis lies on the application of translation solutions—sometimes called strategies or procedures—to specific problems" [15].

Although some instructors who would be instructing the search engine course had performed extensive research in using search engines for preserving the Web, they had never taken a WIR course and had only studied

STUDY ON THE APPLICATION OF WEB INFORMATION RETRIEVAL IN THE TEACHING OF LANGUAGE TRANSLATION

IR independently, therefore had no base materials to develop a comprehensive WIR course. Therefore, a curriculum survey was performed of existing course offerings in WIR. The courses listed in Table 1 were found using word-of-mouth, Google, and referring links from class web pages. It is not meant to be an exhaustive list of all WIR courses, but it is representative of the types of courses available on WIR issues. Several courses were excluded because they only focused on IR with little Web application or were non-technical in nature.

The data from this survey helped in choosing topics that would be more readily accessible to undergraduate students in an endeavor to formulate a comprehensive and effective curriculum. Many courses listed in Table 1 contain extensive websites with syllabi and project guidelines which greatly aided in preparing the curriculum. In some cases, the instructors of the course were contacted for more specifics about the course's content. A core curriculum was then compiled for the IR course based on the curriculum surveyed and on the instructor's research interests. The course had several high-level learning objectives: A basic understanding of how the Web is organized and its fundamental properties. Understand how search engines collect Web content, index it, and present the most relevant results for a given query, including the technical, legal, finance, and social obstacles that make web search very complex. Be aware of practical research skills in the field of WIR. How to elevate has chosen the proper translation materials. Flexibly apply comprehensive skills in translation practice. Perform web search tasks for translation problem solving adeptly. To meet these learning objectives, the course covers (in order) the topics shown in Table II.

Other topics to incorporate into this course include text mining, search interfaces, Semantic Web, latent semantic indexing, collaborative filtering, query log analysis, distributed computing issues, web agents, and machine learning. This curriculum is representative of other WIR courses described in the literature mentioned above.

B. Classroom Evaluation of Students

To examine whether these searching skills for teaching IR are effective in a real educational process, several studies with students of IR courses were conducted. In each of these studies we evaluated several searching skills. This section reports results of five classroom studies performed in the context of a graduate course, "Information Storage and Retrieval," offered annually at the School of Foreign Languages, LIT. The course, which lasted for sixteen weeks during one semester and had 60 participants, provided the setting to collect the study's research data. The basic search skills and Google search syntax, which are the focus of this paper, were evaluated in each study.

All studies of the WIR have the same multi-stage design as presented in Figure 5. The goal of this design was

to assess both objectively and subjectively whether the searching skills are actually working as useful aids to help the students more easily solve for themselves the underlying problems. Knowledge gained between pre-test and post-test was used as an objective measure. Students' answers to a questionnaire about the methods and their implications were used as a subjective measure. The studies of different searching skills were performed during different weeks of the semester. In a given week, the students learned the skills of the subject (Table 2.) during a lecture and the instructor introduced the corresponding searching skill, encouraging them to utilize it for their deeper information literacy in translation that they acquired in the classroom.

The two P-tests are aimed at eliciting data on students' demographic characteristics (gender, age, academic qualifications and working languages) as well as data on their previous knowledge of and experience with both translation and web searching.

Before the lectures devoted to translation related IR skills, the students took a pre-test. Overall, they had little experience in translation (both academically and professionally); hence their level of translation expertise was naturally low. The study's research data shows that the students' web search expertise was initially rather low as well.

The homework assignments given, among other things, specifically required them to use the skills introduced in the lecture for actively solving practical hands-on problems. The role of this homework was to engage the students in active exploration of the information in the context of problem solving. Figure 6 shows a sample of the translation tasks.

BALANCE SHEET		Translation Task Objective : Using the Google "filetype:" Searching Syntax & Bilingual IR	
資產負債表			
as at 31 December 2008 於二零零八年十二月三十一日 (Expressed in Hong Kong Dollars (以港幣列示))			
	Note 附註	2008 \$'000 千元	2007 \$'000 千元
Assets			
Fixed assets	15(b)	206	172
Investments in subsidiaries	17	2,377,599	1,653,359
Interest in associates	18	6,937	580,474
Investments in debt and equity securities	19(b)	571,825	290,845
Deferred tax assets	31(b)	5,549	—
Amounts due from group companies	20(a)	459,651	443,651
Other debtors	23	11,591	51,167
Deposits at banks with original maturity more than three months		24,325	160,531
Cash and cash equivalents	26	251,778	622,159
		3,709,461	3,802,358
Liabilities			
Deferred tax liabilities	31(b)	—	157
Amounts due to group companies	20(b)	1,592,641	1,363,388
Other creditors	34	8,184	9,312
Current taxation	31(a)	—	10,893
		1,600,825	1,383,750
Net assets		2,108,636	2,418,608
Capital and reserves			
Share capital	37	71,086	70,764
Reserves	36(b)	2,037,550	2,347,844
Total equity		2,108,636	2,418,608

Approved and authorized for issue by the board of directors on 18 March 2009. 董事會於二零零九年三月十八日核准及授權發布。

Figure 6. Sample Students' Translation Task

TABLE II.
THE CONTENT OF WIR APPLICATION IN TRANSLATION TEACHING

WIR background	WIR vs. traditional IR; Web search history
Search engine	Types of web search queries; Evaluating search engine efficiency (relevancy, precision, and recall); Google as a powerful search engine
Web crawling	Selection policies (URL characteristics, spider traps, repository freshness); URL normalization ; Sitemap protocol
Indexing	Boolean retrieval model; Synonyms, homographs, and misspellings; Term frequency and inverted document frequency
Ranking	Content-based ranking; Link-based ranking
Other	Link analysis; Wildcard Matching; Search by Image; Multiple file formats; Terminology

STUDY ON THE APPLICATION OF WEB INFORMATION RETRIEVAL IN THE TEACHING OF LANGUAGE TRANSLATION

TABLE III.
SURVEY QUESTIONS FOR THE SUBJECTIVE FEEDBACK ANALYSIS

Number	Questions
Q1	Before I take part in the translation course, I have no ideas of using web recourse effectively for translation.
Q2	The skills are particularly useful for analyzing the research activities which form an integral part of translation processes
Q3	The teaching program on the WIR is very important for learning.
Q4	My perceptions of search success and search satisfaction.
Q5	I think that in the context of the IR course Google should become one of the key engines.
Q6	I would recommend the WIR courses to a friend who is taking a translation course next semester.
Q7	Teaching IR should be a required subject.

To evaluate the results of their learning, the students took a post-test and filled in a questionnaire about the searching skills and their impact. The post-test was identical to the pre-test (Figure 6 as an example). The questionnaire included seven questions concerning a skill or the impact (Table III). A five point Likert scale was used for the questions ranging from 1 (Strongly Agree) to 5 (Strongly Disagree).

The role of the questionnaire was to evaluate student attitudes to the WIR in translation teaching. However, the questionnaire alone was not sufficient for a reliable evaluation. To make sure that the student's attitude, whether positive or negative, is grounded in the student learning experience, we had to check whether student knowledge increased after using these skills flexibly, and examine the correlation between knowledge gain and attitude.

To calculate knowledge gain, which measures an increase in the student's knowledge of the topic, we compared pre- and post-test scores. Because of the minor differences in the number of pre/post-test questions over the span of study semesters, we used normalized pre- and post-test scores, normalized to range from 0 to 100, to calculate the knowledge gain. This allowed evaluation of the effect of the IR practices.

This section analyzes students' subjective feedback on the visualization tools. The distribution of student answers is shown in Figure 7.

Q1 differs from the other questions since it is focused on the student's prior knowledge about the corresponding IR learning. As the data shows, about 80% of the subjects agreed or strongly agreed that they lack knowledge of WIR, even which related to enhancing translation expertise. This result is consistent with the objective analysis in the previous section, where the average pre-test score of the Boolean model was 78.7 out of 100.

The next two questions (Q2 and Q3) assess the usefulness of WIR in the classroom. About 80% to 90% of the students gave positive answers (Strongly Agree or Agree) for both questions. Almost no negative answers were submitted. Moreover, the experience of learning was valued even higher than its usage. This is strong evidence in favor of constructing and promoting the cognitive information search expertise curriculum.

Question 4 asked how easy it was to master the WIR and use the skills in translation work. Overall, the students showed a high level of satisfaction with the after-learning practices (83%). This indicates that the students who were involved in the WIR courses in the first semester may need further improvement.

Questions 5 to 7 asked in different words about the importance and usefulness of the searching skills in the context of an IR course. Students' answers confirmed that the

learning of IR, which helped them with a less well known topic, was most valuable for them. Over 80% of the students agreed that the Google searching syntax should become one of the key course tools and over 90% wanted to recommend the WIR program to students taking the course next semester. The feedback is overwhelmingly positive, with the WIR being considered important and useful by a clear majority of students. Moreover, about 90% of the students agreed or strongly agreed that teaching IR should be a required subject for students majoring in translation.

As described earlier, we measured the pre- and post-test scores for two educational visualization tools for IR: one for the Boolean and one for the vector model. Table IV shows the basic statistics comparing the scores.

We can first observe that there are increases between pre- and post-test scores for both models. Test scores increased by 5.90 and the number of students who maintained the same progress amounts to a larger percentage than pre-test. These differences were statistically significant, which means there was actual knowledge gain after the students learned how to use WIR during the translation process. Thus, it can be hypothesized that searching skills could be especially useful for teaching hard-to-understand topics, where traditional translation methods maybe less effective. The standard deviations of pre-test and post-test scores were just 2.16 and 2.03 respectively. These levels are relatively small, suggesting the starting levels of student knowledge within each topic were quite similar.

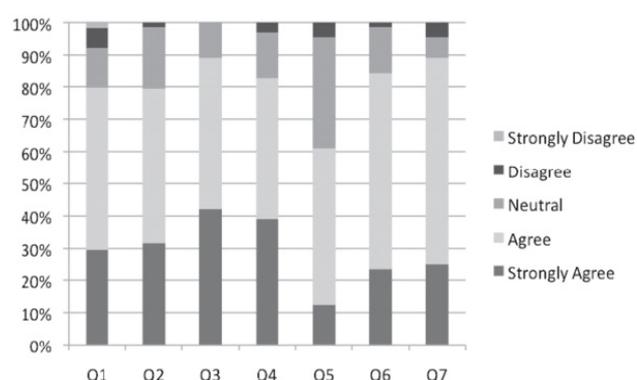


Figure 7. Students Feedback for the WIR in Translation Teaching

TABLE IV.
COMPARISON OF TEST SCORES FOR THE TWO SEARCHING SKILLS

Translation Task	Average Score	Standard Deviation
Pre-test	78.70	2.16
Post-test	87.60	2.03

V. CONCLUSION

The Internet has profoundly changed the way of translators' work. Yet, research into information-seeking patterns of Internet employment among translators has so far been rather rare. The author has carried out a valuable study of WIR practices among translators and translation students, thus contributing to the discussion of translators' WIR skills. This paper also presents some experience in developing a WIR curriculum for teaching and learning translation.

One conclusion is that the students' translation experience reflects the translator's information needs, with the more experienced translators using available resources mainly for producing target texts, and the less experienced ones for understanding source texts. The student participants' searches were often limited to dictionaries, whereas the more experienced participants utilized a wide variety of resources for frequency checks, parallel texts and encyclopedic information. Also, the type of queries used correlated with the participants' level of web search expertise, with the less experienced participants favoring simpler queries with little use of advanced search operators or other methods of query refinement.

Secondly, the above findings clearly have important implications for translation pedagogy. A practical design of the WIR courses is presented in this paper, and the author's experiences in teaching the course have been shared.

Finally, the author argues that WIR skills should be an integral part of translation teaching to promote the translation learners' comprehensive competence.

REFERENCES

[1] NI Chuanbin, Guo Hongjie, Zhao Yong. "On the Scientificity and Feasibility of Web Search Engine Assisted Translating". *Shanghai Journal of Translators for Science and Technology*, 2003, vol.14, no.2, pp.53-55+64.

[2] ZHU Mingju, XIE Shaohua. "Translating Chinese Words and Phrases into English with the Help from Search Engine". *Shanghai Journal of Translators for Science and Technology*, 2003, vol.13, no.1, pp.59-62.

[3] WANG Yong. "Search Engine and Translation". *Chinese Science & Technology Translators Journal*. 2013, vol.21, no.12, pp.28-30+41.

[4] ZHOU Jie. "Search Engine Assisted Translation". *Computer Assisted Foreign Language Education (CAFLE)*, 2007, pp.62-65.

[5] WANG Junli. "The Application of Internet Resources for Translation". *Chinese Science & Technology Translators*. 2014, vol.19, no.1, pp.36-40.

[6] WU Yaowu. "Search Engine's Application in Translation Practice and Translation Studies". *Journal of Henan Institute of Science and Technology*. 2013, vol.24, no.5, pp.44-48.

[7] GE Yi. "Proofreading Functions of Search Engines in the English and Chinese Idioms Translation". *Journal of Jixi University*, 2013, vol. 16, no.5, pp.75-77.

[8] MEI Zhongwei. "The Application of Internet Corpus for English Word Usage Research". *Journal of Changchun University of Science and Technology*, 2012, vol.12, no.1, pp.56-57+59.

[9] LIU Dongfei, CHEN Zhixian, DENG Juntao, YU Jun. "Research and Implementation of Assisted Translation Search Engine". *Computer Engineering*, 2007, vol.9. no.3, pp.216-218.

[10] LIANG Jinzhu, JIA Liping, ZHENG Xiuyan. "Improvement of Translation Efficiency by Using Internet in Translation Teaching". *Journal of Chengdu Textile College*, 2008, vol.31, no.3, pp.58-60.

[11] PACTE.O. "Investigating Translation Competence: Conceptual and Methodological Issues". *Meta*, 2005, vol.50. no.2. pp. 609-618. <http://dx.doi.org/10.7202/011004ar>

[12] PACTE.O. "Building a Translation Competence Model". *Triangulating translation: Perspectives in process oriented research*. 2003, pp.43-66.

[13] GÖPFERICH, S. "Research project TransComp: The development of translation competence". <http://gams.uni-graz.at/fedora/get/container:tc/bdef:Container/get> (19.04.2009) . 2007.

[14] Association of College & Research libraries(ACRL). "Information Literacy Competency Standards for Higher Education". <http://www.acrl.org/ala/mgrps/divs/acrl/standards/standards.pdf>. 2000. [accessed: 23 July 2011].

[15] González Davies, M. *Multiple Voices in the Translation Classroom: Activities, Tasks and Projects*. Manchester: St. Jerome. 2004. <http://dx.doi.org/10.1075/btl.54>

AUTHOR

Lin Zhang is with School of Foreign languages, Luoyang Institute of Science and Technology, Luoyang, 471023, China. (E-mail: maylynn2012@126.com). She is a master of arts majoring in Translation Technology, Corpus Linguistics, and Computational Linguistics.

This paper is partial fruits of the Soft Science Project (No. 152400410540) of Science and Technology Department of Henan Province, and the Education Reform Research Project (No. 2015-JSJYYB-126) of Education Department of Henan Province. Submitted 06 February 2016. Published as resubmitted by the author 15 March 2016.