Line 'Em Up: Advances In Alignment Technology And Their Impact On Translation Support Tools[1]

Elliott Macklovitch **Marie-Louise Hannan**

macklovi@citi.doc.ca **hannan@citi.doc.ca**

Centre for Information Technology Innovation (CITI)
1575 Chomedey Blvd., Laval (Quebec)
CANADA H7V 2X2

Abstract

We present a quantitative evaluation of one well-known word alignment algorithm, as well as an analysis of frequent errors in terms of this model's underlying assumptions. Despite error rates that range from 22% to 32%, we argue that this technology can be put to good use in certain automated aids for human translators. We support our contention by pointing to certain successful applications and outline ways in which text alignments below the sentence level would allow us to improve the performance of other translation support tools.

1. Introduction

Considerable effort has been invested in the past few years in developing what has been called a "new generation" of translation support tools based on parallel text alignment (Brown et al. [2]; Gale and Church [8]; Isabelle et al. [10]). One of the questions we can ask at this stage, then, is: How well do these translation support tools (TST's) work? More specifically, can recent progress in word-level alignment make a significant contribution to their performance?

In an attempt to address these questions, we designed a small-scale experiment to evaluate the performance of one well-known word alignment algorithm. The results of this evaluation, carried out on a bilingual corpus of French and English texts, point to certain weaknesses associated with the model, and our analysis suggests some explanations for these weaknesses. The investigation itself was motivated by a need for finer-grained text alignments than those at the sentence level, which would permit us to push the performance of existing translation support tools even further, and allow us to imagine novel ones.

In the following sections, we will introduce and clarify some of the basic notions related to text alignment and translation analysis in order to situate our study within its larger theoretical context; then discuss progress in text alignment from the sentence to the word level; and, finally, present our experimental design and an analysis of results obtained.

1. We would like to thank George Foster, to whom we are indebted for his patient explanations of the statistical subtleties of alignment technology. Thanks also go to Pierre Plamondon, for providing us with output from TM-Align and input on its operation. Of course, neither is responsible for any of our errors of interpretation.

2. Translation Analysis

2.1 Definition of Terms and Concepts

The term “translation analysis”, introduced by Isabelle et al. [10], can be defined as a set of techniques for automatically reconstructing the correspondences between segments of a source text and the segments of its translation. Translational equivalence must be seen as compositional in some sense, with the global correspondences between source and translation being analyzable into sets of finer correspondences. The challenge for translation analysis is to discover the links between increasingly finer segments, motivated by the fact that existing translations contain a wealth of information. In order to take advantage of this valuable resource, correspondences must be made explicit. This process of analysis in turn facilitates the development of new kinds of automatic translation aids.

The concept of alignment is more difficult to pin down. It has been used to designate both a process and the result of that process. According to the Collins Cobuild dictionary [15], if you align something, you “place it in a certain position in relation to something else, usually along a particular line or parallel to it.” A textual alignment usually signifies a representation of two texts which are mutual translations in such a way that the reader can easily see how certain segments in the two languages correspond or match up. The size of the alignment can vary considerably in theory, from a particular textual unit to the entire text. However, the term “alignment” can also refer to research efforts into finding practical methods for obtaining such aligned texts.

The product of an alignment at the level of the entire text is also known as a bi-text. Harris [9] is credited with coining this term to designate the juxtaposition of a translation’s source and target texts on the same page or screen, in a way that can be stored for future consultation. In the next subsection, we will give a brief overview of the history of text alignment, beginning with the first proposals for recycling past translations.

2.2 Brief History of Alignment

The idea of electronically storing past translations in a bilingual format in the hopes of reusing their contents was first proposed by Melby [13]. The idea was to construct a bilingual concordance by having a data operator manually input the text and its translation, thus creating a valuable reference tool for translators. Harris’ discussion [9] of the uses for bi-text is motivated by the same desire to capitalize on existing resources.

An interest subsequently arose in automatically constructing such bilingual databases on a much larger scale, with initial work focusing on reconstructing the links between a source text (ST) and a target text (TT) on the sentence level. This level of resolution presented a greater challenge to automatic alignment than coarser textual units (such as paragraph or section), since these are almost always in a 1:1 relationship, and thus discovering their correspondence is more straightforward than at the sentence level, where one-to-many correspondences are not uncommon.

The first published algorithms for aligning sentences in parallel texts were proposed by Gale and Church [8] and Brown et al. [2]. Based on the observation that the length of a text and the length of its translation are highly correlated, they calculate the most likely sentence correspondences as a function of the relative length of the candidates. As these methods do not appeal to any outside sources of information, such as bilingual dictionaries, they are considered to be language-independent. The major drawback of this method is that once the algorithm has accidentally misaligned a pair of sentences, it tends to be unable to correct itself and get back on track before the end of the paragraph. Used alone, length-based alignment algorithms are therefore neither very robust nor reliable.

Ensuing work in sentence alignment suggested various ways of dealing with these robustness and reliability issues. Simard et al. [14] propose a method for coupling the length criterion with the notion of cognateness[2] to boost the algorithm's performance. Another word-based refinement method, in which the preferred sentence alignment is the one that maximizes the number of systematic word correspondences, is proposed by Kay and Roscheisen [11]. Debili and Sammouda [7] build on this method by using a bilingual dictionary to guide the initial hypotheses on word correspondences, thereby reducing the algorithm's search space. Each of these methods appeals to the idea of establishing certain guideposts or word "anchors" which will help the alignment algorithm in its progression through the text. The authors of the cognate method report a gain in robustness at little computational cost.

Although the results obtained by some of the above methods have been quite accurate when tested on relatively clean[3], large corpora, they remain "partial" alignments in that a good deal of the correspondence information remains hidden at a finer level of resolution (below the sentence level). Despite this obvious limitation, several useful applications for sentence-level alignments have been found, including bilingual terminology research (Dagan et al. [6]) and novel translator's aids such as TransSearch, a bilingual concordancing tool (Isabelle et al. [10]), and TransCheck, a tool for automatically detecting certain types of translation errors (Macklovitch [12]).

The problems of sentence-level alignment, if not entirely resolved, are fairly well understood. What about translation correspondences below the sentence level? As Debili and Sammouda [7] point out, as we descend the text hierarchy, from chapter to section, section to paragraph, paragraph to sentence, sentence to word and word to character, there seems to be a breaking point at the level of the sentence. The order of elements is much less likely to be preserved across translation below this level, just as the 1:1 relationship between elements becomes increasingly rare. Keeping this kind of difficulty in mind, we will now examine recent work in the area of text alignment.

2.3 Recent Progress

2.3.1 Character Alignment

As a more robust alternative to sentence-level alignment, Church [4] proposes a method for aligning text based on character matches within source and target texts. Following Simard et al. [14], this method makes use of an operational definition of cognates to hypothesize links between the texts. Its basic assumption is that cognates of words in the immediate vicinity of a given word x will be found among the words in the vicinity of its translation $f(x)$. Each text (source and target) is concatenated into a single sequence of characters and passed to a program which calculates a scatterplot (called a *dotplot*). The program then determines the number of cognates between the two texts, statistical smoothing processes are applied to

2. "Cognates are pairs of words in different languages which, usually due to a common etymology, share phonological or orthographic properties as well as semantic properties, so that they are often employed as mutual translations." (Simard et al. [14])
3. Texts containing a lot of graphics, tables or floating footnotes, as well as those that have been converted to electronic form by optical character recognition are among the types of texts which must undergo a semi-automatic pre-editing stage in order to clearly establish certain boundaries on which the alignment algorithms depend.

the results, and finally an optimal alignment path is calculated. This path corresponds to a global alignment of the characters in the two texts. It is important to note that the alignment of characters is not a goal in itself, but only a method for getting at an overall alignment of the entire text. The author reports very accurate results for *char_align*, with offset from the correct alignment often being minimal. This process does away with the need for finding sentence and/or paragraph boundaries, a task which can be quite difficult for noisy input texts.

2.3.2 Word Alignment

As part of their efforts to develop a statistical machine translation system, Brown et al. [3] outline a series of five increasingly complex translation models that can also be used to establish word alignments. The authors define an alignment between a pair of strings (in the source and target languages, which happen to be English and French in this particular case) as: "an object indicating, for each word in the French string, that word in the English string from which it arose." (*ibid., p. 266*) All five models make use of this notion of alignment to calculate the overall likelihood of some target string being the translation of a given source string. Stated very roughly, the translation models allow us to calculate a probability score for all possible configurations of word connections between the source and target strings. The best overall alignment between a pair of strings consists of the most probable connections between the words that each comprises. Since the total number of configurations can be astronomical, even for strings of moderate length, dynamic programming techniques are used to select the best one efficiently.

The statistical translation models are quite technical, so we will only attempt to outline their general properties in layman's terms here; the mathematically fearless reader is referred to Brown et al. [3] for an explanation of the intricacies of the underlying statistical calculations. All of the models are based on the notion that every target string, T, is a possible translation of a given string of source words, S. In this approach, there are no correct or incorrect translations; only more or less probable ones. Thus the translation model probability, $Pr(T|S)$, or the conditional probability of the target string given the source string, must be calculated. External resources such as a standard, bilingual dictionary are not accessed; all of the translational probabilities which the model uses to predict target words are induced from its training corpus solely by statistical methods. In this sense, the approach can be said to have minimal linguistic content.

One way of understanding the increasing complexity of the five models is to examine what the probability of an alignment depends on. In Model 1, the probability[4] of an alignment depends only on the probabilities assigned to its constituent bilingual word pairs. The same probability in Model 2 depends on the probabilities assigned to the word pairs, plus the probabilities assigned to a connection between the positions occupied by each pair of words. Thus, words occurring at the beginning of the French string are more likely to be connected to words near the beginning of the English string; in this way, the importance of relative position is recognized.

In addition to these parameters, Models 3, 4, and 5 also depend on the words connected to a source word, with Model 3's calculations depending on the number of words generated by each source word, and Models 4 and 5 depending on the number and the identity of these words. As mentioned above, these later models are much more complex, making use of statistical approximation techniques, and, since our study is concerned with an implementation of Model 2, we will not go into further detail here.

4. The probability of an alignment is more accurately described as the joint probability of an alignment with the target sentence, but this is proportional to the alignment probability when the target sentence is fixed.

All the models outlined above require aligned sentences as input. An alternative word alignment program is proposed by Dagan et al. [5], in which the input is instead a rough global alignment of the two texts (the output of *char_align*), and the program used is an adapted version of Brown et al.'s Model 2. The most probable alignment is chosen from within a set of relevant connections, using a method which requires fewer parameters than the original model, and hence is applicable to shorter, noisier texts. An evaluation of this implementation was carried out by the authors, the results of which will be discussed in Section 3, below.

2.4 Potential Benefits of Word Alignment for TST's

Recent progress in alignment technology raises the possibility of improving the performance of existing translation support tools and of developing new ones. Despite the theoretical difficulties presented by the problem (as discussed in Section 2.3.2), we maintain that word alignment is a valid research goal which would allow for even more useful translation support tools than are currently available or possible with sentence alignment.

For example, the TransSearch and TransCheck systems briefly alluded to above would stand to gain from advances in word alignment technology. Their performance is currently limited by the relatively coarse resolution of the alignments on which they are based. At present, the output of TransSearch highlights[5] only the word or words submitted as a query; in the case of monolingual queries, the addition of a reliable word alignment component would allow target-language equivalents to be highlighted as well. As for TransCheck, the level of noise currently observed in the output could be reduced considerably with the establishment of more accurate word-level connections. Instead of flagging aligned sentences that happen to contain a prohibited pair of words or phrases, the system would only flag those sentences in which the pair of expressions was determined to be a mutual translation.

In the concluding section, we will briefly describe other innovative TST's that are based on recent advances in alignment technology. In order to assess these and other potential applications of the new word alignment algorithms, we need to get an idea of just how well they perform. We therefore turn to our experiment and the issue of evaluation.

3. How well do current word alignment algorithms work?

The only published quantitative evaluation of current word alignment algorithms that we are aware of is found in Dagan et al. [5]. The authors actually describe two evaluations of their *word_align* program. The first was conducted on a 160 000-word corpus taken from the bilingual Canadian Hansards; errors were sampled at sentence boundaries and compared with the "true" alignments, as established by a panel of human judges. The second evaluation was performed on a 65 000-word technical manual in English and French: 300 tokens were selected using an elaborate sampling procedure, and their alignments were compared to the best corresponding position in the target text, again as established by a human judge. On the first evaluation, the authors state that "in 55% of the cases, there is no error in *word_align*'s output" (p.7); and on the second, "for 60.5% of the tokens the alignment is accurate" (p.7). If we restate these same results in inverse fashion, this means that the program's end-of-sentence alignments were inaccurate in

5. By "highlights", we mean that the word(s) appear in bold characters on the screen or when printed out.

45% of the cases, and that 39.5% of the sampled words were not correctly aligned with their target equivalent.

Before we consider various applications for this new word alignment technology, it is important for us to establish whether these results can be corroborated by other implementations of similar word alignment algorithms. Moreover, we would like to understand the main reasons for incorrect word alignments, to determine whether it will be possible to significantly improve these programs' performance.

To help us answer these questions, we undertook a small-scale study of word alignment outputs produced by our own alignment program, known as TM-Align. This program is based directly on Model 2, the second in the series of statistical translation models defined by Brown et al. [3]. It was trained on a bilingual corpus of 74 million words drawn from the Canadian Hansards, and includes a vocabulary of 92 129 word forms in English and 114 517 word forms in French. The test corpus submitted to TM-Align was composed of a distinct set of aligned sentence pairs from Hansard, with a maximum length of 40 words and exhibiting 1:1 alignment between the English and the French; it totalled about 115 000 words of English and 125 000 words of French.

An example of TM-Align output is given in Figure 1. The output has been reformatted so that the links between numbered word positions are shown graphically as connections between words. Recall that the algorithm is directional: for each word in the French string, the program proposes a word in the English string that gives rise to it, but not vice versa; this is why several words in the English string are not connected to anything. In this particular example, the program proposes that the source of the first French word "les" is the first English word "there"; and so on. When the algorithm can find no likely source word in the English, it connects the French word with a null word which is inserted by convention at the beginning of each English sentence; in Figure 1, this is represented by linking the French word to a series of question marks.

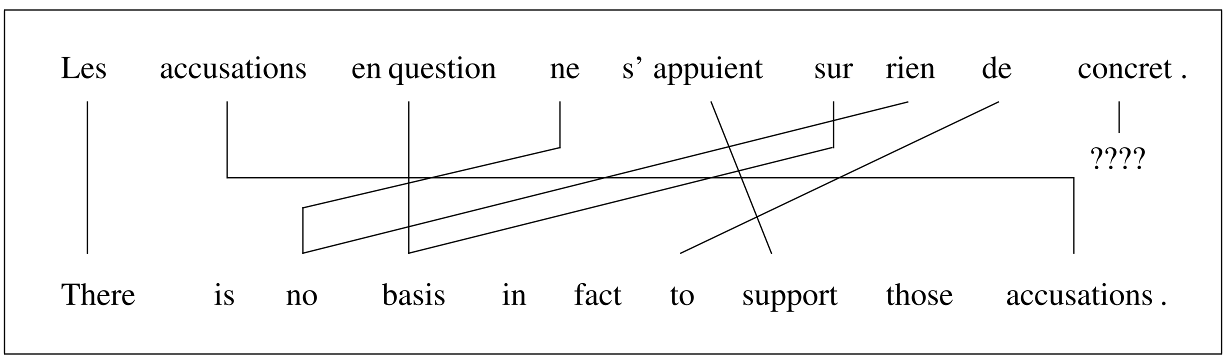

Figure 1: An example of TM-Align Output

Since it is a demanding and time-consuming task for a human to evaluate such word alignments, we decided to extract a sample of 50 sentence pairs from our test corpus, yielding a total of 935 French words. For each French word in these sentences, we asked ourselves whether, as human translators, we agreed with the English word that TM-Align proposed as its source. So, in the example in Figure 1, we disagreed with the source that TM-Align proposes for the first French word, "les", as well as for the expression, "en question"; nor could we accept the linking of the French word "concret" to the null word, which is tantamount to claiming that this word has no source in the English. The results of our evaluation of the 935

French words in the sample are summarized in Table 1 below.[6]

Table 1: Alignment Evaluation Results

	Number of French tokens in sample	Number of incorrect alignments proposed by TM-Align	Percentage error
All words	935	296	32%
Function words	418	181	43%
Non-function words[a]	517	115	22%

a. In addition to nouns, verbs, adjectives and adverbs, these include numbers and punctuation.

Perhaps the first thing to notice about these results is that they are significantly better than those cited from Dagan et al. [5]: they report an error rate of 39.5% for a 300-word sample which does not include any function words.[7] The overall error rate in our sample is 32%, and this drops to 22% when function words are excluded. Several factors have probably contributed to produce this difference. For their evaluation, the corpus on which the system trained was "noisy" and only 65 000 words long; our training corpus, in contrast, was clean Hansard text over 100 times that size. Moreover, the input to their *word_align* program was not composed of pairs of aligned sentences, but rather a single global alignment of the whole text, as produced by their *char_align* program. In a sense, we simplified the job for our TM-Align program by providing it with input we knew to be correctly aligned at the sentence level.

When we recall how little explicit linguistic knowledge these programs incorporate, it's hard not to be impressed by their level of success. Before we attempt to analyse the underlying causes for our alignment program's errors, therefore, it may not be amiss to illustrate a few of its strengths. For those accustomed to the literalness of the texts produced by classic machine translation systems, some of the connections between words that are automatically established by programs like TM-Align can be quite surprising. (All the examples given below are taken from our small test sample. The words underlined in each sentence pair were properly aligned by the program.)

As Brown et al. [3] remark, "number is not invariant across translation" (p.286):

(i) [Comment pouvons nous mener de <u>bonnes</u> <u>négotiations</u> commerciales...
How can we have a <u>proper</u> <u>negotiation</u> on trade when... #850]

Of course, number is not the only grammatical property that may vary across translation. TM-Align is often capable of linking words belonging to different syntactic categories:

6. In the Table and following discussion, we refer to the disagreements between the alignments proposed by the program and the judgements of the human translator as errors committed by the program. This may be somewhat presumptuous on our part, since the judgements being asked of the human are not always self-evident.
7. "Empirically, we found that both high and low frequency words caused difficulties and therefore connections involving these words are filtered out. The thresholds are set to exclude the most frequent function words and punctuations [sic], as well as words with less than 3 occurrences." (Dagan et al. p.5)

(ii) [... si le Canada a modifié sa position au cours des négotiations... #2100
... whether there has been any Canadian movement on its position in the negotiations...]

(iii) [... la présidence hésite à rejeter purement et simplement le rapport en question .
... the chair is reluctant to reject outright the said report . #600]

Sometimes, the program may even connect a pronoun and the fully specified coreferential noun phrase in the other language:

(iv) [Le ministre reconnaît maintenant que cette façon de procéder était fort compliquée, ...
He now admits that it was very complex, ... #1300]

Although these few examples do not suffice to illustrate the full power and flexibility of statistical word alignment algorithms, it seems quite clear that connections like those in (ii-iii) would be difficult to establish using a standard bilingual dictionary.

4. The causes of word alignment errors

We will now examine the more frequent types of word alignment errors found in our test sample and correlate them, where possible, with the major assumptions underlying our implementation of Model 2.

As we saw above, a statistical model of translation can be viewed as a function Pr(T|S), which is intended to approximate the probability that a human translator will produce a sequence of target language words, T, when presented with the string of source language words, S. Since there are an infinite number of sentences in any language, the only feasible way to estimate the probability of two arbitrary strings S and T being mutual translations is to rely on the fact that translation is compositional: the translation of all but a few formulaic sentences is a (more or less complex) function of the translation of their parts. What statistical models of translation must do, therefore, is find some way of expressing the relation between S and T in terms of **sub-events** frequent enough that their probabilities *can* be reliably estimated from a training corpus. Combining these probabilities, called the model's parameters, in a way that will predict the likelihood of T given S requires making certain simplifying assumptions.

Of these simplifying assumptions, the one that is perhaps the easiest to grasp is that each (French) target word can have no more than one English word as its source. Obviously, this is an incorrect assumption about translation in general, and it is built into the model solely for reasons of computational tractability: it reduces the number of possible alignments from $2^{(Ls*Lt)}$ to $(Ls+1)^{Lt}$, where Ls and Lt are the length (in number of words) of S and T respectively. Examples (v-vi) below illustrate two common types of alignment errors that are attributable to this assumption. In the first, the program cannot connect the French noun “jeunesse” with the English phrase “young people”; in the second, a simple form of the French verb cannot be linked to the compound English progressive. (Again, the underlined words correspond to the actual alignments produced by the program.)

(v) [La jeunesse a besoin d' espérer .
Our young people deserve that hope . #200]

(vi) [Bien des provinces font plus de progrès ...
Many of the provinces are making more progress ... #250]

As discussed above, the principal distinction between Models 1 and 2 is that the latter takes word order into account in establishing its connections: all things being equal, Model 2 tends to favour alignments linking words in parallel positions. This may not always yield the desired results, however, as can be seen in example (vii) below.[8]

(vii) [... d'après la fédération <u>canadienne</u> des étudiants :
... from the Canadian <u>federation</u> of students : #300]

A third, more technical assumption of Model 2 may be stated as follows: Given the source text S and an alignment, the probability of a given target word at a particular position in T depends only on the word in S to which it is connected; it does not depend on any of the neighbouring words in T, nor on any of the positions of the words being connected. What this assumption of target word independence means, in effect, is that the system has absolutely no knowledge of target language structure.[9] We can see the consequences of this assumption in example (viii) below.

(viii) [... il a arraché avec son hélicoptère 16 marins russes <u>du</u> pont de leur navire .
... his helicopter plucked 16 Russian sailors off the deck <u>of</u> their ship . #750]

The verbs in example (viii) govern certain prepositional complements: in English, the valency pattern is "pluck X *off* Y", while in French it is "arracher X *de* Y". Due to the assumption of target word independence, however, TM-Align cannot take these patterns into account, and seeks to establish the source of the French preposition "du" in isolation.

As this last example illustrates, the representations afforded by this approach for expressing connections between words do not allow us to relate phrasal units in the two languages in any natural way. What we would like to be able to say in cases like (viii) is that while word *x* is not generally translated as *y*, within these particular phrases *y* **is** a proper translation of *x*. The problem is not just with Model 2; none of the translation models of Brown et al. [3] allow for what they call "general alignments" of phrases to phrases, although they do of course recognize that these may be required for some translations. In part, this is due to the asymmetry of the models, i.e. the fact that every target word must be linked to its source in the other language, though the same is not attempted in the opposite direction. But this is not the whole story.

Consider the objects of the English preposition "about" and the French preposition "de" in example (ix) below:

(ix) [... bombé le torse, fiers de leurs initiatives .
... beating their chests in pride about what they have done . #1900]

Where the French uses a simple noun phrase, the English has a headless relative clause; and yet this seems to be a perfectly acceptable translation. Ideally, we would like our model to show that these two noun phrases correspond, although the only word-level correspondences within the phrases are between "leurs" and "they". Unfortunately, the only way to indicate that two phrases are mutual translations within this word alignment scheme is to connect every word in one phrase to every word in the other. Such

8. If the order of "Canadian" and "federation" is inverted in (vii), TM-Align does correctly connect "canadienne" and "Canadian".

9. This is actually deliberate, since Brown et al. [3] intend the translation component embodied by Model 2 to work in conjunction with a separate target language model within their larger statistical MT system.

connections are clearly not warranted here. This, then, is a serious inadequacy inherent in the flat representations of Brown et al. [3]: they cannot simultaneously express correspondences between higher-level units while at the same time indicating that some of the words within those units are mutual translations and others are not. For this, a richer kind of hierarchical representation would be required.

The same deficiency in the model's expressive power is illustrated by the slightly more complex example in (x) below. What interests us here, though, is the fact that the system does not manage to link the two main verbs "permettrai" and "allowed".

(x) [Toutefois, je ne permettrai pas qu' elle l' oublie tant qu' elle n' y aura pas répondu .
However , as long as it stays there she is not going to be allowed to forget it . #500]

How is it that TM-Align fails to link "permettrai" and "allowed" in this example? The answer is simple: the translation probabilities inferred from the training corpus relate not words, but word *forms*; and no co-occurrences of these particular inflected forms were observed in this training corpus. A straightforward solution to this problem would be to add a morphological analysis component to the system, so that all the different forms of a word could be related to the same lexeme. Brown et al. [3] propose just this change at the end of their paper, saying that it should significantly improve their models' statistics. (p. 295)

Actually, the models of Brown et al. [3] already include certain simple morphological transformations: "We have restored the *e* to the end of *qu'* and have ... analyzed *des* into its constituents, *de* and *les*. We commit these and other petty pseudographic improprieties **in the interest of regularizing the French text**." (p.284-5; emphasis added) The authors refer to these modifications as transformations, and technically speaking that is exactly what they are. Of course, regularizing the distribution of linguistic forms was precisely the reason for which transformations were originally introduced into the apparatus of linguistic description back in the '50s. And so the question that immediately arises is: Why stop with morphology?

Consider, in this regard, example (xi) below:

(xi) [Le ministre prendra il immédiatement les mesures voulues ...?
Will this minister do the right thing now ...? #2200]

TM-Align errs in connecting the pronoun "il" to the English verb "do" in this example. But what *should* the program have connected "il" with here? The two obvious choices are either the null word or the English noun "minister", though neither is entirely satisfactory. The French pronoun shares the same referent as the English noun, and this, in the case of other pronouns, is sufficient to justify a connection between the two. But in this particular sentence pair, it is not the English noun which actually *gives rise to* the French pronoun; rather, the pronoun is inserted to mark the fact that the French sentence is a question. Should "il" then be linked to the English null word, since it certainly doesn't correspond to any other word in the English sentence? Given the target word independence assumption and the fact that a French pronoun may be linked to an English noun in other contexts, it is difficult to see how this could be done. The fact that neither of these alternatives is entirely satisfactory suggests that the question may be ill-conceived. Perhaps this "il" should not be linked to anything in the English sentence – not even the null word – because its source is not in the English sentence at all, but in the construction of an interrogative French sentence. Brown et al. [3] suggest replacing the inflected variants of a verb with its canonical form plus an abstract feature (which they call an "ending"), indicating its person, number and tense values in a given sentence. Similarly, in cases like (xi), if the French text were regularized by removing "il" and replacing it with an abstract question marker, the model's statistics, and hence its predictive power, would also be improved.

5. Conclusion

There is an obvious objection that may be raised regarding our attempt to quantify and analyse word alignment errors produced by a system like TM-Align: Why bother? If, in response to the inadequacies of Model 2, Brown et al. [3] have developed Models 3, 4 and 5, why not perform our tests on an implementation of these more powerful models? There are a number of answers to this objection. For one thing, some of the inadequacies of Model 2 are also present in the later models; this is true of the single source-word assumption, as well as the deficiencies due to the flat representation scheme. For another, there is the fact that the more powerful models are also computationally more costly. Perhaps this is why Dagan et al. [5] base their own *word_align* program on Model 2.

In any case, our intention was not to prove that systems like TM-Align are not up to the task of expressing the full range of translation correspondences that may be found in two texts; we knew this beforehand. Rather, we have attempted to illustrate some of the strengths and weaknesses of these statistical models of word alignment, suggesting that richer, more abstract representations will eventually be required in order to attain the ultimate goals of translation analysis. On the other hand, we would certainly not want to suggest that we need await near-perfect word alignments in order to apply the existing technology in useful and innovative ways. On the contrary, we contend that current word alignment algorithms, if applied judiciously to well-defined sub-tasks in translation, already allow for exciting new types of translation support tools. We will conclude by briefly illustrating with two examples.

The *Termight* system described in Dagan & Church [6] is designed to assist translators and lexicographers in building glossaries from previously translated texts. It extracts candidate terms from the source text, which are reviewed and confirmed by a terminologist in an interactive session. The program then aligns the source and target texts at the word level, identifying candidate translations for each retained source term; again, these must be confirmed by the user before being added to the final glossary. As the authors point out, the mapping between the two texts is partial, since *word_align* skips words that cannot be aligned at a given confidence level; moreover, the proposed target term translations are not always correct. Given the interactive design of the system, however, 100% accuracy on both monolingual and bilingual tasks is not absolutely indispensable. Even if one word in three is not aligned correctly, *Termight* still manages to double the rate at which glossaries were previously compiled.

Another innovative TST is the TransTalk system, a dictation machine designed especially for translators. TransTalk does not actually perform any word alignments, but uses statistically derived translation probabilities for the words in a given source sentence to predict a set of French words from which the translator is likely to compose his dictated translation. This dynamic vocabulary reduces the search space for the voice recognition module and significantly improves its performance. See Brousseau et al. [1] for more details.

These, and the other systems alluded to above, demonstrate that the performance of current alignment algorithms, although far from perfect, is certainly sufficient to support the development of what is justifiably called a new generation of translation support tools. And this may just be the tip of the iceberg. Further improvements in alignment technology are certain to result in new and even more powerful automated aids for human translators.

References

[1] Brousseau J., Drouin C., Foster G., Isabelle P., Kuhn R., Normandin Y., Plamondon P., *French Speech Recognition in an Automatic Dictation System for Translators: the TransTalk Project*, in **Eurospeech '95**, Madrid, Spain, September 1995.

[2] Brown P., Lai J., Mercer R., *Aligning Sentences in Parallel Corpora*, in **Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics**, Berkeley CA, 1991.

[3] Brown P., Della Pietra V., Della Pietra S., Mercer R., *The Mathematics of Statistical Machine Translation: Parameter Estimation*, in **Computational Linguistics**, 19:2, pp. 263-311, 1993.

[4] Church K., *Char_align: A Program for Aligning Parallel Texts at the Character Level*, in **Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics**, Columbus, Ohio, 1993.

[5] Dagan I., Church K., Gale W., *Robust Bilingual Word Alignment for Machine Aided Translation*, in **Proceedings of the Workshop on Very Large Corpora**, Columbus, Ohio, 1993.

[6] Dagan I. and Church K., *Termight: Identifying and Translating Technical Terminology*, in **Proceedings of EACL**, 1994.

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

[8] Gale W. and Church K., *A Program for Aligning Sentences in Bilingual Corpora*, in **Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics**, Berkeley CA, 1991.

[9] Harris B., *Bi-text: A New Concept in Translation Theory*, in **Language Monthly**, no. 54, p.8-10, March 1988.

[10] Isabelle P., Dymetman M., Foster G., Jutras J-M., Macklovitch E., Perrault F., Ren X., Simard M., *Translation Analysis and Translation Automation*, in **Proceedings of the Fifth International Conference on Theoretical and Methodological Issues in Machine Translation**, Kyoto, 1993.

[11] Kay M. and Roscheisen M., *Text-Translation Alignment*, in **Computational Linguistics**, 19:1, p.121-142, 1993.

[12] Macklovitch E., *Using Bi-textual Alignment for Translation Validation: the TransCheck System*, in **Proceedings of the First Conference of the Association for Machine Translation in the Americas** (AMTA-94), Columbia, Maryland, 1994.

[13] Melby A., *A Bilingual Concordance System and its Use in Linguistic Studies*, in **Proceedings of the Eighth LACUS Forum,** p.541-549, Hornbeam Press, Columbia SC, 1981.

[14] Simard M., Foster G., Isabelle P., *Using Cognates to Align Sentences in Parallel Corpora*, in **Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation** (TMI-92), Montreal, Canada, 1992.

[15] Sinclair J. et al., eds. **Collins Cobuild English Language Dictionary**, Collins: London, 1987.