

Using Bi-textual Alignment for Translation Validation: the TransCheck System

Elliott Macklovitch¹
Centre for Information Technology Innovation

ABSTRACT

We describe the first prototype version of TransCheck, a system for automatically detecting certain types of translation errors that is based on the notion of bi-text, or aligned corpora of translated texts. We analyse the preliminary results obtained from applying TransCheck to five lengthy samples of published translations, discuss some of the problems that currently lie beyond the system's scope, and conclude by briefly sketching our plans for extending the first TransCheck prototype.

1. Introduction

Translators, like all people who draft a lot of texts, will commonly run a spelling or style checker on their translations before delivering them, in the hope of detecting simple kinds of errors they may have overlooked. Because these writing aids are monolingual, however, they are entirely useless in detecting even the most flagrant kinds of translation errors, since these are bilingual in nature and depend on relations between two texts in different languages. This paper will describe a novel kind of writing aid that is specifically designed to detect translation errors, errors like those reproduced in Table 1 at the end of the paper. All the examples listed in Table 1 are authentic: they are taken from the Canadian Hansards, and so the errors they contain were actually committed by translators who are among the most qualified in Canada.²

The errors in Table 1 were detected automatically by a translation checking system called TransCheck. TransCheck is one of a family of new translation support tools being developed at the CITI (see reference [1]), all of which are based on bi-textual alignment algorithms like those originally proposed by Gale and Church [2] and Brown et al. [3]. What these alignment algorithms do, in a word, is to calculate, with a high degree of accuracy, the links between corresponding segments in a source text and its translation, regardless of the size of the two texts.³ Following Harris [4], a corpus of translated texts aligned in this way is often referred to as a *bi-text*. In the references just cited, the aligned segments of a bi-text are generally taken to be sentences, but in principle linkages between larger or smaller units are also possible; and, of course, one sentence in the source can be explicitly linked with two or more sentences in the target, and vice versa.

1. Please see the acknowledgments at the end of this paper.

2. In their defence, it must be said that these translators work under enormous pressure to produce the previous day's parliamentary debates in time for publication the next morning. Under these conditions, and given the volume of texts we scanned, what is really surprising is the relative paucity of the errors we detected in the Hansards.

3. For details on the CITI's current alignment algorithm, which has a success rate of over 98%, see Simard et al. [7].

In a number of recent CITI publications (including [1], [5] & [6]), it has been argued that the emergence of the concept of bi-text allows for the development of a whole new generation of translation support tools. Perhaps the most evident of these is a bilingual concordancing tool, like the CITI's TransSearch system. A standard concordancing program is a kind of database retrieval system that will search for, and display in its context, every occurrence of a queried word or expression as it is found in some database of texts. What a *bilingual* concordancing tool like TransSearch does, in addition, is to display alongside each such occurrence the sentence or sentences that constitute its translation, as previously calculated by the alignment programs. This (conceptually) simple linking of source and target language segments turns out to have tremendous consequences, however: it allows translators to readily access, via TransSearch's elegant graphic interface, all the riches lying dormant in their past production. And past translations, as Isabelle [5] has pointedly observed, contain more solutions to more translation problems than any other available resource.

Now as mentioned, the queries that a user can submit to TransSearch may be either simple words or complex expressions; and they can be either monolingual or bilingual. For example, a user can ask the system to display all the sentences in its database that contain the word "gridlock"; or he can ask to see all the aligned pairs in which the English sentence contains "gridlock" and the corresponding French sentence contains the word "impasse." When the user wants to verify a proposed translation, this is the kind of bilingual query he will usually submit. But as we discovered when we first began experimenting with TransSearch, there is nothing to prevent the user from submitting bilingual requests in which the two elements are *not* proper translations of one another. For example, the user could request that TransSearch display all aligned pairs in which the English member contains the word "library" and the French member contains "librairie". In this case, if any were found, they would constitute examples of a mistranslation, since "library" and "librairie" are deceptive cognates (*or faux amis*, as they are more colourfully called in French): though etymologically related and still morphologically similar, one can never serve as the translation of the other, because they have come to have completely disjoint meanings. (French "librairie" corresponds to English "bookstore.")

Suppose we were to compile an extensive list of such deceptive cognates. Suppose, furthermore, that we could take a draft translation we had just completed and have it automatically aligned with its source text; and that we also had some automatic procedure for applying our list of deceptive cognates to the resulting bi-text as TransSearch batch requests. Suppose, finally, that we had an interface for reviewing and editing the output of these requests. Then what we would wind up with, in effect, is a simple translation checker. Simple, in the sense that it could verify a draft translation for one obvious kind of translation error: deceptive cognates. And in fact, this is exactly how TransCheck was first conceived and developed at the CITI - as an extension of our bilingual concordancing system.

2. Which errors?

Complete deceptive cognates like "library//librairie" are a good starting point from which to launch work on a translation checker, because, by definition, the words that form these pairs can never be mutual translations. But that is not the sole criterion for inclusion in a translation checker; if it were, then the program's bilingual lexicon (or anti-lexicon, as the database of prohibited couples might more accurately be called) would contain an enormous number of entries. Obviously, the number of words that can **never** be mutual translations is infinitely greater than the number of those pairs that are permitted as possible translations. To take but one outlandish example, I am quite confident that the English word "very" can never be translated in French by the word "courgette"; and yet, we wouldn't want to include this pair in TransCheck's anti-lexicon. What distinguishes "library//librairie" from "very//courgette" is the fact that the former pair is far more **likely** to appear in an English-French translation, precisely because the two words' morphological similarity can easily mislead a translator into overlooking their semantic dissimilarity. More generally, the kinds of problems we want our program to check for are just those common types of errors that are most likely to occur in translated texts, or have already been attested by translation revisers as occurring frequently in the texts they proofread.

To help us compile an inventory of such errors for our first TransCheck prototype, we consulted a number of well-known reference works on English-French translation problems, which are listed in the bibliography as [8-12]. In all, we extracted descriptions of over 2800 translation problems, which we transcribed and stored in simple database records that included fields for Incorrect French, corresponding English, Correct French, Error Type, Source, and Comments. Another program then took these static descriptions and converted them into TransSearch batch requests of the following general form:

(i) $e(\text{word}+/\text{cat}) \neq f(\text{word}+/\text{cat})$

where "e(word)" designates one or more words of English, "f(word)" one or more words of French, "cat" designates a morpho-syntactic category, and the optional "+" indicates the base form of the word plus all its inflected variants. The braces signify the optionality of the source language member.⁴

When the e(word) and f(word) are single, legitimate members of their respective vocabularies and are etymologically related, the request corresponds to a deceptive cognate. But of course, they need not always be etymologically related; in the reference works we consulted, we found numerous examples of prohibited translations between pairs that were not similar in form or derived from a common root. Adopting the terminology of one of our authorities, these are called translation improprieties, and they receive exactly the same formal description in our database as the deceptive cognates. "Cabaret", for example, is a proper word of French, but apparently it is not an acceptable translation for the English term "tray"; "cabaret" signifies a kind of bar, and the

4. The "!=" operator does not actually appear in the TransSearch requests, but is included here to highlight the fact that the English - French couples are not possible translations. Several of the database fields mentioned earlier also do not appear in the formal request, but are provided as supplementary information during the TransCheck editing session.

correct French translation for "tray" is "plateau". So if our translation checker found two aligned segments in which the English contained "tray" or "trays", and the French contained "cabaret" or "cabarets", it will in all likelihood have detected a mistranslation - assuming, that is, that "cabaret" is not the translation of some other word in the English sentence. (We return to this question below. For other examples of improprieties detected in the Hansard, see Table 1, (v-vii)).

In working through the references cited in [8-12], we came across other types of translation problems which we decided to include in our TransCheck prototype. For languages that are as closely in contact as English and French are in Canada, illicit borrowings are a very common source of translation errors. For our purposes, a borrowing is defined simply as an English word in a French text, or vice versa. In actual fact, things are not quite so simple: all languages borrow constantly, and in some cases, the authorities are not in agreement as to whether or when a given foreign form has been naturalized. In the vast majority of cases, however, there is no controversy. For words like "lunch", "cool", or the verb "backer", all the translation checker has to do is detect their presence in a French text to be able to flag a borrowing; it is not essential here to determine their lexical category or to even verify the presence of the corresponding form in the aligned English segment.⁵ In fact, one could ask whether it is necessary to add entries for such illicit borrowings to the TransCheck database at all; seeing that they are not part of the French lexicon, a monolingual spelling checker should be able to detect them as unknown words. However, what the entries in the TransCheck database can do that a spelling checker cannot is to provide the user with information on the correct French form which should be used in place of each illicit borrowing. It is one thing to discover that "briefing", for example, is not a proper word of French, and quite another to know what the correct term for this concept is. (Again, see Table 1, (xi-xiv) for more examples of illicit borrowings.)

Calques are another common type of source language interference. These are multi-word expressions that are translated literally from the source language, word for word, in a way that produces an unacceptable result in the target. In some cases, the calqued expression may be perfectly correct in its syntactic form, as well as being semantically transparent, but simply not correspond to the accepted or standardized term in the target language. Thus, "certificat de naissance" is composed of correct French words that are combined in accordance with the standard pattern for French nominal compounds; the only problem is that French uses another term for designating what is called a "birth certificate" in English: "acte de naissance". In other cases - for example, "à la journée longue" - the calqued expression does contravene the rules of French grammar. The temporal adjective "long" does not usually follow the noun it modifies in French, and the correct way of translating the English "all day long" is "à longueur de journée". In both types of calques, just as with the borrowings discussed above, it is not always necessary to verify the presence of the corresponding English phrase in the aligned segment for TransCheck to be able to

5. This is not to suggest that the detection of borrowed forms is always straightforward. Among the complications we encountered, there is the problem of generating the morphological variants of non-French forms like "backer"; as well as the problem of borrowed forms that happen to coincide with a perfectly legitimate TL word, e.g. "pin" used incorrectly in French to signify a broach, versus the correct French word "pin", which refers to a type of tree.

flag an error. In fact, as a general rule, whenever the problematic word or expression can be unambiguously identified without recourse to the aligned segment in the other text, monolingual queries are probably preferable, since there is always a risk, in a relation as free as translation, of the specified expression not being present in the source language, and hence of an error being omitted as the result of an unmatched query. (See Table 1, (viii-x) for more examples of caiques.)

3. Preliminary Results

In order to test our first TransCheck prototype, we applied it to five 100-thousand word samples of translated text: three taken from the Canadian Hansards and two from operational manuals used at the Department of National Defense.⁶ Of course, TransCheck is intended to be used on *draft* translations, and in all these cases, the samples being validated were published and polished translations. Therefore, one cannot attach a great deal of significance to the absolute number of errors detected; no more than one would consider evaluating the usefulness of a spelling checker from the results obtained by running it on published monolingual texts. Moreover, this was a first prototype we were developing, in which we wanted to evaluate a certain number of hypotheses; for example, the possibility of identifying all multi-word expressions using only monolingual queries. The retention of such hypotheses also introduces an element of distortion into the data on TransCheck's performance.

Still, it was important for us to get some idea of the kinds of errors TransCheck could detect, as well as the level of noise it produced - if only to be able to assess the general viability of a translation checker that uses a **sentence alignment** program to detect prohibited **word correspondences**. Table 2 below summarizes the results of TransCheck's output on the five 100-thousand word test corpora.

Table 2:

Corpus	total hits	true errors	noise	success rate
Hansard 1	63	60	3	95%
Hansard 2	33	28	5	85%
Hansard 3	50	48	2	96%
DND 1	76	52	24	68%
DND 2	60	31	29	52%

The first thing to notice about these results are the impressive success rates that TransCheck posted on the three Hansard corpora and, to a lesser extent, on the first DND corpus. What this tells us, in other words, is that the great majority of potential errors flagged by TransCheck in these

6. One of the DND manuals deals with the tactical employment of infantry battalions, the other with the training and employment of snipers.

corpora turn out to be "true" errors - in the sense that they correspond precisely to the language problems described by our authorities. The reasons for the system's poorer performance on the DND manuals will be discussed shortly. What needs to be stressed here, however, is that of all the cases of noise in the five corpora, no more than one or two are attributable to the coarseness of the system's alignments, and none to misaligned source and target language sentences. This confirms the earlier results reported in Isabelle et al. [1] and would seem to justify the claim made there that a translation checker based on a sentence alignment program and a part-of-speech tagger may well provide a sufficient platform for real-life applications.

There are several other aspects of the results in Table 2 that merit comment; among them, the significant drop in the total number of hits between Hansard1 and Hansard2. This was caused by the fact that we neglected to specify English as the source language on the first Hansard sample; so that a large number of the errors detected in that corpus actually appear in the speeches of French Members of Parliament. Since we were more interested in the errors introduced by French translators than in those produced by French-speaking MPs, we set the source-language parameter to English for the remaining four test samples.⁷ The total number of hits in those samples would be considerably higher if the source language parameter was left unspecified.

Another constant that emerges from the analysis of the test samples is this: the majority of the errors detected in a given text derive from a relatively small number of problems described in the TransCheck database. Or put another way, most of the queries submitted by the system find no match in the translated text. We mentioned that the current TransCheck database contains about 2800 entries; in each of the final four test samples, no more than 25 of these produced hits (although these weren't necessarily the same queries from one corpus to the next; and, of course, a given query could produce several matches within one text.) Even when TransCheck is applied to rougher draft translations, I would not expect this situation to change significantly. The reason has to do with the nature of the TransCheck database, which seeks to describe as many of the most likely and attested translation errors as possible, regardless of domain or text type. Any given translation, however, will normally contain only a fraction of these; and since even in the roughest of drafts, many more segments are rendered correctly than incorrectly, it is only natural that most of the TransCheck queries will go unmatched.

Of the relatively small number of TransCheck entries that do result in matches, there are a few in each sample corpus that tend to reoccur, sometimes quite frequently. In the Hansards, for example, the term "caucus" is repeatedly flagged, and in the DND manuals, the pair "deception//*déception*". These are not slips, then; their repetition suggests a serious difference of opinion between at least one of our translation authorities and the translators who actually produce these texts. And in fact, if we consult other references (like TERMIUM, for example), we find that both these terms are accepted as being correct. This kind of divergence between authorities raises a thorny problem for the development of an automatic translation checker, and the attitude we have

7. While in principle, a translation checker may be bi-directional, TransCheck has a definite English-to-French bias, due principally to the fact that the reference works we consulted focus almost exclusively on problems of French usage.

had to adopt may be summarized as follows: it is not up to us to decide. We do not claim to be experts on correct French usage, much less prescriptivists, and we take no position on the errors identified by the authorities we consulted. For every error that is currently flagged by the system, a reference is provided to the book in which that problem is described. If eventual users do not agree with that authority's opinion, then clearly they must be able to neutralize the TransCheck entry, in the same way that the users of style checkers can deactivate certain rules or whole rule classes. More generally, TransCheck is a tool that will help enforce translation norms - in *the plural*; because there surely is no single translation standard that applies across-the-board to all domains and types of text. For this reason, it will certainly be more difficult for the developers of a translation checker to provide their users with a core database than it is, say, for the developers of a spelling checker. But one thing seems certain: if working translators are ever to accept such a system, they will have to be able to modify and customize the contents of its database so that it reflects their own translation standards, or that of their various clients.

As it turns out, the principal source of the noise in the TransCheck output for the two DND samples can be attributed to a small number of domain-dependent examples like "deception", which, incidentally, is proscribed by all our authorities as a deceptive cognate and yet has been standardized by the Canadian Armed Forces as the correct term for describing the action of deceiving or misleading the enemy.⁸ The other major source of noise can be traced to certain inadequacies in TransCheck's tokenizing and lexical modules: for example, the abbreviation "s/off" (for "sous-officier", or NCO) is incorrectly split into two units and "off" flagged as a borrowing. We should be able to correct these problems without too much difficulty.

4. Extending the current prototype

The data in Table 2 tells us something about the quality of TransCheck's output, in terms of the ratio of noise to true errors detected. Unfortunately, we do not as yet have any data on the system's silence, or the proportion of true errors TransCheck fails to detect. In order to obtain such data, we would want to run the system on a benchmark translation in which all errors had previously been identified by an expert; this would allow us to tabulate how many of these the system was able to detect and how many it omitted. Still, from the months of preparatory work that went into developing the first prototype, and from the analysis of the preliminary results described above, we do know (albeit, informally) that the current version does incorrectly flag some segments that contain no errors; and, more importantly, that there are many types of translation problems that cannot be captured within its rudimentary descriptive model.

As mentioned above, we are aware that aligning two texts at the sentence level may not always be adequate; occasionally, the aligned sentences will contain the prohibited pair of words or expressions, but the two will not be mutual translations.⁹ It's obvious what must be done to

8. Another example is provided by the pair "instructor//instructeur", again permitted only in the military domain.

9. See Table 1, (xv) for an example of such noise. The prohibited pair in this case are the nouns "grant//octroi", since the latter can only designate the action of granting, not the thing granted. The problem is, the French sentence also contains the noun "subvention", which is a proper translation for "grant".

resolve this kind of problem: we need to obtain finer bi-textual alignments, ones that go beyond the sentence to the phrase and word level. Progress is being made in this direction, both at the CITI and elsewhere. (See, in particular, Dagan et al. [14])

Next, consider the numerous translation problems whose description requires more contextual information than the basic part-of-speech tagging that TransCheck currently provides. Take, for example, the French verb "débuter": it can be used to translate the English verb "start", but only when it is intransitive. For TransCheck to be able to detect just those instances of "débuter" that are used incorrectly, therefore, the system would have to be extended to include some kind of syntactic analyser capable of distinguishing such valency patterns. And of course, exactly the same reasoning applies to semantic constraints on argument types. The French verb "tourner", for example, can only take direct objects that belong to the class of ships, according to one of our authorities, unlike its English cognate, which permits all kinds of vehicles. Again, if TransCheck is to be able to distinguish the correct from the incorrect occurrences of the French verb, then obviously it will have to incorporate some kind of semantic analyser.

Indeed, if Martin Kay is right in his assertion that "there is nothing that a person could know, or feel, or dream, that could not be crucial for getting a good translation of some text or other,"¹⁰ then in principle there is no limit to the linguistic and extra-linguistic knowledge that a translation checker would have to access in order to be able to detect **all** possible translation errors. To revise a translation requires no less knowledge than to produce a first draft; and a translation checker that aims to fully emulate a human reviser will have to attain an advanced level of understanding that goes well beyond the literal text to encompass all the intelligence that a qualified translator brings to his task. Such a system, quite obviously, is not about to burst upon the market, and undoubtedly will not appear before we see the coming of fully automatic high quality machine translation. There is, however, one major difference between partial approximations of an ideal translation checker, like the modest TransCheck prototype described in this paper, and a partial approximation of an ideal machine translation system. A less than exhaustive translation checker, like an incomplete spelling checker, can still render valuable service: in both cases, those errors that the system does manage to automatically detect will improve the quality of the final text. It is not at all obvious that current MT systems, which necessarily function on an incomplete understanding of the text to be translated, can render comparable service to a human translator. A partial approximation of a translation that is generated by an MT system will often require more time and effort to clean up and knock into shape than it takes for a human translator to produce a correct version from scratch. See Isabelle [15], where this argument in favour of translator tools over classic MT is developed in more detail.

One question that might legitimately be raised in regard to the relatively straightforward examples of syntactic and semantic complementation mentioned in the preceding paragraph is whether these problems require a bilingual translation checker to be detected, or whether a unilingual French grammar checker might not do just as well. After all, when the authorities label

10. From the Foreword to Hutchins and Somers [13].

the transitive use of "contribuer" an "anglicisme", for example, what they are really doing is informing us about how the problem arose: as a result of French speakers borrowing the English verb pattern. Whatever its origin, however, it remains essentially a problem of French usage whose automatic detection does not absolutely require recourse to the source of a translation; one could easily imagine exactly the same problem arising in a monolingual French document, where no translation was involved. In fact, it is not entirely unnatural to consider all the problems handled by TransCheck from the point of view of a unilingual target language checker, even the false cognates like "library//librairie." In this view, the French word "librairie" has two meanings, a correct and an incorrect one. When the word appears in a translation, we can use the source text to help distinguish between the two: if "librairie" translates the English word "library", i.e. when it has the same meaning as that word, it is being used incorrectly. But in essence, the source text is just one way of getting at the word's meaning, and other ways are imaginable.

In practice, however, the question raised above is somewhat moot.¹¹ If we had a complete and calculable generative grammar of some language at our disposal, then perhaps we could use it to automatically validate the correctness of any text in the language in terms of the well-formed forms which that grammar enumerates, and the complement of that set, the ill-formed ones.¹² Needless to say, no such grammar exists; and until it does, perhaps the best we can do is to explicitly describe some of the errors that frequently appear in the language. TransCheck attempts to do just that, particularly for those errors that are likely to arise as a result of the contact between two languages; and for each such error, it proposes a probable correction.

The extensions to TransCheck's linguistic model alluded to above will require a substantial research effort, especially if the results of the syntactic and semantic analysis are to achieve anything like the accuracy and robustness of the part-of-speech tagging that the system currently employs. In the shorter term, there are less far-reaching extensions to the system that we are planning to implement. These include the automatic verification of numerical expressions between source and target texts;¹³ the flagging of major unit omissions in the translated text (i.e. sentences and paragraphs); as well, perhaps, as the verification of terminological consistency. These will be more fully documented in forthcoming updates on the TransCheck project.

11. For one thing, the commercial French grammar checkers that I am aware of currently verify little more than grammatical agreement, e.g. between a verb and its subject; and their performance even here is far from stellar.

12. Though even this grammar couldn't help us detect the sentences of the language that *are* well-formed, but don't correctly convey the author's intended meaning.

13. See Table 1, (xvi) for one example of this type of error which we later hope to be able to detect with TransCheck.

5. Conclusion

This paper began by asserting that the advent of bi-text has made it possible to envisage a new type of writing aid for translators, one which is capable of detecting translation errors that are inherently beyond the scope of monolingual spelling or grammar checkers. In the course of the discussion, this claim has been qualified somewhat. We have seen several types of translation errors, including illicit borrowings and ungrammatical calques, which could in principle be detected, though perhaps not easily corrected, by a monolingual writing aid. For other types of translation problems, in particular those in which the target language text remains grammatically correct but does not convey the same meaning as its source language counterpart, a bi-textual validation tool like TransCheck seems to offer the best solution currently possible. We have argued that even a partial approximation of such a translation checker could be useful to the human translator or reviser, and we have sketched our short and long term plans for extending the capabilities of the first TransCheck prototype.

Acknowledgments

The TransCheck system described in this paper is the joint product of a number of people who work at the CITI, whose contributions I want to acknowledge here. They are Pierre Isabelle, Director of the CITI's machine-aided translation group, Monique Chevalier, George Foster, Marie-Louise Hannan, Francois Perrault, Xiaobo Ren, Michel Simard and Caroline Viel. I express my gratitude to all these colleagues for the excellent work they have done. None is of course responsible for any errors I may have inadvertently introduced in this report

References

1. Isabelle, Pierre; Dymetman, Marc; Foster, George; Jutras, Jean-Marc; Macklovitch, Elliott; Perrault, Francois; Ren, Xiaobo; and Simard, Michel. *Translation Analysis and Translation Automation*, in **Proceedings of the Fifth International Conference on Theoretical and Methodological Issues in Machine Translation**, Kyoto, 1993.
2. Gale, William; and Church, Kenneth. "A Program for Aligning Sentences in Bilingual Corpora," in **Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics**, pp. 177-184, Berkeley, California, 1991.

3. Brown, P.; Lai, J.; and Mercer, R. "*Aligning sentences in parallel corpora*," in **Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics** , pp.169-176, Berkeley, California,1991.
4. Harris, Brian. "*Bi-text: A New Concept in Translation Theory*," in **Language Monthly**, no. 54,1988.
5. Isabelle, Pierre. "*Bi-text: Toward a new Generation of Support Tools for Translation and Terminology*," CITI Technical Report, 29 pages, 1992. Also published in French in **Meta**, no. 37:4, pp.721-737,1992.
6. Macklovitch, E. "*Corpus-Based Tools for Translators*," in **Proceedings to the 33rd Annual Conference of the American Translators Association**, pp. 317-328, San Diego, California, 1992.
7. Simard, Michel; Foster, George; and Isabelle, Pierre. "*Using Cognates to Align Sentences in Parallel Corpora*," in **Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation**, pp. 67-81, Montreal, 1992.
8. Colpron, Gilles. **Dictionnaire des anglicismes**, Éditions Beauchemin Ltée, Laval (Qué.), Canada, 1982.
9. Dagenais, Gérard. **Dictionnaire des difficultés de la langue française au Canada**, Les Éditions Françaises Inc., Boucherville (Qué.), Canada, 1984.
10. De Villers, Marie-Éva. **Multidictionnaire des difficultés de la langue française**, Éditions Québec/Amérique, Montreal, Canada, 1988.
11. Rey, Jean. **Dictionnaire sélectif et commenté des difficultés de la version anglaise**, Éditions Ophrys, Paris, 1984.
12. Van Roey, Jacques; Granger, Sylviane; and Swallow, Helen. **Dictionnaire des faux-amis français-anglais**, Duculot, Paris, 1988.
13. Hutchins, John; and Somers, Harold. **An Introduction to Machine Translation**, Academic Press, 1992.
14. Dagan, Ido; Church, Kenneth; and Gale, William. "*Robust Bilingual Word Alignment for Machine Aided Translation*," in **Proceedings of the Workshop on Very Large Corpora**, Columbus, Ohio, 1993.
15. Isabelle, Pierre. "*Machine-Aided Human Translation and the Paradigm Shift*," in **Proceedings of the Fourth Machine Translation Summit**, Kobe, Japan, 1993.

Table 1:

	Source	Translation
(i)	... the overall long term survival of any type and stage of childhood cancer le taux de survie à long terme, pour tous types et à tous les stades de cancer ... (19/05/92)
(ii)	... to ensure progress toward a world free of injustice, discrimination and prejudice.	... notre recherche d'un monde exempt d'injustice, de discrimination et de préjudice. (10/12/91)
(iii)	Since the passage of that infamous bill in 1987, Bill C-22 ...	Depuis l'adoption de l'infâme projet de loi C-22 en 1987... (17/11/92)
(iv)	In this country ... with our reputation for affluence, with out reputation for humanity dans notre pays reconnu ... pour son affluence et son empressement à aider ... (10/05/93)
(v)	Also, the polls say that regardless of the position of the Tories in England before and during the campaign – and it was up and down and neck and neck for awhile...	... quelle qu'ait été la position du Parti Conservateur en Angleterre avant et durant la campagne, et on peut dire que la position des partis a beaucoup fluctué, que ces derniers ont été nez à nez pendant un certain temps... (10/04/92)
(vi)	They ... drive their children to practices and games.	Ils ... conduisent leurs enfants à des pratiques et à des joutes. (12/02/93)
(vii)	It is about \$190 going over \$200 depending upon how much money they earned ...	Il est d'environ 190\$ et peut dépasser les 200\$, dépendant du montant gagné ... (15/02/93)
(viii)	... the destruction of private property cannot be tolerated under any circumstances.	... la destruction de la propriété privée ne peut être tolérée en aucun temps. (5/05/92)
(ix)	... profits from a used equipment sale recettes d'une vente d'équipement de seconde main ... (4/06/90)
(x)	...his interest rate projections ... are on a calendar year basis.	... les taux d'intérêt sont basés sur l'année de calendrier. (4/04/90)
(xi)	It is also an essential condition for the confidence ... that will ensure a strong recovery.	Il s'agit également d'un prérequis au retour de la confiance ... (25/02/92)
(xii)	You are resting on technicalities.	Vous vous arrêtez à des technicalités. (12/03/92)
(xiii)	... many of whom were historic supporters of the Conservative party dont beaucoup ont toujours été des supporters du Parti conservateur ... (21/03/91)
(xiv)	A real rip-off occurred during the long weekend.	Lorsqu'il y a des longues fins de semaine, on assiste à un véritable racket ... (25/05/93)
(xv)	Canada Employment has recently announced a section 25 grant to help it upgrade its facilities ...	Emploi Canada a récemment annoncé l'octroi à cette entreprise d'une subvention en conformité de l'article 25, afin de l'aider ... (27/04/93)
(xvi)	The European Economic Community's research and development project has a budget of \$17 million spread over four years ...	Le projet de recherche et de développement de la Communauté économique européenne bénéficie d'un budget de 70 millions de dollars réparti sur quatre ans... (17/04/86)