[*Translating and the Computer 27*, November 2005 [London: Aslib, 2005]

# Automatic Detection of Translation Errors: The TransCheck System

Graham Russell
Interactive Language Technologies,
NRC Institute for Information Technology
RALI, DIRO, Université de Montréal
russell@iro.umontreal.ca

Among the applications of language technology to human translation, little attention has been paid to error detection. This article discusses the issue from three perspectives. In general terms, it considers the nature of translation errors and what might be involved in detecting them. On a more specific level, it identifies a number of error types which meet the twin criteria of feasibility (errors that can be detected with sufficient reliability using practical means) and utility (errors that are sufficiently troublesome to be worth addressing). Concretely, the paper presents TransCheck, a prototype system for translation error detection, describing its structure and function, and discussing its application in the translation process.

## 1 Background

Translation has long been an archetypal application area for computer-implemented language technology. It is one of the most visible and challenging of linguistic activities, and current trends are conspiring to enhance its economic and cultural significance. Machine translation is, for certain purposes, in everyday commercial use and in addition remains an important research topic, while other less ambitious but often more practical tools have been widely adopted.

The translator's workstation (Picchi et al. 1992; Macklovitch 1993: more recently Bashkansky and Oman 1998) represented one of the earliest attempts at providing translators with task-specific support. Such systems typically integrate dictionaries, terminological databases, text editors, and spelling and grammar checkers. Current leaders in the field of software support for translators, translation memory (TM) systems essentially combine a text editor with automated access to a database of previously translated sentences. Searchable translation archives, exemplified by the RALI's TransSearch system (Macklovitch et al.. 2000), can be viewed as a less automated variant of TM in which the user retains full control over the query and is able to view results in a broader context (Macklovitch and Russell, 2000).

The fundamental task performed by both classes of system is to provide access to relevant information external to the translation itself. Although TM systems may help in avoiding certain errors during translation (but see section 4.5.2 below), they lack any capacity for detecting errors within an already translated text, and as such are at best only a partial solution to the translation error problem. Similarly, searchable archives are of little direct help in avoiding errors, since results returned depend on the content of a query, and users are unlikely to submit queries on every possible cause of error.

Globally, translation involves many activities in addition to the generation of target-language text: preparation, work-flow management and quality assurance also deserve assistance. Both classes of device mentioned above are intended to contribute chiefly to one part of the translation process, and as we have seen, address the question of errors only tangentially. There is also scope for computational support in the other phases; research and preparation may be aided by translation archives and terminology resources, and specialized document-handling tools exist for distribution and tracking of large translation projects, but there are few tools specifically aimed at the quality control domain.

Those tools which do exist tend to concern themselves chiefly with the target text; spelling, grammar and style checkers are obviously as useful in the translation domain as they are in other areas of professional writing, but there are many types of translation error which they cannot hope to detect. The main thesis of this article is that specialized devices for the detection of translation errors are both necessary and, with some restrictions, practical. One such device is the TransCheck system under development jointly by National Research Council Canada and the RALI group of the Université de Montréal.

General issues relating to translation errors are discussed in section 2, where the range of factors involved in translation quality judgments is illustrated, and the difficulty of general-case translation error detection is shown. Section 3 describes a number of types of error which nevertheless hold out some prospect of satisfactory automatic detection. The design and operation of TransCheck are covered in section 4, together with the elements of language technology employed.

At this point, it may be wise to state that the concern of this article is the detection of translation errors, and not their correction. In some cases, an error may be sufficiently circumscribed to allow a correction to be proposed, but this is not a focus of current work.

## 2 Translation Quality and Translation Errors

### 2.1 Factors in Translation Quality

The evaluation of translations, whether produced conventionally or by MT systems, is hampered by the fact that perceived quality is both highly subjective and influenced by a wide range of factors. Not all of these factors are directly related to what is normally thought of as the central property of translation, namely the preservation (or adaptation, where necessary) of meaning. The following, drawn in large part from Massion (2002), lists some areas of interest to consumers of translation:[1]

(i) target language spelling, grammar and style;

(ii) layout: correct numbering of pages, figures, footnotes, etc; correct format and contents of index, table of contents, etc.;

(iii) correct and consistent use of technical terms:

(iv) completeness: all (relevant) source-language content translated:

(v) consistency of reference to objects and concepts;

(vi) language-specific punctuation conventions (e.g. presence of space before colon, semicolon, question mark, etc. in French);

(vii) appropriate preservation of meaning;

(viii) adherence to client-specific stylistic requirements;

(ix) preservation of XML etc. markup, translation of attributes where appropriate;

(x) correct transcription of numbers, accounting for language-specific conventions.

As with all such lists, some details are debatable. Nevertheless, several points are worth noting in connection with these headings. First, as mentioned above, most are not in fact specific to translation; the issues of grammar, style, spelling and punctuation are just those that arise with any writing in the target language, while most professional writers are subject to clients' or employers' style and terminology guidelines. Problems in ensuring the accuracy of metatextual structures such as numbering and indexes arise in all forms of academic or technical writing: various solutions exist, of differing degrees of adequacy. Second, the relative importance of the different headings varies with the context; consistency and completeness are more vital in technical than in literary translation, for example.

### 2.2 Automatic Evaluation of Translation Quality

In recent years moves have been made, especially in the United States, to bring research and development in machine translation into the quantitative evaluation paradigm that has long been used for automatic speech recognition and information retrieval. Within these initiatives, a variety of automatic evaluation methods have been proposed, the aim being to permit the rapid and objective comparison of many systems based on their performance when confronted with relatively large amounts of input.

1. The cited document is a 'white paper', sales literature promoting D.O.G. GmbH's 'ErrorSpy' program (`http://www.dog-gmbh.de/`). Even so. the list does not appear unreasonably skewed in favour of that product's characteristics.

In general, these approaches make use of one or more correct 'standard' translations of a test text: these are of course not available to developers of the systems being evaluated. Evaluation is carried out by submitting the test text to the system in question and comparing the result with the standard. In practice, the comparison amounts to a measure of text similarity, the underlying supposition being that correct translations of the test text will tend to resemble the unseen standard.

Since current automatic evaluation methods assume the existence of a known correct standard translation and yield a global measure of similarity to the standard, they are of little relevance in a context where no such standard exists and where in any case it is desirable to specify the location and type of an error.

### 2.3 The Difficulty of Translation Error Detection

Viewed in very abstract terms, the problem of translation is that of producing a text that will count as a member of the class of possible translations of the original. This characterization obviously skates over some important details — 'count as' depends on who does the counting, and 'possible translation' is a notoriously ill-defined concept — but our aim here is less to define the notion of translation than to suggest parallels between the process of producing a translation and that of identifying errors therein. The latter, at a similarly abstract level, can be viewed as stating that the target text does not count as a possible translation of the original; moreover, to be of any use, it must explain why. In both cases, a fully general solution lies beyond the capability of current automation.[2]

The most familiar and successful type of natural language error-detection tool is the spelling corrector. It may be instructive to consider how this problem differs from the one at hand. First, the correct spelling of a given word is typically unique, and can either be stored in a dictionary or specified in a few simple rules. Second, it is possible to construct a reasonably accurate model of the process giving rise to errors. This may reflect the user's intention (perhaps approximated by a statistical language model), confusion resulting from phonological factors, likelihood of typing errors due to keyboard layout, and so on.

Clearly, neither of these considerations holds for translation error detection. Apart from exceptional cases such as technical terms, there exist an indeterminate number of translations for a source-language expression. Some will be used more frequently than others, of course, but it does not seem possible to predict exactly which equivalents will occur; the more complex the expression, the more difficult it becomes to do so. Practical machine translation systems may restrict themselves to a narrow range of possible realizations for any word or phrase — monotonously unvarying output and failure to find *le mot juste* are perhaps the least of their problems — but fully adequate error detection demands the ability to recognize whatever strategy the translator has chosen to employ.

2. To flesh out this skeletal argument a little further, one could imagine a situation in which a hypothetical translation error detector served as the basis for a MT system which generated random target texts and rejected all those with too high a rate of errors: overall system performance would then depend entirely on the error detector.

The MT analogy would be a system furnished with dictionaries, transfer mechanisms or their equivalent able to produce all and only the possible renderings of its input.

The process leading to translation errors is equally complex, and far more obscure than with spelling errors. Again, there are exceptions: 'false friends' can be explained, and to some extent predicted, by appealing to discrepancies between phonology, orthography and semantics, for example. In general, however, it seems likely that translation errors can only be fully explained with reference to notions such as local and global text meaning, or translators' intentions and knowledge of their readers' interests.

### 2.4 Prospects for Automatic Error Detection

The foregoing paints a deliberately gloomy picture of the prospects for translation error detection. If the problem really is at least as hard as translation itself, what can be done? One route to an answer is to return to the twin analogies of machine translation and spelling correction: in a nutshell, do there exist classes of translation error which can be detected by means resembling those applicable to spelling errors rather than those required by high-quality machine translation? We have already indicated two such classes: errors of terminology and 'false friends'; in the following section we consider these more closely and add some others.

## 3 Feasible Error Types

### 3.1 Introduction

We now turn to the question of which kinds of translation error we can expect to detect automatically. We restrict our attention to 'true' translation errors: those which can be detected only by reference to the content of the source text, rather than those arising solely from the form or content of the target text. For the latter, we assume that target-language spelling, grammar and style checkers will be available, and lower-level issues such as the typography of punctuation symbols present no particular theoretical challenge.

A second criterion in the choice of error types is utility. Following discussions with potential users of a translation error detector, five classes of error meeting the first condition of practical implementability have been identified. Happily, these have the additional advantage of requiring a broad range of language technology, much of which has been applied successfully to other text-analysis problems.

### 3.2 Terminology

Correct translation of technical terms (e.g. ensuring that English *eye relief* is rendered in French as *dégagement oculaire* in the context of optical instruments) is one of the best-known problem areas in translation. From the translator's point of view, it requires specialized knowledge of the domain, and must also take into account the preferences of individual clients.

| English | French |
|---|---|
| telescopic sight | lunette de visée<br>lunette de tir |
| spotting telescope<br>spotting scope | lunette d’observation |

Table 1: Some terms in English and French.

The issue can be divided into the verification of completeness (ensuring that all occurrences of some source-language term *s* are translated by the target-language term *t*) and coherence (ensuring that *t* is only employed as the translation of *s*). A further constraint is sometimes desirable: when multiple equivalents exist for *s*, consistent use of just one of these may be required throughout a text. Detection of errors in terminology usage is complicated by a range of linguistic phenomena: anaphoric expressions may be acceptable alternates and terms may appear in partial, discontinuous or inflected forms (Jacquemin, 2001). Table 1 shows some example term equivalents for English and French. Each may appear as either singular or plural, and constructions like *lunettes de tir et d'observation* are also possible. A correct English equivalent for this would be *telescopic sights and spotting scopes*; in order to avoid flagging this pairing as an error, the effectively elliptical second conjunct *d'observation* must be recognized as matching *spotting scopes* (note that other instances of *d'observation* in e.g. *poste d'observation* 'observation post' must not be so recognized). Detection of errors must therefore be based on more than a simple list of equivalents.

### 3.3 Negative Terminology

One of the pitfalls of translation, and something with which most learners of a foreign language are confronted at an early stage, is the confusion of form and meaning illustrated in table 2. Here, phonetic and orthographic similarities fail to carry through to the semantic domain, making 'false friends' of e.g. 'library' and 'librairie'. This class of error is often thought of as too elementary to be worth attending to: surely no professional translator would fall into such an obvious trap? Yet there are good reasons for addressing it, apart from those given in section 4.5.2 below.

The deceptive cognates of table 2 are a special case of 'negative terminology', differing from regular terms in that the requirement here is to prohibit, rather than enforce, the

| *e* | *f* | *E(f)* | *F(e)* |
|---|---|---|---|
| actually | actuellement | currently | effectivement |
| eventually | éventuellement | possibly | finalement |
| library | librairie | bookshop | bibliothèque |

Table 2: Some deceptive cognates in English and French. Despite appearances, items in columns *e* and *f* are not correct mutual translations: *E(f)* is a correct translation for *f* and *F(e)* for *e.*

appearance of a specified target-language expression as the translation of some source-language original. This aspect of translation, like many others, is influenced by client preferences. While one client may object to the presence of Anglicisms in a text translated into French, the next may be indifferent to these but hostile to 'Canadianisms' such as the Canadian French use of the verb *barrer* to translate *lock.*

Conceptually, these phenomena are clearly related, and are amenable to similar implementations where the relevant restrictions are suitable encoded. As in the case of terminology verification, the task is complicated by a range of possible linguistic variations; if anything, these play a greater role in the present case, since prohibited equivalents are more likely to involve verbal and arbitrary phrasal constructions.

### 3.4 Names

As far as their translation behaviour is concerned, names fall into three categories:

(i) those which are not to be translated (English/French *New York)*

(ii) those whose translation is compositional or predictable (German *Friedrich der Große,* English *Frederick the Great:* English *New Zealand,* French *Nouvelle-Zélande)*

(iii) those whose translation is non-compositional (French *Aix-la-Chapelle,* German *Aachen;* English/French *Charlemagne,* German *Karl der Große).*

Membership of these categories is of course specific to a particular language-pair. Translation errors involving class (i) names may arise from attempts to adapt them to the target language (giving *Nouveau York,* for example), or errors in transcription; this latter type becomes more likely when the name in question is from a third language unfamiliar to the translator. Class (ii) and (iii) names raise obvious difficulties; many such correspondences may be considered part of the translator's background knowledge but some will need to be researched. At first sight, detection of errors involving names appears to be a simple matter, all that is required being one list of forms to be preserved and another pairing source and target language equivalents. However, this is insufficient.

First, it is not practical to list names exhaustively: new names are created constantly, by the formation of new commercial companies, for example. However, it does appear feasible to list names whose translation differs, since these seem to be the exception: they are less common (most of the myriad place-names in the world are invariant) and names which are newly coined or whose referent suddenly becomes prominent are by default invariant.

A second difficulty arises in connection with the translation of names. Certain names fall into more than one of the above classes, depending on their referent: English *London* must be rendered in French as *Londres* when it refers to the capital of the United Kingdom, but must be preserved when it refers to the town in Ontario; English *Mexico* translates as French *la Mexique, Mexico* in French referring to what is known in English as *Mexico City,* etc. Moreover, this behaviour is time-dependent: *New Zealand* could reliably be translated as *Nouvelle-Zélande* without considering its referent only until the emergence of the racing yacht of that name, which remains *le New Zealand*.[3] Some personal names also exhibit this behaviour: English *Socrates* translates as French *Socrate* when it names the fifth-century Athenian philosopher, but remains *Socrates* when it names the 1980s Brazilian footballer.

The treatment of names is connected with the issue of transliteration; except for unusual cases such as scholarly publication, words from languages normally written using a different script from that of the host text are transformed into the 'native' script. Names are by far the most common class of such words. So the surname of the current President of the Russian Federation, naturally written in Cyrillic characters in Russian text, appears as *Putin* in English documents. Transliteration is both approximate and normally indicative of pronunciation: since the two languages will rarely share the relevant parts of their phonological inventory, a representation is chosen which suggests a suitable pronunciation, namely one which is adapted to the sounds of the second language. A further complicating factor is the frequent existence of multiple transliteration schemes for a given language pair; different romanization systems for Japanese and Chinese are a case in point. When translating from Russian into English, the central concern is to ensure that all transliterations within the text are consistent, i.e. in accordance with the same set of conventions. This too can be considered part of a translator's normal expertise.

Where a source text contains words transliterated from a third foreign-script language, the situation is more complex. When an English document mentioning President Putin is to be translated into French, the representation of that name must be adapted to follow the conventions regarding the transliteration of Russian words in French. As noted above, these are largely governed by the phonological patterns of the language; *Putin* thus corresponds to *Poutine*, again indicating an approximation to the Russian pronunciation, this time French-specific.

What this implies for the detection of name errors in translation is that a fully general matching mechanism must take into account the possibilities not only of literal identity and lexically stored equivalences, but also of a potentially much wider range of correspondences governed by properties of third languages unrelated to the source or target.

### 3.5 Paralinguistic Expressions

Errors can arise in the transcription of a wide range of 'paralinguistic' material: expressions denoting numbers, sums of money, dates, product codes, serial numbers, and so on. While these may not be translated in the normal sense, they are nevertheless manipulated by the translator (and so subject to error) and often of crucial importance. Moreover, they require unusual attention from human proofreaders. Certain kinds of textual annotation also fall into this category. When translating a text containing XML markup, it will normally be necessary to preserve tags, for example.

Although perhaps not strictly a matter of translation, various special classes of paralinguistic expression require conversion according to culturally specific conventions.

3. I am grateful to Gilles Gamas for bringing this phenomenon to my attention.

Examples include the syntax of expressions denoting numbers and sums of money (English *1,234.56* is equivalent to French *1 234,56*, German *1.234,56* and Swiss *1'234,56*), dates (*12/7/03, 7/12/03, 2003-07-12*), etc. These phenomena are familiar to those concerned with product localization. They also overlap with conventional translation when they occur as part of, or cooccur with, number and date phrases: it will sometimes be necessary to treat *four million, 4 million* and *4,000,000* as equivalent, and to establish that any of these may correspond to French *4 000 000.* Similarly, it should be possible to treat English *Monday, October 24, 2005*, *Monday 24th October 2005*, etc. as equivalent to French *lundi 24 octobre 2005.* Fortunately, expressions of this kind can be described in terms of a limited vocabulary and a simple and relatively self-contained grammar.

There are of course other kinds of conventional textual feature which often enter into judgments of translation quality. Many French readers have strong opinions concerning the presence or absence of a space before 'tall' punctuation symbols like the colon, semicolon and question and exclamation marks, use of the 'oe' ligature, or accents on upper-case letters, for example. But these are more in the nature of stylistic constraints on the target text than translation phenomena: except in very unusual circumstances they can be verified without reference to the source.

### 3.6 Omissions and Insertions

A final class of translation error involves the non-translation of some material in the source text, a phenomenon which surfaces as the omission of what would be the corresponding portion of the target text. Detecting omissions raises problems not present with other error types; in brief, even when an omission has been reliably detected, it is not possible in general to determine whether it is an error (i.e. whether it alters the sense of the text in an unacceptable fashion). To do so would require not only a full semantic analysis of source and target texts, but also insight into the potential reader's requirements.

Spurious insertions can be regarded as the inverse of omissions: passages included in the target text which lack support from the source text. It may seem unlikely that a translator would fall victim to this type of mistake, but it is not difficult to imagine careless manipulation of word-processor controls having this effect. And both omissions and insertions may be introduced when a large translation project is handled by numerous translators. Melamed (1996) and Russell (1999) discuss this class of error in some detail. Melamed suggests that length can be used to distinguish erroneous from acceptable omissions: “Intended omissions are seldom longer than a few words, while accidental omissions are often on the order of a sentence or more” (p. 766). This criterion is clearly fallible — a single missing *not* may modify the sense of a text far more than an entire paragraph which may be redundant for the target-language reader — but it is a reasonable starting point.

Related to this issue is that of non-translation, where portions of the source text appear to have been copied verbatim into the target. This can arise from incorrect operation at the level of the individual translator, who might skip a section when overwriting source text in a word-processor, either inadvertently or with the intention of returning to it later.

Or it may result from errors in assembling the completed target text, perhaps because a file has been misnamed or wrongly classified. Here too, there may be legitimate reasons for preserving the language of the source text; it may be accompanied by a target-language gloss, for example.

## 4 The TransCheck System

### 4.1 Introduction

TransCheck has its origins in work done during the early 1990s at the Canadian federal government's CITI research institute. Macklovitch (1994) discusses the design and operation of such a system, conceived as an extension to the TransSearch translation archive and concordancer, while Jutras (2000) describes a dedicated proof-of-concept implementation. The current version is based on a complete redesign along the lines illustrated in the following section.

TransCheck is being developed in a joint project by the Interactive Language Technologies group of National Research Council Canada[4] and the RALI research group at the Computer Science Department of the Université de Montréal.[5] Total effort is approximately 24 person-months.

The remainder of this section presents the design of the system and the machinery it employs. In addition, it briefly discusses the framework within which TransCheck is intended to be used, and comments on the perceived necessity of this kind of tool in relation to translators' expertise and other kinds of translation technology.

### 4.2 Architecture

The TransCheck system is structured as shown in figure 1. It consists of a 'core' and a number of components. The function of the core is to accept an aligned text, to support evaluation of the text by the components, and to record errors. The components are responsible for error detection and the auxiliary operations of preprocessing and error reporting.

Input consists of a source/target text pair: these are preprocessed in order to identify word and sentence boundaries, and then aligned to create a sequence of 'regions' (see section 4.3.1). Much of the functionality of the system involves searching for translational equivalents, and this stage serves to restrict the area where the search must be performed.

Error-detection components are implemented as dynamic libraries communicating via a public interface; they typically employ external data resources such as rule sets and dictionaries, and are responsible for compiling, loading and saving these where necessary.

This design is proposed for reasons of flexibility. Users are able to select the components they wish to apply at run-time without paying the penalty of unwanted processing.

---

4. `http://iit-iti.nrc-cnrc.gc.ca/about-sujet/ilt-tli_e.html`
5. `http://rali.iro.umontreal.ca/`

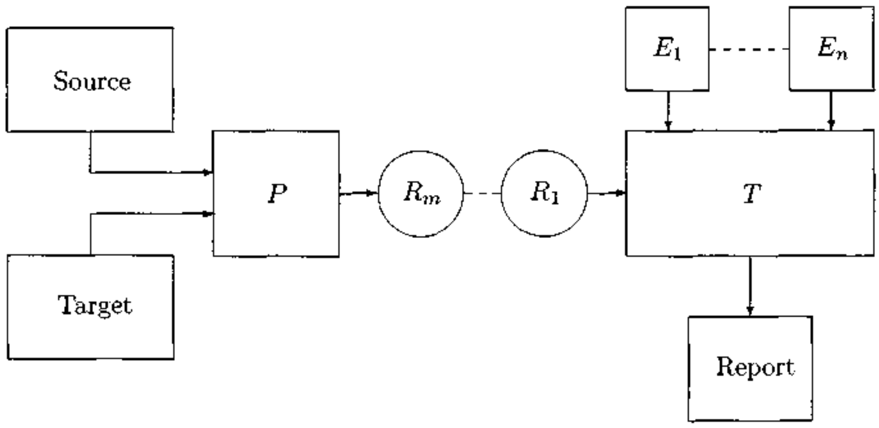


Figure 1: TransCheck architecture. Source and target texts are read by the preprocessor *P*, combined, and split into alignment regions $R_1 \ldots R_m$. Each region is submitted to the core *T* which applies the error detection components $E_1 \ldots E_n$ ; any errors detected in that region recorded for subsequent in an error report.

It is also easy to adapt to different tasks: terminology resources can be changed, or specialized gazetteers added with minimum disruption. In addition, the design is as far as possible language-independent. Adaptation to a new language-pair will obviously demand new lexical data to be loaded into the relevant detectors, and a small amount of preprocessing data to be provided. Finally, it is extensible, in the sense that any future developments in language technology which meet the specification of a TransCheck component can be incorporated with minimal disruption.

Error reports take the form of a sequence of pairs of locations representing the position of the error in the source and/or target text (not all error reports need refer to both texts), together with information concerning the error in question. One other consequence of the modular structure is that this information can be presented to the user in a variety of guises. Possibilities include a table summarizing error-types and frequencies, an annotated copy of the target text (perhaps in a dynamic web page) or exploitation by an interactive 'debugger' which allows the user to inspect putative errors and correct those which turn out to be true. The current version of TransCheck provides the second of these.

### 4.3 Language Technology in TransCheck

#### 4.3.1 Region Alignment

The error types listed above can all be viewed as violations of constraints on the target-language translation *t* of some source-language expression *s*. This perspective clearly presupposes the ability to identify the relevant (*s*, *t*) pairs. In order to facilitate this, the source and target texts are aligned in such a way that the corresponding pairs can be assumed to cooccur within a relatively narrow region: typically this consists of a small number of sentences from the source and target texts (note that translation is not generally a one-to-one relation between sentences). The result of a perfect alignment is a structure consisting of a sequence of regions; each region pairs a subsequence *S* of the source text with a subsequence *T* of the target text in such a way that the translation of *S* is wholly contained within *T*, and, conversely, *T* translates nothing which lies outside *S*. Satisfactory region alignment is thus a crucial element in the operation of the system. The method employed in the current version of TransCheck is that of Simard et al. (1992). This makes use of the relative length of source and target text segments, together with cues based on word similarity. Recall that one of the error types targeted by TransCheck is the deceptive cognate exemplified by the pairs in table 2, which display similarities of just the kind exploited by the aligner. Given the expected presence of deceptive cognates in texts handled by the system, the question arises whether they will interfere with alignment accuracy. However, preliminary experiments have not shown a moderate number of such errors to detract from alignment performance in practice.

### 4.3.2 Shallow Syntax

Many of the error types described above involve multi-token expressions. Patterns describing these are represented within error-detection components as finite-state machines compiled from external specifications. Since there may be many such patterns, most of which will not be relevant to any one region, they are indexed so that matches are only attempted with potentially applicable patterns. Successful matches cause input file positions corresponding to the start and end of the matched subsequence to be recorded together with other information for inclusion in an error report.

### 4.3.3 Morphology and Dictionaries

Detection of e.g. terms within aligned texts requires, among other things, the ability to match inflected variants of expressions in terminological data. Thus, both *function key* and *function keys* should match the term 'function key'. Rather than process the input texts themselves, TransCheck expands term specifications to include the necessary variants; at present, this is controlled by an explicit flag associated with words which are to be expanded, but other solutions can be envisaged. Variants are created by the application of morphological dictionaries and are converted into alternate paths in the finite-state representation mentioned in section 4.3.2.

### 4.3.4 Word Alignment and Translation Models

In some cases, alignment at the level of sentences is insufficient. Since many of the error-types addressed by TransCheck concern lexical units, it is helpful to establish a (partial) alignment at a finer level of detail. In general terms, we would like to know not just that *s* and *t* both occur within the same region, but that on some specific occasion *t* functions as the translation of *s*. This notion is approximated using a statistical translation model to link source/target word pairs.

Word alignment also plays a role in the treatment of omissions. Some omissions extending over one or more entire segments can be recognized by examining the output of the region-level aligner. Finer-grained omissions are also detected, albeit more tentatively. The method employed here is an extension of that proposed by Russell (1999); briefly, it searches for areas of the target text which are significantly shorter than would be expected on the basis of the source, and which do not contain expected equivalents of source-text items as predicted by the translation model mentioned above.

#### 4.3.5 Named-Entity Identification

Name recognition is carried out using techniques developed for purposes other than translation: locating so-called named entities plays an important role in many monolingual text analysis tasks. TransCheck assumes that an expression recognized as a name is invariant unless a translation is found for it in a user-supplied gazetteer.

The transliteration problem described in section 3.4 is not addressed in the current version of the system. However, this is an active research area (Knight and Graehl, 1997) and suitable mechanisms may be incorporated at a later stage.

### 4.4 Usage Scenarios

TransCheck is designed for the processing of (nominally) complete translations: a translator might submit a document to the system after completing a first draft, or before passing it on for revision: a revisor or proofreader might use it in order to identify simple errors before proceeding with a detailed review; it might form part of the quality control process of a translation purchaser, either an end client or a subcontracting agency.

To return briefly to section 2.3’s comparison with spelling correctors, while this class of program can be applied to a complete text, it is more commonly employed interactively, and indeed is typically run as a background process while a text is being composed, so that errors can be flagged as they occur. In some respects it might be attractive to provide TransCheck with this type of real-time detection by integrating it with a word-processor. However, this would mean significant modifications to existing alignment methods: in order to function in the same way as a spelling corrector, a usable fully-interactive version of TransCheck would require the ability to:

(i) identify the prefix of the source text to be aligned with the currently extant target text;[6]

(ii) accommodate frequent additions to the target text as the translation is created;

(iii) accommodate occasional modifications at earlier points within the target text;

(iv) avoid changing its decisions when supplied with new information by (ii) and (iii):

6. It is of course quite possible that the latter will not correspond to any continuous prefix of the source text, in which case the difficulty of this task increases.

(v) produce results relevant to the passage currently being translated, rather than some earlier passage to which the translator is no longer paying attention.

These problems could be circumvented by imposing appropriate working methods on the user (a system might require checking to be initiated manually on a specified portion of the texts, for example), but it is far from clear that this would be preferable to the current arrangement. If a suitable incremental, deterministic and low-latency aligner were to become available, the design of the system would allow it to be used straightforwardly.

### 4.5 Utility of TransCheck

#### 4.5.1 Translation as an Expert Activity

One objection sometimes raised to proposals such as TransCheck is that expert translators do not make mistakes, or at least do not make mistakes of a kind that could readily be detected by automatic means. On this view, experience and rigorous revision procedures reduce errors to a negligible level. The force of this objection is diminished by the following observations:

(i) The system is not intended solely for the use of expert translators; non-specialists may also benefit from its assistance.

(ii) The system is intended to operate as part of the translation/revision process, rather than be applied to the final result of that process. The appropriate comparison is therefore with the error rate obtaining in draft and unrevised translations. It is difficult to estimate the frequency of such errors, since translators tend not to preserve drafts and in any case are naturally reluctant to reveal work which they consider to be less than perfect.

(iii) Professionally translated text is evidently not free from elementary translation errors: a search in the Canadian Hansards (records of the proceedings of the Federal parliament) at TSRali.com shows instances of precisely the deceptive cognates given in table 2, for example.[7]

#### 4.5.2 Translation Memory

In theory, translation memory systems represent repositories of verified translation equivalences. It is tempting to exploit this property by applying them to the error detection problem. However, such systems, at least in their current form, are likely to be of only incidental help in avoiding translation errors:

(i) In order to achieve an acceptable match rate. TM systems ignore, or can be made to ignore, many of the phenomena which form the focus of TransCheck's attention: names, paralinguistic material, etc.

7. `http://www.tsrali.com`

(ii) Omissions or insertions which extend beyond a single sentence are beyond the scope of TM systems operating at the sentence level.

(iii) While the 'repetition processing' aspect of TM may assist in assuring terminological consistency, users remain free to perform arbitrary substitutions.

(iv) Given the low rate of matching achieved on many types of text, the user may never see a correct translation of some problematic input, even when this exists in the database.

Other translation tools are relevant to the detection or avoidance of errors. Terminology processors are able to detect specifically terminological errors, and when used in a 'pretranslation' mode can retrieve approved equivalents for source-language terms and insert them into the text. Disciplined use of archives, glossaries and other information resources will also help. But none of these possesses the twin characteristics of the TransCheck approach: integrated detection of multiple error types in a package that can be applied independently of the text-creation phase of the translation process.

## 5 Conclusions and Prospects

Some applications of language technology benefit from the existence of an established evaluation infrastructure, with standardized test data, measurement techniques and so on. Unfortunately, this is not the case with translation error detection. Although individual elements of TransCheck (word alignment, specialized grammars or name finding) can be tested in isolation, without this type of support it is difficult to draw strong conclusions about their contribution to overall system performance.

Related to this issue is a matter that has been touched on earlier: readily available translation corpora of the kind used in developing alignment-based translation technology reflect neither the frequency nor, necessarily, the type of error which occur in the texts for which TransCheck is intended. Naturally, one can introduce errors into development or demonstration texts, but this is hardly respectable from an engineering point of view. At the time of writing, a beta-testing phase is about to begin, in which the system will be installed in the offices of a translation agency. It is hoped that this will, among other things, result in a more accurate picture of error distribution in real texts.

As much of the discussion above suggests, and unsurprisingly in view of its Canadian origins, TransCheck is being developed primarily with reference to English and French. Adaptation to other language-pairs will involve at least the provision of language-specific data used by the preprocessor for sentence segmentation and the replacement of morphological dictionaries for term expansion and specialized grammars for paralinguistic expressions (we assume that users are in a position to supply suitable terminology data, etc.) In more complex cases, it will also be necessary to adopt a different approach to preprocessing: languages such as Chinese, Japanese and Korean, in which some or all word boundaries are not indicated explicitly, may require more complex analysis. Alternatively, the internal matching mechanisms could be modified to work on a character rather than a word level.

## References


Bashkansky, G. and U. Oman (1998). Monolingual translator workstation. In D. Farwell, L. Gerber, and E. Hovy (Eds.), *Machine Translation and the Information Soup: Proceedings of the Third Conference of the Association for Machine Translation in the Americas,* Volume 1529 of *Lecture Notes in Artificial Intelligence*, pp. 136-149. Berlin: Springer. Langhorne PA. October 28th-31st 1998.

Jacquemin, C. (2001). *Spotting and Discovering Terms through NLP*. Cambridge MA: MIT Press.

Jutras, J.-M. (2000). An automatic reviser: The TransCheck system. In *Proceedings of the 6th Applied Natural Language Processing Conference and the 1st Meeting of the North American Chapter of the Association for Computational Linguistics,* Volume 1, pp. 127-134. Seattle, 29th April-4th May 2000.

Knight, K. and J. Graehl (1997). Machine transliteration. In *Proceedings of the Thirty-Fifth Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*, pp. 128-135.

Macklovitch, E. (1993). PTT-3: A third version of the CITI’s workstation for translators. Technical report, Centre for Information Technology Innovation, Laval, Quebec, Canada.

Macklovitch, E. (1994). 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,* pp. 157-168. Columbia, October 5th-8th 1994.

Macklovitch, E. and G. Russell (2000). What's been forgotten in translation memory? In J. S. White (Ed.), *Envisioning Machine Translation in the Information Future: Proceedings of the 4th Conference of the Association for Machine Translation in the Americas (AMTA 2000),* Volume 1934 of *Lecture Notes in Artificial Intelligence*, pp. 137-146. Berlin: Springer. Cuernavaca, Mexico, October 10th-14th 2000.

Macklovitch, E., M. Simard, and P. Langlais (2000). TransSearch: A free translation memory on the World Wide Web. In *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC),* Volume 3, pp. 1201-1208. Athens, May 31st-June 3rd 2000.

Massion, F. (2002). *Automatisierung der Qualitätssicherung von Übersetzungen mit ErrorSpy.* D.O.G. Dokumentation ohne Grenzen GmbH.

Melamed, I. D. (1996). Automatic detection of omissions in translation. In *Proceedings of the 16th International Conference on Computational Linguistics,* pp. 764-769. Copenhagen, August 5th-9th 1996.

Picchi, E., C. Peters, and E. Marina (1992). A translator's workstation. In *Proceedings of the 15th [sic] International Conference on Computational Linguistics,* Volume 2, pp. 972-976. Nantes, July 22nd-28th 1992.

Russell, G. (1999). Errors of omission in translation. In *Proceedings of the 8th International Conference on Theoretical and Methodological Issues in Machine Translation,* pp. 128-138. University College, Chester, 23rd-25th August 1999.

Simard, M., G. Foster, and P. Isabelle (1992). Using cognates to align sentences in bilingual corpora. In *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-92)*. Montréal, pp. 67-81.