

BILINGUAL CONCORDANCERS: A NEW TOOL FOR BILINGUAL LEXICOGRAPHERS

Lucie Langlois
University of Ottawa

ABSTRACT

Concordances extracted from aligned bi-texts have become an extremely important tool for the bilingual lexicographer. This paper will show in a concrete way how bi-concordances are actually used in a bilingual dictionary project.

1. INTRODUCTION

The Bilingual Canadian Dictionary Project, also called the BCD¹, has as its primary objective the preparation of the first comprehensive Canadian French-English English-French general dictionary. Due for publication in 2003 in both paper and CD-ROM formats, it will contain a total of 80,000 entries. The BCD aims to be different from other English/French dictionaries in that it will present English and French as they are used in Canada, including words and word senses that are particular to Canadian usage. Since our typical users will be translators and writers, our entries will give these sophisticated users more equivalents for each headword, as well as more free combinations and collocations than existing bilingual dictionaries do.

Another way in which the BCD hopes to improve on existing dictionaries is by its substantial use of corpora². BCD lexicographers use both unilingual and bilingual corpora. These corpora are searched using concordance generating programs, which we call concordancers, the output of which are consulted by lexicographers in producing and revising entries.

BCD lexicographers use various corpora in a multitude of ways. First, they analyze unilingual source language dictionary entries to get a feel for the semantic range of the word being worked on. They then check unilingual source language concordances to see if the semantic divisions suggested by the unilingual dictionaries actually reflect reality. This is not always the case; for instance, there have been many instances where the corpus data suggested more sense divisions than were actually found in existing dictionaries. Next, the lexicographer looks at bilingual dictionaries to see which equivalents have been proposed by other lexicographers. Finally, bilingual concordances are examined to validate and add equivalents.

Bilingual corpora are used at different points of the lexicographic process. Although transfer dictionaries will not be mentioned explicitly here, techniques and tools that are useful in producing bilingual dictionaries should also be helpful in setting up dictionaries for machine translation systems. This paper deals solely with the use of bilingual corpora and bilingual concordances in bilingual lexicography. More specifically, it focuses on one particular type of

¹ Funded by the Social Sciences and Humanities Research Council of Canada, the project includes three universities (University of Ottawa, Université de Montréal and Université Laval) and employs between 25 and 30 research assistants at any given time.

² A complete account of the use of monolingual concordances in monolingual lexicography can be found in Sinclair (1987). Church and Gale (1991) suggest ways in which bi-concordancers could be useful in bilingual lexicography.

concordancer, namely one searching bi-text, to use Harris' (1988) term, to produce concordances in which a match consists of a source language segment together with its translation. Bilingual concordance generating programs will be referred to as "bi-concordancers" and their output, "bi-concordances".

The bi-concordancer we have been using at the BCD since 1995 is TransSearch³, which was developed by the Computer-Aided Translation group of the Centre for Information Technology Innovation (CITI). This program allows us to query our bi-text⁴ to generate bi-concordances.

2. CASE STUDIES

Presented here are various ways in which bi-concordances can be used in bilingual lexicography.

2.1 When bilingual dictionaries propose few equivalents

Every translator, at one time or another, has needed to consult a bilingual dictionary, only to put it away a few minutes later, no further ahead. Not only did the dictionary not propose a usable equivalent, it didn't even hint at other options. And this is the case not only for complex words, but also for relatively simple ones, as the following examples illustrate.

2.1.1 *Automatically*

An analysis of the entries for automatically in three well-known bilingual English/French dictionaries⁵ shows that few equivalents are proposed, either in the equivalents list or in the translation of examples. For instance, although all three give *automatiquement* as the main equivalent, only one proposes a second equivalent, namely *machinalement*. However, one did provide *d'office* in the translation of a free combination⁶ and two included the legal term automatically void (*nul de plein droit*).

At this point, the BCD lexicographer might be tempted to only put *automatiquement* as an equivalent for automatically. However, a search of the BCD's bi-textual corpus reveals other possible translations. In fact, in 25% of the 322 cases, automatically is rendered by something other than *automatiquement*. For instance, an adverb-adjective transposition is often used, where the meaning of the adverb automatically is carried by the adjective *automatique* in French, as in the following example which TransSearch helped us extract.

to pay dues automatically:
des contributions automatiques

This is not to say, of course, that the adjective *automatique* should be listed as an equivalent for the adverb automatically, but it could be presented as a possible translation of a free combination. The

³ For more information on TransSearch, see Macklovitch (1992) and Simard, Foster and Perrault (1993).

⁴ Our bi-text is made up of 3 years of the aligned Canadian Hansard (debates of the Canadian House of Commons), containing 21.6 million English words and 24.1 million French words.

⁵ Dictionnaire Hachette-Oxford, Grand Dictionnaire Larousse and Dictionnaire Robert & Collins Senior.

⁶ A *free combination* is a phrase in which the headword is used without any special syntactic or semantic restraints.

analysis of the bi-concordances also reveals that the adverbial phrase *de façon automatique* is sometimes used to translate automatically. This equivalent is not as common as *automatiquement*, but it could also be listed as an option. So, if we were to include to pay dues automatically in the free combinations section, the entry could be enriched in the following way:

to pay dues automatically:

contribuer automatiquement, payer les primes par versement automatique, verser les primes de façon automatique

Another way that House of Commons translators avoid *automatiquement* is by using synonyms such as *forcément* and *nécessairement*. While these equivalents cannot be used to translate automatically in all cases, a well-chosen example would allow the user to get a feel for the contexts in which these partial synonyms could replace *automatiquement*. The following example shows how three different bi-concordances produced by TransSearch were merged in the final BCD entry:

when the economy is weak, government automatically takes in less in taxes:

lorsque l'économie est au ralenti, le gouvernement perçoit automatiquement/nécessairement/forcément moins de taxes

It might be argued that many an experienced translator would automatically (!) think of these translations. However, it could be shown that many others would not, especially novice translators or translators working out of their mother tongue. Consequently, presenting more sophisticated and varied examples in a bilingual dictionary gives the users a glimpse of the wealth of possibilities in the target language and, more importantly perhaps, encourages them to think of other original ways to translate a given word.

2.1.2 Oriented

Seeing that we are a something-oriented, something-conscious and something-friendly society, it is no surprise that compound adjectives containing the words oriented, conscious and friendly often surface in texts dealing with contemporary issues. If defining these adjectives poses somewhat of a problem, translating them is an even more complex task, since their translation often depends on the actual context. We will examine the case of "something-oriented".

Of the three bilingual dictionaries consulted, only one actually gave equivalents for oriented, namely *axé sur* and *orienté vers*; furthermore, most of the 10 examples and/or collocations listed for oriented were translated word-for-word using these equivalents.

An analysis of the bi-concordances produced by TransSearch revealed a wealth of information ready to be used in a bilingual dictionary⁷. For instance, service-oriented was translated in the following ways, depending on its context:

service-oriented Canadians:

les Canadiens accordent plus d'importance au service

⁷ It must be mentioned that the translations are not always brilliant. For example, a native Francophone would never spontaneously say *une société tournée vers ses clients* which was found in the Hansard to translate client-oriented company.

service-oriented organizations/corporations:
des organismes/entreprises de service(s)
service-oriented jobs:
des emplois du secteur tertiaire

Here are some other examples extracted from the Hansard which could well be incorporated into our entry for oriented:

export-oriented industry:
une industrie qui exporte partout dans le monde
export-oriented goods:
des produits destinés à l'exportation
family-oriented group:
un groupe à vocation familiale
growth-oriented economy:
une économie en pleine croissance
business-oriented groups :
des groupes d'affaires

This is not to say that all of these pairs should necessarily form part of the entry. Further research in unilingual corpora to corroborate the actual frequency of these combinations in both the source and target languages is essential. However, these outputs do suggest alternative equivalents for oriented.

2.2 When the lexicographer wants to identify or confirm Canadian usage

Even when existing bilingual dictionaries do provide a number of acceptable equivalents for a given head word, they rarely include Canadianisms, i.e. typically Canadian words or words with Canadian senses. As their inclusion is essential in a Canadian dictionary, the BCD lexicographer must use Canadian corpora, like Textum⁸ and the Hansard, to identify and confirm Canadian usage. Of course, the simple presence of a word or word sense in the Hansard does not a Canadianism make. These candidates must be further researched in unilingual Canadian corpora and the results compared to those found in non-Canadian corpora. The next two examples exemplify what we mean.

2.2.1 Leadership

A native Canadian Francophone who looks up the entries for leadership in the three bilingual dictionaries mentioned previously will notice a major omission. The equivalents listed are *direction*, *leadership* and *dirigeants* which are certainly correct, but no mention is made of the Canadianism *chefferie*. This, however, begs the question: is *chefferie* really an acceptable equivalent for leadership?

⁸ Textum, our unilingual corpus, is made up of 11 different sub-corpora (5 French, for a total of 100 million words and 6 English, totalling 210 million words). Three of the French sub-corpora are Canadian, the other two being from France, and 4 of the English corpora are Canadian, the others being American. We also have a British corpus which has not yet been added to Textum but which is also used for comparative purposes.

To clarify this matter, the lexicographer queried the bilingual Hansard so as to obtain only pairs of sentences containing leadership on the English side and *chefferie* on the French side. There were, indeed, many such concordances, showing that, in Canada at any rate, *chefferie* is used as an equivalent for leadership. More significant, however, is the fact that, when this query was further restricted to English as the source language, *chefferie* never once appeared in the translated French sentences, suggesting that *chefferie* is used only by native French speakers in oral speech and not by translators. To corroborate this, the lexicographer must query the unilingual French Canadian corpora to see if this word is indeed restricted to speech (which it isn't) and then compare these results to those extracted from the European sub-corpora to establish whether *chefferie* is a bona fide Canadianism (which it is).

Whether *chefferie* should be included in the equivalents list for leadership is a matter of editorial policy which goes beyond the scope of this paper. However, this example serves to show how TransSearch's flexible search engine enabled the lexicographer to focus in on a phenomenon that might otherwise have gone unnoticed.

2.2.2 Elder

The case of elder is interesting. All three bilingual dictionaries consulted when drafting this entry were in agreement: when elder is used to designate an older person, its French equivalent is *aîné* and, when one is referring to the older members of a tribe or a Church, the word *ancien* is appropriate. Not convinced by what the dictionaries asserted, the lexicographer responsible for this entry decided to consult the bi-textual version of the Hansard.

She queried the corpus twice - once restricting the search in order to single out pairs of sentences where elder(s) appeared in the English sentence and where *ancien(s)* was used in the French and, the second time, limiting the search to pairs containing elder(s) and *aîné(s)* in the English and French sides of the bi-concordances respectively. Comparing the results, she was able to confirm that both *aîné* and *ancien* are used to refer to older members of Indian tribes. Comparison of the results obtained in European French corpora versus Canadian French corpora proved that *aîné*, in this sense, is indeed a Canadianism.

Again, TransSearch and the nature of our bi-textual corpus enabled the BCD lexicographer to write a better dictionary entry that truly reflects Canadian usage.

2.3 When an equivalent proposed by bilingual dictionaries needs confirmation

It happens regularly in bilingual lexicography that an equivalent found in a dictionary does not ring quite true, either because the equivalent seems semantically inadequate or because the lexicographer is not completely sure of the meaning of the source language lexical item itself and thus cannot be sure if the equivalent proposed is correct. Here again, bi-concordances can be quite helpful, as the following example illustrates.

2.3.1 *Pie in the sky*

Not all Francophones are familiar with the expression pie in the sky. So, it stands to reason that a Francophone lexicographer would not necessarily be in the best position to evaluate the validity of the equivalents proposed in the various bilingual dictionaries, namely *ce sont des promesses en l'air*, *ce sont de belles promesses/paroles* and *c'est de l'utopie*.

However, using TransSearch, it is possible to confirm that *belles promesses* and *promesses en l'air* can indeed be used to translate pie in the sky. Furthermore, since bi-concordances show equivalents in their full sentential context, the lexicographer was able to see that the sentences containing pie in the sky and *de belles promesses* are not always structured the same way in English and in French. Here are two examples, which could be incorporated into the actual entry to illustrate the differences in structure:

they are not looking for pie in the sky:

ils ne veulent pas qu'on les abreuve de belles promesses

what we have seen in this budget is more pie in the sky:

ce budget contient encore d'autres belles promesses en l'air

2.4 When a collocation requires a translation

As the last example illustrates, the inclusion of collocations, which show idiomatic usage in a language, is considered essential in a good dictionary entry. This is especially true for bilingual dictionaries since, in many cases, collocates (words that combine in a collocation) cannot be translated word for word into another language. In this section, using two verb-noun collocations, we will show how bi-concordances can help to provide idiomatic translations for collocations.

2.4.1 *To take action*

The multi-word lexical item to take action is a typical verb-noun collocation in that it does not allow free substitution (for instance, one could not say to take act or to acquire an action), nor can it be translated literally into French (*prendre une action*).

The bilingual dictionaries consulted all proposed *agir* and *prendre des mesures* as translations of to take action. While both translations are suitable, they are by no means the only possible ones. Indeed, when TransSearch was asked to show all pairs of aligned sentences where to take action was not translated by one of the two equivalents set forth in the bilingual dictionaries, these amounted to nearly 40% of all the examples. Translators rendered to take action in a multitude of ways, including *intervenir*, *passer à l'action*, *adopter des mesures* and *réagir*, to list but a few.

All these possible translations will not necessarily find their way into our entry for action. It is impossible for a bilingual dictionary to list all possible equivalents. However, TransSearch does give the lexicographer a range of possibilities from which to choose, including some that would not have come to mind right off the bat.

2.4.2 To make history

While all bilingual dictionaries are in agreement on the translation of to take action, this is not the case for the expression to make history. Two of the bilingual dictionaries give *entrer dans l'histoire* as an equivalent, while the third suggests *être historique* (which literally means to be historical). The latter is obviously a poor general translation since it cannot be used to translate to make history in all contexts. A good translator would never say, for example, *aujourd'hui, nous avons été historiques* to render today, we made history.

The bi-concordances produced by TransSearch confirm that *entrer dans l'histoire* is definitely an idiomatic way of rendering to make history. However, they also propose translation alternatives, including *créer un précédent* (= to create a precedent), *pour la première fois dans l'histoire* (= for the first time in history), *écrire une page d'histoire* (= to write a page of history). These phrases can further be checked in a unilingual corpus before the lexicographer decides which to include in the final entry.

2.5 When an equivalent for an expression cannot be found in bilingual dictionaries

Fixed expressions are particularly difficult for bilinguals to deal with, since they cannot be translated word-for-word nor can their meaning easily be paraphrased. They differ from collocations in that, being proverb-like, they are quite rigid in form. Often, fixed expressions are not even mentioned in bilingual dictionaries, perhaps because of their problematic nature. Consequently, the translator faced with the task of rendering an unknown expression into another language may waste precious time looking it up in different sources in the hopes of locating, if not an equivalent, at least its meaning.

The BCD will include as many fixed expressions as possible. Ideally, the bilingual lexicographer will give an equivalent target-language expression for the source-language expression, and not just an explanation. The following examples show how bi-concordances can help the lexicographer do just that.

2.5.1 To make a silk purse out of a sow's ear

This expression was not found in the bilingual dictionaries consulted. However, an equivalent French saying *faire de la farine d'un sac de son* was found in our bi-text using TransSearch. Furthermore, one translator also used *faire des miracles* (= to perform miracles), which is a paraphrase for the English expression. Since the expression *faire de la farine d'un sac de son* is rare⁹ and its usage cannot be confirmed, it would probably not appear in the equivalents list in the BCD. However, *faire des miracles* gave the lexicographer some leads. Looking up *miracle* in *Le Bouquet des expressions imagées*, different expressions were found that could, pending further research in unilingual corpora, be included as equivalents for to make a silk purse out of a sow's ear, amongst them *vouloir prendre la lune avec les dents* (= to want to capture the moon with one's teeth) and *chercher midi à quatorze heures* (= to look for noon at 2pm).

⁹ It was not found in any of our unilingual French corpora, nor was it listed in the *Dictionnaire des expressions et locutions figurées* or in *Le Bouquet des expressions imagées*.

2.5.2 To be way off base

Colloquial expressions are certainly amongst the hardest expressions to deal with. This is especially true of those which are based on regional or local realities, in this case, baseball. The bilingual dictionaries offered *il n'y est pas du tout* and *dérailler*, which are certainly equivalent in meaning, but not in register. The bi-concordances extracted by TransSearch suggest more stylistically equivalent translations, such as *faire patates*, *se tromper royalement* and *se fourvoyer carrément*, which, again, could easily find their way in the BCD entry for "base", pending confirmation in unilingual sources.

3. CONCLUSION

We have attempted to show that bi-concordances represent an important source of information for bilingual lexicography, giving lexicographers options that they would not have had otherwise. Actually, it is TransSearch's flexibility¹⁰ and ease of use¹¹, enabling the BCD lexicographers to tailor queries to their specific requirements, that makes it a tool of great value in locating single words and multi-word expressions as well as their translations.

However, bi-concordances, like monolingual concordances, must be used prudently. The lexicographer can easily be overwhelmed by the amount of information delivered by the bi-concordancers and, consequently, lose precious time sorting and analyzing them. Furthermore, as is the case with monolingual concordances, the information retrieved is corpus specific. In the case of the Hansard, for instance, the word house is translated far more often by *chambre* and *séance* than by its usual equivalent *maison* because the corpus consists exclusively of the debates of the Canadian House of Commons¹². It must also be pointed out that the source language texts in the Hansard are spoken texts that have been transcribed (and corrected to some extent) whereas the target language texts, having been written by professional translators, are often of a higher register than the original.

Bi-concordances have the additional disadvantage of being based on translations. There is much controversy about putting a translation and its original on an equal footing. Debates on this question have been raging in the field of translation studies for many years¹³. While it is now generally admitted that a translation is indeed a text in its own right¹⁴, it is also true that, however excellent they may be, translators who work under tremendous pressure and time constraints do

¹⁰ TransSearch's search patterns can be imposed on either language, or both languages at the same time; complete boolean logic (*AND*, *OR* and *NOT* operators) is available and the output can be case sensitive. Since TransSearch uses comprehensive morphological dictionaries, it is possible for the lexicographer to search for all possible forms of a given lemma with just one query. It can easily search for complex multi-word items because the user can specify the maximum distance separating two words, which is essential in locating multi-word lexical units that can be separated by several words. Queries can further be restricted to aligned pairs where either French, English, or both, is the source language.

¹¹ TransSearch's intuitive graphical interface enables users (particularly beginners and occasional users) to formulate complex queries rather easily, in either English or French.

¹² The French equivalent for House of Commons is *Chambre des Communes*.

¹³ Schopenhauer (1992) has gone so far as to say that "A library of translations resembles a gallery with reproductions of paintings. Take translations of authors from antiquity: they are as obvious a surrogate as *chicorée* for coffee".

¹⁴ As lexicographer Baker (1993) stated, "It is difficult to understand... why translated texts have been regarded as no more than second-hand and distorted versions of Year texts". She goes on to say that if the texts are translated by native-speakers of the target language, the translation should be viewed as neither inferior nor superior to other communicative events.

make mistakes. Therefore, after reviewing the equivalents provided by the bi-concordancers, the lexicographer must research them further in a unilingual source, such as a target-language corpus or dictionary, to verify the accuracy of the information. While bi-concordancers are not the be-all-and-end-all of corpus lexicography and certainly do not replace the lexicographer's intuition, they do enable lexicographers to tap into information that they might otherwise have missed. Not all the potential equivalents provided by TransSearch will necessarily be included in the BCD. But that is not the point. What is important here is that the use of a bi-concordancer does allow lexicographers to enrich bilingual dictionary entries in a way that was impossible up to now and that, with little additional effort.

4. BIBLIOGRAPHY

4.1 Works Consulted

- BAKER, Mona, (1993). "Corpus Linguistics and Translation Studies - Implications and Applications" in *Text and Technology - in Honour of John Sinclair*. Baker, Mona, Francis, Gill, and Tognini Bonelli, Elena (Eds). John Benjamins, Philadelphia/Amsterdam, pp. 233-252.
- CHURCH, Kenneth Ward and GALE, William A (1991). "Concordances for Parallel Texts" in *The Seventh Annual Conference of the UW Centre for the New OED and Text Research*, Oxford, England.
- HARRIS, Brian, (1988). "Bi-Text: A New Concept in Translation Theory" in *Language Monthly*, No. 54, pp. 8-10.
- LANGLOIS, Lucie (1995). *Rapport sur l'utilisation de TransSearch en lexicographie bilingue*. Technical report, University of Ottawa (langlois@balzac.sti.uottawa.ca).
- MACKLOVITCH, Elliott (1992). "Corpus-based Tools for Translators". *Proceedings of the 33rd Annual Conference of the American Translators Association*, San Diego, California.
- SCHOPENHAUER, Arthur (1992). "On Language and Words" in *Theories of Translation ~ An Anthology of Essays from Dryden to Derrida*. Schulte, Rainer and Biguenet, John (Eds), The University of Chicago Press, Chicago and London, pp. 32-35.
- SIMARD, Michel, FOSTER, George F. and PERRAULT, Francois (1993). *TransSearch: A Bilingual Concordance Tool*, Technical Report, CITI (1575, boul. Chomedey, Laval, Québec, H7V 2X2).
- SINCLAIR, J.M. (Ed) (1987). *Looking up: An Account of the COBUILD Project in lexical computing and the development of the Collins COBUILD English Language Dictionary*. HarperCollins Publisher, London.

4.2 Dictionaries Consulted

- Collins-Robert French-English English French Unabridged Dictionary*, Third Edition, HarperCollins Publishers, 1993, Glasgow.
- French-English/English-French Unabridged Dictionary*, Larousse, 1993, Paris.
- Oxford-Hachette French Dictionary*, Oxford University Press, Oxford, 1994.
- DUNETON, Claude and CLAVAL, Sylvie. *Le Bouquet des expressions imagées*. Éditions du Seuil, Paris, 1990, 1375p.
- REY, Alain and CHANTREAU, Sophie. *Dictionnaire des expressions et locutions*. Les usuels du Robert, Paris, 1979, 946 p.