Bi-Textual Aids for Translators

Pierre Isabelle

CITI
1575 Chomedey Blvd, Laval, Quebec, H7V 2X2
e-mail: isabelle@citi.doc.ca

ABSTRACT

While machine translation can successfully tackle some highly restricted sublanguages, it is in most cases more productive to turn to support tools for human translators. The functions taken over by existing translator's workstations are rather peripheral with respect to the core aspects of the translation task. However, recent developments show that it is possible to automatically produce explicit (partial) representations of the translation correspondences that link pairs of source and target texts. These representations called bi-texts provide the foundation required for the design of support tools that delve deeper into the realm of translation proper, such as: a) a translation memory that can be accessed by various means, including bilingual concordancing; b) translation critiquing tools capable of detecting correspondence errors such as omissions or deceptive cognates; and c) translator-oriented speech recognition systems capable of taking advantage of correspondence contraints with respect to source texts. The outlook for translation support tools is thus highly promising.

1. Introduction

Despite several decades of massive efforts, high-quality machine translation (MT) is still only possible in the case of some very restricted sublanguages such as the one tackled by the TAUM-MÉTÉO system (Isabelle [17]). Moreover, the fact that the resounding success of this system has not been systematically cloned seems to indicate that there are very few simple sublanguages around for which there exists a significant translation volume.

Thus, with the exception of a handful of cases, the current situation is no different from what it was back in 1951, when Bar-Hillel [1] wrote:

> "For those targets in which high accuracy is a *conditio sine qua non*, pure MT has to be given up in favor of mixed MT, i.e., a translation process in which a human brain intervenes. There the question arises: Which parts of the process should be given to a human partner?" (p. 230)

Bar-Hillel's own preference went to approaches in which the human would intervene either before ('pre-editing') or after ('post-editing') the mechanical process, "but preferably not somewhere in the midst of it". The core part of the translation process is still left to the machine.

After four decades of stubborn attempts, the case against extensive pre-editing and post-editing in MT has become overwhelming. On the one hand, no one has yet come up with any kind of practical way of pre-editing general texts so as to guarantee consistently good output in the ensuing MT process. On the other hand, it has repeatedly been demonstrated that it is not cost-effective to resort to human post-editors to salvage the kind of low quality output that current MT systems produce in most situations (see for example Macklovitch [23]). It is therefore only natural that most translation services consider current MT technology as useless, and that MT accounts for only a very marginal share of the translation market. There is no evidence that this situation is about to change.

More than ten years ago, Martin Kay [18] proposed his *translator's amanuensis*, which constitutes a very different answer to Bar-Hillel's question about the optimal division of labor between man and machine:

> "I want to advocate a view of the problem in which machines are gradually, almost imperceptibly, allowed to take over certain functions in the overall translation process. First they will take over functions not essentially related to translation. Then, little by little, they will approach translation itself. The keynote will be modesty. At each stage, we will do only what we know we can do reliably. Little steps for little feet!" (p. 11)

Rather than start from inadequate systems and ask translators to compensate for their flaws, one starts from human translation and looks for ways, however modest, to make machines helpful. It is this down-to-earth approach that the Canadian Workplace Automation Research Center (CWARC) chose to pursue when it started its *translator's workstation* project, back in 1987 (Macklovitch [21], [22]). In its most recent incarnation, the CWARC's workstation provides the translator with a windowing environment where he/she has simultaneous access to a number of tools such as split screen word processing, spelling correction, terminology and dictionary lookup, file comparison, word counting, etc.

This workstation, like most others currently in existence, is still in that early stage of development where most of the functions taken over by the machine have more to do with office automation than with the core aspects of the translation task. Even so, the results obtained with the workstation at the Canadian Translation Bureau, where it has been in use for about three years, have been much better than with MT systems. Workstation users are proud enough of the tool to want to keep it!

Following Kay's proposed scenario, we can now take advantage of this office automation base to provide translators with new tools that will delve deeper into the realm of translation proper. In order to do this, we need some kind of conceptual scheme that provides suitable entry points for technology. We believe that the concept of *bi-text* does provide such as scheme, and opens up a whole range of new possibilities for translation support. In section 2, we introduce this concept. In section 3, we describe how bi-textual representations can be automatically generated. Then, in sections 3, 4 and 5 we explore three different kinds of tools which can be seated onto these representations.

2. The concept of bi-text

What is the single most important characteristic that sets translators apart from other language workers? The obvious answer is that translators work with not one but two texts: a pre-existing source text (ST) and a target text (TT) to be produced in a different language, with the constraint that ST and TT stand in a relation of translational equivalence. Ensuring that this constraint is met constitutes the very crux of the translator's task. Consequently, one would expect translation-specific tools to incorporate some knowledge of translational equivalence.

For translation to be possible at all, translational equivalence must be *compositional* in some sense; that is, the translation of a text must be a function of the translation of its parts, down to the level of some finite number of primitive equivalences (say between words and phrases). Multilingual dictionaries and terminology banks are meant to capture some of these primitive equivalences between different languages. They currently constitute the best examples of translation-specific tools that are available in existing translator's workstations.

However, anyone who has ever tried to translate natural language texts will acknowledge that even with the best existing dictionaries and term banks, translation remains a difficult task. Such lexical resources only describe virtual equivalences. Generally speaking, they enumerate several possible TL equivalents for each SL element, and it is up to the translator to select the right one for his text, according to various contextual factors. Moreover, lexical resources are always incomplete: they invariably fail to exhaust the full range of virtual equivalences.

The only place where one can look for actual equivalences (that is, correspondences) is in existing translations. Thanks to the compositionality principle, the global correspondence between a text ST and it translation TT is normally analyzable into sets of finer correspondences between particular segments of ST and particular segments of TT. As noted by Harris [15], the traditional 'side-by-side' or 'interlinear' layouts commonly used for translations do presuppose a straightforward analyzability of translational correspondences (paragraph-to-paragraph, sentence-to-sentence, etc.). When asked to, any bilingual speaker will be able to point out many if not all of the correspondences between the elements of a source and its translation.

Harris [15], [16] suggests the term *bi-text* to designate any scheme which makes such correspondences explicit. We adopt this term with the following technical definition: a bi-text is quadruple $<T_1, T_2, Fs, C>$ in which T_1 and T_2 are two texts, Fs is a function that analyzes T_1 into some set of elements $Fs(T_1)$ and T_2 into some set of elements $Fs(T_2)$, and C is a subset of the cartesian product $Fs(T_1)$ x $Fs(T_2)$.

This definition raises several important issues. One of them has to do with the nature of the elements produced by the analysis function Fs. One fairly obvious possibility is for Fs to be some kind of syntactic analysis function. In that case Fs will presumably organize each of the texts into some kind of hierarchical structure: texts are made up of sections, sections of paragraphs, paragraphs of sentences, sentences of phrases, phrases of words and words of morphemes.

A related issue is the nature of the correspondence function C. We mentioned that this function must have some degree of compositionality. A hierarchical analysis function lends itself naturally to this requirement, since its output constitutes a natural domain for hierarchical translation correspondences: the translation of a section is made up of the translations of its component paragraphs, the translation of a paragraph is made up of the translations of its component sentences, etc, as in Figure 1.

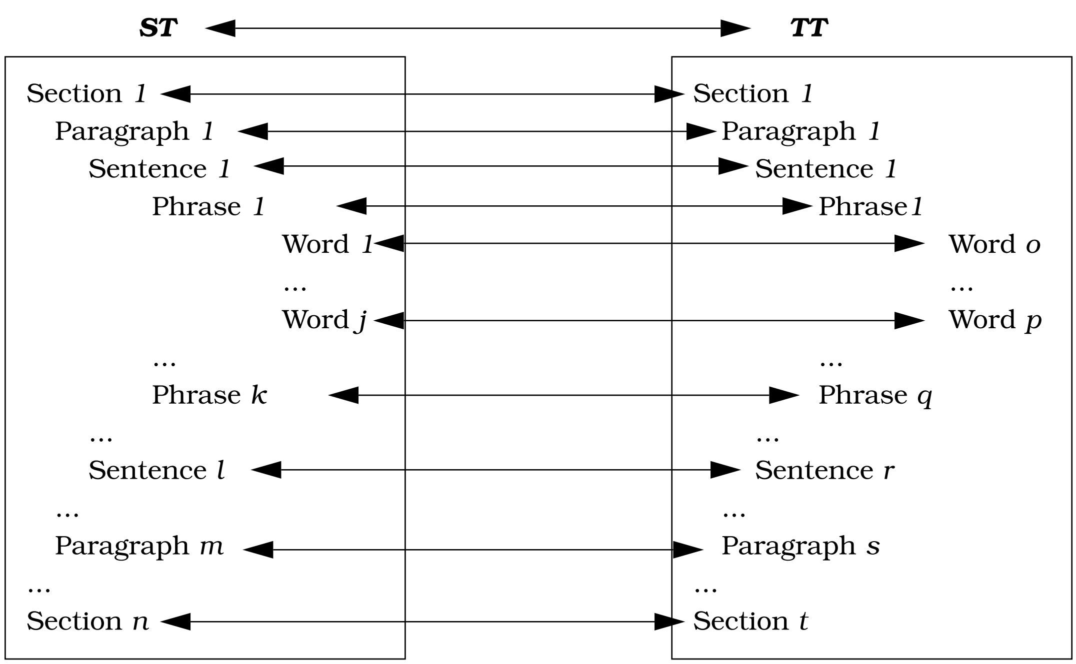

Figure 1: Hierarchical correspondences between source and target texts

However, it is obvious when examining translations that translational correspondences are not always simple one-to-one mappings. This is especially true in the case of lower-rank units. For example, the English word *potato* is usually translated as *pomme de terre,* a sequence of three French words. Hierarchical models can easily deal with cases like this one, since they provide abstract phrasal units between which correspondences can be established: the noun group *pomme de terre* corresponds to the simple noun *potato*.

But it is easy to find cases in which the superficial syntactic structure does not bring out all of the units that are involved in translational correspondences. For example, the discontinuous sequence *ne...pas* is not represented as a unit at that level, but it participates in a translational correspondence with *not* in sentence pairs such as the following:

(1) a) Max n'a pas vu Conrad.

b) Max has not seen Conrad.

Similarly, the discontinous sequence *turn...on* is a unit translated as *alluma* in the following other sentence pair.

(2) a) Max turned the radio on.

b) Max alluma la radio.

Generally speaking, it appears that some translational correspondences can only be expressed through the kind of abstract representations (deep syntactic representations, semantic representations, conceptual representations, etc.) that MT systems need to appeal to for the production of translations. Our knowledge of such representation schemes is far from complete, and it is notoriously difficult to develop algorithms capable of mapping unrestricted natural language texts onto them. Therefore, the construction of a device capable of automatically producing complete bi-texts (that is, bi-texts expressing all of the translation correspondences) for arbitrary pairs of SL/TL texts may prove to be a very hard problem.

Still, there are reasons why the outlook for bi-text production is much brighter than it is for MT systems. First, in contrast with the active linguistic capability required for the production of translations, the reconstruction of translation correspondences in existing translations requires only a passive linguistic capability, which should in principle be easier to characterize. Another reason is that while translation users generally require complete translations, partial bi-textual representations (that is, representations that express only a subset of all the correspondences between ST and TT) can still be very useful, as we will see below. In terms of the kind of hierarchical model discussed above, bi-textual representations can be ranked in terms of a *resolution* criterion. Very low resolution bi-texts will only show correspondences between the highest rank units, such as sections or paragraphs.[1] Resolution increases as we further specify correspondences between lower-rank units: sentences, phrases, words and morphemes. There is no analogue in translation production: one cannot translate a unit (say a sentence) without at the same time translating its components (phrases, words).

Resolution can be taken as a parameter in evaluating particular bi-texts. Another obvious evaluation parameter is *precision*: the proportion of the purported correspondences that are factually correct.

3. Generating Bi-Textual Representations

3.1 Sentence Alignment

To my knowledge, Martin Kay was the first researcher to propose methods for reconstructing correspondences in pre-existing translations, thus enabling the automatic production of what we have called bi-textual representations. It was back in 1984 that I first heard Kay informally outline an algorithm which was later systematically described in Kay & Röscheisen [19]. This algorithm does not aim at discovering all correspondences, but only at producing a correct 'alignment' at the sentence level. What makes the problem difficult, of course, is that the correspondences need not be one-to-one: a sentence of ST can be 'expanded' as two or more sentences in TT; and conversely, several sentences of ST can be 'contracted' into a single one in TT.

Kay & Röscheisen's algorithm proceeds by looking at all possible sentence alignments[2], and selecting the one which maximizes the number of systematic word correspondences that can be hypothesized. For example, suppose that ST contains exactly 10 occurrences of *dog* and TT contains exactly 10 occurrences of *chien*. Then, all other things being equal, sentence alignments which pair occurrences of *dog* and *chien* will be favored. Because sentence and word alignment are mutually dependent, Kay & Roscheisen's algorithm is based on an iterative refinement process.

The authors claim near perfect results on their test corpora (two articles of *Scientific American* with their German translations, and 1000 sentences from the Hansard English/French data). One interesting feature of this approach is that it does not appeal to any evidence external to the texts themselves, such as for example a bilingual dictionary.

However, Catizone & al. [5] claim that when Kay & Röscheisen's method is extended so as to include the use a bilingual dictionary for guiding the initial hypotheses on word correspondences, the search space is drastically reduced.

---

1. Side-by-side translation layouts typically link whole paragraphs only.
2. The range of possible alignments is constrained from the start. Crossing alignments are prohibited and many-to-many alignments are only permitted provided 'many' does not exceed some small n.

Debili & Sammouda [9] propose an alignment algorithm which is different but also based on word correspondences established with the help of a bilingual dictionary.

Brown, Lai & Mercer [3] and Gale & Church [12] address in a different way the same problem of aligning the sentences of parallel texts. They both propose methods which are based on the simple observation that the length of a text and the length of its translation are highly correlated. These methods also have the advantage of using no external evidence. From the computational point of view, they are much less expensive than Kay & Roscheisen' s, and they do surprisingly well on the Hansard data. However, since they do not look at the contents of the sentences that they pair, these methods appear to be less reliable and less robust. Once a length-based algorithm has accidentally misaligned two sentences, it tends to misalign the remainder of the paragraph.

Simard, Foster & Isabelle [26] look at yet another criterion on which to base sentence alignments. They observe that 'cognateness', that is, the proportion of cognate words, is highly correlated with translation. Reasonably reliable operational approximations of the notion of cognate word can easily be defined (e.g. in terms of shared prefixes), and cognateness can therefore be tested without using any external evidence. The authors report that when used as the sole criterion, cognateness does not produce very good results. However, they claim that when used in conjunction with the length criterion, cognateness improves precision and robustness without drastically increasing the computational cost.

3.2 Word Correspondences

As we have seen, some sentence alignment algorithms (Kay & Röscheisen's, Debili & Sammouda's) work by hypothesing some of the word level correspondences in the source and target texts. These word correspondences could perhaps be used to produce higher-resolution bi-texts, but the authors say very little on their coverage and precision.

Brown, Lai & Mercer's sentence alignment mechanism is only the first step of a procedure aimed a estimating the parameters of the stochastic MT system described in Brown & al. [2]. In order to estimate the parameters of a probabilistic transfer dictionary for this system, they then need to make explicit the word correspondences found in their corpus of sentence pairs (a portion of the Hansard data). They do this by means of a particular version of the EM algorithm (Dempster & al. [10]), which should allow them to obtain complete coverage. However, the authors do not discuss the level of precision of their results.

Gale & Church [13] introduce a method for identifying some of the word correspondences in texts that have already been aligned at the sentence level. They first determine a set word pairs that are strongly associated in the sentence pairs. This is done by applying a χ^2-like statistic to two-by-two contigency tables, and selecting word pairs for which the association is above some threshold. They then use these pairs to mark likely word correspondences in their sentence pairs. The authors claim that when they set the relevant thresholds so as to obtain a coverage of 60%, the correspondences are correct in 95% of the cases.

3.3 Conclusions

Even though the field of investigation is very recent, it is already possible to automatically produce high-precision low-resolution bi-texts out of pre-existing translations. Although there is still plenty of room for improvement in speed and/or precision, source and target texts can be matched reasonably well down to the level of their component sentences. It is also possible to calculate word correspondences, but it seems that for the moment one has to compromise either on the coverage or the precision. To our knowledge, the problem of phrase-to-phrase or word-to-phrase correspondences has yet to be addressed, not to mention phenomena like discontinuous constituents. But since research on parallel texts is now receiving more and more attention, we can expect to see some rapid progress in these areas.

We will now argue that these developments are of great significance for the future of translator's aids. More specifically, we will claim that they open up the way for at least three types of entirely novel tools oriented towards: 1) translation memory; 2) translation critiquing; and 3) translation dictation.

4. A Corporate Memory for Translation Services

Most translators are routinely faced with difficult translation problems for which existing resources such as dictionaries and term banks provide no ready answer. One would wish that once a solution has been worked out for some problem, it remained available for future reference either by the same individual or his colleagues. Unfortunately, this is by no means the case at this time. Typically, large translation services cannot even guarantee that they will not retranslate from scratch a document that they have already translated before.

Given the staggering volume of translations produced year after year, it is quite obvious that **existing translations contain more solutions to more translation problems** than any other existing resource. Unfortunately, translators can currently derive very little benefit of this fact. In most cases, previous translations are only archived in hardcopy. Even in the few cases where source and target texts are available in machine-readable form, the translators are not equipped with tools capable of efficiently extracting useful information from such archives.

Suppose now that a translation service systematically organises its production into a bi-textual database. By definition, in such a database ST segments are linked with their TT translations. In particular, segments that constitute translation problems are linked with the solutions that were devised for them. Clearly, what this means is that the translation service is now equipped with a structured translation memory.

There are many possible ways to exploit such a corporate memory. In a long-term perspective, some researchers have started exploring the idea that bi-textual databases would provide the foundation for *memory-based* or *analogy-based* or *example-based* approaches to the MT problem (see for example Sato & Nagao [25]).

A less ambitious approach would be to develop systems that, during a manual translation, will automatically retrieve relevant examples in the database, and let the translator decide whether or not he/she will use them. Some commercially available systems such as ALPS TSS and UNITRAN already incorporate some elements of this approach.[3]

Finally, an even less ambitious, but perhaps more universally useful approach is to provide translators with tools that allow them to search the bi-textual database at will. It has already been suggested by several authors that a tool capable of producing *bilingual concordances* would be useful to bilingual lexicographers (see Klavans & Tzoukermann [20], Catizone, Russell & Warwick [5], Church [7]). It is rather obvious that bilingual concordancing would also be useful to translators. For example, upon encountering some occurrence of an expression like *to be out to lunch* or *to add insult to injury* in his English source text, a translator might be hesitant as to an appropriate French equivalent. He/she might also find out that conventional bilingual dictionaries do not provide satisfactory answers. With a bilingual concordancing tool, he/she could then search a bi-textual database in order to retrieve examples of these expressions together with their translations. See Macklovitch [24] for a more detailed discussion of this issue.

Appendix A contains a screendump of the results produced by OCTA, the CWARC's prototype bilingual concordancing system, of a search for English segments containing the discontinuous sequence *insult...injury* in a database consisting of a sentence-level alignment of the 1986 Hansard data. After examining a few examples like this one, translators usually conclude that bilingual concordancing would be very useful to them.

5. Translation Critiquing Tools

Of course, not all translations are equally good. Hence translators will have to exert some caution when they use translation memory facilities. Interestingly, it

3. These systems are not based on the automatic generation of bi-textual representations. Instead, they resort to special-purpose word processors in which the translator will at any time make explicit what segment he/she is translating. They then use this information to create a database of segment pairs which will later be searched for (nearly-) identical source segments.

turns out that the bi-textual approach could well lead to tools capable of helping translators improve the quality of their production.

In recent years, we have witnessed the appearance on the market of text critiquing tools that help writers improve their texts by spotting potential problems in spelling, grammar and even style. Some translators find these tools useful. However, they are not meant to examine translations qua translations. Since they can only examine one text at a time, there is no way they can detect correspondence errors between two texts. The problem of detecting correspondence errors can only be addressed within a framework in which correspondences are explicitly represented, that is, a bi-textual framework.

Viewed as a whole, the problem of assessing translation quality appears to be an extremely complex and vexing issue. But we can nonetheless make some steps in the right direction by isolating some specific properties that translation correspondences are expected to satisfy, and attempting to provide a precise (though possibly partial) characterization of these properties.

One such simple property is that the correspondence between ST and TT should be exhaustive: no parts of the source text should be omitted in the translation. Nevertheless, it is not rare for human translations to err just in this way. Sentences, paragraphs or even complete pages are sometimes overlooked by the translator. In this case, we can hope that alignment algorithms will soon become robust enough to allow good guesses at omissions.

Translations are also expected to be free from source language (SL) interference. When translators work on a pair of closely related languages (such as English and French), interference problems can become very acute. Take for example the problem of *deceptive cognates*. These are pairs of words which, in spite of obvious etymological connections, are no longer semantically equivalent. In the case of ‘complete’ deceptive cognates, the meanings are totally disjoint and direct correspondence is never possible (eg. *definitely/définitivement*, *actual/actuel*, *ignore/ignorer*). In the case of ‘partial’ deceptive cognates, there is some overlap in meaning, so that the correspondence remains possible in certain contexts (eg. *camera/caméra*).

An in-depth study of the problem that was conducted at the CWARC revealed that deceptive cognates are the source of an important number of errors in the Hansard translations. The bilingual concordancing tool mentioned above helped us document hundreds of examples, including some of a rather elementary nature (*library/librarie*, *physician/physicien*). Even though these translations are the work of some of the best translators in Canada, it appears that the time pressure under which they are produced makes linguistic interference harder to control. Tools capable of flagging potential errors could therefore prove extremely useful.

This notion of ‘deceptive cognate’ is not perfectly well-defined. There exists some useful reference works (see for example Van Roey, Granger & Swallow [27]), but their exhaustivity is doubtful. Moreover, in the case of partial deceptive cognates, the range of disallowed correspondences is often fuzzy and subject to dialectal variation.

The sensible thing to do, of course, is to start with the clearer cases. Bi-textual representations should make it easy to pinpoint correspondences involving a fixed set of complete deceptive cognates. The level of noise in retrieving incorrect correspondences will be a function of bi-textual resolution. If correspondences are worked out down to the word level, the noise will be very low. According to our preliminary experiments, even if bi-textual resolution is no finer-grained than the sentence level, the noise might still remain within tolerable bounds.

The problem of partial deceptive cognates is of course harder. However, it seems reasonable to believe that, at least for a subset of them, the kind of probabilistic sense disambiguation methods proposed by Gale, Church & Yarowski [14] could provide a suitable discrimination mechanism.

In all cases, the use of a part-of-speech tagger such as the one described in Church [6] would be likely to improve precision, since the correct characterization of some deceptive cognates (whether complete or partial) requires part-of-speech information.

We suspect that there are many other properties of translation correspondences which could be verified by means of bi-textual representations.

6. A Dictation Machine for Translators

One problem with current translator's workstation is that many translators are reluctant to use keyboards and prefer to dictate their translations. For these translators, a complete and fully integrated workstation environment would have to feature speech recognition.

Unfortunately, speech recognition technology has not yet reached a stage where many translators will view it as a practical alternative. Speech can only be decoded with some reliability provided we place some relatively stringent constraints on the contents of the acoustic signal. Typical set-ups resort to one or many of the following constraints: limited vocabulary and syntax, isolated word input (as opposed to continuous speech), user-specific system training, low background noise, etc.

Back in 1989, an experiment that we conducted on speech-to-speech translation (Cochard, Isabelle & Simard [8]) convinced us that current technology was not suitable for real-life applications: the input needed to be constrained in an overly artificial way. My colleague Marc Dymetman then made the important observation that the situation might well be different if the speech was input in the target language by the translator. For in that case a natural source of constraints exists on the acoustic signal: **the signal is known a priori to encode a text which is the translation of some given source text.**

In principle, we could therefore design a speech recognition system specifically oriented towards translation tasks. Such a system would resort to some kind of (partial or complete) translation model which makes it possible to use the source text as a basis for predicting some features of the spoken translation. For example, from the presence of the word *government* in some English source sentence, the translation model could predict that the corresponding French sentence is likely to contain a spoken realization of the word *gouvernement*. Clearly, such a scheme should make the speech recognition task much more tractable.

Brown & al. [4] independently arrived at the same conclusion. They report on an experiment in which they compared the per-word perplexity of an unaided target-language model with the per-word perplexity of the same target-language model once combined with a translation model. They claim that perplexity drops from 63.61 in the first case to 17.2 in the second case. They conclude that:

> "it is reasonable in view of these results to hope that high accuracy recognition of fluent speech is possible with present day speech technology when the text is constrained to be the translation of a known source language sequence." (p. 10)

A project is currently underway at the CWARC to explore some aspects of this very interesting possibility (Dymetman & al. [11]). This project is closely connected with our work on bi-textuality, in that it encompasses the development of a probabilistic translation model whose parameters are extracted from a large bi-textual database.

7. Conclusions

Given the current state of the art, it is rather exceptional for MT to constitute a practical solution, and tools for supporting the work of human translators generally constitute a more sensible use of technology. Existing translator's workstations mainly offer office automation functions. Core aspects of the translation task have yet to be addressed in a more direct way. We have argued that the notion of bi-text, that is, explicit represensations of the translation correspondences that link a pair of source and target texts, is highly relevant to that end.

The development of algorithms capable of automatically organizing existing translations into bi-textual representations is progressing at a quick pace. These results open up the way to a variety of new tools for human translators. Among these, bilingual concordancing will enable translators to tap the riches of a corporate translation memory made up of bi-textual representations derived from previous translations. A little further down the road, we can envision translation critiquing tools that will help translators detect correspondence errors such as omissions and deceptive cognates in their translations. Finally, there are reasons to believe that in the not-so-distant future, translator’s workstations will come to incorporate specially designed speech recognition systems that incorporate some translation knowledge that will be extracted from bi-textual databases.

We conclude that, in the near term, support tools for human translation are likely to progress at a much faster pace and have a much greater impact on the translation community than classical machine translation systems.

REFERENCES

[1] Bar-Hillel Y., *The State of Machine Translation in 1951*, in **American Documentation***,* vol. 2, 1951, pp. 229-237 .

[2] Brown P., Cocke J., Della Pietra S., Della Pietra V., Jelinek F., Lafferty J., Mercer R., Roosin P., *A Statistical Approach to Machine Translation*, **Computational Linguistics**, 16:2, 79-85, 1990.

[3] Brown P., Lai C., Mercer R., *Aligning Sentences in Parallel Corpora*, **Proceedings of the 29th Meeting of the ACL**, Berkely, 1991.

[4] Brown P., Chen S., Della Pietra S, Della Pietra V., Kehler A., Mercer R., **Automatic Speech Recognition in Machine Aided Translation**, unpublished ms., IBM T. J. Watson Research Center, 1992.

[5] Catizone R., Russell G., Warwick S., *Deriving Translation Data from Bilingual Texts*, in Zernick (ed.) **Lexical Acquisition: Using on-line Resources to Build a Lexicon,** Lawrence Erlbaum, 1992.

[6] Church K., *A Stochastic Parts Program and Noun Phrase Parser for Unrestricted Text*, **Proceedings of the 2nd ACL Conference on Applied Natural Processing**, Austin, 1988.

[7] Church K., Gale W., *Concordances for Parallel Texts*, in **Proceedings of the 7th Annual Conference the UW Centre for the NOED and Text Research**, Oxford, 1991.

[8] Cochard J.L., Isabelle P., Simard M., **IRMA : an Agricultural Market Report Interpreter**, Tech. rep. no. Co28-1/49-1990E, CWARC, Laval, 1990.

[9] Debili F., Sammouda E., *Appariement des phrases de textes bilingues français-anglais et français-arabes*, **Proceedings of COLING-92**, Nantes, 1992.

[10] Dempster A., Laird N., Rubin D., *Maximum Likelihood from Incomplete Data via the EM Algorithm*, **Journal of the Royal Statistical Society**, 39(B), 1977.

[11] Dymetman M., Foster G., Isabelle P., ***Towards Automatic Dictation Systems for the Professional Translator***, Technical Note, CWARC, Laval, 1992.

[12] Gale W., Church K., *A Program for Aligning Sentences in Bilingual Corpora*, **Proceedings of the 29th Meeting of the ACL**, Berkely, 1991.

[13] Gale W., Church K., *Identifying Word Correspondences in Parallel Texts*, **Proceedings of DARPA SLS Workshop**, 1991.

[14] Gale W., Church K., Yarowski D., *Using Bilingual Materials to Develop Word Sense Disambiguation Methods*, **Proceedings of the 4th International Conference on Theoretical and Methodological Issues in Machine Translation**, Montreal, 1992.

[15] Harris B., *Are you Bi-Textual?*, **Language Technology**, #7, p. 41, 1988.

[16] Harris B., *Bi-text, a New Concept in Translation Theory,* **Language Monthly**, #54, p. 8-10, 1988.

[17] Isabelle P., *Machine Translation at the TAUM Group*, in Margaret King (ed.), **Machine Translation Today: The State of the Art**, Edinburgh University Press, 1987.

[18] Kay M., **The Proper Place of Men and Machines in Translation**, CSL-80-11, Xerox PARC, 1980.

[19] Kay M., Röscheisen M., **Text-Translation Alignment**, unpublished ms., Xerox PARC, 1988.

[20] Klavans J., Tzoukermann E., *Linking Bilingual Corpora and Machine Readable Dictionaries with the BICORD System*, **Proceedings of the 6th Annual Conference of the UW Centre for the NOED and Text Research**, University of Waterloo, 1990.

[21] Macklovitch E., *An Off-the-shelf Workstation for Translators*, dans D. Hammond (ed.) **Proceedings of the 30th Annual Conference of the ATA**, Washington DC, October 11-14, 1989, pp. 4891-498.

[22] Macklovitch E., **A Second Version of the CWARC's Workstation for Translators**, tech. report, CWARC, 1991.

[23] Macklovitch E., **Evaluating Commercial MT Systems**, paper presented at the Evaluator's Forum on MT Systems, Ste-Croix, Switzerland, 1991.

[24] Macklovitch E., *Corpus-Based Tools for Translators,* to appear in the **Proceedings of the ? Conference of the ATA**, San Diego, 1992.

[25] Sato S., Nagao M., *Toward Memory-Based Translation*, **Proceedings of COLING-90**, 247-252, 1990.

[26] Simard M., Foster G., Isabelle P. *Using Cognates to Align Sentences in Parallel Corpora*, **Proceedings of the 4th International Conference on Theoretical and Methodological Issues in Machine Translation**, Montreal, 1992.

[27] Van Roey J., Granger S., Swallow H., **Dictionnaire des faux-amis français-anglais**, Paris, Duculot, 1988.

Appendix A

Screendump from OCTA, the CWARC's bilingual concordancing system