Vilnius University Press

Vertimo studijos
2018, vol. 11, pp. 62–75

ISSN 2424-3590 eISSN 2029-7033
DOI: https://doi.org/10.15388/VertStud.2018.5

The Web as a Corpus: A Resource for Translation

Helia Vaezian

Department of English Language,
Khatam University, Tehran, Iran.
vaezian.helia@gmail.com

Abstract. Accessing ready-made corpora may not be always easy. This is especially true for less dominant languages such as Persian for which the number of available corpora is very limited. Moreover, most existing corpora are domain specific, which implies that they supply a limited range of genres and text types. They, thus, may not always contain the information the translator is looking for. Drawing on the World Wide Web as a big corpus, however, is not subject to such limitations. The Web, in fact, can be considered as a very large multilingual corpus containing texts in almost all languages and all text types. The present paper reports the results obtained from a collaborative experience in which undergraduate English Translation students from the Department of Translation Studies of Allameh Tabataba'i University made use of Google search engine and webascorpus web concordancer to extract translationally-relevant data from the Web.
Keywords: corpora, corpora for translation purposes, *webascorpus*, translator training

Interneto kaip tekstyno ištekliai vertimui

Santrauka. Tekstynai vertimo praktikoje naudojami plačiai, bet jų vis dar nėra tiek, kad būtų patenkinami įvairūs vertėjų poreikiai. Be to, rečiau vartojamų kalbų tekstynų apskritai trūksta – persų kalba nėra išimtis. Nors pastaraisiais metais tekstynų sukurta jau nemažai, dauguma apsiriboja kokia nors konkrečia temine sritimi, o žanrų ir tekstų įvairovė labai menka. Todėl vertėjai susiduria su informacijos paieškos problemomis, kurių neišsprendus kenčia jų darbo kokybė. Vienas iš galimų šių problemų sprendimo būdų – atsigręžti į pasaulinį interneto tinklą (*World Wide Web*), kuris gali būti naudojamas kaip didžiulis daugiakalbis tekstynas. Jo kalbų ir visų tekstų tipų ištekliai beveik neriboti. Straipsnyje pristatoma Allameh Tabataba'i universiteto Vertimo studijų katedros anglų kalbos vertimo programos studentų patirtis naudojant *Google* paieškos sistemą ir *webascorpus* internetinę konkordanciją vertimui reikalingų duomenų paieškai internete.
Pagrindiniai žodžiai: tekstynai, Google paieškos variklis, tekstynų panaudojimas vertime, *webascorpus*, vertėjų rengimas

1. Introduction

When language corpora first entered Translation Studies as a discipline, their application was limited to the research on the language of translation and its distinctive features (Baker 1993). In fact, corpora were primarily used by translation scholars to investigate whether and to what extent translated texts differ from either their source texts or from non-translated texts in target language and they eventually lead to "a better understanding of translation phenomenon and helped raise awareness of what is involved in translating" (Zanettin, Bernardini, Stewart 2003, 3).

Over time language corpora found their way into other areas within the discipline. For instance, corpora of different types were used in studies on translation universals (Baker 1993), translator's style and ideology (Baker 2000) and translation evaluation (Bowker 2000). Another area which has greatly enjoyed the benefits of corpora in recent years is *translator education*. In the context of translation classrooms, corpora came to be appreciated as valuable tools for both learners and teachers. Corpora, among other things, were shown to enhance learner's source text understanding (Bowker 1998), their understanding of specialized terms (Gavioli & Zanettin 1997) and their knowledge of different text types (López-Rodríguez and Tercedor-Sánchez 2008). They also proved useful in providing the student translators with unpredictable and incidental learning (Aston, 1999; Zanettin, 2001). Last but not least, corpora were shown to enhance translation student's confidence (Varantola 2003; Monzo 2003) and autonomy (Bowker 2002).

2. Why the Web, Not Corpora?

As the body of literature on corpora reviewed above shows, corpora have a lot to offer to translators. There are, however, certain practical problems regarding the use of corpora by translators. First, the number of existing corpora is limited and the available corpora are mostly limited to a few dominant languages. A translator translating into English, for instance, can easily enjoy the benefits of freely searchable online English corpora such as BNC (British national corpus)[1] and Collins Wordbanks *Online* English corpus[2], while a translator translating into an under-resourced language such as Persian would face a different scenario. In fact, there is simply no freely searchable corpus of the Persian language available to online users. Second, most existing corpora are domain specific and supply a limited range of genres and text types (Fujii 2007). The existing corpora, thus, cannot be used for translation of all text types and genres. Third, the

1 http://www.natcorp.ox.ac.uk/

2 http://www.collins.co.uk/Corpus/CorpusSearch.aspx

existing corpora may not always contain the exact information the translator is looking for. In fact, even a very large specialized corpus may not always contain the information needed to translate texts on the respective specialized subject.

Drawing on the World Wide Web as a big corpus, however, is not subject to such limitations. The Web, in fact, can be considered as a very large multilingual corpus containing texts in almost all languages and all text types. Apart from that, it is available to users around the world. Nevertheless, to see whether the Web can really provide translators with the benefits associated with using corpora in translation, it is necessary to first discuss the nature of the Web as a big corpus.

3. Is the Web Really a Big Corpus?

There are mixed ideas about the nature of the Web as a big corpus. While some scholars persistently stress the benefits of the Web as a big corpus, there are others who still question the nature of the World Wide Web as a real corpus.

According to Sinclair (2005, 1), "[a] corpus is a remarkable thing, not so much because it is a collection of language text, but because of the properties that it acquires if it is well-designed and carefully-constructed". He further clearly attacks the notion of the Web as a corpus by stating that "the World Wide Web is not a corpus, because its dimensions are unknown and constantly changing, and because it has not been designed from a linguistic perspective (ibid: 15). Despite such negative views about the nature of the Web as a corpus, the number of researches dealing with the use of the Web as a big corpus has increased in recent years. As Gatto (2009, 8) states,

> "Notwithstanding doubts concerning the hypothesis of using the web as a corpus, made explicit by one of the founding fathers of contemporary corpus linguistics, linguists from all over the world have been increasingly turning their attention to the web not only as a source of language text for the creation of conventional (well designed and carefully constructed) corpora, but also as a corpus in its own right."

She further continues, "today web itself seems to claim the right of being considered as a corpus by virtue of its very nature as a collection of machine readable and searchable authentic texts, thus opening up new perspectives and offering new challenges" (2009, 8). Kilgarriff and Grefenstette (2003, 2) define corpus as "a collection of texts when considered as an object of language or literary study" and argue that the Web can definitely be considered as a big corpus by this definition.

To be able to provide a convincing answer to the question on the nature of the Web as a big corpus, the following section is devoted to a comparison between the idiosyncrasies of the Web and the features of corpora. Following the approach adopted

by Kilgarriff and Grefenstette (2003) and Gileva (2005), the definition put forward by McEnery and Wilson (1996) is used as the point of departure. According to McEnery and Wilson (1996, 21),

> "In principle, any collection of more than one text can be called a corpus. . . But the term "corpus" when used in the context of modern linguistics tends most frequently to have more specific connotations than this simple definition provides for. These may be considered under four main headings: sampling and representativeness, finite size, machine-readable form, a standard reference."

Though having *finite size* has been mentioned by McEnery and Wilson as one of the main characteristics of corpora (1996), not all corpora conform to this feature. A Monitor corpus defined by Meyer (2002, 15) as "a large corpus that is not static and fixed but that is constantly being updated", for instance, does not conform to this feature. Monitor corpora are in fact open-ended corpora with texts constantly added to them. Such corpora are dynamic in the sense that new texts are constantly added to them and old texts are transferred to archives as new texts are put into the corpus. The Web, in fact, resembles a monitor corpus in that it is an open-ended collection of texts with new texts constantly added to it.

Sampling and representativeness is another feature of corpora based on the definition put forward by McEnery and Wilson (1996). Not all corpora, however, have this feature. As Gileva (2005, 5) states, "It is indeed true that many of the corpora used for literary, linguistic or language-technology studies do not fit into the McEnery-Wilson definition, especially in the part "sampling and representativeness". Kilgarriff and Grefenstette (2003) too disagree with this feature being one of the main characteristics of corpora. According to them, the definition put forward by McEnery and Wilson (1996) answers the question "What is a good corpus?", not "What is a corpus?".

Another basic feature of corpora, as explained by McEnery and Wilson (1996), is that all corpora are in *machine readable form* and there is no doubt that texts on the web are machine readable. The Web, thus, definitely shares this feature of corpora.

Last but not least, based on the definition given by McEnery and Wilson (1996), corpora represent a standard reference to the language varieties they represent. Regarding this feature of corpora, Gileva (2005, 5) states, "although the web cannot really be called a yardstick it may be a very lucrative source of information, which structured in an appropriate way, may present a linguistic playground not worse than that offered by other well-known corpora". This argument simply implies that the Web with its vast amount of data has the potential to provide adequate information about the language(s) under study and it is up to users to make the most of online texts to get the best answers to their questions.

Based on the above arguments, we can claim that the Web more or less conforms to the basic features of corpora and so it can be considered a big corpus. The following section elaborates on a collaborative experience in which undergraduate English Translation students from the Department of Translation Studies of Allameh Tabataba'i University made use of Google search engine and webascorpus web concordancer to extract translationally-relevant data from the Web.

4. Experiment

The present study was carried out at the Department of Translation Studies at Allameh Tabataba'i University in Tehran, Iran. The course chosen for this study was a two-credit course titled 'Translation of Political Texts' which is a compulsory course offered to undergraduate translation students in the last semester of their eight semester translation program. Drawing on the social constructivist approach to translator education put forward by Kiraly (2000), our class was based on cooperation, group learning and learners' autonomy.

Eight sessions were devoted to teaching about and working with the Web as a large corpus using Google search engine and webascorpus Web Concordancer[3]. During the first and the second sessions, the concept of the Web as a corpus was elaborated on and the important search features of the Google search engine, namely phrase search and wildcard search were introduced. The third session was wholly devoted to discussion about the students' first experience with the Google Search Engine. During the fourth session, after the students shared their second experience with Google search engine, a questionnaire on using Google search engine was administrated. In the fifth session, after the students were lectured on Web concordancers in general and webascorpus in particular, they were given an English text to translate into Persian using webascorpus. It is necessary to mention that the two search features of webascorpus, namely, simple and advanced search features were elaborated on in detail in the class. At the beginning of the seventh session, the students shared their second experience with webascorpus, following which a questionnaire on using webascorpus was administrated. In the eighth session, the groups were asked to translate two English texts, one using the Web plus dictionaries and glossaries, the other one using just dictionaries and glossaries. This practice was done to investigate whether there would be any differences between two sets of translation produced, one using dictionaries plus the Web and the other one using only dictionaries, in terms of their scores.

[3] http://www.kwicfinder.com/searchwac.html

In order to compensate for possible differences in groups' translation abilities and any potential text-specific difficulties, groups 1, 2 and 3 were asked to translate text A using the Internet plus dictionaries and text B using only dictionaries, while groups 4, 5 and 6 were asked to translate text A using only dictionaries and text B using the Internet plus dictionaries. Two external raters were then chosen to do the evaluation of the students' translations. The raters were asked to mark the translations according to their usual methods of marking the students' translations. The mean score for each translation was then calculated by adding the scores from each rater and dividing it by two. After the mean scores for all Internet-based and dictionary-based translations were calculated, the mean scores of translations in the two groups were compared and the T test was calculated to see whether the difference between the mean scores of the Internet-based translations and the dictionary-based translations was significant or not.

5. Findings and discussion

Findings on Google Search Engine

The subjects' overall reaction to Google search engine can be described as positive with most students showing interest in learning about Google search engine to extract translationally-relevant data from the Web.

Based on the findings from our questionnaire, the majority of the subjects indicated that they liked using Google search engine to extract translationally-relevant data. Nineteen students further stated that they would like to continue using search engines for translation purposes. Five students further indicated that consulting relevant Persian texts found using Google search engine gave them more confidence in producing their translations. In addition, more than 90% of the subjects responded positively to the statement indicating that Google search engine must be used along with the resources translators normally use. Furthermore, 85% of the subjects responded positively to the statement on the usefulness of the Google search engine to extract translationally-relevant data.

The most common use of the Google search engine for the students, as observed by the teacher and noted by the students, was combining the data on the number of Google search hits with the contextual data from the online documents to make a decision on choosing one equivalence over another. Following this method, the groups based their decision on the number of search hits found by Google and then examined the contexts in which the terms/phrases in question were used to make sure that they were used in the contexts similar to what they had in the respective source texts.

For instance, two groups opted to use "پیمان منع گسترش سلاح های هسته ای" for the English phrase The treaty on the Non-Proliferation of Nuclear Weapons

in their translations as opposed to another possible translation, namely *"معاهده منع گسترش سلاح های اتمی"*. As the students explained, Google found more hits for "*پیمان منع گسترش سلاح های هسته ای*" compared to the other possible translation, i.e. "*معاهده منع گسترش سلاح های اتمی*". The students further noted that the register and the style of the texts in which "*سلاح های هسته ای پیمان منع گسترش*" was used were closer to the register and style of the source text in question.

Yet, in some instances, the students based their decisions solely on the number of search hits found by Google and mentioned that they could not see any meaningful differences in the contexts of the terms in question. For example, in choosing between the two possible Persian translations for the English term globalization, three groups preferred "*جهانی شدن*" over the other possible equivalence, i.e. "*جهانی سازی*" which was less frequent on the Web. As the students mentioned, they could not see any meaningful differences in the contexts in which the two terms were used. In another instance, one group used the number of hits found by Google to make a decision in choosing between *"مذاکرات..... را دنبال کنند"* and "*مذاکرات..... را پیگری کنند*" as translation for pursue negotiations. As the group members indicated, the verb "*پیگیری کردن*" was a more common collocate for the noun "*مذاکرات*" compared to the other verb, i.e. "*دنبال کردن*". Here again the group members indicated that they could not notice any meaningful differences in the contexts in which *"مذاکرات..... را دنبال کنند"* and *"مذاکرات..... را پیگری کنند"* were used.

There was yet another interesting strategy followed by some groups which proved useful. Three groups opted to save the relevant texts found on the Web on their computers and refer to them while doing their translations. As noted by the students, the texts not only provided them with some additional information to better understand the source texts at hand, but also offered interesting terminological information. The students further stated that they were able to identify interesting translation candidates by going through the texts they had found on the Internet. "*غیر هسته ای کشورهای*" for non-nuclear-weapon states, "*کشورهای هسته ای*" for nuclear-weapon states, "*پنج قدرت هسته ای*" for the five acknowledged nuclear-weapon states, "*تسهیلات غیرنظامی صلح آمیز*" for peaceful civil facilities and "*تدابیرحفاظتی*" for safeguards are the translation candidates the students found following the mentioned method.

The search features of the Google search engine, namely, the phrase search and the wildcard search proved useful to the students too. The groups pointed out that drawing on the phrase search and wildcard search, they were able to extract data on translation of certain terms and phrases. For instance, one group was able to confirm the Persian term "*مالیات مستقیم*" as the translation of the English term direct taxation by using the phrase search to examine the context in which the Persian term was used.

Three groups further stated that by using the wildcard search they could identify "پیمان منع گسترش سلاح های هسته ای" for the English term Treaty on the Non-Proliferation of Nuclear Weapons on the Internet. In another example, the students found "کشورهای هسته ای" for non-nuclear-weapon states by drawing on the wildcard search. One of the groups further used Google phrase search to make sure whether "مسابقه" collocates with the term "تسلیحات هسته ای" for the English term Nuclear Arms Race.

There was also an instance in which one group used Google wildcard search to see which adjectives collocate with the noun "سلطه". In this example, the group members were primarily unable to make a decision as for choosing a proper adjective for "سلطه" when translating apparent hegemony. Drawing on the data extracted using the Google wild card search, the group members decided to use سلطه مشهود for apparent hegemony. As they explained, "مشهود" seemed to be a common collocate for the noun "سلطه". Furthermore, drawing on the wildcard search, one group were able to find the adjective "فراگیر" which collocates with the noun "دولت رفاه" in translating the English term a comprehensive Welfare State.

When it came to the quality of translations produced using online data, more than 90% of the subjects responded positively to the statement indicating that translations produced using Google search engine plus paper or electronic dictionaries will be of a higher quality compared to translations produced using only paper or electronic dictionaries. Furthermore, more than half of the students indicated in their comments that using the data retrieved from the Internet via Google search engine can help in producing better translations. Some students further mentioned that translations produced using the online data would be more reliable and natural. As for the reliability of online data, 59% of the subject stated that they generally trust the reliability of the information retrieved through Google search engine and more than 75% of the subjects stated that they always refer to the original webpage to make sure of the reliability of the information retrieved by Google search engine.

It is interesting that more than 60% of the subjects stated that when they cannot make sure of the reliability of the information retrieved by Google search engine, they do not use it in their translations. This can be interpreted as an indication of the success of the approach adopted by the students towards the online data in the sense that they would not use online data with dubious authenticity in their translations. It can further indicate that the teacher has been successful in alerting the students about the reliability and quality of online data.

Findings on Webascorpus

When the students were first lectured on web concordancers, some of them started to question the rationale behind using web concordancers instead of ordinary search

engines such as Google. Even after the differences between web concordancers and ordinary search engines were explained in detail, some students were still reluctant to switch to web concordancers. This situation may probably be due to the students' positive experience with Google search engine and the fact that they were all familiar with Google search engine in general.

In their webascorpus experiences, almost all the groups drew solely on webascorpus simple query and as expected, they all used this feature to see the terms/phrases in question in their contexts. Three groups decided to increase the number of context words shown in webascorpus result page to maximum (1000 words) to have a better overview of the contexts in which the terms/phrases in question were used. The students further used the contextual data to make a decision in choosing between possible equivalences.

For instance, using contextual data, three groups decided to go for "*تندرو*" instead of "*رادیکال*" for the English term radical. As they explained, "*تندرو*" was mostly used in formal writings on politics, which was closer to the register of the respective English source text. Moreover, as the students explained, "*تندرو*" was mostly used in authentic political websites as opposed to "*رادیکال*" which was used in personal web pages with dubious authenticity. In another similar example, two groups used "*جبهه های خلق*" instead of "*جبهه های مردمی*" for popular front.

Yet, some groups based their decisions in choosing one equivalence over another on the number of hits found by webascorpus search engine. For instance, four groups in choosing between "*اصلاح طلبان بورژوا*" and "*اصلاح طلبان طبقه ی متوسط*" for the English term bourgeois reformists decided to use the second one for which the webascorpus search engine found more hits. In another similar instance, three groups went for "*لخبه گرایی*" instead of "*لخبه سالاری*" for the English term elitism. As they explained, Bing (the webascorpus search engine) found more hits for "*لخبه گرایی*".

There were also some instances in which the groups used the number of search hits found by webascorpus search engine to make a decision in choosing between two spelling variants. For instance, one group used the number of search hits found by webascorpus search engine for making decision between "*چپگرا*" or "*چپ گرا*" for leftist. In this example, Bing (webascorpus search engine) found more hits for "*چپ گرا*". Yet, in another similar example, one group drew on the number of Bing search hits to choose between "*گاتانا موسکا*" and "*گائتانو موسکا*" as transliteration of the Italian name Gaetano Mosca. In this example, webascorpus search engine found more hits for "*گائتانو موسکا*". It is necessary to mention that although the groups were encouraged to use the advanced query feature of the webascorpus, none of them drew on it in practice.

In comparing webascorpus to ordinary search engines, namely, Google search engine, the main positive point mentioned by the students had to do with reading

from the webascorpus result page. Most students indicated that reading from webascorpus result page was easier for them compared to reading from Google result page. Some students further mentioned that they liked the fact that the search words were highlighted in the webascorpus result page. Moreover, some groups mentioned that by increasing the number of context words shown in webascorpus result page to maximum (1000 words), they no longer needed to go to the original webpage to check the wider context and this saved them some time.

It is interesting to note that more than 80% of the students responded positively to the statement indicating that translators can extract translationally-relevant information from the Web using webascorpus. This is while when the students were asked to compare webascorpus to ordinary search engines, fourteen students disagreed with the statement indicating that using webascorpus to extract translationally-relevant information from the Web is easier compared to using ordinary search engines and more than 60% indicated that they do not prefer using webascorpus over using ordinary search engines. When it came to their webascorpus experience, more than 65% of the students indicated that they did not like using webascorpus to extract translationally-relevant data and more than half of them indicated that they would not like to use webascorpus to extract translationally-relevant data in future. These findings may suggest that the students in general were more receptive toward Google search engine as opposed to webascorpus.

It is however necessary to mention that some of the students' negative reaction to webascorpus might have been the result of the problems they faced in working with webascorpus. More than half of the students indicated that working with webascorpus was too time-consuming due to the slow Internet speed and the slow speed of webascorpus server.

There is yet another interesting point observed by the teacher regarding the students' use of webascorpus; the students seemed to stay more focused while working with webascorpus compared to the time they were using Google search engine. This might be due the fact that webascorpus has been primarily designed for extracting linguistic information which implies that its interface and its design are all targeted on extracting linguistic information. So, extracting linguistic information from the Web through webascorpus might be more straightforward compared to using Google search engine to extract linguistic data from the Web. The students themselves indicated that reading from webascorpus result page was easier for them compared to reading from Google result page due to the fact that the search words were highlighted in the webascorpus result page.

Web-based versus Dictionary-based Translations

The following tables present the scores for Web-based and Dictionary-based translations respectively. The grading is based on a 0-20 scale.

Table 1. Scores for the Web-based translations

	Rater 1	Rater 2	The mean scores
Group 1 (text A)	16.5	17	16.75
Group 2 (text A)	14	16	15
Group 3 (text A)	15	15.5	15.25
Group 4 (text B)	18	17.5	17.75
Group 5 (text B)	15.5	15	15.25
Group 6 (text B)	17	18.5	17.75

Table 2. Scores for the Dictionary-based translations

	Rater 1	Rater 2	The mean scores
Group 1 (text A)	15.5	16.5	16
Group 2 (text A)	13.5	14.5	14
Group 3 (text A)	16	15	15.5
Group 4 (text B)	17	17.5	17.25
Group 5 (text B)	16	14.5	15.25
Group 6 (text B)	16	15	15.5

To investigate whether there would be any meaningful differences between the scores of our two sets of translations (one produced using dictionaries plus the Web and the other one produced using only dictionaries), the mean scores of translations in two groups were compared and a matched t-test was computed for analysis. The results are as follows:

Table 3. Statistics for the Dictionary-based versus Internet-based translations

Group	(Web-based Translation)	(Dictionary-based Translation)
Mean	16.2917	15.5833
SD	1.2886	1.0567
SEM	0.5261	0.4314
N	6	6
	Two-tailed P value: 0.1076	
	t = 1.9578	

As the results indicate, the mean of the mean scores of Web-based translations (16.29) was slightly higher than the mean of mean scores of the Dictionary-based translations (15.58). However, considering the P value (0.1076) and the significance level (0.05), the difference between the scores of Web-based translations and Dictionary-based translation is not statistically significant.

This may possibly be due to the fact that the subjects in this study had a very limited experience when it came to using the Web to extract translational data. As stated by Teplitz (1991), every time a new way of working is introduced, there would be a learning curve and experience curve effects. The experience curve effect has to do with the relationship between experience and efficiency and states that output improves as tasks are repeated (ibid). It is thus reasonable to expect improvements in the students' use of the Web as a resource for translation as they gain more experience in utilizing it.

6. Concluding Remarks

Accessing ready-made corpora may not be always easy. This is especially true for less dominant languages such as Persian for which the number of available corpora is very limited. Moreover, most existing corpora are domain specific which implies that they supply a limited range of genres and text types. They, thus, may not always contain the information the translator is looking for. This is while the Web is accessible to all users around the world; it contains an abundance of texts in almost all languages of the world and it has texts in a wide range of genres and text types. With these idiosyncrasies, the Web can definitely be a valuable resource for translators.

The subjects in this used Google search engine and webascorpus to extract translationally-relevant data from the Web. The most common use of the Google search engine for the students was combining the data on the number of search hits with contextual data to make a decision on choosing one equivalence over another. Yet, in some instances, the students based their decisions solely on the number of search hits found by Google and mentioned that they could not see any meaningful differences in the contexts of the terms/phrases in question.

The search features of the Google search engine, namely, the phrase search and the wildcard search proved useful to the students too. The students were able to extract data on translation of certain terms and phrases by drawing on these features. There were also some instances in which the groups used Google wildcard search to extract collocational information. The students' use of webascorpus followed the same pattern too; the students drew on the number of search hits by webascorpus search engine and/or used webascorpus simple query to see the terms/phrases in question in their contexts. There were also some instances in which the groups used the number

of search hits found by webascorpus search engine to make a decision between two spelling variants.

The students' overall reaction to using Google search engine to extract data from the Web was positive, while their experience with webascorpus did not seem so favorable to them. When asked to compare Google search engine to webascorpus, most students indicated that using Google search engine to extract translationally-relevant data from the Web was easier for them compared to using webascorpus and further asserted that they would rather use ordinary search engines in future. There was however one interesting point observed by the teacher regarding the students' use of webascorpus versus their use of Google Search Engine; the students seemed to stay more focused while working with webascorpus compared to the time they were using Google search engine. This might be due to the fact that webascorpus has been primarily designed for extracting linguistic information which implies that its interface and its design are all targeted on extracting linguistic information. Extracting linguistics information from the Web through webascorpus, thus, might be more straightforward compared to using Google search engine to extract linguistic data from the Web. It is interesting that the students themselves indicated that reading from webascorpus result page was easier for them compared to reading from Google result page due to the fact that the search words were highlighted in the webascorpus result page.

As the last word, we conclude this paper by stating that in contexts where ready-made corpora are not available, using the Web as a corpus can be a viable option for translators, provided that they know how to extract translationally-relevant information from the Web and assess the authenticity of it.

References

Aston, G. 1999. Corpus use and learning to translate. *Textus* 12, 289-314.

Baker, M. 1993. Corpus Linguistics and Translation Studies: Implications and Applications. *Text and Technology: In honor of John Sinclair.* Baker M. Francis M. and Tognini-Bonelli, E. (Eds.). Amsterdam, Philadelphia: John Benjamins Publishing house. 233-250.

Baker, M. 2000. Towards a methodology for investigating the style of a literary translator. *Target* 12, 241-266.

Bowker, L. 1998. Using specialized monolingual native language corpora as a translation resource: A pilot study. *Meta*, 43 (4), 631-651. https://doi.org/10.7202/002134ar

Bowker, L. 2000. A Corpus-Based Approach to Evaluating Student Translations. *The Translator* 6(2), 183-210. https://doi.org/10.1080/13556509.2000.10799065

Bowker, L. 2002. *Computer-Aided Translation Technology: A Practical Introduction*. Canada: University of Ottawa Press.

Fujii, Y. .2007. Making the Most of Search Engines for Japanese to English Translation: Benefits and Challenges. Asian EFL Journal 23, 41-77.

Gatto, M. 2009. *From body to web: An introduction to the web as corpus*. Bari: Laterza.

Gavioli, L. & Zanettin, F. 1997. *Comparable corpora and translation: a pedagogic perspective*. Paper presented at the first international conference on Corpus Use and Learning to Translate. Bertinoro, 14-15 November 1997.

Gileva, S. .2005. *Using the Web as a Linguistic Tool in Translation Practice*. Available from: http://www.sophista.info/Doku/SGileva- Web_as_linguistic_tool_in_translation_practice.pdf

Kilgarriff, A. & Grefenstette G. 2003. Introduction to the Special Issue on Web as Corpus. *Computational Linguistics* 29 (3), 333-347. https://doi.org/10.1162/089120103322711569

Kiraly, D. 2000. *A social constructivist approach to translator education*. Manchester: St. Jerome.

López-Rodríguez, C.I. and Tercedor-Sánchez, M.I. 2008. Corpora and Students' Autonomy in Scientific and Technical Translation Training. *The Journal of Specialized Translation* 9, 2–19.

McEnery, T. and Wilson, A. 1996. *Corpus linguistics*. Edinburgh: Edinburgh University Press.

Meyer, C. F. .2002. *English Corpus Linguistics: an introduction*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511606311

Monzo, E. 2003. Corpus-based Teaching: The Use of Original and Translated Texts in the training of legal translators. *Translation Journal*, 7(4).

Sinclair, J. .2005. Corpus and Text - Basic Principles. *Developing Linguistic Corpora: A Guide to Good Practice*, ed. Wynne, M. Oxford: Oxbow Books, 1-16.

Teplitz, C. J. 1991. *The learning curve deskbook: A reference guide to theory, calculations, and applications.* New York: Quorum Books.

Varantola, K. .2003. Translators and Disposable corpora. *Corpora in Translator Education*, eds. Zanettin F., Bernardini, S. and Stewart, D. Manchester: St Jerome Publishing, 55-70.

Zanettin, F. 2001. Swimming in words: corpora, language learning and translation. *Learning with corpora*, ed. Aston, G. Texas: Athelstan, 177-197.

Zanettin, F., Bernardini, S. and Stewart, D .2003. *Corpora in Translator Education*. Manchester: St Jerome Publishing.