

Iranian Perspective on Corpora in Translator Training

Helia Vaezian

Faculty of Foreign Languages and Literatures
University of Tehran, Tehran, Iran.

ABSTRACT

Corpora first found their way into Translation Studies as a means to study the language of translation and its distinctive features (Baker, 1993). They were applied, however, by translation researchers in other areas such as translation evaluation and machine translation as well. Translator education is another area which has enjoyed the benefits of corpora in recent years. This paper elaborates on the applications of corpora in translator training and the kind of information they have to offer to student translators in this context. It further explores the benefits of corpus-based translation teaching for student translators in terms of their professional prospects. The paper finally reports the results of an experiment in which an undergraduate class of translation students from the Department of Translation Studies at Allameh Tabataba'i University, Tehran, Iran was introduced to disposable corpora. The problems students faced in building disposable corpora of Persian texts and, later on, in analyzing Persian corpora with corpus analysis tools and finally the solutions adopted to resolve the problems are all explained in detail.

Keywords: *Translator education, Corpus-based Translation Teaching, Disposable Corpora, Corpus Compilation, Corpus Analysis Tools.*

Introduction

The introduction of corpora into Translation Studies dates back to 1993 when Mona Baker first discussed the importance of applying corpora to study the language of translation (Baker, 1993). Baker states that translation as a unique communicative event must be recorded and explored and corpora can serve the purpose well by providing researchers with authentic samples of translational language (ibid). According to Baker, "large corpora will provide theorists of translation with a unique opportunity to observe the object of their study and to explore what it is that makes it different from other objects of study" (1993: 135). Baker further suggests applying corpus tools to investigate four features of translated texts, i.e., simplification, explication, normalization and levelling out (1996). In this context, parallel and comparable corpora of different languages were compiled and used to study the proposed translation universals in different languages. The presence of corpora in Translation Studies gradually extended and they found their way into areas such as translation evaluation and translation aids. In the context of translation evaluation, corpora were shown to provide evaluators with solid criteria to evaluate translations. Bowker (2000) is possibly the first researcher to adopt a

corpus-based approach to translation evaluation. Drawing on the advantages of corpora over the conventional resources available to evaluators i.e., dictionaries, parallel texts, subject field experts and intuition, Bowker shows how corpora “allow evaluators to both identify and correct a larger number of errors and to do so in a more objective way” (ibid.: 206).

Machine Translation systems and Translation Memories (TMs) have also benefited from the existence of corpora. Parallel corpora, for instance, as databases of translations, are an essential part of an approach to machine translation called Example-based Machine Translation (EBMT) (Somers, 2003). Corpora have also been used in studies on translators’ style. An example of corpus-based research on translator’s style is the one performed by Baker (2000) who shows how linguistic habits of two translators differ and how they are reflection of translators’ styles. Translator training is another area which has enjoyed the benefits of corpora in recent years. This paper aims to elaborate on the applications of different types of corpora in the context of translation classrooms. It further draws on an experiment and explains the pros and cons of introducing disposable corpora into a translation classroom in Iran in the context of translation into Persian. Corpora and Translator Education Bernardini (1997) is among the first scholars to propose the integration of corpora into translator training. Drawing on the benefits of corpora for student translators, she suggested that the traditional translation teaching should be complemented with large corpora concordance to let trainees acquire the skills they need to become a (professional) translator (ibid).

According to Bernardini, “educating learners to use comparable corpora as reference tools in their everyday language may result in a better-documented, more accurate as well as more fluent translations” (2004:20). Zanettin, soon after, puts forward the idea of Translator Trainee Workstations which proposes the introduction of corpus-based translation activities into translation curriculum (Zanettin: 1998). According to this author, “a translator trainee workstation comprising a word processor, bilingual corpora and facilities for bilingual concordance together with other resources may constitute a valuable aid in the training of translators” (ibid:626). He further highlights the importance of adding corpus skills into the translators’ professional competence (2002: 10). A similar view is echoed by Maia (2003:52) who says “the methodology involved in making corpora and extracting terminology should be part of the teaching curriculum, not an optional extra, particularly given the pressure on translators to use computer-assisted tools”.

With this ever-increasing interest in using corpora in translator training, many researchers and practitioners within the discipline of Translation Studies have worked on the applications of different types of corpora in translation classrooms in recent years. The respective studies mainly look into the kind of information trainee translators can elicit from corpora and the effect of using corpus data on the quality of translations produced. The following section provides a brief overview of the important research available on the subject, classified based on the corpus type.

Corpora as Translation Aids

General Reference Corpora as Translation Aids

According to Sinclair (1991:17), a general reference corpus is “a collection of material which is broadly homogeneous, but which is gathered from a variety of sources so that the individuality of a source text is obscured unless the researcher isolated a particular text”. He further states that the function of a general reference corpus is to provide thorough information about a language and therefore it must be large enough to provide all the relevant varieties of the language in question (ibid.). One of the significant studies on the applications of TL general reference corpora as a translation aid for student translators is the one carried out by Stewart (2000), who elaborates on uses of the British National Corpus (BNC) as a translation aid for trainee translators translating from Italian into English. The students, in fact, use BNC as a target language corpus to translate tourist brochures from their first language into their second language, i.e., English. The results of Stewart’s research show that using frequency lists, along with concordances, can help trainee translators produce target texts containing natural collocations.

Parallel Corpora as Translation Aids

A parallel corpus as defined by Teubert and Cermakova (2004:122) is “a corpus consisting of original texts in one language and their translations into another language”. Monzo’s study on parallel corpora in translation classrooms is an interesting study on the subject (2003). Monzo introduces an online database fed with the original legal documents and their translations in three languages into the training of legal translators. Furthermore, a tool is developed to be applied by trainees to access corpora and search for the relevant information. According to Monzo, the information provided by the parallel corpora helped students to see how other translators translate and which translation strategies they may adopt when translating different texts. One of the interesting outcomes of this experiment, as observed by Monzo, was the increase in student confidence in translating legal documents (ibid.).

Bilingual Comparable Corpora as Translation Aids

Olohan (2004) defines bilingual comparable corpora as corpora of comparable original texts in two languages. One noteworthy study on the applications of bilingual comparable corpora in translation classrooms is the one performed by Zanettin (2001). He introduces a comparable corpus of English and Italian newspaper texts dealing with the Olympics, along with concordance software, into a translation classroom to study the ways in which the corpus may be useful to undergraduate student translators. The data offered by the comparable corpora, as Zanettin observed, helped students distinguish cognates from false friends and “allowed learners to contrast source and target languages at various levels, from single words and phrases to discourse functions and organizations” (ibid.:185).

DIY Corpora as Translation Aids

Zanettin defines a DIY (Do-It-Yourself) corpus as a “collection of Internet documents or more precisely web pages in HTML format created ad hoc, as a response to a specific text to be translated, which is an open and disposable corpus” (2002a:242). Varantola (2003) has conducted an interesting experiment on the applications of DIY disposable corpora as translation aids in which she asks a group of student translators at the Department of Translation studies, University of Tampere, to use the World Wide Web as a resource to compile disposable corpora of their target language. The main advantage of using disposable corpora in translation, as Varantola states, was to give reassurance to students in the sense that “when relevant corpus information was available, the users often gained reassurance for their strategic decisions as well as the actual lexical choices” (2003:67).

Specialized Monolingual Corpora as Translation Aids

Specialized corpora are defined by Sinclair as corpora usually smaller than general reference corpora which are designed based on various purposes (cited in Pearson, 1998). Bowker’s (1998) study on the applications of specialized corpora as translation aids is one of the most interesting on this subject. Bowker hypothesizes that a specialized monolingual native language corpus can be used as a resource to help students of translation improve subject field understanding and specialized native language competence. Fourteen translation students from the School of Applied Language and Intercultural studies at Dublin City University in Ireland participated in the Bowker’s study. The students were given two texts to translate from their foreign language (French) into their mother tongue (English). The results of the study showed “a general trend towards improved quality translation for the categories of subject field understanding, correct term choice and idiomatic expression” in translations done using the corpus (ibid.:18).

What Information do Corpora have to offer to Translators?

The following section provides a categorization of the kind of information corpora have to offer to translators in general.

Collocational Information

Collocations are defined as “words that appear together with a greater than random probability” (Bowker, 2002:64). The concordance and collocates features in corpus analysis tools can be used to extract collocational information from corpora. The information offered by the concordance and collocates, among other things, can help translators become familiar with the phraseology of a specialized field, learn about the adjectives that collocate with nouns under question, see which words collocate with the verbs under question or simply verify decisions made based on other tools.

Conceptual Information

Understanding the source text and its concepts is the prerequisite for any act of translation. Using the concordance feature of the corpus analysis tools, translators can

gain access to the concepts expressed in the source texts in their contexts. They can further acquire information about the respective concepts from the multiple contexts in which they have been used.

Terminological Information

Terminology refers to the terms of a particular specialized subject. Frequency lists along with concordance and collocates displays of the specialized corpora can be consulted by translators to become familiar with the terminology of the text type under study. Moreover, translators can identify specialized terms by comparing the frequency list of a monolingual specialized corpus against the frequency list of a less specialized corpus (Bowker and Pearson, 2002).

Linguistic Information

Corpora are representations of the language as it is used; they thus can be reliable sources of linguistic information about the language they represent. Monolingual corpora of the source and target languages, for instance, can be consulted by translators to verify their linguistic intuitions as to how certain structures are constructed in the language under study. Bilingual comparable corpora and parallel corpora can further help translators compare and contrast two languages involved to acquire a contrastive knowledge of the respective source and target languages.

Orthographic Information

Corpora also contain orthographic information about the language they represent and can be used by translators to choose between different orthographic variants.

Information on Semantic Prosody

Semantic prosody is defined as “the consistent aura of meaning with which a form is imbued by its collocates” (Louw, 1993:157). Translators can consult corpora of different types to learn about positive, negative or neutral attitudinal meaning associated with the words in question.

Information on Cognates and False Friends

Cognates are words or phrases that have similar pronunciation and meaning in two languages, while false friends are words or phrases in two languages that look or sound alike, but do not convey the same meaning. Corpora of source and target languages can be applied by translators to differentiate cognates from false friends by providing translators with the contexts in which the words in question are used.

Corpora and Student Translators' Professional Prospects

Corpora in general can provide translators with information beyond what conventional resources have to offer to translators and that is the main reason behind their popularity in translation classrooms. Adopting a different approach to the benefits of corpora, this section aims to elaborate on the advantages of receiving corpus-based translation training for student translators in terms of their professional prospects. The discussion starts with a brief elaboration of the current translation market in section 2.3.1

and then goes into the details of what corpora have to offer to student translators in terms of the changing face of the translation market in section 2.3.2.

The Current Translation Market

There is no doubt that the translation market has changed considerably during recent years. The advent of the Internet along with translation software and translation memories have all had their own effects on translation practice and the translation market. Nowadays professional translators, more often than not, work with a variety of texts and texts types, not to mention their different jobs as editors, technical writers and language consultants. As Ulrych (2005:22) states “the idea that professional translators work predominantly in one or two specialist fields is in fact swiftly losing ground as the need for translation expands exponentially in volume and variety”. To succeed in the current competitive market, translators need to expand both their linguistic and extra linguistic skills. Among other things, they need interpersonal skills to get the job and maintain a good relation with clients, subject experts and fellow translators; they need computer skills to work with translation software, translation memories and corpora; they need editing skills to provide their clients with an acceptable final version and/or get editing jobs increasingly referred to translators and, finally, they need an encyclopaedic knowledge of various subjects to work with various texts.

What do corpora have to offer?

Now let us see what a corpus-based translation class has to offer to students in terms of the above-mentioned skills. As Fawcett (1987) has stated the purpose of translator training is to equip trainees with skills transferable to any text, on any subject and a corpus-based translation teaching by its nature can provide trainees with such skills. In a corpus-based translation class, once students learn about corpora, corpus analysis tools and their applications for translation, they can compile and use corpora for any kind of text they may encounter in the future. Making students familiar with DIY corpora, for instance, enables them to build disposable corpora for any type of text they may come across during their training or later on when they enter the translation market. Furthermore, a corpus-based translation class provides trainees with a favorable opportunity to work together and to experience positive cooperation and group work. In such classes, students can use the corpus to assess and revise translations by their peers and they can back up their criticisms of their peers’ translations with convincing evidence from the corpus.

Such practices can help students develop their interpersonal skills and prepare them to deal with future clients and fellow translators. Apart from that, using a corpus to revise and edit a translation is in itself a good practice for students to improve their editing skills, which is in great demand in today’s translation market. Receiving corpus-based training can also be beneficial for those translation graduates who enter other markets such as technical writing and editing. As stated by Bowker and Pearson (2002), LSP corpora can well be used as a writing guide to write in a particular style or to produce technical texts. Another benefit of using corpora to teach translation has to do

with students' computer skills. Working with corpora demands computer literacy and basic computer skills. Exposure to computers and corpora at undergraduate level would help trainee translators acquire basic computer skills during their training and they can later on develop their computer skills depending on the market needs.

Corpora in Translator Education in Iran

In the existing BA curriculum for English translation in Iranian universities, there is no course on corpora and corpus-based translation. This may be partly blamed on the unavailability of ready-made corpora in the Persian language which requires teachers to compile their own corpora to be used in their classes. The other more important reason behind the unpopularity of corpora in translator education in Iran, is that there has been no practical research on the applicability of corpus-based translation in the context of translation classrooms in Iran and definitely without such research, pursuing decision-makers to change the existing curriculum is not easy. With this in mind, a decision has been made to introduce an undergraduate translation class in Iran to corpus-based translation in order to study the pros and cons of corpus-based translation in the context of translation into Persian.

The Experiment

A class of 22 senior undergraduate translation students from the Department of Translation Studies at Allameh Tabataba'i University, Tehran, Iran participated in this experiment in their first semester of the academic year 2009-2010. The course chosen for this study was a two-credit course entitled Translation of Political Texts which is a compulsory course offered to undergraduate translation students at the Allameh Tabataba'i University in the last semester of their eight-semester translation programme. The objectives of the course include familiarizing students with the political text-type and terminology and training them in translation of political texts from English into Persian. With permission from the head of the Department of Translation Studies, the class was held in a computer laboratory which is normally used for courses on computer skills. The room was equipped with 14 computers all connected to the Internet.

The Corpus Used

In this experiment, a decision was made to draw on disposable corpora, i.e., small corpora made by students for specific translation tasks, due to two main reasons. First, the students had no prior knowledge of corpora and they definitely needed training on corpus compilation. Therefore, becoming engaged in compiling disposable corpora was thought to be a good and simple corpus building experience for them. Second, finding original Persian texts in some genres, including political genres, is not an easy task given that a lot of Persian texts which may seem original are in fact translations from other languages. With this in mind, the researcher led herself to believe that compiling a representative corpus of Persian political texts, which would cover all the sub-genres the students were about to work on for this course, was not feasible within the limited time the researcher had for this study. It is worth mentioning that the students were free to

choose the type of disposable corpora they wanted to compile and work with. In other words, the students were free to draw on either Persian monolingual corpora or comparable corpora of English and Persian political texts, based on their needs and the texts they were dealing with.

Corpus Analysis Tools

The corpus analysis tool used in this experiment was Wordsmith 5. This new version of Wordsmith promotes itself as a tool capable of analyzing texts in many languages including Persian; however, the results for Persian are not promising. The problem with using this tool on Persian corpora is that its concordance features do not work properly on Persian texts. It however can provide the statistics for type, token, mean sentence length, and type token ratio, for Persian texts. It is while the other corpus tools such as Antconc, TextStat and aConCorde has more or less the same shortcoming when it comes to Persian texts. The reason we decided to use Wordsmith 5 was simply based on the fact that it is currently the only available corpus tool that supports Persian, not to mention the fact that its statistics are solid and it has been used by many other researchers. It must be pointed out that there are two software programs namely Sharp Text Analyzer tool and Oxford Concordancer tool which have been specifically reprogrammed to analyze Persian corpora in the Department of Foreign Languages of the Institute for Humanities and Cultural Studies in Iran. These two corpus tools, however, are not available to the public yet.

Methodology

Drawing on the social constructivist approach to translator education (Kiraly, 2000), the methodology adopted in this experiment was based on cooperative learning and learners' empowerment. During the first session, the objectives of the experiment and what was expected from the students were made clear to them. The students were told that they would be the first group of students in the university ever involved in such an experiment and that their cooperation and support is essential for the experiment to run well. Some students though seemed a little reserved when they heard about corpora and corpus analysis tools. In order to minimize the students' stress and encourage more cooperation, they were told that their final scores for the course would be determined based on their cooperation, their homework and their attendance. Furthermore, it was decided that all the assignments and homework would be emailed to the students. It was hoped that this would indirectly encourage the students to be more open to computers and the Internet.

Grouping Students

In order to create a better cooperative environment, the students were asked to work in groups. At first, the group members' and moderators' duties were explained; the students were then asked to form groups of three to five and choose a moderator for their groups. As agreed, moderators were to act as intermediaries between the teacher and the rest of the group members. The teacher thus would email assignments to the moderators,

who would subsequently forward it to their group members. Moderators were also responsible for following up and emailing the teacher the completed assignments.

Assignments

The corpus-related assignments the students were given included building disposable corpora for an English text on politics and, at the later stage, using the corpora to translate the English texts into Persian.

Teaching about corpora and Corpus Analysis Tools

Overall, three sessions were allocated to familiarize students with different types of corpora and teach them how to work with corpus analysis tools. In the first session, the students were lectured on the applications of corpora in Translation Studies, following which different types of corpora were introduced. During the second session, the students were familiarized with the corpus construction process. The resources they could draw on to obtain texts for inclusion in their corpora, the file format they needed to save the texts in and all the technical issues were explained to them in detail. Issues including saving Persian texts in Plain Text format and using Unicode encoding were also elaborated on. The Wordsmith tool was then introduced. As their homework, the students were asked to build DIY corpora for an English text on Social Democracy (Appendix 1). During the third session, the students used Wordsmith on their Persian and English DIY corpora to translate the English text collaboratively in the class.

Corpus Building Experience

As their first corpus building assignment, the students were given an English text on Social Democracy (Appendix 1) to work on. The text was emailed to moderators who subsequently forwarded it to their group members. The instructions were explained to the students and, as agreed, each group was to compile either a target monolingual corpus or bilingual corpora of the source and target languages. Each group member was supposed to contribute at least two texts for inclusion in the corpora and email them to the moderators, who then had to evaluate the texts for relevance to the source text at hand. Moderators then would build the corpus and email it to the teacher. After all the groups submitted their disposable corpora, a questionnaire (Appendix 2) was distributed to gain a better understanding of the students' corpus building behavior. The questionnaire mainly asked students about the resources they had used, the strategies they had adopted, and finally the problems they had faced during the corpus compilation process. It is necessary to mention that, as most students preferred to build and work only on monolingual corpora of Persian texts, the questionnaire administered focused on compiling corpora of Persian texts only. The students' answers to the questionnaire are summarized in the following sections.

Resources used to retrieve texts in Persian

Questions 1 and 2 of the questionnaires asked students about the resources and the search engines they used to collect Persian texts for inclusion in their disposable corpora. The answers given by the subjects are summarized in tables 1 and 2.

Table 1: Percentage of students' use of different resources to retrieve the Persian texts

Which resources did you use to retrieve texts for inclusion in your Persian disposable corpus?				
Web	CD-ROMs	Books	Online databases	Other resources
85%	-	-	15%	-

Table 2: Percentage of students' use of different search engines to retrieve the Persian texts from the Web

If you have used the Web, which search engines did you use to retrieve the relevant Persian texts?				
Google	Yahoo	Alta vista	msn	Others
85.71%	14.28%	-	4.76%	-

As table 1 and 2 suggest, the Web was the main resource the students used to obtain Persian texts for inclusion in their corpora and the most common search engine was Google with 85.71%. Yahoo and msn came next with 14.28% and 4.76% respectively. It is interesting to note that the online resources were the only resources used by the students to obtain texts. This may be due to their limited access to resources such as books and CD-Roms, the limited time they had to prepare their corpora, and the ease of use of online resources.

Problems faced in finding suitable Persian texts on the Web

Question 3 asked students about the problems they faced using the Web to find Persian texts for their disposable corpora.

Table 3: Percentage of the different Problems students faced in finding suitable Persian texts on the Web

If you have used the Web, what problems, if any, did you face finding suitable Persian texts on the Web?			
Identifying original Persian texts	Identifying reliable Persian texts	Identifying relevant Persian texts	Others
32.24%	30.28%	28.57%	15%

As table 3 shows, the students' main problem in finding suitable Persian texts on the Web was identifying original Persian texts. This problem has to do with the fact that a large number of political Persian texts available on the Web are in fact translations from other languages and thus inappropriate for inclusion in corpora of the Persian language. The next problem students encountered most was identifying reliable Persian texts on the Web. Many students noted that a lot of Persian texts they found on the Web were either from personal WebPages or other resources with dubious authenticity. Another problem students faced was that of finding Persian texts relevant to the source text at hand. As noted by students, finding original and authentic Persian texts relevant to their source text was not easy. Under the "Others" category, some students complained about the filtered

Persian political websites which further restricted their access to suitable texts for their corpora.

Strategies to identify relevant Persian texts

Question 4 asked students about the strategies they employed to identify relevant Persian texts for their corpora.

Table 4: Percentage of different strategies students used to identify relevant Persian texts

How did you identify Persian texts for inclusion in your corpus?			
Source text title only	Reading the source text and selecting potential subject keywords	Reading the source text and selecting potential translation problems	Others
31.03%	44.82%	24.13%	-

As table 4 shows, drawing on the keywords of the source text was the students' most popular strategy for identifying the relevant Persian texts. The next strategies students employed were drawing on source text title and potential translation problems respectively.

Technical Problems in building Persian Corpora

Question 5 asked students about the technical problems they faced in building their disposable corpora of Persian texts.

Table 5: Percentage of the different technical problems students faced in building Persian Corpora

What technical problems, if any, did you face building your Persian disposable corpus?			
Downloading Persian texts from the Web	Saving and reading Persian texts in Plain text format	Sending and receiving Persian texts through email	Others
9.09%	72.72%	18.18%	-

As table 5 suggests, saving and reading Persian Texts in Plain text format was by far the students' major problem. This was due to the encoding issue students needed to consider while saving Persian texts in Plain Text format. It is interesting that though students received instruction on how to save Persian texts in Plain Text format, many of them failed to follow the instructions in practice, which caused them further technical problems. The next problems students faced most was sending and receiving Persian texts through email. This problem had to do with the fact that some students did not follow the instruction for sending the texts as attachments in plain text format. Some students, for instance, sent their attachments in docx format (Microsoft Word 2007) which caused problems for those who had the older version on their computers that is Microsoft Word 2003.

Comments by Students

In the last part of the questionnaire, the students were asked to comment on their corpus building experience. The general patterns which emerged from the students' comments are discussed below. More than half of the students (12 out of 22) indicated that building disposable corpora of Persian text was time consuming and difficult, mostly because of lack of suitable Persian texts on the Web.

"Making corpus is a really hard and time-consuming activity, especially in the Persian language"

"The practice was really frustrating because we had to spend a lot of time looking for suitable Persian texts"

"I think building corpora is not practical for every translation because we need to go through the long and difficult process of finding reliable original Persian texts"

"I found it really difficult due to the fact that the number of Online Persian texts is really limited"

It is interesting to point out that 5 students indicated that building the English corpora was easier for them compared to compiling the Persian corpora.

"Making a corpus in English was easier because there were a lot of English texts on the Web"

"There are very few online original political texts in Persian and so building Persian [corpus] was more difficult compared to building English corpus"

"In English, it was much easier to make the corpus because there were more online English texts"

Working with Wordsmith

After building their corpora, the students started translating the source text with their disposable corpora. Since the students did not have access to Wordsmith tools at home, they did their translations collaboratively in the class. The next section provides an overview of the problems students faced during the process and the solutions adopted to resolve the problems.

Problems related to using Wordsmith on Persian Texts

The very first problem students faced had to do with opening their Persian corpora with Wordsmith tools. It seemed that the Concord and Wordlist features of Wordsmith could not read Persian corpora saved with names containing Persian characters. This problem was resolved when we changed the names of Persian corpora and used either numbers or English characters in naming them. The next major problem students faced in using Wordsmith on Persian texts was that they could not read the concord displays of their Persian corpora as Persian sentences and even words were mixed up in the displays. Figure 1 shows the concordance display for the search word "سوسیال" from a 13000 thousand word Persian corpus on social democracy compiled by one of the groups. It is necessary to mention that here Persian language was set as the default language for the Concord feature.

Word #		N
12,242	موسیل	1
12,153	به قصد	2
12,130	میت باز قصد	3
12,031	برقصد اصطلاح	4
11,859	آن به را می دهی	5
10,351	نایب - روح	6
10,271	موسیل	7
10,216	رشد	8
9,967	موسیل	9
9,776	موسیل	10
9,448	موسیل	11
8,230	موسیل	12
8,221	موسیل	13
8,182	موسیل	14
8,121	موسیل	15

Figure 1: Concordance display for the search word "سوسیل" with Persian language set as the default language for Concord

As shown in Figure 1, the Persian words and sentences were jumbled in the concordance display of the Wordsmith Tools. The first solution adopted to resolve this problem was to change the language set for Concord. In fact, when the students set English as the default language for Concord and opened their Persian corpora with it, the displays were more readable and the sentences were less disordered. Figure 2 shows the concordance display for the search word "سوسیل" from the 13000 thousand word corpus of Persian texts on social democracy with English set as the default language for the Concord feature.

N	Concordance	Set	Tag	Word #	W	on	W	en
1	شلی است به گریه و درد قصد و موسیل	12,242	410	6%	0	9%		
2	به راه قصد گلی حوی به و موسیل	12,153	408	5%	0	9%		
3	بازگشت باز قصد برای موسیل	12,130	407	5%	0	9%		
4	باز قصد اصطلاح به یا و موسیل	12,031	404	6%	0	7%		
5	آن به را می دهی موسیل	11,859	401	7%	0	6%		
6	نایب و روح به یا و موسیل	10,351	352	1%	0	4%		
7	موسیل	10,271	350	7%	0	7%		
8	رشد	10,216	348	8%	0	2%		
9	موسیل	9,967	340	2%	0	0%		
10	موسیل	9,776	335	8%	0	9%		
11	موسیل	9,448	323	7%	0	6%		
12	موسیل	8,230	278	1%	0	6%		
13	موسیل	8,221	278	4%	0	6%		
14	موسیل	8,182	278	7%	0	6%		
15	موسیل	8,121	278	4%	0	6%		

Figure 2: Concordance display for the search word "سوسیل" with English set as the default language for Concord feature

As figure 2 shows, the concordance display was more readable, but some sentences and words were still jumbled. Moreover, irrespective of the default language, the sentences were jumbled up as the students moved the cursor on the display. Another solution adopted was to save concord displays as an excel spreadsheet and read from the excel page. Figure 3 shows the concordance display for the search word "سوسیل" from the 13000 thousand word corpus of Persian texts on social democracy saved as an excel spreadsheet with English set as the default language for the Concord feature.

Figure 3: Concordance display for the search word "سوسیل" from a 13000 thousand word corpus of Persian texts saved as excel spreadsheet

The concordance displays from Persian corpora which were saved as excel spreadsheets were free from the problems faced earlier and were totally readable. However, there were still some drawbacks. First, the search words were not highlighted in the displays saved as excel spreadsheets; so finding the search words and following the patterns in which they were used required more time and energy. Second, with the concordance displays saved as excel spreadsheets students could not refer to the wider contexts of the search words. To resolve this problem, some students preferred to increase the characters Concord would save per entry. This strategy seemed quite successful with students having access to a reasonable number of context words for the search words under question.

Overall impression of working with corpora and Wordsmith Tools

After the students handed in their translations done with the help of corpora, they were asked to answer an open-ended question (What did you like and/or dislike about your corpus-based translation experience?) to find out about their general opinion on using corpora and Wordsmith tools in translation into Persian. The general patterns emerged from students' comments are discussed below.

Students' Frustration When Working with Wordsmith Tools On Persian Corpora

From the students' responses to the open-ended questionnaire, it became clear that the majority of students did not like this experience due to the numerous problems they encountered in the process of working with corpora. The general negative points students stated about their experience are as follows:

- Using Wordsmith for analyzing Persian corpora was time consuming (18 students out of 22)
- “Opening and reading Persian corpora with Wordsmith was really challenging and time consuming”
- “Working with Wordsmith was really time consuming as we had to go through a complicated process for each search word to save the concordance in excel format”
- “It took me a lot of time to save the concordance displays of corpora in a readable format and I do not think I like to do it again”

“It takes a lot of time because for each search word we have to make the concordance display and save it and finally read from it”

- Finding useful information from the concordance displays (of Persian corpora) was difficult (16 students).

“Looking for the search words and getting information from their contexts was really problematic, because the search words were not highlighted and we had to look for it”

“The concordance displays were messy and the ones saved in excel were Ok but still the search words were not highlighted”

“As a Persian translator, I do not like to work with wordsmith and Persian corpora since it really is not practical for our language”

Apart from that, 15 students stated that they would rather use Wordsmith for English corpora only.

Positive Comments about the Experience

As mentioned above, it seemed that most students did not like the experience due to the problems they faced while working with Wordsmith on Persian corpora. They, however, mentioned a few positive points about their experiences as follows:

- It is good because we can make use of our corpora when we do not have access to the Internet and online texts (10 students out of 22).
- Corpora contained useful terminological information that cannot be found in other resources (4 students).

Discussion

This study was the first attempt to draw on corpora in the context of translation classrooms in Iran which aimed at investigating the pros and cons applying corpora in this context. One of the main problems in using corpora in translation classrooms in Iran has to do with the fact that there are few ready-made corpora in the Persian language available. This situation implies that the teachers who decide to use corpora in their classes have to either compile the corpus they want to use before the beginning of the class or instead share a corpus building experience with their students by drawing on disposable corpora. In this study, the researcher opted for the second option which seemed more practical considering the researcher’s time constraints. The main problem in compiling disposable corpora in the Persian language, as noted by students, was that finding suitable Persian texts on the Web, as the main resource used to build disposable corpora, was not easy. This is due to several reasons. First of all, the Persian language is somewhat underrepresented on the Web. Second, a lot of Persian texts on the Web (especially in genres like science and politics) are translations from other languages which implies that they cannot be included in monolingual corpora of Persian. Third, a lot of Persian texts on the web are not reliable and/or authentic.

These problems, as noted by subjects in this study, complicated and slowed down the process of building Persian corpora. Some students further stated that building disposable corpora in the English language was much easier for them compared to

building disposable corpora in Persian due to the abundance of English texts on the Web. One of the main points that can be derived from the corpus building experience in this study is that students need to invest more time and effort into building Persian corpora. Students, in fact, need to be alerted to the potential problems of building disposable corpora in Persian and the strategies they can draw on to succeed in the process. Teaching students about different strategies to identify relevant texts, sensitizing them to the quality of texts retrieved, and finally, introducing them to other resources they can use to collect texts for their corpora, are among the issues teachers must address before embarking on a corpus building experience with their students. Another equally important issue is special training for using corpus analysis tools on Persian corpora.

As mentioned earlier, there are some corpus tools which have been reprogrammed to be able to analyze Persian corpora; but they are not available to the public yet. With this in mind, Wordsmith 5 was used as the corpus analysis tool for this research. Though this version of Wordsmith Tools promotes itself as a tool capable of analyzing Persian, its Concord feature does not work properly on Persian corpora. The best solution adopted to resolve this problem in this experiment was saving concordance displays as an excel spreadsheet. This strategy seemed successful; however, there were some drawbacks. First, the search words were not highlighted in the displays saved as an excel spreadsheet which made it difficult to locate them in the displays. Second, with the concordance displays saved as an excel spreadsheet, students could not refer to the wider contexts of the search words. To resolve this problem, some students preferred to increase the characters Concord would save per entry.

This strategy seemed quite successful with students having access to a reasonable number of context words for the search words under question. Last but not least, saving concordance displays as an excel spreadsheet for each search word and reading from it, as noted by students, was very time consuming and frustrating. The main point that can be derived from this experience is that without a corpus analysis tool working properly with the Persian language, incorporating Persian corpora into translation classrooms must be done with caution. The long and complicated procedure students have to follow before they can finally read from the concordance displays of Persian corpora may be frustrating and overwhelming for them. As in this study, the students' frustration was obvious from their feedback and their lack of interest in repeating the experience. Despite the problems faced, our cooperative learning experience seemed quite successful with students actively involved in class activities and group work. The online communication between the teacher and the moderators, and subsequently the rest of the group members, ran well and the students seemed to enjoy receiving and sending their assignments through email.

Final Remarks

Despite the recent popularity of corpora in translator education in some universities (mostly in the West), there is no course yet on corpora and corpus-based translation in undergraduate translation curricula in Iranian universities. In this paper, we have described an attempt to incorporate disposable corpora into a translation class on

Translating Political Texts at Allameh Tabataba'i University in Iran. The corpus building experience proved difficult with students struggling to find suitable Persian texts on the Web. Extracting data from Persian corpora was no different with Persian sentences jumbled up in the concordance displays created by Wordsmith tools. It is hoped that this situation will improve with the recent increasing interest in corpus-based studies in Iran which revolves around compiling different corpora in the Persian language and developing/reprogramming corpus analysis tools for Persian corpora. Such improvements would hopefully put an end to the problems associated with working on Persian corpora and pave the way for the introduction of corpus-based courses into translator education in Iran.

References

- Baker, Mona (1993). 'Corpus Linguistics and Translation Studies: Implications and Applications', in Mona Baker, Gill Francis and Elena Tognini-Bonelli (eds) *Text and Technology: In Honour of John Sinclair*, Amsterdam & Philadelphia: John Benjamins, 233-50.
- (1996) 'Corpus-based Translation Studies: The Challenges the Lie Ahead', in Harold Somers (ed.) *Terminology, LSP and Translation: Studies in Language Engineering*, in Honor of Juan C. Sager, Amsterdam: John Benjamins, 175-86.
- (2000). 'Towards a Methodology for Investigating the Style of a Literary Translator'. *Target* (12): 241-66.
- Bernardini, Silvia (1997). 'A Trainee Translator's Perspective on Corpora', Paper presented at *Corpus use and learning to translate*. Bertinoro. Available at <http://www.sslmit.unibo.it/introduz.htm> (last accessed April 2008).
- (2004) 'Corpora in the Classroom: An Overview and Some Reflections on the Future Developments', in John Sinclair (ed) *How to Use Corpora in Language Teaching*, Amsterdam & Philadelphia: John Benjamins, 16-36.
- Bowker, Lynne (1998) 'Using Specialized Monolingual Native Language Corpora as a Translation Resource: A Pilot Study'. *Meta* 43 (4): 631-51.
- (2000). 'A Corpus-Based Approach to Evaluating Student Translations', *The Translator* 6(2):183-210.
- (2002). *Computer-Aided Translation Technology: A Practical Introduction*, Canada: University of Ottawa Press.
- and Jennifer Pearson (2002). *Working with Specialized Language: A Practical Guide to Using Corpora*, London & New York: Routledge.
- Fawcett, Peter (1987). 'Putting Translation Theory to Use', in Keith Hough and Ian Mason (eds) *Translation in the Modern Language Degree*, London: CILT, 31-18.
- Kiraly, Donald (2000). *A Social Constructivist Approach to Translator Education: Empowerment from Theory to Practice*, Manchester: St. Jerome.

- Kubler, Natalie (2003). 'Corpora and LSP Translation' in Federico Zanettin, Silvia Bernardini and Dominic Stewart (eds) *Corpora in Translator Education*, Manchester: St Jerome, 25- 42.
- Louw, Bill (1993). 'Irony in the Text or Insincerity in the Writer? The Diagnostic Potential of Semantic Prosodies' in Mona Baker, Gill Francis and Elena Tognini-Bonelli (eds) *Text and Technology: In Honour of John Sinclair*, Amsterdam & Philadelphia: John Benjamins, 233-50.
- Maia, Belinda (2003). 'Training Translators in Terminology and Information Retrieving using Comparable and Parallel Corpora' in Federico Zanettin, Silvia Bernardini and Dominic Stewart (eds) *Corpora in Translator Education*, Manchester: St Jerome, 43-53.
- Monzo, Esther (2003). 'Corpus-based Teaching: The Use of Original and Translated Texts in the Training of Legal Translators', *Translation Journal* (7): 4. Available at <http://accurapid.com/journal/26edu.htm> (last accessed May 2008).
- Olohan, Maeve (2004). *Introducing Corpora in Translation Studies*, London & New York: Routledge.
- Pearson, Jennifer (1998). *Terms in Context*, Amsterdam/Philadelphia: John Benjamin Publishing Company.
- Sinclair, John (1991). *Corpus, Concordance, Collocation*, Oxford: Oxford University Press.
- Somers, Harold (2003). *Computers and Translation: A Translator's Guide*, Amsterdam & Philadelphia: John Benjamins.
- Stewart, Dominic (2000). 'Conventionality, Creativity and Translated Text: The Implications of Electronic Corpora in Translation', in Maeve Olohan (ed.) *Intercultural Faultlines. Research Models in Translation Studies 1: Textual and Cognitive Aspects*, Manchester: St. Jerome, 73-91.
- Teubert, Wolfgang and Anna Cermakova (2004). 'Directions in Corpus Linguistics', in Michael A.K. Halliday, Wolfgang Teubert, Colin Yallop and Anna Cermakova (eds) *Lexicology and Corpus Linguistics: An Introduction*, London: MPG Book , 113-67.
- Ulrych, Margherita (2005). 'Training Translators: Programmes, Curricula, Practices', in Martha Tennent (ed.) *Training for the New Millennium*, Amsterdam & Philadelphia: John Benjamins, 3-33.
- Varantola, Krista (2003). 'Translators and Disposable Corpora', in Federico Zanettin, Silvia Bernardini and Dominic Stewart (eds) *Corpora in Translator Education*, Manchester: St Jerome, 55-70.

- Zanettin, Federico (1998). 'Bilingual Comparable Corpora and the Training of Translators', in Sara Laviosa (ed.) *Meta* 43(4) Special Issue: The Corpus-based Approach: A New Paradigm in Translation Studies: 616-30.
- (2001). 'Swimming in Words: Corpora, Language Learning and Translation', in Guy Aston (ed.) *Learning with Corpora*, Houston, Texas: Athelstan, 177-97.
- (2002a). 'DIY Corpora: The WWW and the Translator', in Belinda Maia, Jonathan Haller and Margherita Ulrych (eds) *Training the Language Services Provider for the New Millennium*, Porto: Faculdade de Letras, Universidade do Porto, 239-48.
- (2002b). 'Corpora for Translation Practice' in Elia Yuste-Rodrigo (ed.) *Language Resources for Translation Work and Research, LREC 2002 Workshop Proceedings*, Las Palmas de Gran Canaria, 10-14.