Industry Canada
Centre for Information Technology Innovation

Can Terminological Consistency be Validated Automatically?

Elliott Macklovitch

To appear in the Proceedings of the IV[es] Journées scientifiques,
Lexicommatique et Dictionnairiques, organized by AUPELF–UREF,
Lyon, France, September 28-30, 1995.

Laval
December 1995

This document reports on research carried out at the Centre for Information Technology Innovation (CITI). The views expressed are strictly those of the author.

Également disponible en français.

Catalogue No. Co28-1/125-1995E
ISBN No. 0-662-24011-1

Can Terminological Consistency Be Validated Automatically?

Elliott Macklovitch[1]

1. Introduction

It often happens in translation services that lengthy texts have to be divided up among several translators, some of whom may be freelancers who work outside the service. In such situations, it is generally the reviser's job to piece together the parts translated by different people and to ensure that the resulting final text is coherent. One particularly arduous aspect of this job is to see to it that the terminology of the final text is consistent, or uniform. Intuitively, it is quite clear what we mean by terminological consistency here: each terminological unit should receive the same translation throughout the final text, so that readers are not unduly confused.

Terminological consistency is generally accepted as being one property of a good translation, and of course the situation described above is not the only one in which it comes into play. At the CITI, we are currently developing a novel kind of machine-aided system that is specifically designed to support human translators in the revision process by validating certain properties of a translated text. The system is called TransCheck, and in its first prototype version it is capable of detecting some of the more frequently occurring types of translation errors, including deceptive cognates, calques, illicit borrowings, and other sorts of translation improprieties. It is important to note that most of these errors are generally beyond the reach of monolingual writing aids such as spelling or grammar checkers, precisely because translation errors are *bilingual* in nature and depend upon relations that exist between two texts in different languages. TransCheck, on the other hand, *can* detect these errors, because the system was specially designed to handle this particular kind of parallel text. To be more precise, TransCheck seeks to reconstitute part of the human translation process by **automatically aligning** two texts; that is to say, the system attempts to explicitly link various segments in a source language text with what it automatically determines to be the corresponding segments in its target language translation. For a detailed description of the first TransCheck prototype, see (Macklovitch 1994); the following gives a general idea of how users might employ such a translation checker.

Before TransCheck can verify any properties of a translation, the source and target language files must be submitted to the system for alignment. The actual algorithms that the program uses to automatically calculate the correct correspondences between the two texts need not concern us here. Suffice it to say that when the resolution of the alignments does not go beyond the level of the sentence, the program is highly accurate;[2] furthermore, it is capable of handling cases where a sentence in one language is translated by two or even three sentences in the other language, and vice versa. Following (Harris 1988), we will call the output of such an alignment program a "bi-text". Now suppose that a reviser wanted to validate a draft translation before sending it out to the

1. Please see the acknowledgements on page 14.
2. For more details on the CITI's current alignment algorithms, see (Simard et al. 1992).

client, to ensure that it was free of source language interference; this too is a generally accepted property of a good translation. S/he could call upon TransCheck to help do this, in much the same manner that monolingual writers commonly use a spelling checker to ensure that their texts are free of spelling errors. The CITI's first TransCheck prototype incorporates a database of approximately 2800 prohibited translation pairs, including many of the classic examples of deceptive cognates, like "library//librairie" and "deception//déception". Concretely, what the system does is take each of the entries in this database and apply it in turn to the bi-text produced by the alignment program. If it finds any SL segment containing "library", for example, that is aligned with a TL segment containing "librairie", it flags that aligned pair for the reviser's attention. During a subsequent editing session, the reviser reviews all the flagged segments and makes any necessary corrections to the target text, including those that the system itself suggests, drawn from each database entry. (A screen dump of a TransCheck editing session appears on the following page.)

This basically is how TransCheck operates to detect cases of source language interference in a draft translation. We are in the process of working on a number of extensions to the first TransCheck prototype that will hopefully allow the system to automatically detect the omission of major textual units and to verify the correct transposition of various types of numerical expressions. Another important question that we are currently exploring is whether a tool like TransCheck could be of help in verifying the terminological consistency of a draft translation. It is the results of these first experiments on terminology that I will be reporting on here. Again, the basic idea is quite simple and actually quite similar to the way in which cases of SL interference are detected. Suppose we adopt as our starting point the naive definition of terminological consistency given above, namely that each occurrence of a designated source term must be translated quite literally as a specified target term. Suppose too that the reviser is able to enumerate the terms s/he wants checked for consistency before beginning the revision of the draft translation. This might take the form of a text-specific glossary in which each entry is a simple term equivalence statement – nothing more than "e-term1 = f-term1". A bi-text would be produced from the source text and the draft translation, as before. The system would then convert the entries in the reviser's term equivalence glossary into a series of TransCheck queries and apply each in turn to the bi-text. Those aligned segments found to contain one of the specified source terms BUT NOT the corresponding target term would once again be flagged for the reviser's attention.

In order to test this idea, we decided to undertake a small-scale feasibility study, although at the outset, our expectation was that this schema would be altogether too simplistic to allow for the development of an operational term checking system. But where exactly would it fail? What are the kinds of problems it would run up against and what is their relative importance? Which of these problems would be amenable to short-term solutions and which would have to await the results of long-term research? To help us answer questions like these, which are crucial if we are to eventually develop an operational term consistency checker, we decided to implement a rudimentary version of the schema outlined above within the current TransCheck prototype, and to apply it to a number of authentic translations.

TransCheck : a screen dump

```
*TransCheck Text*
Buffers   File   Edit   Help
That is all I ask.                          Je n'en demande pas plus.

If you have a marketing and sales guy      Le représentant des ventes et de la
to sell Dash-8s, where does he get         commercialisation des Dash-8, de qui
his direction right now?                   reçoit-il ses directives en ce
                                           moment?

The history of the boundary, the A-B       L'histoire de l'établissement de la
line and this body of water which is       frontière, de la ligne AB et de ce
sovereign and internal to Canada, is       plan d'eau, qui fait partie des eaux
one of a trail of bullying, broken         intérieures canadiennes, est marquée
promises, treachery, deception, lies       par l'intimidation, les promesses
and treason, and it carries on to          brisées, la trahison, la déception
this very day.                             ainsi que le mensonge, et ça se
                                           continue jusqu'à ce jour.

As well, the 1987 CPP changes              Les modifications apportées au RPC en
included the payment of two benefits       1987 permettaient, en outre, le
--%%-Emacs: *TransCheck Text*      (TransCheck)--66%---------------------------
<Couple id=357>
<Error no=357.1>
  Compiled query: f(déception+/NomC\*) e(deception+/NC\*)
  Français incorrect: déception+/NomC
  Anglais:            deception+/NC
  Français correct:   tromperie ; duperie
  Type d'erreur:      faux ami complet
  Source:             Van Roey
</Error>

<Text>
CoupleInfo: Couple ID = 357,
            Document = "text",
            Source = English
  tokens[L1] = {13567}
  tokens[L2] = {15913}
The:                DT                     L':                Déte-ddéf-fémi-sing
history:            NC                     histoire:          NomC-fémi-sing
of:                 PP                     de:                Prép
the:                DT                     l':                Déte-ddéf-masc-sing
boundary:           NC                     établissement:     NomC-masc-sing
,:                  PC                     de:                Prép
the:                DT                     la:                Déte-ddéf-fémi-sing
A:                  ZZ                     frontière:         NomC-fémi-sing
-:                  PC                     ,:                 Punc-pccm
--%%-Emacs: *TransCheck Info*      (TransCheck Narrow)--Top--------------------
Garbage collecting...done
```

2. The Feasibility Study

2.1 Methodology

For our feasibility study, we sought to obtain a number of texts from different domains, each with two versions of the target translation: a preliminary or draft version, and a final revised version. (Recall that TransCheck is meant to flag potential errors in a *draft* translation.) It turns out, however, that it is not as easy as one might imagine to obtain authentic draft translations, however these are defined. When asked for texts to be used in experiments on error detection, translators and translation services are understandably reticent to hand over their unfinished products, even when they are told that it is strictly for research purposes and reassured that their identities will be kept anonymous. As a result, we were forced to make certain compromises in our methodology, particularly with regard to the first of the texts that we analysed.

For each of the four source texts we did obtain, we selected twenty-five of the most frequent or salient terms, with the help of an in-house program for candidate term extraction called F-TERM. F-TERM is based on ideas first proposed in (Justeson & Katz 1993). It operates on a text that has previously been assigned part-of-speech tags and extracts sequences of words that correspond to a syntactic definition of a multi-word term: for English, this is basically a noun phrase stripped of its determiner and consisting of a string of nouns and/or adjectives ending in a noun and followed by an optional prepositional phrase. Again, it should be emphasized that what the program produces is a list of *candidate* terms, which are sorted by the frequency of their appearance in the text; most of these, at least at the top of the list, do turn out to be valid terms, however. On the other hand, not all the terms in a text are found in the list. For one thing, F-TERM ignores single-word terms; for another, automatic tagging problems can lead to the inclusion of sequences that are not well-formed noun phrases and to the omission of others that are. Most importantly, F-TERM has no notion of what distinguishes a non-lexical (or descriptive) NP from a bona-fide term, apart from the literal repetition of the expression in the text.[3]

We located the translation of each selected source term in the final version of the corresponding target language text. When the TL text contained conflicting equivalents for the same source term, we selected the most frequently occurring target term, occasionally appealing to TERMIUM, the Canadian government's well-known term bank, to help us arbitrate. These translations reflect part of the reviser's decisions on the proper terminology for the text, and they were formalized as the simple term equivalency statements and converted into TransCheck queries, as outlined above. TransCheck could then scan the bi-text produced from each source text and its draft translation, and flag all cases in which a source term was not rendered **exactly** as the specified target term, or an inflectional variant of that target term.[4] Finally, by analysing the system's output and comparing the potential inconsistencies flagged by TransCheck with the terminology of the final version, we hoped be able to get a clearer idea of the major difficulties facing this type of approach.[5]

---

3. See (Justeson & Katz 1993) for an interesting discussion of repetition as a diagnostic criterion of term status, and (Hannan 1995) for a critical discussion of the limits of syntactically-based approaches to automatic term extraction.
4. An inflectional variant here signifies the plural form of nouns, plural and feminine forms of French adjectives, etc.

2.2 The texts

Given our initial difficulty in obtaining draft translations, we decided to proceed with our first test on term consistency checking using an 80-thousand word army manual on sniper training and deployment, for which we had the English original and a final French version but no preliminary draft translation. Moreover, this was a manual that had actually been published in both official languages, and so presumably the terminology in the final French version had already been checked for consistency. These may at first appear to be formidable obstacles; however, at this point, we were more interested in the types of noise[6] that TransCheck would generate when called upon to verify term consistency than in any isolated errors the reviser of this text may have overlooked.

The twenty-five terms that we selected for the Sniper text are listed in the first column of Table 1, which appears at the end of this paper. The third column of the Table gives the target term (TT) that corresponds to each source term in the final French translation. Column 2 indicates the total number of occurrences of each source term (ST), and column 4, the number of times which that source term is NOT translated exactly as the target term. Taking the first entry in the Table, for example, we see that "sniper" appears a total of 1277 times in the text, and that of these, 105 occurrences are not translated as "tireur d'élite". The remaining five columns in Table 1 provide a breakdown of these cases in which ST =//= TT. "Head only" refers to cases in which a multi-word target term is truncated, so that only the head word is used instead of the full term; see (1) below for examples. The next column, "Pron", is for cases in which the French text employs a pronoun or other kind of anaphor instead of the full target noun phrase; see (2) below for examples. The column headed "ST not trans" refers to cases in which the source term is not actually translated in the target text: either it is entirely omitted, or in some cases the translation provides a paraphrase instead of an equivalent term; see (3) below for an example of each. The "Alt.TT" column is for those cases in which the French text employs an alternative term to the one given in the third column. In a few instances in the Sniper text, these may in fact be true terminological inconsistencies, but, as we would expect in a published manual, they are not very numerous; some possible examples are given in (4) below. The final column, "Other", is for all cases that do not fall into the preceding categories. These include instances of coordination and other grammatical constructions which fragment the target term (see 5 below for examples); system noise that arises due to tagging problems; and a variety of minor spelling or typing errors that are detected by TransCheck but do not really qualify as *terminological* inconsistencies.

(1a) ... the **sniper** moves his head back and forth =>
... le **tireur** déplace la tête vers l'avant ou vers...

5. Although we were aware right from the outset that the draft translations had not been subjected to a systematic revision, and hence that the final versions might well contain terminological inconsistencies that had not been corrected.
6. The term "noise" is employed here in a sense similar to that commonly used in document retrieval to designate results that are not pertinent to a query, in our case because they do not correspond to a true translation error.

(1b) The Unertl **telescopic sight** is a fixed 10 power... =>
La **lunette** Unertl grossit 10 fois et ...

(2a) The **sniper** places his hand to his chest... =>
Il place la main contre sa poitrine...

(2b) **Snipers** will infiltrate enemy areas... =>
Ces derniers infiltrent les secteurs de l'ennemi...

(3a) IDENTIFYING **SNIPER** TARGETS =>
IDENTIFICATION DES CIBLES

(3b) Adjustment for **eye relief** should be made by... =>
L'ajustement de la distance entre l'oeil et la lunette se fait en...

(4a) The sniper can insert a pad on the **ghillie suit**... =>
Le tireur d'élite peut insérer un coussin dans sa **tenue de camouflage**...

(4b) ...which is supported from underneath by the top of the **trigger**... =>
...qui est supportée du dessous par la tête du **percuteur**...

(5a) direct or **indirect fire** => **tirs** direct ou **indirect**

(5b) Determining correct **eye relief** => Déterminer le **dégagement** correct **de l'oeil**

We refer to the next text we analysed as the Dairy text; it came from the Department of Agriculture and is an economic analysis of the competitiveness of the Canadian milk and dairy products industry. The translation had been contracted out to a private sector service bureau; before being sent on to the client department, however, the text underwent a summary revision known as quality assurance, which was carried out by a senior translator in the government's own translation bureau. Although the outside service bureau did not specify how many translators had actually worked on the text, there are strong indications that at least two were involved: there is a change in the font style part way through the word processing file delivered by the service bureau; this coincides, moreover, with a change in the target terminology for a number of source terms. One of the concerns of the government reviser, therefore, was to ensure that in the final French translation the terminology employed was consistent throughout. Table 2 at the end of this paper lists the twenty-five source terms we selected for this text, along with their TL equivalents and the breakdown of all the cases of non-correspondence.

The third text, which we refer to as PIBD, is part of a manual from the Department of External Affairs describing departmental policy and procedures for organizing business promotion events, under the government's Program for International Business Development. The original English, which was approximately 16 thousand words long, divides into two distinct parts: a well-structured description, written in standard bureaucratese, of the program and its procedures; this is followed by a series of disjoint appendices, containing government forms, multiple choice questionnaires, sample contracts, etc. Here, we know for a fact that the two parts were assigned to different translators within the government's own translation bureau. The source terms we selected for this text and their target term equivalents appear in Table 3.

The final text we analysed, which we refer to as Cluster, was only 4235 words long, but was still assigned to two translators, because the time allowed for its translation was very short. The text came from the Department of Industry, Science and Technology, and is a call for proposals for a particular type of economic study to be conducted in western Canada, based on the concept of cluster analysis. Those parts of the RFP which describe the work to be done (the deliverables, dates, selection criteria, etc.) are intended for the general public, and are not very technical; but the text also provides some background information on the theory of cluster analysis, and so indirectly refers to a relatively specialized area of economics. Table 4 lists the source terms we selected for the Cluster text, along with their target term equivalents and the breakdown of the non-correspondences.

Before we examine the results TransCheck produced on these four texts, there is an important question of definition which we should address regarding our use of the word "term". At the outset of this study, we assumed that the units which a reviser would want to verify in a draft translation would all be bona-fide terms, in the technical sense of that word, i.e. elements of a specialized vocabulary describing the fundamental concepts and relations within a specialized domain. This assumption was not entirely warranted, however. One of the things we discovered in working with the reviser who performed the quality assurance on three of our texts is that the term/non-term distinction is not critical for the job of ensuring a consistent translation. Many of the units which the reviser had wanted to verify for consistency in these draft translations did turn out to be terms (with records in TERMIUM), but others did not; while in other cases still, the terminological status of the units was difficult to determine. In fact, the distinction between specialized terms and words of the general vocabulary in any given text is not always clear, even to a terminologist.[7] The reader should bear this in mind when examining the lists of source terms in Tables 1-4, since some of these – the acronyms, for example, which occur so frequently in the PIBD text – may not at first appear to be terms. We make no claims about the terminological status of the 100 units we selected for our feasibility study. Just as Martin Kay has defined translation as *what translators do*, so we would like TransCheck to support and assist revisers in what they do. Insofar as term consistency checking is concerned, our reviser informed us that the overriding goal was to avoid reader confusion; whether the units that need to be rendered uniformly in order to attain this goal are bona-fide terms or not is more or less incidental. For lack of a convenient alternative, however, we shall continue to employ the word "term" in this paper, although we would prefer not to be held to its technical definition.

3. Results

Table 5 on the next page presents a synthesis of the results which are tabulated separately for the four texts in Tables 1-4. Notice that the figures given here correspond to the total number of

7. Though, presumably, the distinction should be clear to an expert in the domain of the text. See (Ahmad et al. 1994) for an interesting discussion of the difficulties involved in defining term-hood in the context of automatic term extraction.

occurrences in each column of Tables 1-4, generated by the twenty-five pairs of terms that were selected for each of the Sniper, Dairy, PIBD and Cluster texts.

Let us focus, to begin with, on the figures in the third column of Table 5, which give the overall non-correspondence rates for each text, or the proportion of source term tokens that were not translated exactly as the target terms specified in that text's equivalence glossary. Notice that at 8.5%, the non-correspondence rate for the Sniper text is significantly lower than the non-correspondence rates for the other three texts. Recall, however, that the French version of the Sniper text submitted to TransCheck was actually a published translation which had presumably been verified for term consistency; whereas in the case of the other three texts, the system was verifying a preliminary translation. Hence, this discrepancy in the non-correspondence rates is not really an anomaly; indeed, if our hypothesis about how the texts were translated is correct, this is more or less what we would expect to find. Notice, moreover, that the non-correspondence rates on the three texts other than Sniper are relatively constant: 30% on both the Dairy and the Cluster texts, and 24% on PIBD.

Table 5 reveals another significant difference between the results obtained on the Sniper text and those of the other three translations: the number of cases in which a target term is rendered by a pronoun or other anaphor is much higher in Sniper (35 occurrences under "Pron." versus 11 in total for the other three texts); and the same is true, though to a lesser extent, for the number of occurrences of "Head only", or target term truncation (29% of non-correspondences in Sniper, versus 19%, 18% and 0% in the Dairy, PIBD and Cluster texts respectively.) Why this should be

Table 5: Cumulative results for the four texts:

Text / # words	# ST tokens	ST =//= TT # / %	Head only # / %	Pron. # / %	Alt. TT # / %	ST not trans # / %	Other # / %
Sniper / 80,000	1799	153 / 8.5%	44 / 29%	35 / 23%	10 / 6.5%	54 / 35%	10 / 6.5%
Dairy / 16,500	547	165 / 30%	31 / 19%	4 / 2%	36 / 22%	86 / 52%	8 / 4.8.%
PIBD / 16,00	367	87 / 24%	16 / 18%	5 / 5.7%	46 / 53%	10 / 11%	10 / 11%
Cluster / 4,235	211	64 / 30%	0	2 / 3%	35 / 55%	17 / 27%	10 / 16%
Total / 116,735	2924	469 /16%	91 / 19%	46 / 9.8%	127 / 27%	167 / 36%	38 / 8%
Total excl. Sniper	1125	316 / 28%	47 / 15%	11 / 3.5%	117 / 37%	113 / 36%	28 / 8.8%

so is not altogether obvious. It may have to do with the fact that the term “sniper” occurs so frequently in that text: hence, recourse to such anaphoric devices avoids the awkward repetition of “tireur d’élite”, allowing for a target text that is lighter and somewhat more varied.[8] Or perhaps the explanation has more to do with pragmatics: there is only one possible referent for “sniper” in this text (the prototypical trainee to whom this manual is addressed), whereas truncated terms in the other texts may be more ambiguous in the entities they refer to.[9]

Turning now to the other categories in Table 5, we observe that the proportion of non-correspondences attributable to alternative terms (Alt.TT) is much higher in the PIBD and Cluster texts than it is in the Sniper or Dairy texts. As it happens, PIBD and Cluster are the least technical of the four texts we analysed; that at least is the distinct impression one has in reading through these source texts. Actually, we may be able to corroborate this impression using the outputs of our F-TERM program. F-term, recall, locates sequences of words that correspond to a syntactic definition of a multi-word term. Following (Justeson & Katz 1993), we shall assume for the sake of this argument that most of the multi-word sequences which reoccur verbatim at least twice in a text are in fact terminological units. In the Cluster text, for example, F-TERM identifies 377 multi-word term candidates, of which only 49 occur more than once in the text. The total number of tokens corresponding to these 49 candidate terms is 125; and if we add to these the 43 occurrences of the single-word term “cluster”, which, like the word “sniper”, is the most frequent term in its text, we arrive at a total of 168 term tokens in a text that is 4235 words long. Dividing the total number of words in the text by the number of candidate term tokens should give us a very rough correlate of term density: in the case of the Cluster text, 25.21 words of text per term. By way of comparison, the same calculation for the Dairy text, which is perhaps the most technical of the four we analysed, yields a much lower ratio of words per term: 9.42. The figure for the Sniper text is 19.14; and for the PIBD text, which was even less technical than Cluster, 32.60. As discussed in section 2 above, we are aware that the results produced by F-TERM are not entirely reliable; all we need to assume here, however, is that F-TERM’s weaknesses are constant across various texts. These figures, then, do not have any absolute value; but they do seem to correlate with the *relative* term density of the four texts we analysed. Now term density is undoubtedly one element that contributes to the impression of text’s “technicity”, and so we may be tempted to postulate an inverse correlation between it and the tendency to allow for greater term variability in a text. But overall term density is certainly not the crucial factor in determining when a reviser will decide that a particular alternative term is acceptable in the context of a given translation. Our analysis seems to suggest that the terms most susceptible to this kind of alternation are those that are general and non-technical in nature, whose referent will normally be obvious to all readers. For example, “la province” tends to be an acceptable equivalent for “gouvernement provincial” in most contexts, and in fact is permitted by the reviser in the PIBD text. Again, where there is no risk of confusing

8. This does not necessarily mean that “tireur d’élite” will occur less frequently than the word “sniper”; in fact, the opposite is true in this text. Anaphor-antecedent patterns seem to be defined language-internally and frequently are not transposed directly from the SL to the TL text. The awkward repetitions I am referring to are those the translator would want to avoid, according to the stylistic imperatives of the target text.
9. This last suggestion is due to Pierre Isabelle.

the reader, this kind of variation may actually produce a more readable target text. Elsewhere in the PIBD draft, however, the terms “project manager” and “reporting officer” are both occasionally rendered as “agent responsable”; here, the reviser felt it necessary to correct the alternative and maintain three distinct terms, since these various functions could well be assumed by different persons.

Another of the results in Table 5 that calls for comment is the high proportion of cases in the Dairy text in which the source term is not actually translated but is either omitted or paraphrased in the target: “ST not trans” accounts for 52% of the non-correspondences in that text. Analysing the segments that TransCheck flagged in the preliminary translation and comparing these to the final French version, we did come across a number of glaring omissions. One example is given in (6) below (where a single arrow introduces the draft translation and a double arrow the final translation.)

(6) ...Canadian consumption of **milk fat and milk proteins**...is relatively similar to that of the U.S. ...

--> ...la consommation canadienne est comparable à celle de...

=> ...la consommation individuelle canadienne de **matière grasse et de protéines du lait** est comparable à celle...

In many other cases, however, the problem is not so much due to an oversight on the part of the translator as to the verbosity of the source text. A notable characteristic of the English version of the Dairy text is the preponderance of lengthy nominal compounds; and in many of these, the modifiers that precede the head noun are either superfluous or can easily be inferred from the context, and so are not essential to the meaning of the phrase. In such cases, the translator may decide to omit that part of the complex noun phrase, producing a target text that is both lighter and clearer. One simple example is given in (7) below. The TL term for “industrial milk” is “lait de transformation”, which by definition is raw milk that is going to be processed into cheese, yogurt, etc.; the inclusion of the adjective “raw” is therefore redundant in (7). TransCheck, of course, has no way of knowing this; and finding an occurrence of “raw milk” that is not aligned with “lait cru”, it signals a potential term inconsistency – mistakenly, it turns out.

(7) ...for increased shares of industrial **raw milk** supplies =>

...accroître leur part du marché de l’approvisionnement en lait de transformation...

This raises an important question: How do *we* know which of the non-correspondences flagged by TransCheck in the preliminary translations correspond to real terminological inconsistencies and which correspond to “false positives”[10] that arise because of a deficiency in our approach? Initially, we had hoped to be able to answer this question by comparing the flagged segments with the terminology of the final translations; but, as we have already pointed out, the preliminary translations did not undergo systematic revision, but only a process of quality assurance that aims to correct the most flagrant errors. The reviser who performed the quality

10. “False positive” is a term that comes from the domain of diagnostic procedures, and is defined as an individual whose lab data or scores classify him in a reference group or diagnostic category he does not belong in, especially because of an imperfection in the methodology.

assurance was quite candid on this point: she admitted that the final translations might still contain certain inconsistencies; unfortunately, she did not have the time or tools that would have allowed her to do a more thorough job.

Nevertheless, on the basis of our detailed analyses of these translations, there *are* certain inferences about TransCheck's performance which can be drawn from the data in Table 5. For example, it seems quite evident that nearly all of the non-correspondences tabulated under "Head only" and "Pron." are attributable to false positives, or system noise. Virtually none of these were corrected in the final French translations. Together, the two categories account for nearly 30% of all the segments flagged by TransCheck; and if we add to these the cases of coordinated terms tabulated under "Other", they can safely be said to account for about one case of non-correspondence in three.[11] On the other hand, a much larger proportion of the non-correspondences tabulated under "Alt.TT" and "ST not trans." *were* corrected by the reviser, presumably because they represented true terminological inconsistencies. Together, these last two categories account for between 63-73% of all the potential errors flagged, or approximately two cases out of three. Now notice that there would appear to be a rather fundamental difference between the cases of non-correspondence in these two large groupings of our taxonomy. Those non-correspondences in the former group are all concerned with variations to the **form** of the designated term. This is true of the simplification of a complex term by reducing it to its head, or by combining it with another term under coordination, and (less obviously) by replacing with a pronoun or other anaphor. In contrast, those potential inconsistencies that the system flags due to the omission of a term, or its replacement by a paraphrase or an alternative term concern more than just the superficial form of the term; they are conditioned by its redundancy in context, or by the synonymy of the term with a proposed paraphrase or alternative – questions that have more to do with the **meaning** of the term. We shall return to this distinction in the concluding discussion below.

4. Conclusion

In (Bédard 1986, especially Chapter 2), the author criticizes what he calls "l'obsession des équivalents", or the tendancy to mechanically reproduce in the target text the equivalents of all the terms found in the source. This, he argues convincingly, invariably results in inferior translations. The translator's responsibility is not to the literal wording of the source text but to its intended meaning, and to properly render this, s/he should not hesitate to make use of his or her professional judgement. Insofar as the technical terminology of the text is concerned, the translator must not feel bound to reproduce "les équivalences directes ou toutes faites". In order to create a target text that is both intelligible and natural in the target language, s/he may on occasion be required to modify or abridge certain terms, omit other terms that appear in the source text, and even coin new terms.

---

11. There are other syntactic constructions that could also be added which, like coordination, break up a multi-word expression that TransCheck seeks to recognize as a continuous string of characters. When an English term is pre-modified, as in (5b) above for example, the corresponding modifier may intervene between the head noun and its complement in French. Such cases are surprisingly infrequent, however.

> "Any object or situation can always be described in more than one way, and technical writing is no exception to this general rule. The terms employed by the writer, no matter how technical or exact, are not necessarily the only ones he or she could have used. Corollary: for the same reasons, it follows that the translator is not forced to employ direct translational equivalents in order to get his message across."
> (Bédard 1986, p.31; my translation)

Bédard's arguments would seem to run directly counter to the automated approach to term consistency checking that has been presented here; for our TransCheck implementation certainly *is* based on a literal 1:1 transposition of terms in the source and target texts – at least for those terms specified by the reviser in the term equivalence glossary. It is important to recall, however, that TransCheck is a **translation support tool**, not a machine translation system; its function is not to impose terms on the reviser, but only to assist him or her in validating a preliminary translation. As such, it is the reviser, and not the system, who always has the last word. S/he will have to decide whether to ignore or to accept each potential term inconsistency flagged by the system, and in the latter case, how to modify the target term. On the other hand, the approach to term consistency checking embodied by TransCheck does assume that, for certain types of texts at least, a considerable degree of terminological uniformity is desirable. If this were not the case, i.e. if eventual users found themselves consistently ignoring the majority of potential errors flagged by the system, it would not take long for them to abandon the system.

In the preceding Results section, we noted that for the 100 terms selected from our four sample texts, TransCheck generated an overall non-correspondence rate of between 16% and 28% (depending on whether the Sniper text is included in the calculation.) We also observed that, while our methodology did not allow us to determine precisely what proportion of these non-correspondences represented true inconsistencies, approximately one third of the segments flagged by the system could be assumed to be false positives, i.e. noise that arises because of TransCheck's failure to recognize certain formal variations to target terms. To these must be added an indeterminate number of omissions and alternative terms which the reviser may decide, for various reasons, not to correct. In short, if this prototype version of TransCheck were to be placed in the hands of users as is, it would mean that in at least one case out of three, the system would be asking the user to verify potential inconsistencies that s/he would not want to modify or correct. Is this a noise level that users would be prepared to tolerate? It is difficult to say; in part, the answer depends on how much time and embarrassment the system would be able save users on the other potential errors it brought to their attention.[12] But one thing is certain: it would definitely facilitate the acceptance of automated term checking if we could somehow reduce the incidence of noise that the system currently generates. This may in fact be possible, particularly for those cases that are due to variations in a term's superficial form. It is quite clear what has to be done here: the conditions of complete formal identity in TransCheck's definition of terminological consistency have to be relaxed so as to allow the system to recognize at least some of these formal variants as valid instances of the fully specified target term. Whether we will be able to do so without

12. My impression is that users do tolerate a noise level that is often this high with standard monolingual spelling checkers.

inadvertently exempting any of the true terminological inconsistencies the system currently flags remains to be seen. However, it would be a much more difficult task to conceive and implement strategies that would allow the system to distinguish between acceptable and unacceptable cases in which a target term has been omitted, or replaced by an alternative term or a paraphrase. Indeed, one could ask whether it is not somewhat abusive to consider these as instances of *terminological* consistency, even when they happen to yield acceptable translations. Be that as it may, our attitude in these cases is that they are best left to the human reviser, so that s/he can examine each in turn and make the appropriate decision in each case.

Our goal at the outset of this study was to determine whether or not it is feasible to verify the terminological consistency of draft translations with a tool like TransCheck, using essentially the same approach that the system employs to detect problems of source language interference. We were aware, of course, that in implementing a naive definition of terminological inconsistency and applying it to real texts, we were certain to encounter examples where the form of the target term would not correspond exactly to that designated in the term equivalence glossary, but which would nevertheless be acceptable. Some of the recent work on automatic term extraction details the possible variations that monolingual terms may undergo; see in particular (Daille 1994). The question was not *whether* such phenomena exist, but rather what types of variation to terminological units actually occur in authentic translations, and what is their relative frequency. Hopefully, this study has provided some data that helps begin to answer this question, allowing us to identify a subset of problems that may be amenable to short-term solutions.

There are (at least) two weaknesses in our approach to term consistency checking that have not as yet been mentioned. The first concerns the implicit directionality of the algorithm; i.e. the fact that TransCheck begins its search for a specified pair of terms from a detected occurrence of the source term and only then verifies the aligned target segment for the presence of the target term. But, of course, source terms are also subject to the same variations in form as target terms: they too can be ellipted, omitted, pronominalized, mis-spelled, etc. And in each such case, TransCheck's literal pattern matching strategy will cause the system to pass over the aligned segment without bothering to verify the target term, thereby reducing the system's coverage. One possible solution to this problem which immediately comes to mind would be to convert the entries in the term equivalence glossary into bi-directional queries; i.e. to take each "e-term1 = f-term1" statement and have the system search the bi-text for any aligned pair in which the specified f-term appears without the corresponding e-term. There are a number of difficulties with this strategy, however. One has to do with the fact that a translation is often more explicit than its source text. We mentioned in Note 8, for example, that the term "tireur d'élite" actually occurs more frequently in the target text than "sniper" does in the source. This suggests that reverse queries would result in even higher noise levels than the original queries. More fundamentally, however, these reverse queries may not necessarily correspond to the terminological equivalences that the reviser originally wanted to have verified. In stipulating that a given "e-term1" must be translated as "f-term1", the reviser may not have meant to exclude the possibility that other e-term's could also have f-term1 as their target language equivalent. But that is just what the reverse query "f-term1 BUT NOT e-term1" would flag.

Another general problem with our approach to term checking is the extent to which it relies on the reviser to furnish the system with an explicit statement of all the term equivalences to be verified. No doubt, situations will often arise where the reviser simply does not have the time to do this, but would still like to have the translation validated for terminological consistency before sending it out to the client. This is possible for errors of SL interference, because TransCheck incorporates a pre-existing database of attested translation interdictions. Perhaps the same could be done for term consistency checking when the text to be validated clearly belongs to a well-defined domain for which a reliable term glossary already exists.

Another interesting possibility has recently been raised in (Dagan and Church 1994). Their *Termight* system extracts candidate terms in a source text (like our F-TERM), and then goes on to automatically identify a likely target equivalent in the translation, based on a word alignment program that they have developed. *Termight* could be adapted to term consistency checking, the authors suggest, by having the system draw the user's attention to pairs of terms for which the system finds more than one target equivalent for any given source term. Notice that these will include all the formal variants that currently cause TransCheck to flag false positives. What is more, the user would not want to overlook those cases in which all the occurrences of a given source term have been consistently translated by the same but **incorrect** target term; we did, in fact, encounter several such examples in the PIBD text. While relying on the reviser to furnish the term equivalence glossary may have its drawbacks, it does ensure not only the consistency of the terminology being verified, but also its correctness.

ACKNOWLEDGEMENTS

A number of colleagues in the CITI's Machine-aided Translation group were kind enough to provide me with assistance, both technical and linguistic, during the course of this study: Pierre Isabelle, the group's director, Michel Simard, Marie-Louise Hannan, François Perrault, and Jean-Marc Jutras. I gratefully acknowledge their support, though of course none is responsible for any errors of detail or judgment contained herein. Special thanks also go to the anonymous reviser who helped me obtain the translations that were analysed, and who was patient enough to answer my many questions.

REFERENCES

AHMAD, Khurshid et al. (1994) : "What is a term? The semi-automatic extraction of terms from text," M. Snell-Horby & F. Pochhacker (eds.), *Translation Studies: An Interdiscipline*, Selected Proceedings of the Vienna Conference, pp.9-12.

BÉDARD, Claude (1986) : *La traduction technique : principes et pratique*, Montréal, Linguatech.

DAILLE, Béatrice (1994) : *Approche Mixte pour l'extraction de terminologie : statistique lexicale et filtres linguistique*, Doctoral thesis in computer science, Université Paris VII.

DAGAN, Ido and Ken CHURCH (1994) : "Termight: Identifying and Translating Technical Terminology," in the Proceedings of EACL.

HANNAN, Marie-Louise (1995) : *The Use of Semantic Classes for Automatic Terminology Extraction*, Master's thesis in linguistics, Université de Montréal.

HARRIS, Brian (1988) : "Bi-text: A New Concept in Translation Theory," in *Language Monthly*, no.54.

ISABELLE, Pierre et al.(1993) : "Translation Analysis and Translation Automation," in the Proceedings of the *Fifth International Conference on Theoretical and Methodological Issues in Machine Translation* (TMI-93), Kyoto, pp.201-217.

JUSTESON, John and Slava KATZ (1993) : "Technical Terminology: some linguistic properties and an algorithm for identification in text," *Technical Report #RC 18906 (82591)*, IBM T.J. Watson Research Center, Yorktown Heights, New York, 13 p.

MACKLOVITCH, Elliott (1994) : "Using Bi-textual Alignment for Translation Validation: the TransCheck System," in the Proceedings of the *First Conference of the Association for Machine Translation in the Americas* (AMTA-94), Columbia, Maryland, pp. 157-168.

SIMARD, Michel, G. FOSTER et P. ISABELLE (1992) : "Using Cognates to Align Sentences in Parallel Corpora," in the Proceedings of the *Fourth International Conference on Theoretical and Methodological Issues in Machine Translation* (TMI-92), Montréal, pp.67-81.