A STATISTICAL APPROACH TO LANGUAGE TRANSLATION

P. BROWN, J. COCKE, S. DELLA PIETRA, V. DELLA PIETRA,
F. JELINEK, R. MERCER, and P. ROOSSIN

IBM Research Division
T.J. Watson Research Center
Department of Computer Science
P.O. Box 218
Yorktown Heights, N.Y. 10598

ABSTRACT

An approach to automatic translation is outlined that utilizes techniques of statistical information extraction from large data bases. The method is based on the availability of pairs of large corresponding texts that are translations of each other. In our case, the texts are in English and French.

Fundamental to the technique is a complex glossary of correspondence of fixed locutions. The steps of the proposed translation process are: (1) Partition the source text into a set of fixed locutions. (2) Use the glossary plus contextual information to select the corresponding set of fixed locutions into a sequence forming the target sentence. (3) Arrange the words of the target fixed locutions into a sequence forming the target sentence.

We have developed statistical techniques facilitating both the automatic creation of the glossary, and the performance of the three translation steps, all on the basis of an alignment of corresponding sentences in the two texts.

While we are not yet able to provide examples of French / English translation, we present some encouraging intermediate results concerning glossary creation and the arrangement of target word sequences.

1. INTRODUCTION

In this paper we will outline an approach to automatic translation that utilizes techniques of statistical information extraction from large data bases. These self-organizing techniques have proven successful in the field of automatic speech recognition [1,2,3]. Statistical approaches have also been used recently in lexicography [4] and natural language processing [3,5,6]. The idea of automatic translation by statistical (information theoretic) methods was proposed many years ago by Warren Weaver [7].

As will be seen in the body of the paper, the suggested technique is based on the availability of pairs of large corresponding texts that are translations of each other. In particular, we have chosen to work with the English and French languages because we were able to obtain the bi-lingual Hansard corpus of proceedings of the Canadian parliament containing 30 million words of text [8]. We also prefer to apply our ideas initially to two languages whose word order is similar, a condition that French and English satisfy.

Our approach eschews the use of an intermediate mechanism (language) that would encode the "meaning" of the source text. The proposal will seem especially radical since very little will be said about employment of conventional grammars. This omission, however, is not essential, and may only reflect our relative lack of tools as well as our uncertainty about the degree of grammar sophistication required. We are keeping an open mind!

In what follows we will not be able to give actual results of French / English translation: our less than a year old project is not far enough along. Rather, we will outline our current thinking, sketch certain techniques, and substantiate our optimism by presenting some intermediate quantitative data. We wrote this somewhat speculative paper hoping to stimulate interest in applications of statistics to translation and to seek cooperation in achieving this difficult task.

2. A HEURISTIC OUTLINE OF THE BASIC PHILOSOPHY

Figure 1 juxtaposes a rather typical pair of corresponding English and French sentences, as they appear in the Hansard corpus. They are arranged graphically so as to make evident that (a) the literal word order is on the whole preserved, (b) the clausal (and perhaps phrasal) structure is preserved, and (c) the sentence pairs contain stretches of essentially literal correspondence interrupted by fixed locutions. In the latter category are [I rise on = je souleve], [affecting = a propos], and [one which reflects o n = pour mettre en doute].

It can thus be argued that translation ought to be based on a complex glossary of correspondence of fixed locutions. Included would be single words as well as phrases consisting of contiguous or non-contiguous words. E.g., [word = mot], [word = propos], [not = ne ... pas], [no = ne ... pas], [seat belt = ceinture], [ate = a mange] and even (perhaps) [one which reflects on = pour mettre en doute], etc.

Translation can be somewhat naïvely regarded as a three stage process:

(1) Partition the source text into a set of fixed locutions.

(2) Use the glossary plus contextual information to select the corresponding set of fixed locutions in the target language.

(3) Arrange the words of the target fixed locutions into a sequence that forms the target sentence.

This naïve approach forms the basis of our work. In fact, we have developed statistical techniques facilitating the creation of the glossary, and the performance of the three translation steps.

While the only way to refute the many weighty objections to our ideas would be to construct a machine that actually carries out satisfactory translation, some mitigating comments are in order.

We do not hope to partition uniquely the source sentence into locutions. In most cases, many partitions will be possible, each having a probability attached to it.

Whether "affecting" is to be translated as "a propos" or "concernant," or, as our dictionary has it, "touchant" or "emouvant," or in a variety of other ways, depends on the rest of the sentence. However, a statistical indication may be obtained from the presence or absence of particular guide words in that sentence. The statistical technique of decision trees [9] can be used to determine the guide word set, and to estimate the probability to be attached to each possible translate.

The sequential arrangement of target words obtained from the glossary may depend on an analysis of the source sentence. For instance, clause correspondence may be insisted upon, in which case only permutations of words which originate in the same source clause would be possible. Furthermore, the character of the source clause may affect the probability of use of certain function words in the target clause. There is, of course, nothing to prevent the use of more detailed information about the structure of the parse of the source sentence. However, preliminary experiments presented below indicate that only a very crude grammar may be needed (see Section 6).

3. CREATING THE GLOSSARY, FIRST ATTEMPT

We have already indicated in the previous section why creating a glossary is not just a matter of copying some currently available dictionary into the computer. In fact, in the paired sentences of Figure 1, "affecting" was translated as "a propos," a correspondence that is not ordinarily available. Laying aside for the time being the desirability of (idiomatic) word cluster - to - word cluster translation, what we are after at first is to find for each word f in the (French) source language the list of words $\{e_1, e_2, ..., e_s\}$ of the (English) target language into which f can translate, and the probability $P(e_i \mid f)$ that such a translation takes place.

A first approach to a solution that takes advantage of a large data base of paired sentences (referred to as 'training text') may be as follows. Suppose for a moment that in every French / English sentence pair each French word f translates into one and only one English word e, and that this word is somehow revealed to the computer. Then we could proceed by:

1. Establish a counter $C(e_i, f)$ for each word e_i of the English vocabulary. Initially set $C(e_i, f) = 0$ for words e_i. Set J = 1.
2. Find the Jth occurrence of the word f in the French text. Let it take place in the Kth sentence, and let its translate be the qth word in the Kth English sentence $E = e_{i_1}, e_{i_2}, \ldots, e_{i_n}$. Then increment by 1 the counter $C(e_{i_q}, f)$.
3. Increase J by 1 and repeat steps 2 and 3.

Setting $M(f)$ equal to the sum of all the counters $C(e_i, f)$ at the conclusion of the above operation (in fact, it is easy to see that $M(f)$ is the number of occurrences of f in the total French text), we could then estimate the probability $P(e_i \mid f)$ of translating the word f by the word e_i by the fraction $C(e_i, f)/M(f)$.

The problem with the above approach is that it relies on correct identification of the translates of French words, i.e., on the solution of a significant part of the translation problem. In the absence of such identification, the obvious recourse is to profess complete ignorance, beyond knowing that the translate is one of the words of the corresponding English sentence, each of its words being equally likely. Step 2 of the above algorithm then must be changed to

2'. Find the Jth occurrence of the word f in the French text. Let it take place in the Kth sentence, and let the Kth English sentence consist of words $e_{i_1}, e_{i_2}, ..., e_{i_n}$. Then increment the counters $C(e_{i_1}, f), C(e_{i_2}, f), ..., C(e_{i_n}, f)$ by the fraction $1/n$.

This second approach is based on the faith that in a large corpus, the frequency of occurrence of true translates of f in corresponding English sentences would overwhelm that of other candidates whose appearance in those sentences is accidental. This belief is obviously flawed. In particular, the article "the" would get the highest count since it would appear multiply in practically every English sentence, and similar problems would exist with other function words as well.

What needs to be done is to introduce some sort of normalization that would appropriately discount for the expected frequency of occurrence of words. Let $P(e_i)$ denote the probability (based on the above procedure) that the word e_i is a translate of a randomly chosen French word. $P(e_i)$ is given by

$$P(e_i) = \sum_{f'} P(e_i \mid f') P(f') = \sum_{f'} P(e_i \mid f') M(f')/M \tag{3.1}$$

where M is the total length of the French text, and $M(f')$ is the number of occurrences of f' in that text (as before). The fraction $P(e_i \mid f) \,/\, P(e_i)$ is an indicator of the strength of association of e_i with f, since $P(e_i \mid f)$ is normalized by the frequency $P(e_i)$ of associating e_i with an average word. Thus it is reasonable to consider e_i a likely translate of f if $P(e_i \mid f)$ is sufficiently large.

The above normalization may seem arbitrary, but it has a sound underpinning from the field of Information Theory [10]. In fact, the quantity

$$I(e_i; f) = \log \frac{P(e_i \mid f)}{P(e_i)} \tag{3.2}$$

is the mutual information between the French word f and the English word e_i.

Unfortunately, while normalization yields ordered lists of likely English word translates of French words, it does not provide us with the desired probability values. Furthermore, we get no guidance as to the size of a threshold T such that e_i would be a candidate translate of f if and only if

$$I(e_i; f) > T \tag{3.3}$$

Various ad hoc modifications exist to circumvent the two problems. One might, for instance, find the pair e_i, f with the highest mutual information, eliminate e_i and f from all corresponding sentences in which they occur (i.e. decide once and for all that in those sentences e_i is the translate of f!), then re-compute all the quantities over the shortened texts, determine the new maximizing pair e_i', f' and continue the process until some arbitrary stopping rule is invoked.

Before the next section introduces a better approach that yields probabilities, we present in Figure 2 a list of high mutual

information English words for some selected French words. The reader will agree that even the flawed technique is quite powerful.

4. A SIMPLE GLOSSARY BASED ON A MODEL OF THE TRANSLATION PROCESS

We will now revert to our original ambition of deriving probabilities of translation, $P(e_i \mid f)$. Let us start by observing that the algorithm of the previous section has the following flaw: Should it be "decided" that the qth word, e_{i_q}, of the English sentence is the translate of the rth word, f_{j_r}, of the French sentence, that process makes no provision for removing e_{i_q} from consideration as a candidate translate of any of the remaining French words (those not in the rth position)! We need to find a method to decide (probabilistically !) which English word was generated by which French one, and then estimate $P(e_i \mid f)$ by the relative frequency with which f gave rise to e_i as "observed" in the texts of paired French / English sentence translates. Our procedure will be based on a model (an admittedly crude one) of how English words are generated from their French counterparts.

With a slight additional refinement to be specified in the next section (see the discussion on position distortion), the following model will do the trick. Augment the English vocabulary by the NULL word e_0 that leaves no trace in the English text. Then each French word f will produce exactly one 'primary' English word (which may be, however, invisible). Furthermore, primary English words can produce a number of secondary ones.

The provisions for the null word and for the production of secondary words will account for the unequal length of corresponding French and English sentences. It would be expected that some (but not all) French function words would be killed by producing null words, and that English ones would be created by secondary production. In particular, in the example of Figure 1, one would expect that "reflects" would generate both "which" and "on" by secondary production, and "rise" would similarly generate "on." On the other hand, the article "l'" of "l'Orateur" and the preposition "a" of "a propos" would both be expected to generate a null word in the primary process.

This model of generation of English words from French ones then requires the specification of the following quantities:

1. The probabilities $P(e_i \mid f)$ that the ith word of the English dictionary was generated by the French word f.
2. The probabilities $Q(e_j \mid e_i)$ that the jth English word is generated from the ith one in a secondary generation process.
3. The probabilities $R(k \mid e_i)$ that the ith English word generates exactly k other words in the secondary process. By convention, we set $R(0 \mid e_0) = 1$ to assure that the null word does not generate any other words.

The model probability that the word f generates e_{i_1} in the primary process, and $e_{i_2},\ldots,e_{i_k}$ in the secondary one, is equal to the product

$$P(e_{i_1} \mid f)\ R(k-1 \mid e_{i_1})\ Q(e_{i_2} \mid e_{i_1})\ Q(e_{i_3} \mid e_{i_1}) \ldots Q(e_{i_k} \mid e_{i_1}) \qquad (4.1)$$

Given a pair of English and French sentences E and F, by the term generation pattern \$ we understand the specification of which English words were generated from which French ones, and which secondary words from which primary ones. Therefore, the probability P(E,\$ | F) of generating the words of E in a pattern \$ from those of F is given simply by a product of factors like (4.1), one for each French word. We can then think of estimating the probabilities $P(e_i \mid f)$, $R(k \mid e_i)$, and $Q(e_j \mid e_i)$ by the following algorithm at the start of which all counters are set to 0:

1. For a sentence pair E,F of the texts, find that pattern \$ that gives the maximal value of P(E,\$ | F), and then make the (somewhat impulsive) decision that that pattern \$ actually took place.
2. If in the pattern \$, f gave rise to e_i, augment counter $CP(e_i, f)$ by 1; if e_i gave rise to k secondary English words, augment counter $CR(k, e_i)$ by 1; if e_j is any (secondary) word that was given rise to by e_i, augment counter $CQ(e_j, e_i)$ by 1.
3. Carry out steps 1 and 2 for all sentence pairs of the training text.
4. Estimate the model probabilities by normalizing the corresponding counters, i.e.,

$$P(e_i \mid f) = CP(e_i, f)/CP(f) \quad where\ \ CP(f) = \sum_i CP(e_i, f)$$

$$R(k \mid e_i) = CR(k, e_i)/CR(e_i) \quad where\ \ CR(e_i) = \sum_k CR(k, e_i)$$

$$Q(e_j \mid e_i) = CQ(e_j, e_i)/CQ(e_i) \quad where\ \ CQ(e_i) = \sum_j CQ(e_j, e_i)$$

The problem with the above algorithm is that it is circular: in order to evaluate P(E,\$ | F) one needs to know the probabilities $P(e_i \mid f)$, $R(k \mid e_i)$, and $Q(e_j \mid e_i)$ in the first place! Fortunately, the difficulty can be alleviated by use of iterative re-estimation, which is a technique that starts out by guessing the values of unknown quantities and gradually re-adjusts them so as to account better and better for given data [11].

More precisely, given any specification of the probabilities $P(e_i \mid f)$, $R(k \mid e_i)$, and $Q(e_j \mid e_i)$, we compute the probabilities P(E,\$ | F) needed in step 1, and after carrying out step 4, we use the freshly obtained probabilities $P(e_i \mid f)$, $R(k \mid e_i)$, and $Q(e_j \mid e_i)$ to repeat the process from step 1 again, etc. We halt the computation when the obtained estimates stop changing from iteration to iteration.

While it can be shown that the probability estimates obtained in the above process will converge [11,12], it cannot be proven that the values obtained will be the desired ones. A heuristic argument can be formulated making it plausible that a more complex but computationally excessive version [13] will succeed. Its truncated modification leads to a glossary that seems a very satisfactory one. We present some interesting examples of its $P(e_i \mid f)$ entries in Figure 3.

Two important aspects of this process have not yet been dealt with: the initial selection of values of $P(e_i \mid f)$, $R(k \mid e_i)$, and $Q(e_j \mid e_i)$, and a method of finding the pattern \$ maximizing P(E,\$ | F).

A good starting point is as follows:

A. Make $Q(e_j \mid e_i) = 1/K$, where K is the size of the English vocabulary.

B. Let $R(1 \mid e_i) = 0.8$, $R(0 \mid e_i) = 0.1$, $R(2 \mid e_i) = R(3 \mid e_i) = R(4 \mid e_i) = R(5 \mid e_i) = 0.025$ for all words e_i except the null word e0. Let $R(0 \mid e_0) = 1.0$.

C. To determine the initial distribution $P(e_i \mid f)$ proceed as follows:

(i) Estimate first $P(e_i \mid f)$ by the algorithm of Section 3.

(ii) Compute the mutual information values $I(e_i; f)$ by formula (3.2), and for each f find the 20 words e_i for which $I(e_i; f)$ is largest.

(iii) Let $P(e_0 \mid f) = P(e_i \mid f) = (1/21) - \varepsilon$ for all words e_i on the list obtained in (ii), where ε is some small positive number. Distribute the remaining probability ε uniformly over all the English words not on the list.

Finding the maximizing pattern \$ for a given sentence pair E, F is a well-studied technical problem with a variety of computationally feasible solutions that are suboptimal in some practically unimportant respects [14]. Not to interrupt the flow of intuitive ideas, we omit the discussion of the corresponding algorithms.

5. TOWARD A COMPLEX GLOSSARY

In the previous section we have introduced a technique that derives a word - to - word translation glossary. We will now refine the model to make the probabilities a better reflection of reality, and then outline an approach for including in the glossary the fixed locutions discussed in Section 2.

It should be noted that while English / French translation is quite local (as illustrated by the alignment of Figure 1), the model leading to (4.1) did not take advantage of this affinity of the two languages: the relative position of the word translate pairs in their respective sentences was not taken into account. If m and n denote the respective lengths of corresponding French and English sentences, then the probability that e_{i_k} (the kth word in the English sentence) is a primary translate of f_{j_h} (the hth word in the French sentence) should more accurately be given by the probability $P(e_{i_k}, k \mid f_{j_h}, h, m, n)$ that depends both on word positions and sentence lengths. To keep the formulation as simple as possible, we can restrict ourselves to the functional form

$$P(e_{i_k}, k \mid f_{j_h}, h, m, n) = PW(e_{i_k} \mid f_{j_h}) \;\; PD(k \mid h, m, n) \tag{5.1}$$

In (5.1) we make the 'distortion' distribution $PD(k \mid h,m,n)$ independent of the identity of the words whose positional discrepancy it describes.

As far as secondary generation is concerned, it is first clear that the production of preceding words differs from that of those that follow. So the R and Q probabilities should be split into left and right probabilities RL and QL, and RR and QR. Furthermore, we should provide the Q -probabilities with their own distortion components that would depend on the distance of the secondary word from its primary 'parent'. As a result of these considerations, the probability that f_{j_h} generates (for instance) the primary words e_{i_k} and preceding and following secondary words $e_{i_{k-1}}$, $e_{i_{k+2}}$ would be given by

$$PW(e_{i_k} \mid f_{j_h}) \; PD(k \mid h,m,n) \; RL(2 \mid e_{i_k}) \;\; RR(1 \mid e_{i_k}) \tag{5.2}$$

$$QL(e_{i_{k-3}}, 3 \mid e_{i_k}) \; QL(e_{i_{k-1}}, 1 \mid e_{i_k}) \; QR(e_{i_{k+2}}, 2 \mid e_{i_k})$$

Obviously, other distortion formulations are possible. The purpose of any is to sharpen the derivation process by restricting the choice of translates to the positionally likely candidates in the corresponding sentence.

To find fixed locutions in English, we can use the final probabilities QL and QR obtained by the method of the previous section to compute mutual informations between primary and secondary word pairs,

$$IR(e;e') = \log \frac{QR(e' \mid e)}{P(e')} \tag{5.3}$$

and

$$IL(e';e) = \log \frac{QL(e' \mid e)}{P(e')}$$

where $P(e') = C(e')/N$ is the relative frequency of occurrence of the secondary word e' in the English text ($C(e')$ denotes the number of occurrences of e' in the text of size N), and QR and QL are the average secondary generation probabilities,

$$QR(e' \mid e) = \sum_i QR(e', i \mid e) \tag{5.4}$$

and

$$QL(e' \mid e) = \sum_i QR(e', i \mid e)$$

We can then establish an experimentally appropriate threshold T, and include in the glossary all pairs (e, e') and (e', e) whose mutual information exceeds T.

While the process above results in two-word fixed locutions, longer locutions can be obtained iteratively in the next round after the two-word variety had been included in the glossary and in the formulation of its creation.

To obtain French locutions, one must simply reverse the direction of the translation process, making English and French the source and target languages, respectively.

With two-word locutions present in both the English and French parts of the glossary, it is necessary to reformulate the generation process (4.1). The change would be minimal if we could decide to treat the words of a locution (f, f') as a single word $f^* = (f, f')$ rather than as two separate words f and f' whenever both are found in a sentence. In such a case nothing more than a recoding of the French text would be required. However, such a radical step would almost certainly be wrong: it could well connect auxiliaries and participles that were not part of a single past construction. Clearly then, the choice between separateness and unity should be statistical, with probabilities estimated in the overall glossary construction process and initialized according to the frequencies with which elements of the pair f, f' were associated or not by secondary generation when they appeared in the same sentence.

Since the approach of this section was not yet used to obtain any results, we will leave its complete mathematical specification to a future report.

6. GENERATION OF TRANSLATED TEXT

We have pointed out in Section 2 that translation can be somewhat naïvely regarded as a three stage process:

(1) Partition the source text into a set of fixed locutions.
(2) Use the glossary plus contextual information to select the corresponding set of fixed locutions in the target language.
(3) Arrange the words of the target fixed locutions into a sequence forming the target sentence.

We have just finished arguing in Section 5 that the partitioning of source text into locutions is somewhat complex, and that it must be approached statistically. The basic idea of using contextual information to select the correct 'sense' of a locution is to construct a contextual glossary based on a probability of the form $P(e \mid f, \psi[F])$ where e and f are English and French locutions, and $\psi[F]$ denotes a 'lexical' equivalence class of the sentence F. The test of class membership would typically depend on the presence of some combination of words in F. The choice of an appropriate equivalence classification scheme would, of course, be the subject of research based on yet another statistical formulation. The estimate of $P(e \mid f, \psi[F])$ would be derived from counts of locution alignments in sentence translate pairs, the alignments being estimated based on non-contextual glossary probabilities of the form (5.2).

The last step in our translation scheme is the re-arrangement of the words of the generated English locutions into an appropriate sequence. To see whether this can be done statistically, we explored what would happen in the impossibly optimistic case where the words generated in (2) were exactly those of the English sentence (only their order would be unknown):

From a large English corpus we derived estimates of trigram probabilities, $P(e_3 \mid e_1, e_2)$, that the word e_3 follows immediately the sequence pair e_1, e_2. A model of English sentence production based on a trigram estimate would conclude that a sentence $e_1, e_2, \ldots e_n$ is generated with probability

$$P(e_1, e_2)\ P(e_3 \mid e_1, e_2)\ P(e_4 \mid e_2, e_3) \ldots P(e_n \mid e_{n-2}, e_{n-1}) \qquad (6.1)$$

We then took other English sentences (not included in the training corpus) and determined which of the $n!$ different arrangements of their n words was most likely, using the formula (6.1). We found that in 63% of sentences of 10 words or less, the most likely arrangement was the original English sentence. Furthermore, the most likely arrangement preserved the meaning of the original sentence in 79% of the cases.

Figure 4 shows examples of synonymous and non-synonymous re-arrangements.

We realize that very little hope exists of the glossary yielding the words and only the words of an English sentence translating the original French one, and that, furthermore, English sentences are typically longer than 10 words. Nevertheless, we feel that the above result is a hopeful one for future statistical translation methods incorporating the use of appropriate syntactic structure information.

REFERENCES

[1] L.R. Bahl, F. Jelinek, and R.L. Mercer: A maximum likelihood approach to continuous speech recognition, **IEEE Transactions on Pattern Analysis and Machine Intelligence,** PAMI-5(2):179-190, March 1983.

[2] J.K. Baker: Stochastic modeling for automatic speech understanding. In R.A. Reddy, editor, **Speech Recognition,** pages 521-541, Academic Press, New York, 1979.

[3] J.D. Ferguson: Hidden Markov analysis: An introduction. In J.D. Ferguson, Ed., **Hidden Markov Models for Speech.** Princeton, New Jersey, IDA-CRD, Oct. 1980, pp. 8-15

[4] J. McH. Sinclair: "Lexicographic Evidence" in, **Dictionaries, Lexicography and Language Learning** (ELT Documents: 120), editor R. Ilson, New York: Pergamon Press, pp. 81-94, 1985.

[5] R.G. Garside, G.N. Leech and G.R. Sampson, **The Computational Analysis of English: a Corpus-Based Approach,** Longman 1987.

[6] G.R. Sampson, "A Stochastic Approach to Parsing" in, **Proceedings of the 11th International Conference on Computational Linguistics** (COLING '86) Bonn 151-155, 1986.

[7] W. Weaver: **Translation** (1949). Reproduced in: Locke, W.N. & Booth, A.D. eds.: **Machine translation of languages.** Cambridge, MA.: MIT Press, 1955.

[8] Hansards: **Official Proceedings of the House of Commons of Canada, 1974-78,** Canadian Government Printing Bureau, Hull, Quebec Canada.

[9] L. Breiman, J.H. Friedman, R.A. Olshen, and C.J. Stone: **Classification and Regression Trees,** Wadsworth and Brooks, Monterey, CA, 1984.

[10] R.G. Gallager: **Information Theory and Reliable Communication,** John Wiley and Sons, Inc., New York, 1968.

[11] A.P. Dempster, N.M. Laird, and D.B. Rubin: Maximum likelihood from incomplete data via the EM algorithm, **Journal of the Royal Statistical Society,** 39(B):1-38, 1977.

[12] A.J. Viterbi: Error bounds for convolutional codes and an asymtotically optimum decoding algorithm, **IEEE Transactions on Information Theory,** IT-13:260-267, 1967.

[13] L.E. Baum: An inequality and associated maximization technique in statistical estimation of probabilistic functions of a Markov process, **Inequalities,** 3:1-8, 1972.

[14] F. Jelinek: A fast sequential decoding algorithm using a stack, **IBM T. J. Watson Research Development,** vol. 13, pp. 675-685, Nov. 1969.

Mr. Speaker , I rise on a question of privilege
Monsieur l'Orateur , je souleve la question de privilege

affecting the rights and prerogatives of parliamentary committees
a propos des droits et des prerogatives des comites parlementaires

and one which reflects on the word of two ministers
et pour mettre en doute les propos de deux ministres

of the Crown.
de la Couronne.

FIGURE 1
ALIGNMENT OF A FRENCH AND ENGLISH SENTENCE PAIR

eau	water
lait	milk
banque	bank
banques	banks
hier	yesterday
janvier	January
jours	days
votre	your
enfants	children
trop	too
toujours	always
trois	three
monde	world
pourquoi	why
aujord'hui	today
sans	without
lui	him
mais	but
suis	am
seulement	only
peut	cannot
ceintures	seat
ceintures	belts
bravo	!

FIGURE 2
A LIST OF HIGH MUTUAL INFORMATION FRENCH-ENGLISH WORD PAIRS

	WHICH		QUI	
1.	qui	0.380	who	0.188
2.	que	0.177	which	0.161
3.	dont	0.082	that	0.084
4.	de	0.060	.	0.038
5.	d'	0.035	to	0.032
6.	laquelle	0.031	of	0.027
7.	ou	0.027	the	0.026
8.	et	0.022	what	0.018

	THEREFORE		DONC	
1.	donc	0.514	therefore	0.322
2.	consequent	0.075	so	0.147
3.	par	0.074	is	0.034
4.	ce	0.066	then	0.024
5.	pourquoi	0.064	thus	0.022
6.	alors	0.025	the	0.018
7.	il	0.025	that	0.013
8.	aussi	0.015	us	0.012

	STILL		ENCORE	
1.	encore	0.435	still	0.181
2.	toujours	0.230	again	0.174
3.	reste	0.027	yet	0.148
4.	***	0.020	even	0.055
5.	quand	0.018	more	0.046
6.	meme	0.017	another	0.030
7.	de	0.015	further	0.021
8.	ne	0.014	once	0.013

FIGURE 3 (PART I)
EXAMPLES OF PARTIAL GLOSSARY LISTS OF MOST LIKELY WORD TRANSLATES AND THEIR PROBABILITIES

Note: *** denotes miscellaneous words not belonging to the lexicon.

	PEOPLE		GENS	
1.	les	0.267	people	0.781
2.	gens	0.244	they	0.013
3.	personnes	0.100	those	0.009
4.	population	0.055	individuals	0.008
5.	peuple	0.035	persons	0.005
6.	canadiens	0.031	people's	0.004
7.	habitants	0.024	men	0.004
8.	ceux	0.023	person	0.003

	OBTAIN		OBTENIR	
1.	obtenir	0.457	get	0.301
2.	pour	0.050	obtain	0.108
3.	les	0.033	have	0.036
4.	de	0.031	getting	0.032
5.	trouver	0.026	seeking	0.023
6.	se	0.025	available	0.021
7.	obtenu	0.020	obtaining	0.021
8.	procurer	0.020	information	0.016

	QUICKLY		RAPIDEMENT	
1.	rapidement	0.508	quickly	0.389
2.	vite	0.130	rapidly	0.147
3.	tot	0.042	fast	0.052
4.	rapide	0.021	quick	0.042
5.	brievement	0.019	soon	0.036
6.	aussitot	0.013	faster	0.035
7.	plus	0.012	speedy	0.026
8.	bientot	0.012	briefly	0.025

FIGURE 3 (PART II)
EXAMPLES OF PARTIAL GLOSSARY LISTS OF MOST LIKELY WORD TRANSLATES AND THEIR PROBABILITIES

EXAMPLES OF RECONSTRUCTION THAT PRESERVE MEANING:

would I report directly to you?
I would report directly to you?

now let me mention some of the disadvantages.
let me mention some of the disadvantages now.

he did this several hours later.
this he did several hours later.

EXAMPLES OF RECONSTRUCTION THAT DO NOT PRESERVE MEANING

these people have a fairly large rate of turnover.
of these people have a fairly large turnover rate.

in our organization research has two missions.
in our missions research organization has two.

exactly how this might be done is not clear.
clear is not exactly how this might be done.

FIGURE 4
STATISTICAL ARRANGEMENT OF WORDS BELONGING TO ENGLISH SENTENCES