The Use of Corpora and Concordancing in Translation

Prima Mallikamas
Faculty of Arts, Chulalongkorn University

Abstract

This paper examines the roles, both theoretical and applied, which corpora[1] play in the field of translation. The main emphasis of the paper is on the use of corpora evidence in applied translation studies, which include the teaching of translation in translator training and in foreign language learning. The paper also looks at the use of corpora as valuable resources for translators. Different types of corpora and their applications are reviewed in detail. In the end, the paper discusses the key success factors of corpus-driven translation.

1. Introduction

Although the use of corpora in the subject field of translation began only a few decades ago, there have been remarkable developments, and today corpora are playing an increasingly important role in this field. One of the examples of early research work involving the use of corpora is the study of Vanderauwera (1985). Her corpus comprises 50 Dutch novels translated into English. The searching and analysis of the corpus were done manually, and therefore, the work was laborious and highly time-consuming. Although Vanderauwera's corpus is relatively very small compared to electronic corpora generally used at present, her work has shed new light on translation research, particularly in terms of approach and methodology.

---

[1] Corpora (plural form of corpus) are collections of linguistic data, either written texts or transcriptions of recorded speech, which are available for analysis.

Sinclair (1992), a leading figure in corpus linguistics, refers only briefly to the use of corpora in translation studies, focusing more on the impact of corpora on translator performance and machine translation. It is Baker (1993) who clearly points out that the availability of corpora and the application of corpus techniques have direct bearing on both theoretical and descriptive translation studies. Electronic corpora and new corpus-based research methodology, including the use of powerful concordancing tools[2], provide solutions to major problems in carrying out translation research, such as human limitations in handling large translated text corpora and a lack of appropriate statistical methods to investigate translational norms, as mentioned in Toury (1978), and thus unavoidably heavy reliance on introspective methods, as pointed out in Holmes (1988).

Aside from profound effects on what Holmes calls "pure translation studies" (1988, p.71), which include the two subdivisions mentioned above, the use of corpora has contributed greatly to "applied translation studies" (1988, p.77). The applied aspect discussed in this paper deals with three main branches: translation resources, teaching of translation in translator training, and foreign language learning.[3] Corpora, especially those of specialized texts, can enhance understanding of the source text as well as the quality of translation work. Translating skills can also be developed through the use of corpus-based investigation strategies. Additionally, corpora of translated texts provide abundant data which can be highly beneficial to foreign language learning. The impacts of corpora application in these three branches will be discussed more comprehensively in section 3.

2. Corpora and translation studies: theoretical aspect

It is generally accepted that the two basic assumptions underlying "conventional" translation studies are the primacy of the source text and the notion of equivalence. It was only recently that both the status of source text compared to translated text and the earlier classification of equivalence, which was mainly semantic and formal, were undermined. Translation theorists suggest many new concepts of "meaning" in translation, most of which serve to undercut the idea that to translate is basically to transfer meaning from the source text to the target text. Firth (1968) proposes an approach which shifts away from the interpretation of meaning based on structures and systems of thought to structures and systems in the context of situation. According to this approach, expressions in two different

---

[2] Concordancing tools are computer programs used to search for words or other linguistic entity in corpora and display them in their contexts.

[3] According to Holmes (1988), applied translation studies cover four branches: the teaching of translation in foreign language learning and in translation training, translation policy, translation aids, and translation criticism

languages can confirm their equivalence in meaning because they have a correspondence in their uses. Such studies that focus on contextual and functional aspects have led to a move away from introspective to empirical method made possible only with access to a substantial amount of real data. This, then, can be seen as the first step in support of corpus-based translation studies.

Another two important perspectives which strongly affect the status of the source text and the notion of equivalence are Even-Zohar's polysystem theory and Toury's notion of norm, both developed in the 1970's. In polysystem theory, literature is viewed as a hierarchical and dynamic structure of systems. This structure itself is in turn part of a larger cultural polysystem which comprises many other polysystems such as politics and religion. In other words, there is a strong interdependence among the various systems and subsystems. Translated literature and the original are also subject to this pattern of relationship. As a result, "the status of translated literature is elevated to the point where it becomes worthy of investigation as a system in its own right" (Baker, 1993, p.238). Toury (1985) consolidates the new status of the translated text, claiming that the significance of a translation can be felt only because it is perceived as an utterance in the target language system in which it belongs. As entities separate from the source texts, translated texts can serve, therefore, as objects of study for translation theorists wishing to explore the features of translated texts i.e., what makes them different from other texts in the same language and so on.

This also allows translation theorists to describe what translation is and how it works.

In 1978, Toury introduced a three-layered model comprising competence, norms and performance. While competence is defined as the level of description which yields all options available to translators in a given context and performance as the subset of options actually chosen by translators, norms represent an intermediate level between the two and constitute a further subset of the selected options. Forming the typical body of translated texts in a language and cultural system, they serve as objects of analysis. The notion of norms not only reflects one important aspect of the nature of translation work--that it mainly concerns decision-making--but also strongly promotes the role of corpora in translation studies. According to this concept, we can no longer study translation as individual works; instead, we must investigate a corpus of authentic translations in order to identify norms of translational behaviour and normal patterns of translation. These include the types of strategies translators in a given language and cultural stystem tend to choose on a regular basis.

Another important issue in translation studies involving the application of corpora concerns the universals of translation. Baker (1993) describes them as "features which typically occur in translated text rather than original utterances and which are not the result of interference from specific linguistic systems" whether source or target language. They are, for example, explicitation, disambiguation and simplification, avoidance of repetitions present in the

source text and preference for conventional 'grammaticality.' Other typical features of translated texts, i.e. deviations from the target language system and norms caused by influence from the source language can also be investigated (Schmied and Schaffler, 1996). Access to electronic corpora and corpus techniques enable translation scholars to examine these features of translated texts efficiently and reliably.

From the above, it is obvious that polysystem theory, the notion of norms, and the universals of translation have contributed to many major changes in theoretical and descriptive translation studies. They weaken the primacy of source text and point to the need to reassess the nature of equivalence between the source and target texts. Based on these frameworks, a target-oriented approach has been introduced. Its emphasis is on the significance of translated texts, which, through access to corpora and the application of corpus techniques, are viewed as objects of study. Translation theorists also realize that emphasis in the discipline has shifted from prescription to description. Most importantly, these perspectives justify the fact that corpora of both original and translated texts, together with corpus-based methodology, are indispensible tools for translation scholars conducting research. Tymoczko (1998) clearly summarizes that "corpus translation studies change in a qualitative as well as a quantitative way both the content and the methods of the discipline of translation studies."

3. Corpora and translation studies: applied aspect

As stated earlier, applied translation studies in this paper cover three main branches:

translation resources, translator training, and the teaching of translation in foreign language learning. The roles which corpora play in each branch are discussed below.

3.1. Corpora as translation resources

The use of corpora as translation resources has become one of the most significant practices in applied translation studies. Corpora and concordancing are considered valuable tools for translators. They improve both productivity and the quality of translation in many respects, for example, subject field understanding, proper term choice, idiomatic expression, register, and stylistic variation. Compared to conventional resources such as dictionaries (monolingual or bilingual; general or specialized) and other non-electronic documents, corpora and concordancing lead to better understanding of the source language texts and a higher ability to produce translations that read as naturally as texts originally written in the target language.

Corpora play an important role in both translation from L2 to L1 and vice versa. In the former case, translators who have an opportunity to consult an electronic corpus of reference texts can increase their productivity and are likely to make fewer mistakes. In the latter case, the role of corpus analysis proves to be even more significant as it can enhance translation quality with regard to correct term choice, idiomatic expressions, collocation patterns and so forth. The improvements are particularly apparent in the translation of domain-specific texts where understanding of the subject field is one crucial factor in achieving accurate and fluent work. As a result, in translating specialized texts, a

translator has an even greater need for reliable resources in order to check the language used by target language experts, no matter whether that target language is the translator's L1 or L2. Ingrid & Michael

Friedbichler (1997a) sum up the predominant applications of corpora in the translation of general and domain specific texts in the following table:

Table 1. Applications of corpora in translation of general and domain specific texts

	Language for general purposes	Language for specific purposes
L1	confirm opinion by corpus evidence	check register preferred by target language expert
L2,3,..	settle translation problems at phraseologic, semantic, syntactic, etc. levels	study source language meaning with the help of contexts

In addition, it is obvious that corpora prove beneficial to translators in the early stages of their work as they enhance comprehension of the source text. Corpora and concordancing are also very powerful tools which strongly support the process of translation. For instance, by searching a target language corpus, translators can obtain information on common authentic usage of terms, collocations and phrases which cannot be found in conventional resources. Finally, at the editing stage, target language corpora and concordancers can help translators polish their work as they enable them to "get a quick overview of the most common expressions (frequency) in the output page and of instantly cross-checking the complete source text of any concordance line of interest. Concordancers can also be used to search previous translations and reference texts for reusable terms, fixed phrases, idioms and many other items" (Ingrid & Michael Friedbichler, 1997b).

More importantly, since these tasks can be carried out at high speed, translators are able to work efficiently and save a considerable amount of time.

3.2. Corpora in translator training

Corpora play a major role in translator training, particularly in the development of translation skills through the use of corpus-based investigation strategies. It is widely accepted that translation depends heavily on mastery of two languages; therefore, it is basically concerned with the acquisition of general linguistic knowledge. However, Bernadini (1997) proposes that translation should be viewed as an LSP (language for specific purpose) situation of a non-standard type. This is because "translating requires the development of specialized capacity in the form of a number of skills and strategies which are specific to the translating process, rather than the discipline which the text to be translated belongs to."

According to Bernadini, there are three broad sets of translator capacities, namely awareness, reflectiveness and resourcefulness. She explains that first of all, translators should be aware that they have to go beyond single words in a text and see through language to the ways in which messages are conveyed. It is necessary for translators to see language as a network of related choices influenced by and influencing the cultural system in which it belongs. They should also know rules of usage and the implications when these rules are violated. Since meaning is not constructed linearly, translators have to see both intertextual and intratextual relationships in order to produce an appropriate translation. Second, translators need to be reflective. They must have knowledge of strategies and procedures for text analysis, which include interpretation of textual and contextual clues, interpretation of implications such as illocutionary intent inferred from the texts, and sensitivity to writing strategies. Finally, translators should be resourceful. They need to know how to search for knowledge by themselves and how to exploit the knowledge they acquire to solve new and unexpected problems.

Bernadini claims that analysis of a large corpus can help develop the three capacities discussed above. During the training process, as translators learn to use corpus-based investigation strategies, such as observing, asking questions, forming hypotheses, searching, analyzing, refining hypotheses, finding possible answers and drawing conclusions, they begin to look for more precise solutions rather being satisfied with approximates. They become suspicious of what may seem obvious at first. Since corpus analysis focuses translators' attention on both intratextual and intertextual observations, it seems to provide a means of exercising and refining a set of cognitive instruments which would be less likely to develop if the focus were on intratextual analysis of single texts, as is normally practiced in conventional translation tasks. In addition, corpus analysis seems to promote translators' capacity to interpret texts critically, and this helps to build up their knowledge of what are (or are not) typical utterances in a language. The use of corpora also increases translators' lexical competence since they are more exposed to authentic language use. As for the development of resourcefulness, translators learn to make use of their knowledge and capacity to solve problems they may encounter while performing translation tasks with available resources, and to improve the corpus use skills required for more extensive investigations.

3.3. Corpora in the teaching of translation and foreign language learning

The use of translated text corpora and concordancing tools for translation activities can help translation and foreign language learners develop competence and increase their cross-cultural and cross-linguistic awareness. As mentioned earlier, it is necessary for translators to be competent in both the source and target languages. To a large extent, then, they share a common need with foreign language learners, which is to achieve a high degree of second language proficiency. Corpora of translated texts and concordancing can be used effectively to develop learners' competence in both translation teaching and foreign

language learning. They provide abundant data for contrastive analysis in meaning, grammatical systems, syntactic structures and strategies used in the two languages. By means of such analysis, the corpora can illustrate how the same message is conveyed in two separate linguistic and socio-cultural settings. In other words, corpus-based investigation techniques employed with bilingual text corpora enable learners to study the use and function of language cross-culturally.

The contribution of bilingual corpora and concordancing to improving learners' cross-cultural and cross-linguistic awareness is in allowing learners to go beyond syntactic and semantic levels in trying to produce a text in another language system. For instance, a concordaning tool makes it possible for them to compare sentences which appear to be semantically and syntactically equivalent but which differ in the likelihood of occurrence. Baker (1998) illustrates this by giving the example of the Portuguese expression of thanks, "muito obrigado/a," which, considering their semantic and syntactic equivalence, might be directly translated into "much obliged" in English. However, evidence from translated corpora shows that while "muito obrigado/a" is a normal way of expressing thanks in Portuguese, "much obliged" is rare and has more restricted usage than other common expressions like "thanks a lot," "thank you so much," and so forth. It is obvious that corpus evidence provides learners with both frequency information and plentiful examples of actual use of language in its varying contexts. These are crucial factors leading to a higher ability to produce target language that sounds natural and is appropriate to social and cultural context.

In terms of pedagogy, the application of corpus techniques has set a new direction for how we teach, i.e. teaching methods and classroom activities. Johns (1991b) calls a methodology based on the use of corpora and concordancing data-driven learning. Based on this concept, Murison-Bowie (1996) points out there are two approaches that may be used to carry out corpus-based tasks, namely the bottom-up approach proposed by Johns (1991a) and the top-down approach. Both can be applied effectively in the learning of translation and a foreign language.

Johns describes three phases of inductive reasoning with corpora via the bottom-up approach: observation, classification and generalization. One good example of how this approach works in translation and language teaching is given in Zanettin (1998). He asks his students to use a bilingual comparable corpus (which he defines as a collection of texts composed independently in the respective languages and put together on the basis of similarity of content, domain and communicative function) of news articles about France to investigate how proper names are used in this particular genre in English and Italian. They begin with a search for the proper name *Mitterand*. The corcordance lines of all instances of this proper name in English and Italian are observed, showing some differences in the way *Mitterand* is referred to in these two languages. The Italian texts seem to call him often by his full or last name while the English texts prefer to call him by a title, i.e. President or Mr.

A further search which yields a KWIC (Key Word In Context) concordance output[4] reveals that the adjective *francese* and the noun *presidente* are used quite commonly in the Italian texts whereas the equivalent words are rarely used in the English texts. In addition, when a full text is scanned, different structural positions of these references can be classified. English articles, for example, use the name *Mitterand* without a title only in headlines. In the body of the articles, he is introduced as *President Mitterand*, and often referred to later on as *Mr. Mitterand* or *the President*. On the contrary, he is commonly first mentioned as *François Mitterand* or *il Presidente Mitterand*, and subsequently called *il presidente*, *il presidente francese* or simply *Mitterand* in Italian articles. By comparing the evidence from bilingual comparable corpora, patterns of reference used in the two languages can be generalized and they can serve as stylistic models to refer to when writing in a foreign language.

In contrast with the top-down approach, the bottom-up approach involves hypothesis testing and deductive reasoning. An example of how this approach can be used in translation tasks also comes from Zanettin (1998). In his experiment, small bilingual comparable corpora on "post transfusion hepatitis C" are used in the teaching of specialized text translation from Italian into English. When the students see the phrase "biopsia epatica" in the Italian text, they hypothesize that it should be translated into the English medical term "hepatic biopsy." This is seemingly unproblematic. However,

after testing their hypothesized translation against the corpus evidence by searching for "biopsy," no single instance of "hepatic biopsy" was found while there were 39 occurrences of "liver biopsy." The students, therefore, need to refine their hypothesized translation and carry out a search again to confirm their choice. A search for "hepatic vs. epatico/a" and "liver vs. fatego" and comparison of collocations between the words in each pair assure that they have made the right choice as in English "hepatic" occurs only with generic terms such as disease, failure and lesion. By contrast, the Italian adjective "epatico" occurs mostly in compounds with the word "fegato".

We may conclude from the above that evidence from corpora constitutes a valuable resource for translators as both general and specialized corpora aid the comprehension of the source language texts and improve production skills. Corpus-based investigation strategies also help translators develop capacities which are specifically required for translation work. In addition, the use of corpora and concordancing software for translation activities can develop competence and help learners gain a more profound understanding of the languages and the cultures involved. It encourages both inductive and deductive learning and promotes a learner-centered approach via data-driven learning activities, shifting the emphasis from final product to the learning process. More importantly, corpus-based tasks help learners to build up their potentiality for observation and analysis, and enhance their linguistic ability.

---

[4] Example of KWIC concordance output : have turned him off, he coolly turned **round** and sai that my father was a geo uth where the sun shone all the year **round** and he could ride half th day w lue sky; smooth and dimpled slopes as **round** as cones, bare but for the green

4. Types of corpora and their applications

According to Baker (1995), the corpora compiled for translation purposes can be classified into 3 types. **Comparable corpora** comprise a collection of texts originally written in one language as well as other texts translated into that language. Ulrych (1997) suggests that this type of corpora be called "monolingual comparable corpora." **Multilingual corpora** contain sets of original texts written independently in two or more languages with no reference to translated texts and put together on the basis of similar design criteria such as content, genre, communicative function etc. Ulrych refers to them as "bi/multilingual comparable corpora." **Parallel corpora** consist of original source texts in one language and their translated versions in one or more other languages. Ulrych uses the term "bi/multilingual parallel corpora" to refer to this type of corpora. Applications of each type of corpus are discussed below.

4.1 Comparable corpora (monolingual comparable corpora)

According to Baker (1993, 1995), comparable corpora have a significant contribution to make to theoretical and descriptive translation studies but they have no clear role in applied areas. With an access to this type of corpus, we can explain the nature of translated texts as a mediated communicative event. This type of corpus, together with a concordancing tool, allows us to identify universal features of translation which are not the result of interference from the source or target languages involved as well as other typical features of translated texts i.e., deviations from the target language system and norms caused by the influence of the source language.

Simplification, for example, is considered a typical feature of translated texts. The fact that translated texts display a high type-token ratio, low lexical density and low sentence length (all calculated by a concordancing software such as WordSmith Tools) compared to other texts in the same language supports the simplification hypothesis. More details on this feature can be found in Baker (1995) and Laviosa-Braithwaite (1996) who investigated a sub-section of English Comparable Corpora (ECC) and found four patterns of lexical use in translated versus original texts. Since these patterns occur in both the Newspaper and the Narrative Prose Subcorpora of ECC, they may prove typical of translational English in general. Another universal feature of translation which can be tested via the use of concordancing tools and comparable corpora is explicitation. Baker (1996) points out that increased text length and a disproportionate number of explanatory vocabulary words and conjunctions can support the explicitation hypothesis. As for the claim that translated texts are more grammatically conventional than their source or other target language texts, collocational patterns used in comparable corpora can be investigated to prove this hypothesis (Toury, 1980).

Although it has been widely accepted that this type of corpus is mainly used in theoretical aspects, it has recently started to play a role in applied areas. Pearson (2000) explains how comparable corpora can be used to evaluate student translations. She says that one of the most difficult tasks for teachers of specialized translation is to

justify the corrections they make on students' work, particularly when it is grammatically and semantically correct but still does not read as naturally as the text originally written in the target language. She claims that a comparable corpus can be of great help to teachers as they provide the evidence they need to explain why they have made such corrections and recommendations.

In order to assess the benefits of using monolingual comparable corpora to evaluate student translation output, Pearson carried out an experiment in which a small English corpus of scientific texts on genetic transfer was compiled. This corpus is used to compare term choices made by the students whose task is to translate a French text on this topic into English by consulting conventional resources such as dictionaries and parallel documents with the choices made by authors writing the original English articles. It was found that most students tend to retain the construction of the source text terms as a result of source language interference. By searching the comparable corpus, teachers can verify student choices. Not only can they point out which of the proposed translations are actually used in the corpus and therefore prove to be more suitable choices, but they can also give conclusive explanations to the students.

4.2 Multilingual corpora (bi/multilingual comparable corpora)

For many translation scholars including Baker (1993), Shuttleworth and Cowie (1997), multilingual corpora have contributed mainly to applied translation studies. Shuttleworth and Cowie (1997, p.109) state that this type of corpus "has been profitably used in translator training as it can be exploited to provide useful insights into the typical means employed by two or more languages to express similar meanings."

The research work conducted by Zanettin (1998) discussed in 3.3 above is one good example showing how bilingual comparable corpora can be used to explore texts of a particular genre in two different languages and to learn about terminology and content of a specific-domain text. He illustrates that English-Italian comparable corpora of news articles provide learners with a means of testing the relationship between linguistic items which they perceive as holding some kind of similarity or equivalence, such as proper names and cognates. In addition, Zanettin finds that access to the hepatitis corpus helps learners to acquire knowledge about a specific subject and to understand the source text and produce a better translation. In particular, it can be used to identify technical terms in the target language and to compare their collocates with those of the corresponding terms in the source language to investigate varying patterns. This enables translators to produce translation work that satisfies target language norms.

Carol et. al (2000) also demonstrate how we can make use of bilingual comparable corpora of specialized texts to look for natural language examples of L2 lexical equivalents of a given word or expression in L1. Assuming that word meaning can be drawn from the context and words used in a similar way in domain specific corpus will be semantically similar, they propose that "if it is possible to establish equivalences between several items contained in two different contexts, there is a high probability that the two contexts

themselves are to some extent similar." Exactly what they are trying to do is to isolate the set of contexts in the L2 corpora which is most likely to provide L2 correspondences to the L1 input. Given a particular term found in the texts in L1, they aim to identify contexts which treat the same argument in the second language texts. To achieve this objective, they try to isolate the vocabulary related to that term in the L1 corpus based on the hypothesis that the word will be surrounded by a similar vocabulary in L2. In this way, bilingual comparable corpora should enhance the ability of translators and language learners to produce a natural sounding target language text, which is one of the main goals of every translation and language teaching program.

4.3 Parallel corpora (bi/multilingual parallel corpora)

According to Baker (1995, 1998), parallel corpora which are aligned[5] either on a word-by-word, sentence-by-sentence or paragraph-by-paragraph basis have contributed to both theoretical and applied translation studies. She points out that practical applications of this type of corpus include materials writing, computer-aided translator training, and improvement of machine translation systems. Zanettin (forthcoming) also mentions that parallel corpora have been extensively used as a basis for bi/multilingual terminology databases and thesauri. Their most important contribution to the theoretical aspect of the discipline is that they support a shift of

emphasis from prescription to description. They allow us to explore how translators handle difficulties of translation in practice and to use the findings to set up models for trainee translators. They also provide information on language-pair specific translational behavior and help to postulate equivalence relationships between lexical items or structures in the source and target languages. They enable us to study translational norms or normal patterns of translation and typical features of translated texts i.e., deviations from the target language system and norms caused by influence of the source language.

In addition to their significant role in translation studies, bi/multilingual parallel corpora can be used effectively in language teaching. Carol et. al (2000) state that translation exercises are still valuable as a means of improving learners' written competence. To the researchers, the translation must be linguistically correct and achieve equivalence at the level of style, register, cultural and social context, etc. The learners, therefore, "need to be able to directly access real world translation data so that they can be exposed to different renderings of a given word or expression according to variations in the communicative function of a particular text." This is the reason why parallel corpora can serve as a valuable tool in second language teaching. Since parallel corpora comprise real-world texts from which data on in-context translation equivalents can be extracted, they are most likely to provide useful data for the second

---

[5] Example of aligned parallel text on sentential basis:
English: The curtains don't go with the carpet.
French: Les rideaux ne vont pas avec le tapis.

language learner, who is mainly looking for information on ways in which a given word or phrase can be translated acceptably in another language. Procedures for how the parallel corpus system developed by Carol et. al works in detail are available at http://www.ilc.pi.cnr.it/dbt/pisystem.html .

Barlow (2000) also argues strongly for the use of parallel corpora in language teaching. Taking a schema-based approach to grammatical description, he proposes that learning a second language involves some use of first language schemas as templates for creating schemas in the second language, and for forming completely new schemas in the second language as well. Through the use of parallel texts, language learners are allowed to directly investigate the main correspondences between particular words and structures in two languages. The students' exposure to this information helps them form new schemas and appropriate schema-meaning links.

Parallel corpora can be used in language teaching at many levels. Barlow (2000) points out that "parallel texts can be exploited by beginning students to create an awareness, at least, for the feel of a second language and also to obtain some concrete knowledge of correspondences" (p.113). Parallel corpora can help dispel learners' misperception that there is a word-for-word

equivalence between different languages. He gives an example of the English noun "head," which most learners would automatically translate into "tête" in French. However, with our knowledge of the English schema associated with "head" and the links to different meanings of this term, we need to teach learners of French that it is not always the case that English "head" can be translated as "tête" in French. Analysis of parallel corpora provides clear evidence of the many-to-many nature of French and English correspondences.[6] In addition, a parallel concordance search using parallel concordancing tools such as ParaConc (Barlow, 1995) gives important information on relative frequency of different constructions and collocations. The output from the bi-directional search (L1-L2 and L2-L1) clearly shows strong and frequent correspondences, giving very good indication of the preferred structures in each language.

As for more advanced learners, Barlow (2000) states that what a parallel text has to offer is not only equivalences but also contexts, both of which help to increase learners' knowledge of the range of word meanings and collocations. He says "advanced learners need to deepen their knowledge of words and phrases: to understand not just the main meaning or most common meaning of a word,

---

[6] A selection of French correspondences to the English noun 'head' given in Barlow (2000) are for example:

1. a sprinkler head	la tête d'un extincteur
2. Acting Head	chef intérimaire
3. army and its head	l'armée et son chef
4. price on his head	sa tête prix
5. head of the delegation	chef de la délégation

but to understand a range of meanings and to know how context in terms of discourse and genre provide clues to the appropriate meaning" (Barlow, 2000, p.114). One good example of this is a study of the translations of the French subjunctive in English and the translations of the English mandative "should" in French from English-French parallel corpora of two different text types: press and learned prose, conducted by Serpollet (2000).

Besides the three major types of corpora discussed above, there is another type of corpus which is contributing significantly to applied translation studies, that is **monolingual specialized target language corpora**. It consists of texts in a specialized subject field written in the target language of a translation task. (The target language may be either the translator's native tongue or second language). Bowker (1998) illustrates how this type of corpus can be used as a translation resource.

Bowker (1998) points out that two major types of errors are commonly found when correcting student translations, in particular those of domain specific texts. The first type is comprehension errors caused by a lack of understanding of the subject field. The second is production errors resulting from a lack of competence in the specialized target language. She claims that access to specialized target language corpora is a possible solution to these problems as they provide the students with resources which can help them acquire background knowledge and with examples of correct usage in context. To prove her claim, she carried out an experiment in which she compared two French to English translations

on optical scanners produced by a group of translator trainees. One was carried out with the use of conventional resources including a selection of general bilingual dictionaries and specialized lexicographic and non-lexicographic resources; the other with a selection of general bilingual dictionaries and the aid of a specialized target language corpus. A concordancing program called WordSmith Tools was used to search and analyse the corpus. The findings reveal that the quality of corpus-aided translations is higher in respect of subject field understanding, correct term choice and idiomatic expression.

5. Key success factors

The above section, in which different types of corpora and their applications are reviewed in detail, clearly illustrates that language corpora and concordancing programs have gained increasing popularity in this subject field because they have proved highly beneficial for both theoretical and applied aspects of translation studies. However, the success of corpus-driven translation, especially in the teaching of translation in translator training and foreign language learning, which is the main emphasis of this paper, depends on various factors.

The first factor involves the concordancing tools used. Since there are many types of corpora, suitable concordancing software is needed in order to carry out the tasks, which include searching for, sorting, retrieving and calculating linguistic data. For monolingual corpora, there have been no monolingual corpus analysis tools developed specifically for

translation applications so far. Therefore, the tools that most translators normally use in order to extract information from monolingual corpora are generic corpus analysis tools basically aimed at language teaching and linguistic research. At present, the most widely used generic concordancing software for Windows is WordSmith Tools developed by Mike Scott (1996). This powerful software contains various useful features which enable users to generate and manipulate word lists, concordances and collocations.

As for parallel bi/multilingual corpora, ParaConc, developed by Michael Barlow, is considered a simple parallel text concordance program also available in a Windows version. It contains all the features typical of corpus analysis tools, but the output of the search is displayed in windows – how many windows depends on the number of parallel texts the user is working with. For example, if the user is working with two parallel texts, lines containing each instance of the search term in the first language are shown either in sentence or KWIC form in one window, while the corresponding lines in the second language are displayed in another window. In this way, a translation, if any, for each occurrence of the keyword in the first language can be examined along with the original lines.

The second factor concerns the methodology used to search for information from a corpus. Stevens (1995) points out that concordancers are not tools that can be used without proper preparation. As a consequence, translation students and teachers have to be aware of the methodological considerations underpinning the use of the tools. Users also need to know

how to search for the item or pattern they want to investigate. Teachers, therefore, should provide students with sample tasks or exercises so as to help them learn how to carry out corpus searches effectively. To be more specific, the learners need to know the techniques that can be used to identify and extract information from the corpus. The techniques described in Bowker (2000) are, for example, identifying important terms by using the Word List feature to get a list of the most frequently occurring words, and investigating a linguistic hunch by using the corpus as a testbed to check if a hypothesized translation is actually the suitable one. The feature used in the latter technique is the wildcard search, such as 'writ*' for all forms of 'write, writes, writing, written,' which are possible translations of the French verb 'écrire.'

The third factor is interpretation of the output. Gavioli (1997) emphasizes the importance of this factor, noting that while computer concordancing produces samples of actual language occurrences, the samples may not always provide immediate answers to students' queries. Learners can obtain the answers only through interpretation and categorization. Therefore, they should be introduced to the process of interpretation and categorization and what this process leads to. Murison-Bowie (1996) sums up the three ideas above saying that in order to use a corpus and concordancer as effective tools in the learning process, it is necessary for the students to receive training consisting of the following elements:

a) The mechanics of using concordancers and knowing which functions to use for what purpose;

b)   The means for conducting searches (or searching techniques), single or sequential, in the most effective way (as several attempts using different search words might be needed before appropriate solutions can be identified);

c) Guidance on what kinds of problems can be effectively resolved by the use of corpora and concordancing, and

d) Guidance on interpreting evidence.

The last key success factor involves the corpus which will serve as the source of input data to be used in the translation process. First of all, it is necessary that corpus data be carefully selected to suit the objective; in other words, whether it is to be used in translation research, in translation teaching or as a translation resource. This means that there must be a needs analysis for the corpus selection, i.e. a general corpus vs. a domain specific corpus, a monolingual corpus vs. a bi/multilingual corpus, and a comparable corpus vs. a parallel corpus. The issue of representativeness must also be taken into consideration, especially when general translation corpora are used in conducting theoretical and descriptive translation research. This is because we need to make sure that the corpus data adequately represent the target population we want to study.

Ideally, translation students and trainee translators should be able to acquire and explore corpora for their individual objectives. However, there are certain limitations; for example, quite a few corpora compiled for translation purposes cannot be accessed by the public. A possible solution to this problem is to purchase commercial corpora or to develop one's own corpora either by using optical character recognition

devices (OCR) to scan written text documents to compile electronic databases or by downloading electronic texts from such inexhaustible resources as the World Wide Web (WWW). As for parallel corpora, which need to be aligned before they are put into use, they may become more readily available in the future, as more translators have begun to use translation memory tools to build up such resources. In any case, it seems to me that there are enough publicly available corpora at present to allow teachers to introduce corpus-based activities in the translation classroom.

Based on my experience teaching a graduate course in business and economics translation, I have found that monolingual specialized target language corpora can be particularly useful resources for students. They can help the students choose correct terms, and by providing information about common word collocations, the corpora enable students to use proper idiomatic expressions which make their translation work approach the naturalness of texts originally written in the target language. Yet despite these advantages over conventional specialized dictionaries, there are two major drawbacks in using such corpora in translation classes. One is that it is time consuming. Because students must search and interpret their findings, they are likely to prefer to use conventional resources which are less complicated and more familiar to them when working under the pressure of time constraints. The other major drawback is related to two of the key success factors mentioned earlier. First, students must have a corpus that actually represents the domain and text type they are working on; otherwise, they would not find the terms they are

searching for. In addition, I found it essential to train the students on how to search for items they wish to investigate. In other words, they should familiarize themselves with the concordancing programs and they must know the strategies and techniques for making a search if they are to make the best use of the corpora.

6. Conclusion

It can be seen from all the above that corpora and concordancing tools have had and will continue to have a profound impact on both the theoretical and applied aspects of translation studies. Their use provides crucial and abundant data for descriptive translation as well as valuable evidence which helps improve not only productivity and quality of translation but also translators' capacities. In addition, the use of translated text corpora and concordancers can help translation and foreign language learners both to develop competence and to increase their cross-cultural and cross-linguistic awareness. More importantly, with the use of corpus-driven activities, trainee translators are introduced to an advanced method of consulting a novel type of translation resource which has proved more useful than conventional tools. Language students are also introduced to an innovative approach to language study which promotes exploration and discovery. Although some practical limitations to corpus-driven translation activities remain, language learners and trainee translators should be given an opportunity to benefit from the fruitful and stimulating activities offered by this approach and to exploit this new tool as best they can.

The Author

Assistant Professor Prima Mallikamas graduated with a B.A. in English (1st Class Honors) from Chulalongkorn University, and obtained an M.A. in English from the University of Wisconsin-Madison. She has been teaching at the Faculty of Arts, Chulalongkorn University, for over 10 years. Her fields of interest include English teaching, corpus linguistics, and translation. Some of her publications are "Translation as Language Teaching Technique" and "Applications of Corpora and Language Teaching."

References

Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. In Baker, G. Francis & E.T. Bonelli (Eds.), Text and technology in honour of John Sinclair (pp.233-250). Philadelphia: John Benjamins Publishing Company.

Baker, M. (1995). Corpora in translation studies: An overview and some suggestions for future research. Target, 7(2), 223-243.

Baker, M. (1996). Corpus-based translation studies: The challenges that lie ahead. In H. Somers (Ed.), Terminology, LSP and Translation: Studies in language engineering in honour of Juan Sager (pp. 175-186). Amsterdam and Philadelphia: John Benjamins.

Baker, M. (1998). Routledge encyclopedia of translation studies. London and Newyork: Rouledge.

Barlow, M. (1995, December). ParaConc: A concordancer & texts [Online]. Available: http://www.ruf.rice.edu/~barlow/para-ox.html

Barlow, M. (2000). Parallel texts in language teaching. In P. Botley, T. McEnery & A.Wilson (Eds.), Multilingual corpora in teaching and research (pp.106-115). Amsterdam: Rodopi.

Bernadini, S. (1997, November). A 'trainee' translator's perspective on corpora. [Online]. Available: http://www.ssimit.unibo.it/cultpaps/trainee.htm

Bowker, L. (1998, December). Using specialized monolingual native-language corpora as a translation resource: A pilot study. Meta, 43(4) [Online]. Available: http://www.erudit.org/erudit/meta/v43no4/bowker/bowker.html

Bowker, L. (2000). Towards a methodology for exploiting specialized target language corpora as translation resources. International Journal of Corpus Linguistics, 5(1), 17-52.

Peters, C., Picchi, E., & Biagini, L. (2000). Parallel and comparable bilingual corpora in Language teaching and learning. In S.P. Botley, T. McEnery & B.Wilson (Eds.), Multilingual corpora in teaching and research (pp.73-85). Amsterdam: Rodopi.

Firth, J. R. (1968). Linguistics and translation. In F.R. Palmer (Ed.), Selected papers of J. R. Firth 1952-59 (pp.84-95). London: Longman.

Gavioli, L. (1997). Exploring texts through the concordancer: Guiding the learner. In A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (Eds.), Teaching and language corpora (pp. 83-99). London: Longman.

Holmes, J. S. (1988). Translated! Papers on literary translation and translation studies. Amsterdam: Rodopi.

Friedbichler, I. & M. (1997a, November). The potential of target-language domain-specific corpora for the translator's workbench. [Online]. Available: http://www.ssimit.unibo.it/cultpaps/fried.html

Friedbichler, I. & M. (1997b, March). Open up new horizons with corpus-driven translation. [Online]. Available: http://info.uibk.ac.at/c/c6/c613/friecorp.html

Johns, T. (1991a). Should you be persuaded – two samples of data driven learning materials. ELR Journal, 4, 1-16.

Johns, T. (1991b). From printout to handout: grammar and vocabulary teaching in the context of data-driven learning. ELR Journal, 4, 27-46.

Laviosa-Braithwaite, S. (1996). The English comparable corpus (ECC): A resource and Methodology for the empirical study of translation. An unpublished Ph.D. thesis, UMIST, Manchester, UK.

Murison-Bowie, S. (1996). Linguistic corpora and language teaching. Annual Review of Applied Linguistics, 16, 182-199.

Pearson, J. (2000). Using specialized comparable corpora to evaluate student translations. In B. Lewandowska-Tomaszczyk & P.J. Melia (Eds.), PALC'99: Practical Applications in Language Corpora, Lodz 1999 (pp.541-552).

Schmied, J. & Schaffler, H. (1996). Approaching translationese through parallel and translation corpora. In C.E. Percy, C.F. Meyer & I. Lancasshire (Eds.), Synchronic corpus linguistics: Papers from the sixteenth International Conference on English Language Research on Computerized Corpora, Toronto 1995 (pp.41-56).

Scott, M. (1996). WordSmith Tools. Oxford: Oxford University Press.

Serpollet, N. (2000, October). From French to English and back again: How a grammatical approach to corpus data can be useful to translation studies. Paper presented at the 2nd International Conference on Corpus Use and Learning toTranslate, Bertinoro, Italy.

Shuttleworth, M. & Cowie, M. (1997). Dictionary of Translation Studies. Manchester: St. Jerome Publishing.

Sinclair, J. (1992). The automatic analysis of corpora. In J. Svartvik (Ed.), Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82, Stockholm 1991. (pp. 379-397). Berlin & New York: Mouton de Gruyter.

Stevens, V. (1995). Concordancing with language learners: Why? When? What? CAELL Journal, 6(2), 2-10.

Toury, G. (1978). The nature and role of norms in literary translation. In J.S. Holmes, J. Lambatt & R. van den Broeck (Eds.), Literature and Translation (pp. 83-100). Leuven: ACCO.

Toury, G. (1980). In search of a theory of translation. Tel Aviv: The Porter Institute for Poetics and Semiotics.

Toury, G. (1985). A rationale for descriptive translation studies. In T. Hermans (Ed.), The manipulation of literature: Studies in literary translation (pp.16-41). London and Sydney: Croom Helm.

Tymoczko, M. (1998, December). Meta, 43(4) [Online]. Available: http://www.erudit.org/erudit/meta/v43no4/tymoczko/tymoczko.html

Ulrych, M. (1997). The impact of multilingual parallel concordancing on translation. In B. Lewandowska-Tomaszczyk & P.J. Melia (Eds.), PALC'97: Practical Applications in Language Corpora, Lodz 1997 (pp.421-435).

Vanderauwera, R. (1985). Dutch novels translated into English: The transformation of a "minority" literature. Amsterdam: Rodopi.

Zanettin, F. (1998, December). Meta, 43(4) [Online]. Available: http://www.erudit.org/erudit/meta/v43no4/zanettin/zanettin.html

Zanettin, F. (forthcoming). Swimming in words: Corpora, translation, and language learning. In G. Aston (Ed.), Learning with Corpora.