

# Towards a corpus of corrected human translations

Andrei Popescu-Belis, Margaret King, Houcine Benantar

ISSCO/TIM/ETI, University of Geneva,  
École de Traduction et d'Interprétation  
40 Bvd. du Pont d'Arve  
CH-1211 Geneva 4 – Switzerland  
andrei.popescu-belis@issco.unige.ch  
margaret.king@issco.unige.ch  
benanta4@etu.unige.ch

## Abstract

This section of the workbook describes the test data that is proposed to the participants. The data is part of a broader-scope corpus containing translations produced by students and corrected by their professors. Such a corpus will be used in automatic evaluation of MT systems. This section describes the structure of the corpus and provides some sample data. The full workshop data can be downloaded from: <http://www.issco.unige.ch/projects/isle/mteval-may02/>.

## 1. Introduction

Several automatic measures for MT evaluation have been proposed, and computational tools to carry them on effectively are now available. From Henry Thompson's (1992) proposal to IBM's BLEU, through Niessen et al.'s (2000) proposal and NIST's 2001 MT Evaluation, all of these measures make heavy use of large sets of reference data (or golden standard).

It is indeed acknowledged that, while a unique «correct translation» of a source is insufficient for evaluation (since another perfectly acceptable translation can differ substantially from the first one), the solution may reside in the use of a set of reference translations, which will hopefully encompass the range of possible variations among acceptable translations. Once such a set available, the quality of candidate translations can be judged with respect to it, by automatically computing a similarity distance between the candidate and the set. Evaluation is thus greatly accelerated.

However, producing such resources is quite expensive. A team of professional translators must be hired and asked to translate a number of reference texts. The quality of the reference translations thus produced would be high, but maybe some more simplistic formulations, acceptable from an MT system, would not be present in the corpus, thus biasing the results.

We propose here to build a corpus of translations using translations exams from the Ecole de Traduction et d'Interprétation (University of Geneva). These translations are encoded using markup, together with the corrections made by professors, and most important, with the *grade* that has been decided. We describe below this construction effort, then describe the data that will be used in the LREC 2002 MT Evaluation Workshop.

## 2. Description of the corpus

### 2.1. Structuring the data

One of the principles underlying the encoding of the data is to encode the most part of the information present on the paper version of the exam. This includes mainly the

text produced by each student, the corrections added by the professors grading the exam, and the final grade.

We chose an XML-based annotation format, with one file per translation. Each file has a header containing useful data (except the name of the student, who is never typed in), and a <content> element with the translation. Instead of giving the DTD that was written, here is an example of exam file.

```
<?xml version="1.0"
      encoding="iso-8859-1"
      standalone="no" ?>
<!DOCTYPE exam SYSTEM "exam.dtd">
<exam>
  <header>
    <index>101</index>
    <author>101</author>
    <date>11/02/2002</date>
    <source-language>en</source-language>
    <target-language>fr</target-language>
    <level>2e cycle (years 3-4)</level>
    <exam-title>Traduct. FR/EN</exam-title>
    <comments>Exam graded by two
independent reviewers. This is a non-native
English speaker. Teacher's comments: "Your
style was confident, your English
idiomatic. Only minor mistakes appear in
the flow of your translation. Good work."
  </comments>
    <grade max="6.0" pass="4.0">5.0</grade>
  </header>
  <contents>
    <title-zone>
      <s>...</s>
    </title-zone>
    <p>
      <s>...</s>
      ...
    </p>
  </contents>
</exam>
```

Figure 1. Example of translation header.

Together with the DTD, we also use tools to validate each XML file, as well as a simple XSL file (stylesheet) that extracts the original text and discards the markup (this

stylesheet is used to produce the workshop data described in the next section).

The innovative part of this corpus of «imperfect» translations is the encoding of the mistakes, together with their corrections. This requirement renders the typing of the data a bit more tedious, but increases the value of the resource, since the erroneous fragments of the texts can be discarded (or given a lower weight) when computing the distance between a candidate translation and the corpus.

Several conventions have been used to encode the mistakes and their correction: the `<m>` tag denotes a mistake, and the attributes encode its correction. The ‘t’ attribute encodes the type, as noted by the professor (‘-’ means a fragment to be deleted), while the ‘w’ attribute encodes the replacement string. Missing parts are encoded as an empty `<m/>` element, with `t=»miss»` and `w=»the missing string»`. A sample corrected paragraph is shown below.

```
<p>
<s>Just like you, we feel convinced
that the prevention of drug addiction
<m t="-" w="none">s</m> starts at
home, through <m t="-">the</m> <m
t="miss" w="a good"/> <m t="w"
w="relationship">relation</m> between
adults and children, by strengthening
self-esteem.</s>
<s>The findings of recent studies
clearly show that the earlier the
prevention, the <m t="gr" w="more">
most</m> efficient it is.</s>
</p>
<p>
<s>You do not necessarily need to be a
specialist in drug addiction <m t="-">
s</m> to talk over this issue with
your children.</s>
<s> The most important thing <m t="-"
w="is">lies in</m> dialog, <m t="-">
in</m> attentive listening, <m t="-">
in</m> reciprocal confidence.</s>
</p>
```

**Figure 2.** Translated paragraph and annotated mistakes.

## 2.2. Present state of the corpus

The corpus presented above is still under construction. As members of the Translation Faculty at the University of Geneva, we have been granted access to the written examinations of translations students (anonymized). We are focusing, for this corpus, on pure translations: the students are required to produce, in a limited amount of time and without dictionary, a translation of a piece of text – in general an excerpt from an article or essay, broadly speaking with a «general» vocabulary (through more specific exams, such as law translation, do exist).

Several language pairs are tested for at our faculty. The best represented ones, in terms of number of translations, are translations from English into French. However, given that a majority of researchers focuses on

translation *into* English, we collect also French-to-English translations (less numerous).

The quality level of these translations is quite variable, as well as the difficulty of the source text. A considerable part of the corpus comes from entry-level examinations, but there are also translations from students that are close to graduation; in this case, the source texts are more «difficult» (a notion that must still be quantified).

The corrections are done on the paper version by two graders, teachers of the faculty. Their annotations are by no means standardized, but we attempt to grasp them in the most precise manner using the annotation format described above. The encoding principle is that *stripping a text from its XML annotation must yield exactly the text produced by the candidate*. The consistency and correction of the typed texts are checked by a second annotator, and the validity of the XML mark-up is checked against the DTD using a parser (Xalan-Java).

For the time being, a total of about 50 translations of two texts have been encoded. The public distribution of this data is still under consideration.

## 2.3. Possible uses of the corpus

The construction of this corpus is part of a long-term effort in MT evaluation at ISSCO/TIM/ETI, University of Geneva. The main use of the corpus is as a resource for automatic evaluation, where the cost of the resource lies in typing and encoding the data, rather than asking professional translators to translate a given source text. Given that this is a corpus of «imperfect» translations, we must encode also the corrections that were made by the graders (teachers). This increases the reliability of the corpus when used for automatic evaluation, since the erroneous fragments of the student translations can be discarded or given less confidence. The grades obtained by each translation can also be used to modulate the confidence attributed to each translation.

The corpus can also be used, of course, to extract statistics about the types of translations mistakes, and the correlation between the distribution of mistakes in a translation and the grade scored by that translation. Of course, the corpus could serve also to explore automatic techniques to grade human translations, which differ quite strongly from machine translations (translation quality, proximity to source structures, etc.).

## 3. Description of test data for the workshop

For the present workshop, the organizers provide test data consisting in two sets of translations extracted from the corpus, enriched with machine translations of the same text. The test data is available at the workshop’s site: <http://www.issco.unige.ch/projects/isle/mteval-may02/>.

- The source texts (*10S.txt* and *20S.txt*) are excerpts from two longer essays, originally in French – the source is of course provided, as well as a reference translation for each text (*10A.txt* and *20A.txt*) constructed from the best student translations, using also the teacher’s corrections. Of course, these aren’t meant to be «the perfect translation», but only correct

translations that are close enough to the source text to help evaluators that do not understand French

For each of the two source texts, we provide about a dozen translations in English, some of them by translation students and some by commercial systems available over the Internet. Translations are numbered *101.txt* through *113.txt* and *201.txt* through *213.txt* (three numbers are missing from the second list, for technical reasons). There is no particular order, and in particular *1XY.txt* vs. *2XY.txt* are not necessarily translated by the same translator (human or system).

The human translators were not instructed to use either of the particular varieties of English (British vs. American), hence some slight spelling variations. The systems were simply those made available over the Internet by various providers, as listed for instance on the

following page, compiled by Laurie Gerber: <http://www.lim.nl/eamt/resources/>. We do not wish to disclose the names of the systems that produced the various translations, since the evaluations produced in this workshop do not claim commercial-level reliability.

A sample of the translations produced for the first text (including source and reference) is provided for visual comparison in the table below.

Subject to availability, and depending on decisions that will be made after the time of writing, extra data will be made available at the workshop's website (<http://www.issco.unige.ch/projects/isle/mteval-may02/>), and the participants will be informed as soon as possible about updates.

<i>Source text</i>	
<p>Comme vous, nous sommes convaincus que la prévention des toxicomanies commence dans la famille, dans la relation entre adultes et enfants, à travers le renforcement de l'estime de soi.</p> <p>Les résultats d'études récentes le démontrent clairement : plus la prévention commence tôt, plus elle est efficace.</p> <p>Il n'est pas forcément nécessaire d'être un spécialiste des toxicomanies pour aborder ce sujet avec vos enfants. L'essentiel est ailleurs, dans le dialogue, dans l'écoute, la confiance réciproque.</p>	
<i>Reference translation</i>	
<p>Like you, we are convinced that the prevention of dependence begins at home, through the relationship between adults and children. This is done through reinforcing the child's self-esteem.</p> <p>The findings of recent studies clearly show that the earlier prevention starts, the more efficient it will be. You do not necessarily need to be an expert in drug dependence to talk about this issue with your children. What really matters is talking together, listening to each other, and having mutual confidence in one another.</p>	
<i>Translation 101</i>	<i>Translation 108</i>
<p>Just like you, we feel convinced that the prevention of drug addictions starts at home, through the relation between adults and children, by strengthening self-esteem.</p> <p>The findings of recent studies clearly show that "the earlier the prevention, the most efficient it is."</p> <p>You do not necessarily need to be a specialist in drug addictions to talk over this issue with your children.</p> <p>The most important thing lies in dialog, in attentive listening, in reciprocal confidence.</p>	<p>As you, we are convinced that the prevention of the drug addiction begins in the family, in the relation among adults and children, through the intensification of the respect of one.</p> <p>The results of recent studies demonstrate him(it) clearly: the more the prevention begins early, the more it is effective.</p> <p>It is not necessarily necessary to be a specialist of the drug addiction to approach this subject with your children.</p> <p>The main part is somewhere else, in the dialogue, in the listening, the mutual confidence.</p>
<i>Translation 102</i>	<i>Translation 109</i>
<p>One thing is sure, we both agree: prevention of drug addiction starts at home, through the relationships between adults and children where the self-esteem has to be strengthened.</p> <p>Outcomes of recent studies carried out recently, clearly demonstrate that the sooner the prevention begins, the better and the more successful it will be.</p> <p>You needn't be a specialist in drugs to talk about it with your children.</p> <p>It is necessary to listen to them, you must establish a real dialogue based on reciprocal confidence.</p>	<p>As you, we are convinced that the prevention of the drug addiction begins in the family, in the relation between adults and children, through the intensification of the self-respect.</p> <p>The results of recent studies demonstrate him(it) clearly: the more the prevention begins early, the more it is effective.</p> <p>It is not necessarily necessary to be a specialist of the drug addiction to approach this subject with your children.</p> <p>The main part is somewhere else, in the dialogue, in the listening, the mutual confidence.</p>
<i>Translation 103</i>	<i>Translation 110</i>
<p>Like you, we are convinced that drug prevention begins within the family, in the relationship between</p>	<p>Like you, we are convinced that the prevention of drug-addiction starts in the family, in the relation</p>

<p>grown-ups and children, through the encouragement of self-esteem.</p> <p>Recent studies have clearly shown that the earlier the prevention begins, the more efficient it is.</p> <p>It is not unavoidably necessary to be a specialist in drug addictions to talk about this subject with your children.</p> <p>What matters more is discussion, attentive listening and mutual trust.</p>	<p>between adults and children, through the reinforcement of the regard of oneself.</p> <p>The results of the recent studies show it clearly: the more the prevention starts early, the more it is effective.</p> <p>It is not inevitably necessary to be a specialist in drug-addiction to tackle this subject with your children.</p> <p>Essence is elsewhere, in the dialogue, in listening, reciprocal confidence.</p>
<p><i>Translation 104</i></p> <p>Like you, we are convinced that the prevention of dependences begins at home, through the relationship of parents with their children. This is done through the reinforcement of the child's self-esteem.</p> <p>As recent studies have clearly shown, the earlier prevention starts, the more efficient it will be.</p> <p>You do not necessarily need to be an expert in dependences to talk about this issue with your children.</p> <p>What really matters is talking together, listening to each other, and having confidence in one another.</p>	<p><i>Translation 111</i></p> <p>Like you, we are convinced that the prevention of drug-addiction starts in the family, in the relation between adults and children, through the reinforcement of the regard of oneself.</p> <p>The results of the recent studies show it clearly: the more the prevention starts early, the more it is effective.</p> <p>It is not inevitably necessary to be a specialist in drug-addiction to tackle this subject with your children.</p> <p>Essence is elsewhere, in the dialogue, in listening, reciprocal confidence.</p>
<p><i>Translation 105</i></p> <p>Like you, we are convinced that prevention starts at home: the relationship between parents and children as well as the child's self-esteem are of great importance.</p> <p>Recent studies have shown very clearly that the earlier prevention starts, the more effective it will prove.</p> <p>You do not necessarily need to be an expert in addictions to talk about that issue with your children.</p> <p>Exchanging thoughts, listening to each other as well as mutual trust is much more important.</p>	<p><i>Translation 112</i></p> <p>As you, we are convinced of the prevention of the drug addictions beginning in the family, in the relationship between adults and children, through the reinforcement of the esteem of themselves.</p> <p>The results of recent studies demonstrate it clearly : the earlier the prevention begins, the more efficient it is.</p> <p>It is not inevitably necessary of to be a specialist of the drug addictions to approach this subject with your children.</p> <p>The essential is elsewhere, in the dialogue, in the listening, the reciprocal trust.</p>
<p><i>Translation 106</i></p> <p>Like you, we are convinced that the prevention of drug addiction begins within the family, in the relationship between adults and children, through the reinforcement of self-confidence.</p> <p>Recent study results show this clearly: the earlier the prevention starts, the more efficient it is.</p> <p>It is not completely necessary to be a specialist on drug addiction to discuss this subject with your children.</p> <p>The importance is elsewhere: it is in the discussion, in the listening, in the mutual confidence.</p>	<p><i>Translation 113</i></p> <p>Like you, we are convinced that the prevention of drug-addiction starts in the family, in the relation between adults and children, through the reinforcement of the regard of oneself.</p> <p>The results of the recent studies show it clearly: the more the prevention starts early, the more it is effective.</p> <p>It is not inevitably necessary to be a specialist in drug-addiction to tackle this subject with your children.</p> <p>Essence is elsewhere, in the dialogue, in listening, reciprocal confidence.</p>
<p><i>Translation 107</i></p> <p>As you, we are convinced that the prévention of the toxicomanies begin in the family, in the relation between adults and children, through the reinforcement of the esteem of oneself.</p> <p>The results of recent studies show it clearly: more the prévention begin early, more she is effective.</p> <p>It is not necessarily necessary be a specialist of the toxicomanies to approach this subject with your children.</p> <p>The essential is elsewhere, in the dialog, in the listen, reciprocal confidence.</p>	

**Figure 3.** Excerpt from the test data: source text (French), reference translation, candidate translations from humans and from commercial systems available over the Internet.

The references of the two source texts are the following:

- Excerpts from the brochure «Prévenir ses enfants des problèmes de drogue», Institut Suisse de Prévention de l'Alcoolisme et Autres Toxicomanies (ISPA), 24 p., 1999. (Free, order at <http://www.sfa-ispa.ch>)
- Micheline Centlivres-Demont, «Hommes combattants, femmes discrètes : aspects des résistances subalternes dans le conflit et l'exil afghan» (p.169-182, excerpt at p. 178). In «Hommes armés, femmes aguerries : rapports de genre en situations de conflit armé», Fenneke Reysoo, editor, DDC/Unesco/IUED, Geneva, 2001, 250 p. Proceedings of a colloquium held at the Institut Universitaire des Études du Développement, Geneva, 23-24 January 2001. Available freely at the IUED's press service or at: [http://www.unige.ch/iued/new/information/publications/yp\\_tm\\_hommes\\_armes\\_femmes.html](http://www.unige.ch/iued/new/information/publications/yp_tm_hommes_armes_femmes.html)).

#### 4. References

- G. Doddington. 2002. Automatic Evaluation of Machine Translation Quality Using N-gram Co-Occurrence Statistics. In *HLT 2002, Human Language Technology Conference*, San Diego, CA.
- K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu. 2001. BLEU: a method for automatic evaluation of MT. Research Report, Computer Science RC22176 (W0109-022), IBM Research Division, T.J.Watson Research Center, 17 September 2001. See <http://domino.watson.ibm.com/library/CyberDig.nsf/home>, and search for 'RC22176'.
- H. S. Thompson, ed., 1992. The Strategic Role of Evaluation in Natural Language Processing and Speech Technology. Record of a workshop sponsored by DANDI, ELSNET and HCRC, University of Edinburgh, Technical Report, May 1992.