Translating and the Computer 36

AsLing
27-28 November, 2014
One Birdcage Walk, London

Proceedings
Conference Chairs and Editors of the Proceedings

- João Esteves-Ferreira, Tradulex – International Association for Quality Translation.
- Ruslan Mitkov, University of Wolverhampton.
- Olaf-Michael Stefanov, JIAMCATT - International Annual Meeting on Computer-Assisted Translation and Terminology, United Nations (ret.).

Programme Committee

- Alain Désilets, National Research Council of Canada (NRC).
- David Chambers, World Intellectual Property Organisation (ret.).
- Gloria Corpas Pastor, University of Malaga.
- Estelle Delpech, Nomao.
- David Flip, LRC - Localisation Research Centre (Ireland), CNGL - Centre for Global Intelligent Content (Ireland). Web, University of Limerick.
- Pamela Mayorcas, Fellow of the Institute of Translation and Interpreting (FITI).
- Paola Valli

Conference Manager

- Nicole Adamides

Technical Advisor

- Jean-Marie Vande Walle

Editorial Assistants

- Míriam Urbano Mendaña
- Petya Petkova
Acknowledgements

AsLing wishes to thank and acknowledge the support of the sponsors of TC36
Preface

2014 has been one of profound changes for the Translating and the Computer conference that has managed to maintain over 36 years its character of a unique forum for researchers, developers and users. Bringing together academics involved in language technology research and in teaching translation and terminology with those who develop and market tools for language transformation and both of these groups with practitioners: translators, terminologists, interpreters, and voice-over specialists, whether freelancers or working in translation departments of large organisations, international companies as well as Language Services Providers (LSPs), large and small. At the end of April the Chairs were invited to take over the entire organisation of the event previously run by Aslib. In the midst of many doubts and uncertainties we decided to accept this challenge. To do this, we set up a legal entity, AsLing (Association internationale pour la promotion des technologies linguistiques), a not for profit association, to run the conference and promote the advancement of language technology.

In June the adventure began! Despite the inevitable delays in the call for papers and deadlines that dangerously clashed with the holiday season, we had a generous response. We would therefore like to thank all those who believed in our initiative and provided the worthy submissions contained in this volume. In particular we would like to thank our keynotes who so generously gave us credit. Again we are also grateful to those who agreed to the new opportunity to present Posters and Workshops to extend the programme and provide a rich and interesting opportunity for our delegates. We were very fortunate in convincing the Programme Committee members to step in and provide expert support in selecting the submissions and providing feedback to the authors as well as helping with the conference itself. You will also be pleased to know that the volume has been assembled with care by our attentive editorial assistants under the guidance and supervision of the Editors.

Finally, we would like to say a big thank you to: Nicole Adamides, whom we are delighted to welcome back, Jean-Marie Vande Walle, and naturally our sponsors without whom the conference would not have taken place and consequently these Proceedings would not exist.

Conference Chairs

João Esteves-Ferreira, Juliet Margaret Macan, Ruslan Mitkov, Olaf-Michael Stefanov.
### Programme

**THURSDAY 27 November 2014**

**MORNING Session (incl. Lunch)**

#### Lecture Theatre

<table>
<thead>
<tr>
<th>Time</th>
<th>Speakers</th>
<th>Title</th>
<th>Time</th>
<th>Title – Presenter/Moderator</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:30</td>
<td>Registration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>09:00</td>
<td><strong>Lead Chair &amp; Session Chair</strong></td>
<td><strong>Welcome and Introduction</strong></td>
<td></td>
<td>Everyone is invited to the Lecture Theatre for the conference Opening</td>
</tr>
<tr>
<td>09:15</td>
<td>Kevin Flanagan</td>
<td>Fill in the gaps: what we need from TM subsegment recall</td>
<td></td>
<td></td>
</tr>
<tr>
<td>09:45</td>
<td>Gábor Prószéky</td>
<td>Some thoughts on progress in MT — almost fifty years after the (first?) ALPAC report</td>
<td></td>
<td>Everyone is invited to the Lecture Theatre for the Keynote</td>
</tr>
<tr>
<td>10:35</td>
<td>Break</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:55</td>
<td>Workshop moderators</td>
<td>Workshop infos (8 x 5 min.) and Poster overview</td>
<td></td>
<td>Everyone is invited to the Lecture Theatre for an overview of the Workshops and Poster Presentations to be held in the George Stephenson Room</td>
</tr>
<tr>
<td>12:00</td>
<td>Miguel A. Jimenez Crespo</td>
<td>Beyond prescription: What empirical studies are telling us about localization crowdsourcing</td>
<td></td>
<td>Workshop 1: SDL trados (Gold Sponsor) “Productivity beyond the Translation Memory” - Lydia Simplicio Technical Business Consultant</td>
</tr>
<tr>
<td>12:30</td>
<td>Nasredine Semmar</td>
<td>Using Cross-Language Information Retrieval and Meaning-Text Theory in Example-Based Machine Translation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13:00</td>
<td>Lunch</td>
<td></td>
<td>13:25–13:50</td>
<td>Lunch (incl. 2 Poster sessions)</td>
</tr>
</tbody>
</table>

#### George Stephenson Room

<table>
<thead>
<tr>
<th>Poster</th>
<th>Presenter</th>
<th>Additional Author(s)</th>
<th>Title of Poster Presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Koen Kerremans</td>
<td></td>
<td>Representing intra-and interlingual terminological variation in a new type of translation resource: a prototype proposal</td>
</tr>
<tr>
<td>P2</td>
<td>Hernani Costa</td>
<td>Gloria Corpas Pastor, Miriam Seghirí</td>
<td>iCompileCorpora: A Web-based Application to Semi-automatically Compile Multilingual Comparable Corpora</td>
</tr>
</tbody>
</table>
### THURSDAY 27 November 2014

**AFTERNOON Session (Lunch listed here again)**

<table>
<thead>
<tr>
<th>Time</th>
<th>Speakers</th>
<th>Title</th>
<th>Time</th>
<th>Title – Presenter/Moderator</th>
</tr>
</thead>
<tbody>
<tr>
<td>13:00</td>
<td>Lunch</td>
<td></td>
<td>13:25–13:50</td>
<td>P1</td>
</tr>
<tr>
<td>13:55–14:20</td>
<td>Lunch (incl. 2 Poster sessions)</td>
<td></td>
<td></td>
<td>P2</td>
</tr>
<tr>
<td></td>
<td>&amp; Gilles Falquet</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15:00</td>
<td>Victoria Porro, Johanna Gerlach, Pierrette Bouillon, &amp; Violeta Seretan</td>
<td>Rule-based automatic post-processing of SMT output to reduce human post-editing effort: a case study</td>
<td>15:30–15:50</td>
<td>Workshop 3: MateCat (Gold Sponsor) “Free. A new business model for CAT tools” - Alessandro Cattelan</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15:30</td>
<td>Break</td>
<td></td>
<td>15:30–15:50</td>
<td>P3</td>
</tr>
<tr>
<td>15:50–16:50</td>
<td>Workshop 4: “Top-down or bottom-up: what do industry approaches to translation quality mean for effective integration of standards and tools?” - Joanna Drugan</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16:20</td>
<td>Marion Wittkowsky</td>
<td>Integrating Machine Translation (MT) in the Higher Education of Translators and Technical Writers.</td>
<td>15:50–16:50</td>
<td></td>
</tr>
<tr>
<td>16:50</td>
<td>AsLing Executive Committee</td>
<td>Special Award Ceremony</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17:00</td>
<td>Gold sponsors</td>
<td>Product presentations</td>
<td>17:00–18:00</td>
<td></td>
</tr>
<tr>
<td>17:30</td>
<td>Close of Day 1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Poster Presentations

<table>
<thead>
<tr>
<th>Poster</th>
<th>Presenter</th>
<th>Additional Author(s)</th>
<th>Title of Poster Presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Koen Kerremans</td>
<td>Gloria Corpora Pastor, Miriam Seghiri</td>
<td>Representing intra-and interlingual terminological variation in a new type of translation resource: a prototype proposal</td>
</tr>
<tr>
<td>P2</td>
<td>Hernani Costa</td>
<td>Rohit Gupta, Constantin Orasan</td>
<td>iCompileCorpora: A Web-based Application to Semi-automatically Compile Multilingual Comparable Corpora</td>
</tr>
<tr>
<td>P3</td>
<td>Hanna Bechara</td>
<td></td>
<td>Intelligent Translation Memory Matching and Retrieval Metric Exploiting Linguistic Technology</td>
</tr>
</tbody>
</table>
## FRI

### 28 November 2014

#### MORNING Session (incl. Lunch)

<table>
<thead>
<tr>
<th>Time</th>
<th>Speakers</th>
<th>Title</th>
<th>Time</th>
<th>Title – Presenter/Moderator</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00</td>
<td>Terence Lewis</td>
<td>Getting the best out of a mixed bag</td>
<td>09:00–09:30</td>
<td>“Terminology finding in the Sketch Engine: an Evaluation”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Adam Kilgarriff</td>
</tr>
<tr>
<td>09:30</td>
<td>Angelika Zerfass</td>
<td>Translation Tools are TOOLS. How do we make</td>
<td>10:20–10:40</td>
<td>Everyone is invited to the Lecture Theatre for the Keynote</td>
</tr>
<tr>
<td></td>
<td>Keynote – Day2</td>
<td>the most of them?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:20</td>
<td>Break</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- A. Görög</td>
</tr>
<tr>
<td>11:20</td>
<td>Kurt Eberle</td>
<td>AutoLearn&lt;word&gt;</td>
<td>11:50–12:50</td>
<td>Workshop 6: Kilgray / MemoQ (Silver Sponsor)</td>
</tr>
<tr>
<td>11:50</td>
<td>Erin Lyons</td>
<td>Far From the Maddening Crowd: Integrating Collaborative Translation Technologies into Healthcare Services in the Developing World</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12:20</td>
<td>Antonio Toral, &amp; Andy Way</td>
<td>Is Machine Translation Ready for Literature?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12:50</td>
<td>Lunch</td>
<td></td>
<td>13:45–14:05</td>
<td></td>
</tr>
</tbody>
</table>

### Poster Session

<table>
<thead>
<tr>
<th>Poster</th>
<th>Presenter</th>
<th>Additional Author(s)</th>
<th>Title of Poster Presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>P4</td>
<td>Sabine Hunsicker</td>
<td>Alexandru Ceausu</td>
<td>Machine Translation Quality Estimation Adapted to the Translation Workflow</td>
</tr>
<tr>
<td>P5</td>
<td>Manny Rayner</td>
<td>Alejandro Armando, Pierrette Bouillon, Nikos Tsourakis</td>
<td>A Tool for Building Multilingual Voice Questionnaires</td>
</tr>
</tbody>
</table>
## Session Chair:
David Chambers

### Lecture Theatre

<table>
<thead>
<tr>
<th>Time</th>
<th>Speakers</th>
<th>Title</th>
<th>Time</th>
<th>Title – Presenter/Moderator</th>
</tr>
</thead>
<tbody>
<tr>
<td>12:50</td>
<td>Lunch</td>
<td>Lunch</td>
<td>13:45–14:05</td>
<td>Lunch (incl. 2 Poster sessions)</td>
</tr>
<tr>
<td>14:15</td>
<td>Panel Debate</td>
<td><strong>Toolkit 2020. What have we got? What do we need? How do we get there?</strong></td>
<td>13:45–14:05</td>
<td>Both rooms will be integrated into the panel debate</td>
</tr>
<tr>
<td>15:30</td>
<td>Break</td>
<td>Break</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15:45</td>
<td>Irina Burukina</td>
<td>Translating implicit elements in RBMT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16:10</td>
<td>Mozghan Ghassemiazhgahidim &amp; Tengku Sepora Tengku Mahadi</td>
<td>Losses and Gains in Computer-Assisted Translation: Some Remarks on Online Translation of English to Malay</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16:35</td>
<td>AnneMarie Taravella</td>
<td>Affective Impact of the use of Technology on Employed Language Specialists: An Exploratory Quality Study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17:00</td>
<td>Eduard Šubert, &amp; Ondrej Bojar</td>
<td>Twitter Crowd Translation – Design and Objectives</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17:25</td>
<td>Najeh Hajlaoui</td>
<td>SMT for restricted sublanguage in CAT tool context at the European Parliament</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17:50</td>
<td>ASLING Chairs</td>
<td><strong>Wrap-up announcements &amp; Conference Close</strong></td>
<td></td>
<td>Everyone is invited to the Lecture Theatre for the wrap-up and Closing</td>
</tr>
</tbody>
</table>

### George Stephenson Room

- **Workshop 7:** “Solving Terminology Problems More Quickly with ‘IntelliWebSearch (Almost) Unlimited’”
  - Michael Farrell

- **Workshop 8:** “Teaching the use of computer-aided translation systems in a progressive manner”
  - Jessica Xiangyu Liu

### Poster

<table>
<thead>
<tr>
<th>Poster</th>
<th>Presenter</th>
<th>Additional Author(s)</th>
<th>Title of Poster Presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>P5</td>
<td>Manny Rayner</td>
<td>Alejandro Armando, Pierrette Bouillon, Nikos Tsourakis</td>
<td>A Tool for Building Multilingual Voice Questionnaires</td>
</tr>
</tbody>
</table>
# Table of contents

**Kevin Flanagan**

*Filling in the gaps: what we need from TM subsegment recall*

**Gábor Prószéký**

*Almost fifty years after the (first?) ALPAC report*

**Miguel A. Jimenez Crespo**

*Beyond prescription: What empirical studies are telling us about localization crowdsourcing*

**Nasredine Semmar, Othman Zennaki and Meriama Laib**

*Using Cross-Language Information Retrieval and Statistical Language Modelling in Example-Based Machine Translation*

**Koen Kerremans**

*Representing intra-and interlingual terminological variation in a new type of translation resource: a prototype proposal*

**Hernani Costa, Gloria Corpas Pastor and Miriam Seghiri**

*iCompileCorpora: A Web-based Application to Semi-automatically Compile Multilingual Comparable Corpora*

**Nizar Ghoula, Jacques Guyot and Gilles Falquet**

*Terminology Management Revisited*

**Victoria Porro, Johanna Gerlach, Pierrette Bouillon and Violeta Seretan**

*Rule-based Automatic Post-processing of SMT Output to Reduce Human Post-editing Effort*

**Jerzy Czopik**

*Quality Assurance Process in Translation*

**Rohit Gupta, Hanna Bechara and Constantin Orasan**

*Intelligent Translation Memory Matching and Retrieval Metric Exploiting Linguistic Technology*

**Tom Vanallemeersch and Vincent Vandeghinste**

*Improving fuzzy matching through syntactic knowledge*

**Marion Wittkowsky**

*Integrating Machine Translation (MT) in the Higher Education of Translators and Technical Writers*
<table>
<thead>
<tr>
<th>Author/Editors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joanna Drugan</td>
<td>109</td>
</tr>
<tr>
<td>Top-down or bottom-up: what do industry approaches to translation quality mean for effective integration of standards and tools?</td>
<td></td>
</tr>
<tr>
<td>Terence Lewis</td>
<td>118</td>
</tr>
<tr>
<td>Getting the best out of a mixed bag</td>
<td></td>
</tr>
<tr>
<td>Adam Kilgarriff</td>
<td>129</td>
</tr>
<tr>
<td>Terminology finding in the Sketch Engine: an Evaluation</td>
<td></td>
</tr>
<tr>
<td>Sabine Hunsicker and Alexandru Ceausu</td>
<td>133</td>
</tr>
<tr>
<td>Machine Translation Quality Estimation Adapted to the Translation Workflow</td>
<td></td>
</tr>
<tr>
<td>Andrzej Zydroń</td>
<td>137</td>
</tr>
<tr>
<td>The Dos and Don'ts of XML document localization</td>
<td></td>
</tr>
<tr>
<td>Kurt Eberle</td>
<td>145</td>
</tr>
<tr>
<td>AutoLearn&lt;word&gt;</td>
<td></td>
</tr>
<tr>
<td>A. Görög</td>
<td>155</td>
</tr>
<tr>
<td>Quality Evaluation Today: the Dynamic Quality Framework</td>
<td></td>
</tr>
<tr>
<td>Erin Lyons</td>
<td>165</td>
</tr>
<tr>
<td>Far from the Maddening Crowd: Integrating Collaborative Translation Technologies into Healthcare Services in the Developing World</td>
<td></td>
</tr>
<tr>
<td>Antonio Toral and Andy Way</td>
<td>174</td>
</tr>
<tr>
<td>Is Machine Translation Ready for Literature?</td>
<td></td>
</tr>
<tr>
<td>Manny Rayner, Alejandro Armando, Pierrette Bouillon and Nikos Tsourakis</td>
<td>177</td>
</tr>
<tr>
<td>A Tool for Building Multilingual Voice Questionnaires</td>
<td></td>
</tr>
<tr>
<td>Irina Burukina</td>
<td>182</td>
</tr>
<tr>
<td>Translating implicit elements in RBMT</td>
<td></td>
</tr>
<tr>
<td>Mozhgan Ghassemiazghandi and Tengku Sepora Tengku Mahadi</td>
<td>194</td>
</tr>
<tr>
<td>Losses and Gains in Computer-Assisted Translation: Some Remarks on Online Translation of English to Malay</td>
<td></td>
</tr>
<tr>
<td>AnneMarie Taravella</td>
<td>202</td>
</tr>
<tr>
<td>Affective Impact of the use of Technology on Employed Language Specialists: An Exploratory Qualitative Study</td>
<td></td>
</tr>
<tr>
<td>Michael Farrell</td>
<td>211</td>
</tr>
<tr>
<td>Solving Terminology Problems More Quickly with ‘IntelliWebSearch (Almost) Unlimited’</td>
<td></td>
</tr>
<tr>
<td>Eduard Šubert and Ondřej Bojar</td>
<td>217</td>
</tr>
<tr>
<td>Twitter Crowd Translation — Design and Objectives</td>
<td></td>
</tr>
</tbody>
</table>
Najeh Hajlaoui

SMT for restricted sublanguage in CAT tool context at the European Parliament

Jessica Xiangyu Liu

Task-based Teaching of Computer-aided Translation in a progressive manner

Index
Filling in the gaps: what we need from TM subsegment recall

Kevin Flanagan
Swansea University

ABSTRACT

Alongside increasing use of Machine Translation (MT) in translator workflows, Translation Memory (TM) continues to be a valuable tool providing complementary functionality, and is a technology that has evolved in recent years, in particular with developments around subsegment recall that attempt to leverage more content from TM data than segment-level fuzzy matching. But how fit-for-purpose is subsegment recall functionality, and how do current Computer-Assisted Translation (CAT) tool implementations differ? This paper presents results from the first survey of translators to gauge their expectations of subsegment recall functionality, cross-referenced with a novel typology for describing subsegment recall implementations. Next, performance statistics are given from an extensive series of tests of four leading CAT tools whose implementations approach those expectations. Finally, a novel implementation of subsegment recall, ‘Lift’, is presented (integrated into SDL Trados Studio 2014), based on subsegment alignment and with no minimum TM size requirement or need for an ‘extraction’ step, recalling fragments and identifying their translations within the segment even with only a single TM occurrence and without losing the context of the match. A technical description explains why it produces better performance statistics for the same series of tests and in turn meets translator expectations more closely.

1. Introduction

The segment-oriented nature of Translation Memory (TM) can seem to restrict its usefulness, in ways to which Machine Translation (MT) – in particular, Statistical Machine Translation (SMT) – provides an alternative. Bonet explains that, for the TMs at the DGT, “Many phrases were buried in thousands of sentences, but were not being retrieved with memory technology because the remainder of the sentence was completely different” (2013: 5), and that SMT trained on those memories enabled some of that ‘buried’ content to be recalled. However, TM technology has evolved in recent years, including subsegment recall features that attempt to leverage more content from TM data than segment-level fuzzy matching. In principle, TM subsegment recall – automatically finding phrases within segments that have been translated before and identifying the corresponding translated phrase in the previously-translated segment – should recover all that content. This functionality is described by Zetzsche as “probably the biggest and the most
important development in TM technology” (2014), but in practice, implementations in TM systems vary widely, and fall very short of that level of capability, leading to further observations by Zetzsche that “we are still in the infancy of these developments”, and that “subsegmenting approaches are almost as varied as the number of tools supporting them” (Zetzsche, 2012: 51).

The discussion in this paper is expressed in terms of segment-based TM, that is, TM containing Translation Units (TUs), each containing an easily-demarcated source text (ST) segment – such as a sentence, heading or list item – and its corresponding target text (TT) translation. However, the principal issue for subsegment recall – how to match fragments of segments, and retrieve the translation of a fragment, rather than of the whole segment where it occurs – applies equally to character-string-in-bitext (CSB) TM systems, where STs and TTs are stored in full, since the ST and TT alignment information available is essentially at the same level of granularity, so automatic identification of the translation of a fragment is problematic. For both segment-based and CSB systems, translators can usually prompt a search for a specific fragment – referred to herein as a concordance search – to find occurrences of fragment repetitions. Even so, discounting the time required to do so for all possible fragments (which some CAT tools will attempt automatically), the results show only the larger segment within which the fragment’s translation is found, leaving the translator obliged to spend time and effort scanning through it. To aid discussion of these and other considerations, and since the distinctions between approaches to subsegment recall in different CAT tools are not immediately obvious, Table 1 defines a typology of ‘behaviours’ – different techniques and characteristics – that can be used to describe subsegment recall implementations. These are discussed at greater length in (Flanagan, forthcoming 2015b). A more detailed version of this paper is also available at http://kftrans.co.uk/FillingInTheGaps.pdf. In the next section, the typology will be used to present the views of translators that participated in a subsegment recall survey.

2. Translators’ views

To gauge what functionality translators would like from subsegment recall, a controlled multiple-choice survey was conducted of translators from four groups: the Western Regional Group of the Institute of Translation and Interpreting (ITI); translators registered with Wolfestone, a successful language services agency; the ITI’s French Network, and students on MA in Translation programmes at Swansea University. In all, 91 responses were received, approximately evenly spread across the four groups. Details of questions and responses can be viewed at http://kftrans.co.uk/benchmarks/Home/Survey and are discussed at greater length in (Flanagan, forthcoming 2015b).

In summary, the responses showed a broad consensus with regard to subsegment recall features. Most expect TM-TDB to be available; there is a fairly equal split between wanting DTA/BFE and wanting ACS; VL is not desirable; requiring a TM to be large for subsegment recall is not desirable, and subsegment recall should be available for fragments occurring only once in the TM. The split between those wanting DTA/BFE and those wanting ACS merits examination. As

1. http://www.itiwrg.org.uk
2. http://www.wolfestone.co.uk/
3. http://www.iti-frenchnetwork.co.uk/
ACS, on the face of it, requires more translator time, since the TU has to be manually examined to locate the corresponding fragment translation, why would this be preferred by some over DTA or BFE? I speculate that this is because experienced translators are more aware of the dangers of decontextualisation, and the DTA/BFE response option did not specify whether context is provided. If another option had been available, like the DTA/BFE option but explaining that the translation suggestion was provided by (say) displaying the target segment from the TU with the translation suggestion highlighted, I suspect this response would have been chosen by the majority of respondents.

Having established a baseline for translators’ expectations for subsegment recall functionality, the next section will compare those expectations with actual CAT tool capabilities.

### 3. CAT tool comparison

Table 2 compares the subsegment recall functionality for all CAT tools that provide such a feature and were available at time of writing for trial (or free) use by translators, representing the range of software available to a translator evaluating tools before purchase. A tick means the CAT tool supports the feature, and any term used for it appears below the tick. This gives a high-level view of how varied is the functionality in different CAT tools providing subsegment recall. DTA and BFE implementations merit further examination, since approaches and results vary much more than for (say) the comparatively straightforward TM-TDB feature. Furthermore, the expectations from translators include the ability to recall translations of fragments even if they only occur once in a TM, and without needing the TM to be large. The following section examines this more closely.

<table>
<thead>
<tr>
<th>Behaviour</th>
<th>TM content</th>
<th>Example query</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Use TM like a TBD (TM-TDB)</strong></td>
<td><strong>EN</strong>: Dynamic Purchasing System FR: Système d’acquisition dynamique</td>
<td><strong>EN</strong>: We will define a completely electronic dynamic purchasing system for commonly-used purchases</td>
<td><strong>Highlights ‘dynamic purchasing system’ in query, displays “Système d’acquisition dynamique”</strong>.</td>
</tr>
<tr>
<td><strong>Automatic Concordance Search (ACS)</strong></td>
<td><strong>EN</strong>: A procuring entity may set up a system for commonly-used purchases, <strong>FR</strong>: L’entité adjudicatrice peut mettre en place un système pour des achats d’usage courant.</td>
<td><strong>as above</strong></td>
<td><strong>Automatically highlights ‘commonly-used purchases’, displays complete FR segment from TM where match found</strong>.</td>
</tr>
<tr>
<td><strong>Dynamic TM Analysis (DTA)</strong></td>
<td><strong>as above</strong></td>
<td><strong>as above</strong></td>
<td><strong>Retrieves translation for ‘commonly-used purchases’, i.e. ‘achats d'usage courant’</strong>.</td>
</tr>
<tr>
<td><strong>Bilingual Fragment Extraction (BFE)</strong></td>
<td><strong>as above</strong></td>
<td><strong>as above</strong></td>
<td><strong>Same as DTA, but requires TM content to be extracted beforehand</strong>.</td>
</tr>
<tr>
<td><strong>Decontextualisation</strong></td>
<td><strong>as above</strong></td>
<td><strong>as above</strong></td>
<td><strong>Retrieves translation for ‘commonly-used purchases’, i.e. ‘achats d'usage courant’, but does not show context</strong>.</td>
</tr>
<tr>
<td>Machine Recall (MR)</td>
<td>(as above)</td>
<td>(as above)</td>
<td>Displays translation for 'commonly-used purchases' automatically.</td>
</tr>
<tr>
<td>---------------------</td>
<td>------------</td>
<td>------------</td>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>Assisted Recall (AR)</td>
<td>(as above)</td>
<td>(as above)</td>
<td>Only displays translation for 'commonly-used purchases' when user starts to type it, i.e. types 'a' or 'ac'.</td>
</tr>
<tr>
<td>Variation Loss (VL)</td>
<td>EN: The company can therefore be qualified as a firm in difficulty. FR: C'est pourquoi elle est considérée comme une entreprise en détresse. [...]</td>
<td>EN: The firm in difficulty may benefit from aide. FR: L'entreprise en difficulté peut bénéficier d'une aide.</td>
<td>It is doubtful whether a firm generating profits so quickly can be deemed to be a firm in difficulty. Only one of the translations of 'firm in difficulty' is retrieved.</td>
</tr>
</tbody>
</table>

**Table 1: Typology examples**

<table>
<thead>
<tr>
<th>SDL Trados Studio 2014</th>
<th>TM-TDB</th>
<th>ACS</th>
<th>DTA</th>
<th>BFE</th>
<th>Min TM size</th>
<th>Min. occurrences</th>
<th>Decontextualisation</th>
<th>Recall type</th>
</tr>
</thead>
<tbody>
<tr>
<td>MetaTexis v3.17</td>
<td>✓ ‘use TM as TDB’</td>
<td>-</td>
<td>-</td>
<td>✓ ‘AutoSuggest Creator’</td>
<td>10,000</td>
<td>✓</td>
<td>Yes</td>
<td>AR</td>
</tr>
<tr>
<td>memoQ 2013³ R2</td>
<td>✓ ‘LSC’</td>
<td>✓ ‘LSC’</td>
<td>✓ ‘Muse’</td>
<td>-</td>
<td>(ACS)² (BFE)S</td>
<td>(ACS)No (BFE)Yes</td>
<td>(ACS)MR³ (BFE)AR</td>
<td></td>
</tr>
<tr>
<td>MemSource v3.148</td>
<td>✓ ‘Subsegment match’</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>MR</td>
</tr>
<tr>
<td>Déjà Vu X2⁵ v8</td>
<td>✓ ‘Assemble’</td>
<td>✓ ‘DeepMiner’</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Yes</td>
<td>MR³</td>
<td></td>
</tr>
<tr>
<td>Similis Freelance v2.16</td>
<td>-</td>
<td>-</td>
<td>✓ ‘Glossary’</td>
<td>-</td>
<td>-</td>
<td>Yes</td>
<td>MR</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2: Subsegment recall types by CAT tool**
1. if ‘Guess translation’ activated.
2. Can be configured for just one occurrence, though DTA results less reliable (see later analysis in this paper).
3. No minimum specified, but with few occurrences or only one, results may be poor (see later analysis in this paper).
4. AR suggestions are also available.
5. Déjà Vu X3 was released in February 2014; initial testing indicates this functionality is essentially unchanged.
6. The Concordance Search option “Perform search if the TM lookup returns no results” is not an implementation of ACS.
7. The same ‘LSC’ feature names covers both TM-TDB and ACS when – say - enabling/disabling this functionality, even though they give rise to different behaviours; TM-TDB matches show the translation in the results pane, ACS matches don't.
8. memoQ 2014 was released in June 2014; initial testing indicates this functionality is essentially unchanged.

(Note: Fluency 2013 includes BFE, but this was not functional at time of writing, something the vendor confirmed would be addressed (Tregaskis, 2014). Across Language Server provides BFE functionality, but unlike Personal Edition there is no trial or free version available.)

4. Performance comparison

DTA and BFE subsegment recall implementations in CAT tools are very varied and require close examination to determine how well they meet translators’ functionality expectations. This section presents a suite of tests used to measure their performance in this regard, starting with a TM containing known subsegment fragments and their translations, querying the TM with sentences to translate containing one of the fragments, then checking whether the fragment translation is recalled.

4.1. Data preparation

To select test fragments and their translations for use in such a performance evaluation, a 40,000 TU French-English section of the DGT-TM (Steinberger, Eisele, Klocek, Pilos, & Schlüter, 2013) was processed to select some frequently-occurring fragment pairs, shown in Table 3, along with codes used to refer to them herein. For each fragment pair, 100 ‘fragment-bearing’ TUs (TUs containing the fragment pair) were extracted. A further 10,000 ‘padding’ TUs containing none of the fragments was extracted for creating test TMs. To simulate translating a source text that includes a test fragment also found in a test TM, example sentences – hereafter, ‘queries’ – were created by adapting fragment-bearing TUs. Each query TU was compared to the 10,000 ‘padding’ TUs and the relevant fragment-bearing TUs to ensure that neither French nor English segment constituted a ‘fuzzy match’ with any TU segment.

<table>
<thead>
<tr>
<th>Code</th>
<th>French</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Règlement</td>
<td>Regulation</td>
</tr>
<tr>
<td>1a</td>
<td>Établi</td>
<td>Established</td>
</tr>
<tr>
<td>2</td>
<td>conclut que</td>
<td>concludes that</td>
</tr>
<tr>
<td>2a</td>
<td>État membre</td>
<td>Member State</td>
</tr>
<tr>
<td>3</td>
<td>modifiée comme suit</td>
<td>amended as follows</td>
</tr>
<tr>
<td>3a</td>
<td>les autorités polonaises</td>
<td>the Polish authorities</td>
</tr>
<tr>
<td>4</td>
<td>intégrée dans l'accord</td>
<td>incorporated into the Agreement</td>
</tr>
<tr>
<td>6</td>
<td>Journal officiel de l'Union européenne</td>
<td>Official Journal of the European Union</td>
</tr>
</tbody>
</table>

Table 3: Test fragment pairs
For each query TU, a TM was created for different combinations of padding-TU quantity and fragment-bearing TU quantity; 100, 1,000 or 10,000 padding TUs combined with 1, 100 or 1,000 fragment-bearing TUs, making nine TMs per query TU (and nine further reversed-language-way TMs). Two documents per query TU were created – one containing the French query sentence; the other, the English – and presented for translation by each CAT tool using each of the nine TMs in turn. Subsegment translation suggestions were recorded and scored as described below. Further details of test data and queries, plus discussion of the motivation behind them and their preparation can be found in (Flanagan, forthcoming 2015b).

4.2. Scoring

For these tests, suggestions were scored in terms of precision and recall. Precision is the percentage of words in the suggestion that occur in the expected fragment translation (expressing how much is relevant), while recall is the percentage of words in the expected fragment translation found in the suggestion (expressing how complete the recall is). Where there are several suggestions, these values are averaged. This is discussed further in (Flanagan, forthcoming 2015b).

For certain CAT tools, subsegment recall precision cannot be evaluated, since the interface does not show which part of the source text the suggestion words are meant for. These cases are shown below as ‘Precision unavailable’. The exact procedures for recording results for each CAT tool vary according to their very different approaches; the specifics for each can be found in (Flanagan, forthcoming 2015b). Results show how each tool performed under varying conditions (TM size, number of fragment occurrences, fragment length). Due to their approaches to providing subsegment recall suggestions, results for different tools are not directly comparable, but do show how the variables concerned affect performance in different ways and give some indication of how performance may differ between tools. Results are based on AR suggestions only if the tool does not offer any MR implementation.

4.3. Results

The graphs in Table 4 show recall and precision for each CAT tool tested, averaged over all test queries for that tool (eight English queries and eight French queries), where the X-axis shows fragment frequency (the number of fragment-bearing TUs in the TM), and the different lines show the number of padding TUs in the TM. Note: for Similis, varying volumes of TM padding make no difference to results, so they were all obtained using the same amount of TM padding. For memoQ, DTA subsegment recall was evaluated, since its AR-based BFE implementation cannot be configured to recall fragments with fewer than 5 occurrences. Detailed results for the individual queries and CAT tools can be found at http://kftrans.co.uk/benchmarks.

The averaged results show that very different results are produced by the tools tested. Results from a given CAT tool for individual queries show an interesting lack of consistency. With memoQ, for fragment 3, recall is consistently high, and precision tends to increase with frequency, regardless of padding volume, while for fragment 3a, recall and precision drop sharply as frequency increases, depending on padding volume. Performance is generally comparable when the language direction is reversed, but in some cases differs noticeably. With Déjà Vu X2, performance in individual cases is very varied, with noticeable differences dependent on
language direction. Graphs for Similis tend to be flat – if Similis can recall a fragment suggestion, frequency usually makes no difference to whether it is recalled, generally with consistent precision. However, recall seems to be affected by the grammatical category of the fragment sought (per the results for the two different three-word fragments, for example), so that for certain fragments, no translation suggestion is produced regardless frequency. With SDL Trados Studio, performance overall is quite consistent, with 100% recall usually achieved at a frequency of 10, though it has a large TM requirement, and the implementation is AR rather than MR.

Table 4: Averaged performance statistics by CAT tool
4.4. Discussion

At least some CAT tools provide implementations which – under the right circumstances – provide subsegment translation suggestions with good recall and precision levels, though performance may be inconsistent, with identical texts and data giving different results if the language direction is reversed, for instance. Translators surveyed have some clear preferences about subsegment recall functionality, including wanting it available even for small TMs, and even for fragments occurring only once. Of the DTA/BFE systems tested, Similis had the best average performance under those circumstances, recalling translations of single-occurrence fragments about half the time, with average precision around 60%. However, its BFE methodology decontextualises the translations, arguably aggravating still further a weakness in segment-level TM, and in different circumstances (more fragment occurrences, sufficiently large TM) it can be out-performed by other systems.

Although weaker in other areas, Similis meets the aforementioned preferences better because it is the only system not reliant on statistical analysis or repetitions, instead aligning ‘chunks’ of source and target language segments, “as long as the languages processed are parallel enough for it to do so” (Planas, 2005: 5). The next section presents Lift, a TM system intended to meet translator expectations better by also taking an ‘aligning’ approach, but developed for more consistent results, and with a DTA rather than BFE methodology so as not to decontextualise translations recalled, as well as to reflect TM content changes immediately.

5. Lift

5.1. Overview

Lift is a TM system implementing DTA subsegment recall based on fine-grained alignment of segment pairs, or subsegment alignment. To enable subsegment alignment, Lift uses a range of bilingual dictionary resources to establish tentative lexical alignments, then an iterative hypothesis evaluation algorithm plus some deductive heuristics to produce a hierarchically-arranged set of word span alignments. The alignment algorithm is described in detail and compared with other approaches in (Flanagan, 2014). Figure 1 shows a high-level view of the alignment process. (For the aligned sentence pair, connecting lines show alignments between words, while parallelograms show alignment between spans of words.) The effects of using the optional components are described in (Flanagan, forthcoming 2015a).

During translation, Lift uses a longest-common-substring algorithm coupled with indexing techniques and configurable parameters (such as minimum fragment length and proportion of ‘stop’ words) to match fragments of a query (that is, a sentence to translate) with fragments of TM content, and uses the alignment information to recall and propose the translations of those fragments to the translator. The recall process is described at greater length in (Flanagan, forthcoming 2015a).
5.2. SDL Trados Studio 2014 integration

Lift exposes an Application Programming Interface (API) allowing it to be integrated into CAT tools. An example integration has been developed for SDL Trados Studio 2014. The screen capture in Figure 2 gives an overview of how Lift’s functionality is provided while translating. Figure 3 shows a larger view of the ‘Lift Results’ pane. The top section of the pane shows the sentence being translated, with underlining on all words that have been found in a subsegment match. This allows the translator to see at a glimpse for which parts of the sentence translations have been located, without having to scroll through the list of matches. The matches and their corresponding translation suggestions are shown in a list immediately below. The user can quickly insert a selected suggestion into the target text for the segment by double-clicking on the highlighted text it matches in the sentence being translated (right-clicking highlighted text produces a fly-out display of the translation suggestion, to save scrolling the list of matches if it is not visible). Double clicking an item in the list will also insert the translation. Alternatively, the user can begin typing one of the translation suggestions, then use auto-complete functionality for the rest of it, as shown in Figure 4. To review additional matches, the user can scroll down the list and click an item to see details and context, as shown in Figure 5.

To examine whether this functionality would better meet translators’ expectations, the next section describes the results achieved using Lift for the same suite of tests used above to compare CAT tool performance.
Figure 2: Trados integration overview

Sentence being translated

Figure 3: Lift results pane

Figure 4: Auto-complete functionality
5.3. Performance comparison

In order to compare Lift's subsegment recall performance with that of the CAT tools evaluated above, the same data and suite of tests were used with a Lift installation. The graphs in Table 5 show recall and precision for Lift, averaged over all test queries (eight English queries and eight French queries). For Lift, varying volumes of TM padding make no difference to subsegment recall results, so they were therefore all obtained using the same amount. Detailed results for the individual queries can be found at [http://kftrans.co.uk/benchmarks](http://kftrans.co.uk/benchmarks).

The average results help summarise that with the fragments and TMs described above, Lift recalls their translations regardless of the number of occurrences, with generally very good precision, as well as neither decontextualising the translations nor exhibiting variation loss. The detailed results show that incorrect translation suggestions can be produced when TUs have not been correctly aligned by Lift. An example is discussed in the longer version of this paper mentioned at the end of the introduction.

6. Conclusion

The survey of translators' expectations of subsegment recall functionality found that around half expected functionality corresponding to DTA or BFE per the typology presented above (and I speculate more would do so if it were clear that the implementation would not be decontextualising). In particular, they expected recall to be available even for fragments occurring only once in the TM, and without any requirement for the TM to be large. A test suite was used to analyse subsegment recall capability for a range of CAT tools, showing that performance did not meet translators' expectations well. When used to analyse Lift's capability, results showed it met those expectations much better. Notwithstanding the small number of problematic alignment
cases, these results seem very encouraging. Nevertheless, the suite of tests used involves a limited number of variables and carefully-controlled test data. A wider-ranging evaluation covering English, French, German, Spanish and Welsh, using much more extensive testing, is described in (Flanagan, forthcoming 2015a), where results indicate that performance is also good for those languages and with more comprehensive test cases. Nevertheless, even if controlled experiments suggest that new TM technology performs well when measured using whatever metrics, the success or failure of any attempt to develop and improve TM can ultimately only be judged by providing the developments to translators for real-world use, so that translators themselves can return a verdict.

References

BONET, J. (2013). No rage against the machine. Languages and Translation (6), EU Directorate-General for Translation, Brussels.


Almost fifty years after the (first?) ALPAC report

Gábor Prószéky
MorphoLogic & Pázmány University, Budapest, Hungary

ABSTRACT

The presentation will provide a historical flashback of Machine Translation (MT) by reviewing the significant milestones in its development and will reflect as to what the future has in store. To start with, the early developments after the Second World War II will be outlined. Next, the presentation will elaborate that the addition of morphological, syntactic and semantic knowledge did not lead to expected improvements which in turn, triggered the ALPAC report in 1966.

In the 1990s that landscape significantly changed due to the emergence of large amount of language data (corpora) which offered new opportunities for the rise and deployment of Statistical Machine Translation (SMT). SMT has been recently enhanced by the incorporation of morphological, syntactic and semantic information but the results are still not as good as expected. The presentation will review these recent developments and will reflect as to what are the options for the EU decision makers given that high quality MT is still a desideratum...

1. Some thoughts on progress in MT — almost fifty years after the (first?) ALPAC report

1. Several years after the World War II, a new scientific plan was born: translating from one language to another by the use of modern computing devices. The idea was simple: converting strings of language A into strings of language B, as the cryptographic activity of the war period suggested. Support of the idea of machine translation in the United States at that time was mainly motivated by the Cold War. Decision makers in the US governing bodies were quickly convinced by the new idea. There were not too many languages in their minds: first of all translation of Russian texts into English was in the focus.

   a. When the first real translation algorithms were made, a small modification of the original idea of pure string manipulation was made soon. The reason was very simple: words of Russian have inflections at their ends, thus, an unavoidable module, namely morphological analysis, was added to the basic “string transforming” algorithm. The results of the modified translation systems were, however, still not good enough...
b. Some years later, partly due to research results of the early generative linguists, it became clear that **syntactic structures** of human languages must play an important role in computational language processing. The analysis phase of machine translation, therefore, started to use syntactic modules to replace the simple word reordering technique. The results of the modified translation systems were, however, still not good enough...

c. Some years later, the first attempts toward computational **semantics** arose, and meaning-oriented information was added to some machine translation systems. The argument was easy to understand: high-quality machine translation cannot be made without a sort of "understanding" the string to be translated. Unfortunately, after adding so many linguistically important modules to the basic algorithm, the final results of fully automatic rule-based machine translation systems did not become significantly better.

d. The US Government had been waiting for a rather long time, but the expected **high-level results of machine translation did not come**. This situation led to the birth of the Automated Language Processing Advisory Committee, which published the opinion of its expert members in the famous **ALPAC Report** in 1966.

2. A quarter century later, in the early nineties, another new scientific plan was born: translating with the help of statistical knowledge derived from huge text corpora that were already available at that time. **The idea was simple** again: strings of language A can be translated into strings of language B, if there is enough statistical evidence for it in the corpora. Decision makers in the EU were quickly convinced by the new idea: multilingual Europe has a lot of potential language pairs and this solution would solve the problem of huge amounts of translation tasks.

   a. A step towards application of linguistically motivated modification of the basic algorithm came soon: when some languages use magnitudes more word forms than other languages, something should be done with **morphology**. To solve this problem, factored statistical translation was introduced, and made the basic algorithm a bit more better and a bit more complicated. The results of the modified translation systems were, however, still not good enough...

   b. Syntactic constructions of the languages of the EU are quite heterogeneous, thus structurally different language pairs are rather difficult to use in the statistical machine translation paradigm: some automatic tool was needed making strings of source and target languages formally more similar to each other. This led to the birth of syntactic reordering and other sophisticated **syntax-driven** algorithms added to the basic SMT paradigm. The results of the modified translation systems were, however, still not good enough...

   c. Nowadays, there are several attempts that use some sort of **semantic information** combined with the basic statistical machine translation algorithm. Unfortunately, after adding so many linguistically important modules to the basic
algorithm, the final results of statistical machine translation systems did not become significantly better.

d. The EU decision makers have been waiting for a rather long time, but the expected high-level results of machine translation do not come. The question to be answered is the following: How long do the EU decision makers have to wait yet for the expected high-level results of (statistical) machine translation? And if they don't come very soon, what's the next step?
Beyond prescription: What empirical studies are telling us about localization crowdsourcing

Miguel A. Jiménez Crespo
Rutgers University

ABSTRACT

Translation crowdsourcing represents a new and quickly evolving phenomenon that has attracted the attention of industry experts and scholars alike. During recent years the industry has released a number of publications, mainly case studies and best-practice reports, while academic disciplines such as Computational Linguistics and Translation Studies (TS) have primarily focused on empirical studies. This paper attempts to compare and critically analyze research produced from both perspectives and locate these different approaches within the wider cycle of applied and theoretical/descriptive research. The findings of empirical studies on volunteer motivation and quality in TS will be contrasted with the best practices in the industry. This analysis will show a potential avenue to engage both perspectives to collaborate towards closing the existing research gap.

1. Introduction

During the last two decades, translation has experienced a digital revolution that has given rise to new phenomena and practices, such as different translation technologies or translation crowdsourcing that are reshaping both industry practices as well as societal views and theories of translation (O’Hagan 2013; Jiménez-Crespo 2013a). Over the years, different stakeholders in the study of translation have followed different paths due to diverging objectives. These objectives range from the more prescriptive and applied industry approaches to theoretical or empirical studies. Industry research often appears in response to the rapid development of technologies and the need to quickly adapt to an ever-evolving field. For example, the industry has tried to rapidly understand, harness and exploit the power of the crowd to produce translations (Jimenez-Crespo 2011). This means that industry experts normally produce applied research at a much quicker rate than academic disciplines (O’Hagan 2013). On the other hand, Translation Studies (TS) often trails behind industry research, adopting industrial de-facto models and conceptualizations that result from an applied and prescriptive approach (Jimenez-Crespo 2013a). This paper argues that crowdsourcing represents a prime example of an exciting new phenomenon that can help us assess and understand why the “gap” between both fields exists while simultaneously helping us to be more aware of possible synergies between both fields.
This study is partly motivated by the existing need in the industry to identify best practices for crowdsourcing in the rapidly developing world of crowdsourcing. Desilets and van de Meer indicate that “there is a clear need for a more concise, summative body of knowledge that captures recurrent best practices” (2011: 29). The authors also mention that current practitioners are the most suitable subjects for creating them: “we advocate the building of such a compendium, [...] which could be written collectively by practitioners of collaborative translation.” (Ibid: 29). Prescriptive collections of best practices can be found in the different publications by Desilets (Desilets 2011; Desilets and Van de Meer 2011), as well as work by DePalma and Kelly (2011), among others. It is often the case that the research gap rests on pressure from experts and professionals to convert research findings into applicable “how to” knowledge. Nevertheless, both perspectives do feed into each other. After all, prescriptive practices recommended in industry publications and existing case studies can help develop studies and testable hypotheses in the descriptive branch and develop theoretical models. Similarly, empirical and theoretical research can help shape best practices.

It should be mentioned at this point that according to the canonical map of TS as a discipline (Holmes 1984), research can fall under the Theoretical/Descriptive or the Applied branches. The latter branch focuses its attention on the work of professionals and practitioners, while the Theoretical/Descriptive branch is largely the realm of scholars and researchers. Both branches represent a global cycle in which the transfer of knowledge in both directions represents one of the main engines of evolution of the discipline and the production of knowledge about existing phenomena (Rabadán 2010). That is, both branches feed into each other and therefore help refine theories, models and applied practices. Obviously, the several stakeholders interested in the advancement of research (namely professionals and scholars), can have different objectives, tempos and research agendas, but both can and should cooperate towards a common goal.

The following sections review empirical research into crowdsourcing in TS and related disciplines and connect findings from this research with best practices recommended by the industry.

2. Empirical research into crowdsourcing

Empirical research into translation crowdsourcing has emerged mainly from two related perspectives, (1) Computational Linguistics / Machine Translation and from (2) TS. In the first case, research has focused on the development of workflows and models to harness the knowledge of the crowd (i.e. Shimoata et al. 2001; Morera-Mesa and Filip 2013), sometimes comparing professionals vs. crowdsourced translations to feed MT engines (Zaidan and Calliston Burch 2012). Empirical research in TS has mainly focused on two main research questions: motivation of volunteers to participate in translation initiatives and translation quality in crowdsourced texts. Both research issues, motivation and quality, also appear to be primary concerns in industry publications. For example, Desilets and Van de Meer (2011) indicate in their collection of best practices that emerged after a TAUS 2011 meeting that “[m]otivation issues are most critical in crowdsourcing scenarios, and this is possibly the main reason why it has yet to become widespread.” (p.32). Similarly, DePalma and Kelly indicate that it is necessary to discover volunteer motivations, and organizations need to “keep nurturing them with rewards and incentives” (2011: 401). Motivation also predominantly appears in Common Sense Advisory
publications (DePalma and Kelly 2008; Kelly and Ray 2011). For its part, and despite the enormous importance of translation quality, crowdsourcing quality seems to be less of a concern in industry publications than motivation:

Quality Control issues tend to resolve themselves, provided that enough of the "right" people can be enticed to participate and that you provide them with lightweight tools and processes by which they can spot and fix errors. (Desilets and van de Meer 2011: 41).

In any case, TS studies have also paid less attention to quality issues in crowdsourcing, with a lower number of empirical studies on this topic as the following sections will show.

3. Empirical Research into motivation in TS

Since 2010, a growing number of empirical studies have appeared in TS related to the motivation of volunteers. Theoretically, these studies have mostly departed from what is known as “sociological approaches” to translation (Wolf 2010). According Chesterman topics of interest in sociological approaches relevant for crowdsourcing research are “the social role of the translators and the translators' profession, translation as a social practice...” (2007: 173-174). The main research questions that have been the object of empirical inquiry have been (1) what are the motivations of volunteers, (2) what are their profiles? and (3) how are these volunteers organized? (Orrego-Carmona 2012). This section will focus on the results of volunteer motivation to participate in crowdsourcing initiatives.

The methodologies for these studies are mostly interventionist ones such as online surveys. Studies have focused on motivation to participate in Wikipedia (McDonough Dolmaya 2012), Facebook (Dombek 2013; Mesipuu 2012), TED open translation initiative (Camara forthcoming), Skype (Mesipuu 2012), or non-profits such as the Rosetta Foundation (O´Brian and Schäler 2010). The following table summarizes the studies, initiatives and the number of respondents in the surveys.

<table>
<thead>
<tr>
<th>Researcher(s)</th>
<th>Initiative</th>
<th>N. of subjects in survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>O'Brien and Schäler (2010)</td>
<td>Rosetta Foundation</td>
<td>139</td>
</tr>
<tr>
<td>Mesipuu (2012)</td>
<td>Facebook and Skype</td>
<td>10 each (20 total)</td>
</tr>
<tr>
<td>Dombek (2013)</td>
<td>Facebook / Poland</td>
<td>19 + 20</td>
</tr>
<tr>
<td>Camara (forthcoming)</td>
<td>TED</td>
<td>177</td>
</tr>
</tbody>
</table>

In order to compare the results of these studies with the best practices in the industry, the results from these studies were summarized and critically analyzed. Even when all the above-mentioned studies depart from slightly different perspectives and different theoretical foundations, it was possible to identify similar formulations of survey questions and underlying motives. Most studies separate between two basic notions in existing theories of motivation: the fact that they can be intrinsic or extrinsic (Frey 1997). Intrinsic motivations are those related not to financial compensation or reward but rather to personal enjoyment or a feeling of obligation to a specific community. Examples of this motivation type in the survey questions in studies are “Found the project intellectually stimulating” or “Help make information available in other
languages”. Extrinsic motivations are related to direct or indirect rewards, such as gaining more clients, getting presents or the potential to attract customers.

Comparability was somewhat made difficult by the (1) different measuring scales, such as the likert scale of the O´Brien and Schäler (2010) compared to the multiple choice options in McDonough Dolmaya (2012), or (2) the differences in formulation of the potential motives. It was decided for comparability purposes, to rank the results from all studies numerically and then subsequently aggregate and compare them. The outcome of this analysis yielded three tiers or groups of motivations, from the first tier of motivations, namely those that consistently seem to appear at the top of most studies, to those less important for the volunteers. The first group or main tier of motivations includes exclusively intrinsic motivations such as:

1. Making information in other languages accessible to others.
2. Helping the organization with their mission or a belief in the organizations' principles.
3. Receiving intellectual satisfaction, probably related to the notion of ‘cognitive surplus’ (Shirky 2010).

The second tier as reported by participants includes both intrinsic and extrinsic motivations:

4. The desire to practice the second language.
5. The need to gain professional translation experience or increase one’s reputation

Finally, a range of other motivations that appear consistently at the lower end of the results are:

6. The desire to support less known languages.
7. The satisfaction of completing something for the good of the community.
8. For fun.
9. The desire to be part of a network.

To some extent, it is surprising that the community component of this participation, that is, being part of a network, tends to be at the bottom of the motivations reported by users. This was the main motivation finding in the study of the close community of Skype volunteer translators (Mesipuu 2012). This finding may point to different motivations in cases of open or closed translation communities. It should also be kept in mind that all studies, in tune with findings about motivations in other crowdsourcing and volunteering areas, have concluded that a combination of motives, rather than any single one, underlies volunteer motivation. According to Olohan “volunteers are often motivated by a combination of factors and can be seen as behaving simultaneously altruistically and egoistically” (2014: 19). In any case, the only study that separated between professional and non-professional translators, that of McDonough-Dolmaya, identified that the main difference between both populations is the greater significance of extrinsic motivations for translation professionals, i.e. reputation, attract clients, etc. Another difference of interest between professionals and non-professionals, even when professionals make around 7 to 16 % of volunteers in the studies on Wikipedia and TED talks, is that professionals are also more attracted to initiatives that they perceive to have “greater cultural or symbolic values” or “more prestigious activities” (McDonough-Dolmaya 2012: 188). This brings up the question of the role of professionals in these initiatives. It is often the case that best practices reports indicate
that professionals should be involved where needed (DePalma and Kelly 2011), both by conducting in-house reviews or attempting to motivate them to volunteer. In this case, only those initiatives perceived by professionals with this higher symbolic value or prestige will be able to attract them as uncompensated participants. As the case of LinkedIn showed, requesting exclusively professionals to participate can backfire, but certain filters such as exams or evaluations can help bring to the initiative participants with a sufficient degree what is known as “professional expertise”.

4. Motivation in best practices publications

As previously mentioned, the publications by Common Sense Advisory (i.e. DePalma and Kelly 2009, 2011) and those resulting from the TAUS 2011 meeting (Desilets 2011; Desilets and van de Meer 2011; Collaborative Translation Patterns 2011) are examples of prescriptive best practices developed within the industry. Both publications include a list of similar areas. The TAUS report includes a compendium of the most commonly used decision-making patterns, previously identified issues in the implementation of crowdsourcing during a meeting with industry experts and recommendations for how to best tackle each of them. The areas of interest include the following sections:

1. Planning and scoping
2. Community Motivation
3. Translation Quality
4. Contributor Career Path
5. Community right-sizing
6. Tools and Processes

As far as the Community motivation is concerned, industry experts suggest that these issues can be potentially be solved fostering of intrinsic motivations (12 recommendations out of 13), while only one of them relates to handing out branded products. Only two studies asked translators whether they would be motivated if gifts were handed out. In the case of Skype (Mesipuu 2012), it was found that community events, getting together in beta releases, was a more powerful motivator according to participants than handing out t-shirts or other merchandise. Similarly, in the study for the non-profit Rosetta foundation, merchandise, gifts or monetary compensation came at the bottom of the list in the survey, while intrinsic motivations such as feedback from qualified translators or clients, as well as invitations to events were reported twice as often as free gifts or payments. Additionally, subjects indicated that the least attractive incentives to motivate them in the future were practices such as translator of the month profiles or monthly top-ten lists. Practically all best-practice reports include these types of incentives and to some extent this finding contradicts these recommendations in industry publications. For example, DePalma and Kelly identify the main incentive to motivating participants is to “keep nurturing them [volunteer translators] with rewards and incentives...Something as simple as a certificate can be a powerful form of recognition” (2011: 403). They also indicate the value of “Highlight[ing] and showcase[ing] member contributions.

---

Companies in the article find the “leaderboard” to be an effective tool.\(^2\) However, the study by O’Brien and Schäler (2010) also found out that top lists could be somewhat detrimental to the engagement of “lurkers”, those with little time to volunteer (Nielsen 2006)\(^3\). The authors indicated that:

> Some volunteers […] mention factors that would demotivate them. In particular, turning their activities into a competition by making them bid against each other or simply compete for positions on leadership boards was highlighted... (n.p.)

It is therefore necessary to put our fingers on the existing discrepancies between initial industry practices to motivate volunteers and the opinions of those participating in motivation-related surveys. The practice of including a leaderboard might, nevertheless, be recommended in the industry since it seems directed towards recruiting and retaining the low percentage of highly active participants that volunteer beyond the average weekly average of 2 to 5 hours identified in studies (McDonough-Dolmaya 2012; Camara forthcoming).

5. What about empirical studies on crowdsourcing quality?

The issue of quality in crowdsourcing has been the focus of a number of studies in Machine Translation with the goal of feeding MT engines (i.e. Yan et al. 2014; Zaidan and Callison-Burch 2012). Within TS, two empirical studies have focused on aspects related to quality of crowdsourced translations (Jimenez-Crespo 2013b, forthcoming). The research objectives of these studies are to identify whether crowdsourcing can produce ‘naturally’ sounding translations and whether the actual crowdsourcing process has an impact on the result.\(^4\) In Jimenez-Crespo (2013b) a corpus based approach to research crowdsourcing was taken to research whether the Spanish version of Facebook included the most conventional terminology and phraseology that appeared more frequently in non-translated Spanish social networking sites. The non-translational section of the corpus included all the interactive and navigation segments in the 25 most popular social networking sites in Spain in 2009. The results showed that the localized version of Facebook included the most terminology and phraseology identified in the non-translated sites. This study therefore confirmed that the translation workflow used by this company, as with many other TS theoretical proposals over the years (Jimenez-Crespo 2011), is effective in order to achieve texts similar to non-translated ones. The study also concluded that this method is more effective for producing a localized website that resembles non-translated ones than the average professional localization process.

The best practices repository and the workflow studies by Morera-Mesa (Morera-Mesa and Filip 2013; Morera Mesa 2014) document a range of possible procedures employed to guarantee

\(^2\) DePalma and Kelly (2011) indicate also that it is necessary to remember that all volunteers do not have the same amount of time and it is necessary to recognize them all.

\(^3\) According to Nielsen 90% of participants in crowdsourcing efforts are ‘lurkers’ who never contribute, while 9% contribute a little and 1% of participants account for all of the activity.

\(^4\) Dombek (2013) in her study on the interaction between the Facebook Translate platform and volunteers concluded that subjects expressed the actual configuration of the process was detrimental to their translation process.
quality in crowdsourced translations. Many of them depend on the type of workflow in place. For example, organizations such as Kiva include entry exams, while many others apply automatic reputation management such as Cucumis. The same occurs when translation alternatives are open or closed, such as Asia Online (Morera and Filip 2013). In any case, the implications of the above mentioned studies for the industry point at two facts of interest: (1) Provided that enough volunteers participate in the open voting process implemented by Facebook, translations can result in more natural texts similar to those spontaneously produced in-country, and (2) the process implemented to produce the translations does have an impact on the final configuration of the target texts (Jimenez-Crespo 2013b). In this sense, more research should be conducted to compare different crowdsourcing workflow models and their resulting translations, such as open or closed alternatives followed by voting, translation plus in house revision, publish then revise, etc.

6. Conclusions

The uncertainty of working with volunteers and therefore the dependency on their motivation present interesting challenges that both the industry and TS are attempting to uncover. This paper has argued that this area presents an interesting case for bringing together industry and TS research since both interested parties are working in interrelated areas of a global research cycle, applied and prescriptive vs. theoretical and descriptive, and both feed into each other. Desilets and van de Meer concluded in their paper that “most practices […] are not that different from best-practices which are being used for crowdsourcing in other domains” (2011: 41), and ask whether translation requires a set of best-practices. However, research by McDonough-Dolmaya (2012) identified clear differences in the motivations and types of participants if translation crowdsourcing is compared with studies in Free and Open Software. Other studies have identified a great variety of existing workflows that apply exclusively to crowdsourced translation (Morera-Mesa 2014). This means that the potential to research and identify best practices through the global cycle of research, whether it starts in the applied or the theoretical-descriptive side is wide open. It is hoped that this paper helps bridge the gap between the industry and academic research and starts to provide a foundation for joint research projects in this fascinating area.

References


5 The TAUS report includes: Entry Exam, Peer Review, Automatic Reputation Management, Random Spot-Checking, In-House Revision, Users as Translators, Voting, Transparent Quality Level, Publish then Revise, Refining Redundancy, Open Alternatives, Hidden Alternatives


MESIPUU, M. (2012).“Translation crowdsourcing and user-translator motivation at Facebook and Skype”. Translation Spaces 1:33-53.


Using Cross-Language Information Retrieval and Statistical Language Modelling in Example-Based Machine Translation

Nasredine Semmar
CEA, LIST
Vision and Content Engineering Laboratory
Gif-sur-Yvette, France

Othman Zennaki
CEA, LIST
Vision and Content Engineering Laboratory
Gif-sur-Yvette, France

Meriama Laib
CEA, LIST
Vision and Content Engineering Laboratory
Gif-sur-Yvette, France

ABSTRACT

In this paper, we present the CEA LIST Example-Based Machine Translation (EBMT) prototype which uses a hybrid approach combining cross-language information retrieval and statistical language modelling. This approach consists, on the one hand, in indexing a database of sentences in the target language and considering each sentence to translate as a query to that database, and on the other hand, in evaluating sentences returned by a cross-language search engine against a statistical language model of the target language in order to obtain the n-best list of translations. The English-French EBMT prototype has been compared to the state-of-the-art Statistical Machine Translation system Moses and experimental results show that the proposed approach performs best on specialized domains.

1. Introduction

Parallel corpora are only available for a limited number of language pairs and the process of building these corpora is time consuming and expensive. For example, in the field of news, there are enough corpora, including bilingual, in particular between English and the languages that are economically the most important. In all the other fields available corpora are not sufficient to make statistical machine translation approaches operational.

This paper aims to describe the hybrid approach used in the CEA LIST Example-Based Machine Translation prototype. This approach consists, on the one hand, to use only a monolingual corpus in the target language in order to be independent of the availability of parallel corpora, and on the other hand, to use transfer lexicons and rules to produce translations which are grammatically correct. For each sentence to translate, a cross-language search engine returns a set of sentences in the target language with their linguistic properties (lemma, grammatical category, gender, number and syntactic dependency relations). These properties are combined with translation candidates provided by transfer lexicons and rules. The
result of this combination is evaluated against a statistical language model learned from a target language corpus to produce the n-best list of translations.

We present in section 2 a short state of the art of machine translation. Section 3 presents the main components of the CEA LIST Example-Based Machine Translation prototype and Section 4 describes its implementation. We discuss in section 5 results obtained after translating two types of texts. Section 6 concludes our study and presents our future work.

2. Previous work

There are mainly two approaches for Machine Translation (MT): rule-based and corpus-based [1] [2]. The rule-based approaches regroup word-to-word translation, syntactic translation with transfer rules and interlingua which uses an intermediary language representation which is independent of any of the languages involved in the translation. The corpus-based machine translation approaches regroup example-based MT and statistical-based MT techniques [3]. These two techniques have in common the use of a database containing already translated sentences. Example-based MT uses a process which consists in matching a new sentence against this database to extract suitable sentences which are recombined in an analogical manner to determine the correct translation. The second corpus-based strategy is the statistical approach [4] [5] which consists in searching for a target language string that maximizes the probability that this string is the translation of a source target string (translation model) and the probability that this target language string is a valid sentence (language model). This approach uses frequency of co-occurrence of strings in aligned texts in order to build the translation model and succession of strings (n-grams) in order to build the language model.

Rule-based MT approaches require manual development of bilingual lexicons and linguistic rules, which can be costly, and which often do not generalize to other languages. Corpus-based MT approaches are effective only when large amounts of parallel text corpora are available. Recently, several strategies have been proposed to combine the strengths of rule-based and corpus-based MT approaches or to add deep linguistic knowledge into statistical machine translation (SMT). Koehn et al. [6] proposed an approach which enriches words with additional linguistic information (lemma, Part-Of-Speech and morphological information) within the framework of factored and tree-based translation models. The authors reported that these additional linguistic information improved the performance of the standard approach. Carpuat and Wu [7] have successfully applied word sense disambiguation models to increase statistical machine translation accuracy. Likewise, Wu and Fung [8] used semantic role labels to improve the performance of a translation engine based on Moses. The authors observed a gain of +0.5 BLEU points for Chinese-English translation.

3. The CEA LIST Example-Based Machine Translation Prototype

The hybrid approach used in the CEA LIST Example-Based Machine Translation prototype consists, on the one hand, in indexing a database of sentences in the target language and considering each sentence to translate as a query to that database, and on the other hand, in combining sentences returned by the cross-langue search engine with a statistical model of the target language provided by a bilingual finite state transducer of syntactic structures from the source language into the target language [9]. This prototype is composed of:
A cross-language search engine to extract sentences or sub-sentences of the target language from the textual database which correspond to a total or a partial translation of the sentence to translate;

A bilingual transducer for lexical and syntactic transfer of the sentence to translate into the target language;

A generator of translations which consists, on the one hand, in assembling the results returned by the cross-language search engine and the bilingual transducer, and on the other hand, in choosing the best translations according to a statistical language model learned from the target language corpus.

Figure 1: Main components of the CEA LIST Example-Based Machine Translation prototype

This approach includes a preliminary step for creating the textual database of the target language according to the following phases:

- Collecting the texts of the target language from different sources;
- Segmenting the collected texts into sentences and performing a syntactic analysis on these texts;
- Gathering together the texts of the sentences accompanied by a representation of their syntactic analysis in an inverted file.

3.1. The Cross Language Search Engine

The purpose of Cross-Language Information Retrieval (CLIR) is to find all the relevant documents from a collection of documents that are in a different language from that of the user's query [10]. When using CLIR in machine translation, a document corresponds to a sentence. The role of the cross-language search engine is to retrieve for each user's query translations from an
indexed monolingual corpus. The cross-language search engine used in this study is based on a deep linguistic analysis of the query and the monolingual corpus to be indexed [11]. It is composed of the following modules (Figure 2):

- The linguistic analyzer LIMA [12] which includes a morphological analyzer, a Part-Of-Speech tagger and a syntactic analyzer. This analyzer processes both sentences to be indexed in the target language and sentences to translate in order to produce a set of normalized lemmas, a set of named entities and a set of compound words with their grammatical tags;

- A statistical analyzer that computes for sentences to be indexed concept weights based on concept database frequencies;

- A comparator which computes intersections between sentences to translate and indexed sentences and provides a relevance weight for each intersection. It retrieves the ranked and relevant sentences from the indexes according to the corresponding reformulated query (sentence to translate) and then merges the results obtained for each language taking into account the original words of the query (before reformulation) and their weights in order to score the returned sentences;

- A reformulator to expand queries during the search. The expansion is used to infer from the original query words other words expressing the same concepts. The expansion can be in the same language (synonyms…) or in different language by using bilingual lexicons. The bilingual lexicon used to reformulate (translate) words of the English sentence into French is composed of 244,532 entries which could be single or compound words.

- An indexer to build the inverted files of the sentences to be indexed on the basis of their linguistic analysis and to store these sentences in a database.
The cross-language search engine returns for the sentence to translate and for the retrieved sentences the linguistic information of their words (lemma, Part-Of-Speech tag, gender, number and syntactic dependency relations). The retrieved sentences are modeled by finite state machines (FSM). Each transition of the automaton corresponds to the lemma and its linguistic properties (Part-Of-Speech tag, gender, number...).

3.2. The Bilingual Transducer

Because the indexed monolingual corpus does not contain the entire translation of each sentence, we need a mechanism to extend translations returned by the cross-language search engine. This is achieved by a finite state transducer (FST) which consists, on the one hand, in transforming into the target language the syntactic structure of the sentence to translate, and, on the other hand, in translating its words. This transducer uses the English-French bilingual lexicon of the cross-language search engine to translate words and a set of linguistic rules to transform syntactic structures. These rules create translation hypotheses for the sentence to translate.

3.3. The Generator of Translations

The generator of translations consists in producing a correct sentence in the target language by using the syntactic structure of the translation candidate. A flexor is used in order to obtain the right forms of the translation candidate words. The flexor transforms the lemmas of the words of the target language sentence into their surface (inflected) forms. Linguistic properties (grammatical category, gender, number...) returned by the cross-language search engine are used to produce the right form of the word from its lemma.

4. Implementation of the CEA LIST Example-Based Machine Translation Prototype

The CEA LIST Example-Based Machine Translation prototype proceeds in three steps. The first one extracts sentences or sub-sentences of the target language from the textual database which correspond to a total or a partial translation of the sentence to translate. The second step creates translations from the source language to the target language for the sentence to translate using the bilingual lexicon of the cross-language search engine and a set of syntactic transfer rules. The goal of the third step is to assemble the results returned by the cross-language search engine and the bilingual transducer and to choose the best translations according to a statistical language model learned from the target language corpus.

In order to illustrate the functioning of the CEA LIST Example-Based Machine Translation prototype, we indexed a textual database composed of 1127 French sentences extracted from the ARCADEII corpus [13] and we submitted the English sentence "Social security funds in Greece encourage investment in innovation." for translation.

The CEA LIST EBMT prototype proceeds as follows:

1. The sentence to translate is considered as a query to the French textual database of the cross-language search engine. The result is a list of classes of sentences ordered according to the weight of the semantic intersection between the sentence to translate and the French sentences of the textual database. Retrieved sentences for the two first classes are illustrated in Table 1.
In addition, the cross-language search engine returns for the sentence to translate and for the retrieved sentences the linguistic information of their words (lemma, Part-Of-Speech tag, gender, number and syntactic dependency relations).

2. In order to produce translation hypotheses, the bilingual transducer proceeds in two phases: The first one (Syntactic transfer) consists in transforming syntactic structures of the sentence to translate from the source language to the target language using a set of linguistic rules (Figure 3). The second phase (Lexical transfer) translates in the target language the lemmas of words of the obtained syntactic structures using the bilingual lexicon of the cross-language search engine (Figure 4).

This step could produce an important number of translation hypotheses. This is due to the combination of the syntactic transfer rules and the bilingual lexicon which can contain several
French translations for the same English word. Table 2 illustrates the translations of lemmas of the words of the sentence to translate.

<table>
<thead>
<tr>
<th>Lemmas of words</th>
<th>Grammatical category</th>
<th>Translations of words’ lemmas with the English to French lexicon</th>
</tr>
</thead>
<tbody>
<tr>
<td>the</td>
<td>Definite article</td>
<td>le, la, les...</td>
</tr>
<tr>
<td>fund</td>
<td>Common noun</td>
<td>caisse, masse, capital, fond ...</td>
</tr>
<tr>
<td>of</td>
<td>Preposition</td>
<td>de, en, pour</td>
</tr>
<tr>
<td>security</td>
<td>Common noun</td>
<td>gage, garantie, sécurité...</td>
</tr>
<tr>
<td>social</td>
<td>Adjective</td>
<td>mondain, social</td>
</tr>
<tr>
<td>in</td>
<td>Preposition</td>
<td>entre, dans, en...</td>
</tr>
<tr>
<td>Greece</td>
<td>Proper noun</td>
<td>Grèce</td>
</tr>
<tr>
<td>encourage</td>
<td>Verb</td>
<td>encourager, favoriser, flatter</td>
</tr>
<tr>
<td>investment</td>
<td>Common noun</td>
<td>investissement, placement</td>
</tr>
<tr>
<td>innovation</td>
<td>Common noun</td>
<td>innovation, novation, changement</td>
</tr>
</tbody>
</table>

Table 2: Translations of lemmas of the words of the sentence “Social security funds in Greece encourage investment in innovation.”

3. The translation hypotheses produced by the bilingual transducer are assembled in a lattice with the retrieved sentences returned by the cross-language search engine. This lattice is implemented by using the AT&T FSM toolkit [14]. In order to find the best translation hypothesis from the lattice, a statistical model is learned with the CRF++ toolkit [15] on a monolingual lemmatized corpus in the target language. Therefore, the words of the best translation hypotheses are in their normalized forms (lemmas). In order to generate correct translations, a flexor is used to transform the lemmas of words of the translation hypotheses into their surface forms. This flexor uses the linguistic information (Part-Of-Speech tag, gender, number…) returned by the cross-language search engine for each word of the sentence to translate and the retrieved sentences. Table 3 shows the two best translations produced by the CEA EBMT prototype for the English sentence “Social security funds in Greece encourage investment in innovation.”.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>les caisses de la sécurité sociale en Grèce encouragent l’investissement dans l’innovation.</td>
</tr>
<tr>
<td>2</td>
<td>les fonds de la sécurité sociale en Grèce encouragent l’investissement en l’innovation.</td>
</tr>
</tbody>
</table>

Table 3: The translation results for the English sentence “Social security funds in Greece encourage investment in innovation.”

5. Experimental Results

We evaluate the results of the CEA LIST Example-Based Machine Translation prototype and the state-of-the-art SMT system Moses [16] using the BLEU automatic evaluation metric [17].

In order to constitute the training data, we extracted from the English-French Europarl corpus 750000 sentence pairs to train the Moses translation model and 1000000 French sentences to train the Moses language model and to create the indexed database of the cross-language search engine. We conducted two experiments: In-Domain and Out-Of-Domain. For this, we randomly extracted 500 parallel sentences from Europarl as an In-Domain corpus and 100 pairs of sentences from a corpus related to Banking and Finance (Out-Of-Domain). The goal
of these two experiments is to show the impact of the domain vocabulary on the translation results. Table 4 reports the BLEU scores of the two systems. Note that we consider only one reference for each test.

<table>
<thead>
<tr>
<th>MT system</th>
<th>In-Domain test corpus</th>
<th>Out-Of-Domain test corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEA LIST EBMT prototype</td>
<td>23.28</td>
<td>21.51</td>
</tr>
<tr>
<td>Moses</td>
<td>25.50</td>
<td>14.86</td>
</tr>
</tbody>
</table>

Table 4: BLEU scores of the CEA LIST EBMT prototype and Moses

The first observation is that, when the test set is In-Domain, we achieve a relatively high score BLEU for both the two systems but the score of Moses is slightly better. For the Out-Of-Domain test corpus, it is not surprising that the CEA LIST EBMT prototype performs better than Moses which has obtained a very low BLEU score. This result can be explained by the fact that the test corpus has a vocabulary which is different from the entries of the translation table. On the other hand, it seems that the English-French lexicon used in the bilingual transducer has had a significative impact on the result of the CEA LIST EBMT prototype. It improved its BLEU score with 6.65 points.

6. Conclusion and Future Work

In this paper, we have presented the CEA LIST Example-Based Machine Translation prototype which uses a hybrid approach combining cross-language information retrieval and statistical language modelling. This approach needs only monolingual texts in the target language. The first results of our experiments show that the CEA LIST EBMT prototype performs better than Moses when texts to translate concern specialized domains such as Banking and Finance. In future work, we plan, on the one hand, to consolidate our results by experimenting our approach on other specific domains (tourism, healthcare...), and, on the other hand, to develop machine learning for automatically acquiring syntactic transfer rules used in the bilingual finite state transducer.

Acknowledgment

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 312651 - ePOOLICE

References


ABSTRACT

In this study, terminological variation pertains to the different ways in which specialised knowledge is expressed in written discourse by means of terminological designations. Choices regarding the use of term variants in source texts (i.e. intralingual variation) as well as the different translations of these variants in target texts (i.e. interlingual variation) are determined by a complex interplay of contextual factors of several kinds. For translators, it is therefore important to know the different language options (i.e. variants) that are available when translating terms and to know in which situational contexts certain options are more likely to be used.

To this end, translators often consult bi- or multilingual translation resources (e.g. terminological databases) to find solutions to certain translation problems. Different possibilities are offered in terminological databases to represent and visualise intra- and interlingual variants. In conventional terminology bases, terms in several languages usually appear on concept-oriented term records. This particular way of structuring and visualising terminological data has its roots in prescriptive terminology in which terms are merely viewed as ‘labels’ assigned to clearly delineated concepts (Picht and Draskau 1985). In ontologically-underpinned terminological knowledge bases or TKBs, terminological data tend to be represented in networks comprised of conceptual and semantic relations (Kerremans et al. 2008; Faber 2011; Durán Muñoz 2012; Peruzzo 2013). As opposed to traditional ways of representing terminological data (e.g. on the basis of alphabetically sorted lists, tables or matrices), such networks allow for a flexible and dynamic visualisation of data that may be connected to one another in several ways.

The aim of this article is to reflect on how visualisations of terms, variants and their translations in networks can be improved by taking into account the contextual constraints of the texts in which they appear. To this end, a novel type of translation resource has been developed, resulting from a semi-automatic method for identifying intralingual variants and their translations in texts.
A prototype visualisation of this resource will be presented in which terms, variants and their translations appear as a contextually-conditioned network of ‘language options’. The proposed model derives from the Hallidayan premise that each language option or choice acquires its meaning against the background of other choices which could have been made. The choices are perceived as functional: i.e. they can be motivated against the backdrop of a complex set of contextual conditions (Eggins 2004). Changing these contextual conditions causes direct changes in the network of terminological options that are shown to the user.

1. Introduction

Choices regarding the use of term variants in source texts (i.e. intralingual variation) as well as the different translations of these variants in target texts (i.e. interlingual variation) are determined by a complex interplay of contextual factors of several kinds (Freixa 2006). For translators, it is therefore important to know the different language options (i.e. variants) that are available when translating terms and to know in which situational contexts certain options are more likely to be used.

To this end, translators often consult bi- or multilingual translation resources to find solutions to certain translation problems. However, such ‘structured resources’ never fully cover the wealth of options available in language. By separating terms from their ‘natural environment’ (i.e. the texts in which they appear), a lot of valuable information on which translation decisions should be based is lost. This is why translators also often resort to ‘unstructured resources’: i.e. texts originally written in the source and target languages or previously translated texts.

In a recently conducted study on terminological variation, it is argued why the representation of intra- and interlingual variation in existing multilingual termbases is too restrictive to account for the wealth of potential linguistic options to express units of specialised knowledge (or units of understanding) in source and target texts (Kerremans 2014). Based on this study, a new type of translation resource has been worked out in which intra- and interlingual variation retrieved from parallel texts (i.e. source texts and their translations) is structured according to semantic and contextual criteria.

The aim of this article is to discuss how intra- and interlingual terminological variants in this resource can be visualised in a dynamic and flexible graph to be used by translators. The idea for this type of visualisation further builds on recent initiatives in multilingual cognitive-oriented terminology studies to represent the conceptual organisation of a specialised field as a relational network comprised of units of understanding (denoted by terms in multiple languages) and different types of conceptual relations (see Section 2).

The graph representation in our approach differs from these initiatives in the sense that several contextual parameters will be taken into consideration when visualising intra- and interlingual variants for a given unit of understanding that were retrieved from a corpus of parallel texts (see Section 3). Terminological data in the envisaged graph representation will be dynamically structured as translators will have the possibility to zoom into specific occurrences of terms and translations in selected registers (see Section 4). Apart from summarising the basic principles underlying the prototype, we will also briefly reflect on our future work regarding its implementation (see Section 5).
2. Graph representations of multilingual terminological knowledge

In ontologically-underpinned terminological knowledge bases or TKBs, terminological data tend to be represented in network representations comprised of conceptual and semantic relations (Kerremans et al. 2008). As opposed to traditional ways of representing terminological data (e.g. on the basis of alphabetically sorted lists, tables or matrices), such networks allow for a flexible and dynamic visualisation of terminological data.

Particularly relevant for the present study is the fact that the methodological principles underlying TKBs are increasingly applied to the creation of bi- or multilingual special language resources for translators (Durán Muñoz 2012; Peruzzo 2013).

An example of an advanced implementation of a multilingual TKB is the EcoLexicon1 database (León Araúz et al. 2011). Ecolexicon is targeted towards “different user groups, such as translators, technical writers, environmental experts, etc., who wish to expand their knowledge of the environment for the purpose of text comprehension or generation.”. Given this objective, the database is allegedly primarily concerned with the conceptual organisation of the environmental domain. Descriptions of possible uses or preferences of terms and variants in certain communicative contexts is not provided, which seems to us an important limitation of a multilingual TKB for translators.

We have therefore defined a model for a new type of translation resource that specifically covers the choices that were made by translators when confronted with multiple terminological variants for a given unit of understanding. This translation resource is comprised of semantically and contextually-structured, term-based translation units that were extracted from a multilingual parallel corpus (Kerremans 2014).

3. Structure of the translation resource

Term-based translation units are the primary building blocks of the resource presented in this article. Each translation unit (TU) is further classified according to text-related and semantic categories:

- Text-related categories are properties originally assigned to the bitext (i.e. the combination of a source text and its translation) from which the TU is extracted. Examples of such categories are text type, text source, language, text topic, etc.

- Semantic categorisation involves classifying the English term in the TU according to the ‘concept’ to which it refers in the source text. This means that each term in the source texts is marked with a unique identification code – i.e. a so-called ‘cluster label’. Terms extracted from the source texts that carry this label appear in the same ‘cluster’ of terminological variants (Kerremans 2011).

---

1 http://ecolexicon.ugr.es/en/aboutecolexicon.htm
4. A prototype proposal for visualising intra- and interlingual terminological variants

The proposed graph representation of the translation resource derives from the Hallidayan premise that each language option or choice acquires its meaning against the background of other choices which could have been made. The choices are perceived as functional: i.e. they can be motivated against the backdrop of a complex set of contextual conditions (Eggoins 2004). Changing the contextual conditions causes direct changes in the network of terminological options that are shown to the user. This idea is illustrated by means of Figure 1.

In a bilingual view, the user will first specify a search term in the source language. This will activate in the contextual (i.e. semantic, situational and linguistic) filters different fields that are associated with the search query. In case the search term is connected to multiple conceptual clusters (see Section 3), the user will be able to select the proper cluster label. The result of the search query is visualised in a graph representation. Situational (see Text options) and linguistic (see Lemma and POS options) criteria or filters are used in this example to highlight or deactivate certain connections or nodes in the source and target languages, causing constant changes or shifts in the structure of variants.

<table>
<thead>
<tr>
<th>Bilingual view</th>
</tr>
</thead>
<tbody>
<tr>
<td>[add SL search term or ClusterID]</td>
</tr>
<tr>
<td>Text options</td>
</tr>
<tr>
<td>✓ EU</td>
</tr>
<tr>
<td>✓ Non-EU</td>
</tr>
<tr>
<td>✓ EEA</td>
</tr>
<tr>
<td>Cluster option(s)</td>
</tr>
<tr>
<td>Source language options (English)</td>
</tr>
<tr>
<td>Lemma option(s)</td>
</tr>
<tr>
<td>✓ ...</td>
</tr>
<tr>
<td>POS option(s)</td>
</tr>
<tr>
<td>✓ noun noun</td>
</tr>
<tr>
<td>✓ ...</td>
</tr>
<tr>
<td>✓ ...</td>
</tr>
<tr>
<td>✓ ...</td>
</tr>
</tbody>
</table>
5. Conclusion

In this article we discussed how intra- and interlingual variants, extracted from parallel texts, can be used to populate a new type of translation resource. We discussed how this resource can be represented as an innovative graph. The resource can be perceived as an additional ‘tool’ that can be integrated in a computer-assisted translation (CAT) environment or workflow together with bi- or multilingual termbases and translation memories.

The next step in our research is to turn the ideas and requirements concerning the visualisation part into an actual implementation that could initially be tested in an experimental setting by means of a group of translators.

References


iCompileCorpora: A Web-based Application to Semi-automatically Compile Multilingual Comparable Corpora

Hernani Costa
University of Malaga

Gloria Corpas Pastor
University of Malaga

Miriam Seghiri
University of Malaga

ABSTRACT

This article presents an ongoing project that aims to design and develop a robust and agile web-based application capable of semi-automatically compiling monolingual and multilingual comparable corpora, which we named iCompileCorpora. The dimensions that comprise iCompileCorpora can be represented in a layered model comprising a manual, a semi-automatic and a Cross-Language Information Retrieval (CLIR) layer. This design option will not only permit to increase the flexibility of the compilation process, but also to hierarchically extend the manual layer features to the semi-automatic web-based layer and then to the semi-automatic CLIR layer. The manual layer presents the option of compiling monolingual or multilingual corpora. It will allow the manual upload of documents from a local or remote directory onto the platform. The second layer will permit the exploitation of either monolingual or multilingual corpora mined from the Internet. As nowadays there is an increasing demand for systems that can somehow cross the language boundaries by retrieving information of various languages with just one query, the third layer aims to answer this demand by taking advantage of CLIR techniques to find relevant information written in a language different from the one semi-automatically retrieved by the methodology used in the previous layer.

1. Introduction

The interest in mono-, bi- and multilingual corpora is vital in many research areas such as language learning, stylistics, sociolinguistics, translation studies, amongst other research areas. Particularly in translation, their benefits have been demonstrated by various authors (cf. Bowker and Pearson, 2002; Bowker, 2002; Zanettin et al., 2003; Corpas Pastor and Seghiri, 2009). The main advantages of its usage are their objectivity, reusability, multiplicity and applicability of uses, easy handling and quick access to large volume of data. In detail, corpus linguistics:

- Empowers the study of the foreign language: the study of the foreign language with the use of corpora allows the foreign language learners to get a better “feeling” about
that language and learn the language through “real world” texts rather than "controlled" texts (cf. Gries, 2008).

- Simplifies the study of naturalistic linguistic information: as previously mentioned, a corpus assembles “real world” text, mostly a product of real life situations, which results in a valuable research source for dialectology (cf. Hollmann and Siewierska, 2006), sociolinguistics (cf. Baker, 2010) and stylistics (cf. Wynne, 2006), for example.

- Helps linguistic research: as the time needed to find particular words or phrases has been dramatically reduced with the use of electronically readable corpora, a research that would take days or even weeks to be manually performed can be done in a couple of seconds with an high degree of accuracy.

- Enables the study of wider patterns and collocation of words: before the advent of computers, corpus linguistics was studying only single words and their frequency. More recently, the emergence of modern technology allowed the study of wider patterns and collocation of words (cf. Roland et al., 2007).

- Allows simultaneous analysis of multiple parameters: in the last decades, the development of corpus linguistic software tools helped the researchers to analyse a wider number of parameters simultaneously, such as determine how the usage of a particular word and its syntactic function varies.

Moreover, they are a suitable tool for translators, as they can easily determine how specific words and their synonyms collocate and vary in practical use or even help interpreters speeding up the research for unfamiliar terminology (cf. Costa et al., 2014). Furthermore, in the last decade, a growing interest in bi- and multilingual corpora has been shown by researchers working in other fields, such as terminology and specialised language, automatic and assisted translation, language teaching, Natural Language Processing, amongst others. Nevertheless, the lack of sufficient/up-to-date parallel corpora and linguistic resources for narrow domains and poorly-resourced languages is currently one of the major obstacles to further advancement on these areas. One potential solution to the insufficient parallel corpora is the exploitation of non-parallel bi- and multilingual text resources, also known as comparable corpora (i.e. corpora that include similar types of original texts in one or more language using the same design criteria, cf. EAGLES, 1996; Corpas Pastor, 2001:158).

Even though comparable corpora can compensate for the shortage of linguistic resources and ultimately improve automated translations quality for under-resourced languages and narrow domains for example, the problem of data collection presupposes a significant technical challenge. Moreover, the difficulty of retrieving and classifying such data is considered a complex issue as there is no unique notion of what it really covers and how it can be truly exploited (cf. Skadina et al., 2010:12).

2. Existing Corpora Compilation Solutions

Although this compilation process could be manually performed, nowadays specialised tools can be used to automate this tedious task. By a way of example, BootCaT (Baroni and Bernardini,
2004) was built to exploit specialised monolingual corpora from the Web. It is capable of compiling a corpus through automated search queries, and only requires a small set of seed words as input. This tool has been used, for example, to create specialised comparable corpora for travel insurance (Corpas Pastor and Seghirí, 2009), medical treatments (Gutiérrez Florido et al., 2013), among other narrow-domains. WebBootCat (Baroni et al., 2006) is similar to BootCaT, but instead of having to download and install the application, WebBootCat can be used online. Despite being designed for other purposes, Terminus and Corpógrafo should also be mentioned as examples of web-based compilation tools.

As we can see, several semi-automatic compilation tools have been proposed so far. Nevertheless, these compilation tools are scarce or proprietary, simplistic with limited features, built to compile one monolingual corpus at a time and do not cover the entire compilation process (i.e. apart from compiling monolingual comparable corpora, they do not allow managing and exploring both parallel and multilingual comparable corpora). Thus, their simplicity, lack of features, performance issues and usability problems result in a pressing need to design new compilation tools tailored to fulfil not only translators’ and interpreters’ needs, but also professionals' and ordinary people’s.

3. **iCompileCorpora**

Departing from a careful analysis of the weaknesses and strengths of the current compilation solutions, we started by designing and developing a robust and agile web-based application prototype to semi-automatically compile mono- and multilingual comparable corpora, which we named iCompileCorpora. iCompileCorpora can be simply described as a Web graphical interface that will guide the user through the entire corpus compilation process. Designed and implemented from scratch, this application aims to cater to both novice and experts in the field. It will not only provide a simple interface with simplified steps, but also will permit experienced users to set advanced compilation options during the process.

The dimensions that comprise iCompileCorpora can be represented in a layered model comprising a manual, a semi-automatic web-based and a semi-automatic Cross-Language Information Retrieval (CLIR) layer (see Figure 1). This design option will permit not only to increase the flexibility and robustness of the compilation process, but will also hierarchically extend the manual layer features to the semi-automatic web-based layer and then to the semi-automatic CLIR layer. In detail, the manual layer represents the option of compiling monolingual and multilingual corpora. It will allow for the manual upload of documents from a local or remote directory onto the platform. The second layer will permit the exploitation of both mono- and multilingual corpora mined from the Internet. Although this layer can be considered similar to the approaches used by BootCaT and WebBootCat, it has been designed to address some of their limitations (e.g. allow the use of more than one boolean operator when creating search query strings), and to improve the User Experience (UX) with this type of software. As nowadays there is an increasing demand for systems that can somehow cross the language boundaries by retrieving information in various languages with just one query, the third layer aims to answer this demand by taking advantage of CLIR techniques to find relevant information written in a language different to the one semi-automatically retrieved by the methodology used in the previous layer.
4. Conclusion

This article presents an ongoing project that aims to increase the flexibility and robustness of the compilation of monolingual and multilingual comparable corpora by creating a new web-based application from scratch. iCompileCorpora intends to fulfil not only translators' and interpreters' needs, but also professionals' and ordinary people's, either by breaking some of the usability problems found in the current compilation tools available on the market or by improving their limitations and performance issues. By the end of this project, we intend to make this compilation tool publicly available, both in a research or in a commercial setting.

Acknowledgements

Hernani Costa is supported by the People Programme (Marie Curie Actions) of the European Union's Framework Programme (FP7/2007-2013) under REA grant agreement no 317471. Also, the research reported in this work has been partially carried out in the framework of the Educational Innovation Project TRADICOR (PIE 13-054, 2014-2015); the R&D project INTELITERM (ref. no FFI2012-38881, 2012-2015), and the R&D Project for Excellence TERMITUR (ref. no HUM2754, 2014-2017).

References


Terminology Management Revisited

Nizar Ghoula
University of Geneva
and The Olanto
Foundation

Jaques Guyot
The Olanto Foundation

Gilles Falquet
University of Geneva
and The Olanto
Foundation

ABSTRACT

Large repositories publishing and sharing terminological, ontological and linguistic resources are available to support the development and use of translation. However, despite the availability of language resources within online repositories, some natural languages associations cannot be found (rare languages or non-common combinations, etc.). Consequently, multiple tools for composing linguistic and terminological resources offer the possibility to create missing language associations. These generated resources need to be validated in order to be effectively used. Manually checking these resources is a tedious task and in some cases hardly possible due to the large amount of entities and associations to go through or due to the lack of expertise in both languages. To solve this matter and generate sound and safe content, tools are needed to automatically validate and filter associations that make no sense. Hence, a validation tool is based itself on external resources such as parallel corpora which need to be either collected or created and filtered. To solve these matters we propose a set of tools that generate new terminological resources (myTerm) and a filter them using a parallel corpus generated by another tool (myPREP). We describe our methodology for terminology management and we describe its implementation within an original framework.

1. Introduction

The translation business has considerably changed over the past decade. Smaller full-time teams must translate larger volumes, the difference being distributed over a network of external translators, which are located worldwide. Besides, deadlines are ever tighter and costs must be reduced. As a result, translation workflows are changing in order to automate every possible step: submitting a document for translation, affecting the translation to a translator, performing the translation, performing the quality control steps, sending back the translation to the customer and feeding the CAT tools with the new document pair and/or related segments. Consequently, a complete suite of CAT tools is needed to support every phase of this new workflow. Within this context, the Olanto foundation¹ proposes and publishes Open Source tools for professionals to face these new challenges.

¹ www.olanto.org
The initial goal of the Olanto Foundation is to build and share a complete suite of professional CAT tools:

- a Concordancer (Bitext-based search engine);
- a Statistical Machine Translation Tool;
- a Terminology Database Management System;
- a Translation Memory Management System.

These tools can be integrated within several Electronic Document Management Systems (EDMS). In particular, a cross-lingual search engine, which may be, associated with other existing search tools (typically Lucene or SharePoint). Despite the existence of a considerable number of open source tools in the CAT field, these tools remain complex and their integration incomplete. Thus, these tools do not meet the complete chain of needs commonly expressed by Translation Services and Language Service Providers. Additionally, they generally don't benefit from a robust distribution and support structure and some of them are not really scalable.

- Based on a previous research work on building a repository of multilingual terminological and ontological resources (Ghoula, Falquet, & Guyot, 2010), we identified the following objectives for such a tool:
  - Compatibility of the resources representation models with TBX (basic) (Wright, Melby, Rasmussen, & Warburton, 2010);
  - Ability to manage a large number of terminological resources;
  - Ability to support a large number of standards and formalisms for resources representations (TBX, UTX, DXDT, GlossML, etc.);
  - Availability of XML-based representation models for structured resources that do not correspond to all standards or formalisms (e.g. JIAMCATT²).

One of the latest tools in development by Olanto is the myTerm terminology manager. In myTerm, resources are imported into the terminology manager's repository and attached to a hyper graph where terminological resources from different domains connect languages to each other either directly or by transitivity.

Our main goal is not to generate dictionaries by transitivity but we mainly focus on building a framework and a set of tools for helping translators to validate automatically or semi-automatically their dictionaries using their own corpora or other kinds of corpora.

This approach can also be used to interactively query a large parallel or comparable corpora to compute candidates translation for a given term or multi-word expression. We implemented this idea in the “How2Say” interactive tool. This tool allows finding the possible translations of a specific expression from one language to another. We will describe this framework through the different sections of this paper.

2. Context and research issues

As a consequence of the availability of large repositories publishing and sharing terminological, ontological and linguistic resources on the Web, we notice a significant

² www.jiamcatt.org
improvement in the quality of automatic and semi-automatic translation systems. Nevertheless, these resources are not yet available for all possible combinations of pairs of languages. For example, to run or enhance a translation process from a language A to a language B, there is a need for a dictionary or a terminology associating both languages. Despite the existence of these types of resources for language A and language B within online repositories, none of them may directly associate the pair of languages (particularly in the case of rare languages or non-common combinations, etc.). One way to address this issue consists in using available language resources to generate the missing ones. Hence, automatically deriving terminologies by transitivity has become a common procedure to produce resources for language services.

For example, if we have a glossary EN→FR and another glossary FR→DE, using composition, we can generate a new glossary EN→DE. It is well known that polysemy within both resources can produce associations between pairs of terms that do not make sense. For example, starting from the associations time→temps in EN→FR, temps→Zeit and temps→Wetter, in FR→DE, the composition produces two term associations: time→Zeit and time→Wetter∗ for EN→DE. Consequently, this kind of operations on terminological resources is not completely safe in terms of sense. Therefore, the resulting terminological resource has to be filtered to detect and remove meaningless term associations. Manually checking these resources is a tedious task and in some cases hardly possible due to the large amount of entities and connections to go through or due to the lack of expertise in both languages.

In the context of composing ontology alignments, we encountered the issue of inconsistent mappings, which can be solved using reasoning and combination of confidence measures to filter mappings (Ghoula, Nindanga, & Falquet, 2013). Unfortunately, for terminological resources associations, there are no standards or use cases allowing the application of confidence measures. However, it is possible to use a parallel corpus of (aligned) sentences between both languages to assign a confidence measure to associations between pairs of words. This measure is based on the co-occurrence of both terms in the sentences of the corpus.

In this paper, we describe our approach and present the architecture of myTerm repository and define operations for producing, managing and filtering terminological resources for validating languages associations. We explain in detail the computation of correlation measures that filter term associations based on their co-occurrence.

3. State of the art

New technologies for the Web and knowledge sharing made language resources more accessible. Thus, the volume of data to process for training or evaluating machine translation tools is more important. The scalability issues to process more data and more resources are often revisited and algorithms are becoming more efficient thanks to the rising hardware power and the efficiency of new software and services architectures. Consequently, we can aim for interactive systems that process large corpora (public and private) and offer a real time and flexible interaction with translators. This also opens the door for creating more language services and better interaction between different CAT tools. Multiple tools for creating language resources such as dictionaries by transitivity have been proposed in the literature (Paik, Shirai, & Nakaiwa, 2004), (Zhang, Ma, & Isahara, 2007).
Our concern is about the approaches that use parallel or comparable corpora to validate the result of transitivity. An approach for automatically generating dictionaries between languages was proposed by (Nerima & Wehrli, 2008) in order to reduce the number of linguistic resources used as an input for a multilingual translation system. Since such a system requires a lexical database for each pair of language combination, then for a number of \( n \) languages, there is a need for \( n^2(n-1)/2 \) dictionaries. Even if these dictionaries were available, which is not always the case, the large number of bilingual dictionaries might affect the performances of such a system. Thus, the authors propose to derive a bilingual dictionary by transitivity using existing ones and to check the generated translations in a parallel corpus. The quality of the result relies on multiple parameters such as the quality of the input, which have to be manually validated, the attribution of a preference and the usage of tagging. Consequently, the approach is language dependent and there are needs to have multiple language models for an effective tagging.

Another approach proposed by (Tao & Zhai, 2005) close to our methodology is based on using correlation measures between words within comparable corpora to build a cross-lingual text mining framework that can exploit these bilingual text corpora to discover mappings between words and documents in different languages. This approach is based on the hypothesis that the words that tend to co-occur more frequently in comparable corpora are either translations of each other or related to the same topic. Thus, the authors use comparable corpora to extract associations of words in multiple languages. The authors use the Pearson's correlation coefficient to compute associations between words, which are used to create a similarity score between documents. The correlation measure is combined with information retrieval techniques in order to match documents between languages but does not go further into matching sentences inside the associated documents.

We propose an original approach that is language and corpus independent and a framework for indexing parallel corpora and calculating correlation measures between n-grams at the level of sentences. We also propose a real-time application that retrieves n-grams and their translations from voluminous corpora.

4. Approach

Our approach relies on the validation of new glossaries, generated by transitivity, using parallel and comparable corpora between the two languages. For instance, going back to the example for the introduction we can find in the EN→DE corpus a number of co-occurrences of "time" and "Zeit" in a sentence and its translation that confirm the "time→Zeit" association (« Members shall furnish statistics and information within a reasonable time... » → « Die Mitglieder legen Statistiken und Angaben innerhalb einer angemessenen Zeit ... »), whereas almost no co-occurrence confirms the "time→Wetter" association.

For implementing and preparing a corpus independent framework we assembled different components of the Olanto's suite as described in figure 1. The reference parallel corpora are produced by myPREP, Olanto's text aligner tool. This tool automatically aligns pairs of documents from a multilingual corpus at the sentence level and generates a translation memory in TMX format. Each set of TMX is then indexed using the myCAT indexer (Guyot, Falquet, &
Benzineb, 2006). The indexer generates two types of vectors of values for each term within the corpora:

- \( \text{id}_{j} = (o_1, o_2, \ldots, o_n) \) is the index of the n-gram \( g_j \) for a given corpora \( C \) containing \( n \) documents where \( o_k \) defines the number of occurrences of \( g_j \) within the document \( d_k \).
- \( \text{pos}_{jk} = (p_1, p_2, \ldots, p_m) \) is the vector of positions of the n-gram \( g_j \) within a given document \( d_k \) where \( p_m \) defines the position of the \( m \)-th occurrence of \( g_j \) in the document \( d_k \).

We developed a module that calculates the correlation between two n-grams based on the generated index and used it to build the How2Say web application on top of the indexed corpora. This application let the user enter an n-gram, then it computes, in real-time, the best translation in the target language (the maximal n-grams with the highest correlations), and displays these translations together with example sentences.

In order to automatically validate the generated translation (dictionary entries) we created a module that generates dictionaries by transitivity. Finally we added a component that takes as input the generated dictionary and generates as output for each entry a triple (n-gram, n-gram, correlation).

**Figure 1: Architecture of the bilingual dictionary automatic validation framework**
Each correlation is measured based on a specific corpus. The whole approach is supported by the framework of processing parallel and comparable corpora to compute correlation measures for pairs of n-grams in a given source and target language.

5. Correlation

A correlation-based technique computes a correlation measure between two terms or expressions based on their co-occurrences in aligned sentences. Based on our indexer, the calculation of correlation measures is quite fast. If g1 and g2 are two n-grams (d-grams) in the source and target languages respectively, the similarity between g1 and g2 is obtained as the correlation between the occurrence vectors x and y, where x_i (resp. y_i) = 1 if g1 (resp. g2) occurs in sentence no. i of the source (resp. target) language, and 0 otherwise.

The correlation between x and y is defined as:

\[ r_{xy} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}} \]

Since x and y are binary vectors, \( r_{xy} \) can be reduced to

\[ r_{xy} = \frac{n n_{12} - n_1 n_2}{\sqrt{n n_1 - n_1^2} \sqrt{n n_2 - n_2^2}} \]

where:

- \( n \) is the number of aligned sentences;
- \( n_1 \) is number of sentences in the source language containing g1;
- \( n_2 \) is number of sentences in the target language containing g2;
- \( n_{12} \) is the number of aligned sentences containing g1 in the source language and g2 in the target language (co-occurrences).

![Figure 2: Correlation measures between n-grams within the indexed corpora](image)
6. Testing our approach for validating bi-lingual dictionaries

In order to test the usefulness of the corpora in the process of automatic validation of terms associations within a bilingual dictionary, we conducted two types of experiments.

The first experiment is intended to use a valid bilingual dictionary to test the quality of the used corpora and their relevance to the dictionary based on calculating correlation measures for valid terms associations.

The second experiment is intended to test the usage of correlation measures for the validation of the generated dictionary.

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Size (# of sentences)</th>
<th>Number of languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wikipedia: comparable built using myPREP</td>
<td>1,000,000</td>
<td>3</td>
</tr>
<tr>
<td>MultiUN parallel corpora</td>
<td>69,300,000</td>
<td>7</td>
</tr>
<tr>
<td>DGT2014 parallel corpora</td>
<td>84,561,191</td>
<td>23</td>
</tr>
<tr>
<td>EuroBook parallel corpora</td>
<td>173,200,000</td>
<td>48</td>
</tr>
</tbody>
</table>

Table 1: Parallel and comparable corpora, parsed, indexed and mapped for How2Say

6.1. Evaluating the correlation method

The first step of our approach is to use an existing valid dictionary to test the coverage of the corpora used in the experimentation. We selected the Dictionary “Wiktionary 2008” containing 15’000 entries between English and French. The main characteristic of this dictionary is its general aspect covering multiple domains.

This experiment is used in order to define the terminological signature of corpora and the utility of the approach in general. There are two interpretations of the weak correlation values:

- The corpora's terminology does not support the domain of the dictionary;
- The dictionary contains false associations between terminological entities due to polysemy;

The horizontal axis represents the term number and the vertical axis shows the correlation values. The terms are sorted according to their correlation values. For instance the ONU curve (red) shows that approx. 6500 terms in this corpus have a translation with a correlation higher than 0.1.

Figure 3: Coverage of the used corpora for “Wiktionary” for English and French

The coverage of the used corpora varies from 40% to 65% taken separately. In order to maximize the coverage of the corpora, we used a maximum aggregation of the correlation measures. Thus, the resulting coverage of all the corpora combined for the used dictionary is 75%.

In general, based on this experiment we realized that:

- the correlation is different depending on the corpora;
- a given corpus does not always cover the dictionary;
- a corpus has a specific terminological signature;
- the maximum aggregation of correlation measures allows to enlarge the coverage of a corpus;
- using the correlation-based method we can also determine the correlation between corpora for a given dictionary.

6.2. Validating a transitive bi-lingual dictionary using out method

For the second experiment we used two corpora, EuroBook and MultiUN. These corpora contain the biggest number of entries for German (among the four processed corpora). In order to validate the generated dictionary using transitivity, there are two possibilities for interpreting a weak correlation; the first interpretation is that the entry is not covered by the corpora or that the entry is invalid due to polysemy.

Two types of experimentations have been driven for the validation of the generated dictionary from French to German through English:

- Generate the transitive dictionary FR -> DE and then validate it using the maximum of correlation from both corpora (EuroBook[FR-DE], DGT2014 [FR-DE]). The result of this operation gave a dictionary of 27'183 entries where 11'662 have a not null correlation based on the maximum from both corpora. This is a result of a maximum aggregation of the correlation values;
- A more radical approach is to compose only the dictionary entries from FR to EN, that are covered by the corpora DGT2014[FR-EN], with the dictionary entries from EN to DE that are covered by the corpora DGT2014[EN-DE] and then validate the resulting dictionary FR to DE using the corpora DGT2014[FR-DE]. The resulting dictionary contains only 3'800 terms associations that are considered as valid based on their correlation measures.

![Figure 4: Validation of the FR-DE dictionary using two corpora](image)
The terminological signature of the used corpora is not as general as the used dictionary for transitivity. For each couple of terminological entries there are different possible translations depending on the context. An automatic translation system imposes the translations using a translation memory. In our context, we propose a diversified approach for proposing translations based on correlation measures.

The approach that we propose is flexible and simple; it offers an original and efficient framework for validating transitivity translation (How2Say). While testing this approach we realized that the corpora’s coverage is very important and the results depends highly on it. We created a system to explore expressions and n-grams. This system explores the parallel corpora and classifies the list of n-grams. For each query retrieving an expression in a source language, we classify the corresponding n-grams in a target language using the correlation measure.

**Figure 5: How2Say Interface**

**7. Conclusion**

We propose a framework and a browser (How2Say) offering the possibility of validating transitive language associations based on correlations measures calculated using parallel or comparable corpora. The proposed framework is generic and language independent offering multiple possibilities such as:

- The usage of a sophisticated a query language for finding expressions (“AND”, “OR”);
- The openness for multiple corpora and dynamic support and processing of new corpora;
- multiple options for translating expressions based on their co-occurrence within the corpora.

We evaluated our approach on multiple voluminous corpora. The feasibility of this approach has been proven by experiments and evaluation. We developed an online demo offering a real-
time interrogation of large parallel and comparable corpora supported by correlation measures. The values of correlation measures allow exploring parallel corpora for mining n-gram associations. The correlation measure within this framework is used for automatic validation of associations between expressions from pairs of languages. This approach is a first step for an automatic system of validating glossaries and dictionaries that are created using transitive tools. Parallel corpora use specific terminologies and do not cover all domains.

References


Annex

![Figure 6: myTerm browsing interface](image)
Rule-based Automatic Post-processing of SMT Output to Reduce Human Post-editing Effort

Victoria Porro  
Université de Genève  
FTI/TIM

Johanna Gerlach  
Université de Genève  
FTI/TIM

Pierrette Bouillon  
Université de Genève  
FTI/TIM

Violeta Seretan  
Université de Genève  
FTI/TIM

ABSTRACT

To enhance sharing of knowledge across the language barrier, the ACCEPT project focuses on improving machine translation of user-generated content by investigating pre- and post-editing strategies. Within this context, we have developed automatic monolingual post-editing rules for French, aimed at correcting frequent errors automatically. The rules were developed using the Acrolinx IQ technology, which relies on shallow linguistic analysis. In this paper, we present an evaluation of these rules, considering their impact on the readability of MT output and their usefulness for subsequent manual post-editing. Results show that the readability of a high proportion of the data is indeed improved when automatic post-editing rules are applied. Their usefulness is confirmed by the fact that a large share of the edits brought about by the rules are in fact kept by human post-editors. Moreover, results reveal that edits which improve readability are not necessarily the same as those preserved by post-editors in the final output, hence the importance of considering both readability and post-editing effort in the evaluation of post-editing strategies.

1. Introducción

Since the emergence of the Web 2.0 paradigm, user-generated content (UGC) represents a large share of the informative content available nowadays. Online communities share technical information and exchange solutions to technical issues through forums and blogs. However, the uneven quality of UGC can hinder both readability and machine-translatability, thus preventing sharing of knowledge between language communities (Jiang et al., 2012; Roturier and Bensadoun, 2011).

The ACCEPT project\(^1\) aims to improve the Statistical Machine Translation (SMT) of community content through minimally-intrusive pre-editing techniques, SMT improvement methods and post-editing strategies. The project targets two specific data domains: the technical forum domain, represented by posts in the Norton Community forum, and the medical domain, illustrated by Translators without Borders documents written by health professionals.

\(^1\)[http://www.accept-project.eu/](http://www.accept-project.eu/)
During the first year of the project, we found that pre-editing forum data significantly improves MT output quality (Lehmann et al., 2012; Gerlach et al., 2013a). Further work (Gerlach et al., 2013b) has shown that pre-editing which improves SMT output quality also has a positive impact on bilingual post-editing time. We are now developing post-editing rules intended to reduce post-editing effort, by automatically correcting the most frequent errors before submitting MT output to the post-editor.

This study focuses on the evaluation of the post-editing rules developed for French, and more specifically, on automatic rules designed for monolingual application. In the related literature, there are several studies describing post-editing rules and evaluating them using automatic metrics or fluency-adequacy measures (Guzman, 2008; Valotkaite et al., 2012). However, to our knowledge, few such studies look into the actual use of the modifications produced by rules. We will assess: (1) the impact of the rules on the readability of the MT output and (2) their usefulness during the subsequent manual post-editing phase.

Our study relies on the following hypotheses: (1) the changes produced by our automatic monolingual rules contribute to making the text more readable; (2) automatic post-editing produces useful changes for the post-editing task and reduces technical effort; and (3) readability and usefulness for post-editing do not necessarily go hand in hand.

The paper is organised as follows. In Section 2, we show how post-editing research is performed in ACCEPT and describe the rules developed for French. In Section 3, we describe the experimental setup and provide details about data, tasks and participants. The results are analysed in Section 4, and conclusions and future work are presented in Section 5.

2. Post-editing in ACCEPT

In the ACCEPT project, post-editing rules, as well as pre-editing rules, are developed using the technology developed by one of our project partners, i.e., the Acrolinx\textsuperscript{IQ} engine (Bredenkamp et al, 2000). This rule-based engine uses a combination of shallow NLP components enabling the development of declarative rules, written in a formalism similar to regular expressions, based on the syntactic tagging of the text. A sample rule is displayed in Figure 1.

![Figure 1: Title](image-url)

TRIGGER(80) == [@ne]? @auxFin\textsuperscript{1} [@adv]* @verbInf\textsuperscript{2} -> ($aux, $inf)

Rules can be applied through the ACCEPT portal interface (Seretan et al., 2014) or directly in any forum interface, using specific plugins that allow to check compliance with the rules (ACCEPT D5.6; Roturier et al., 2013).

The ACCEPT partners have so far explored several approaches to post-editing: manual vs automatic, monolingual vs bilingual (ACCEPT D7.2 and D2.4; Mitchell et al., 2013). For French texts machine-translated from English, we have focused on automatic monolingual rules for various reasons. Surface errors abound in machine-translated French texts. These errors seem a good target for source-independent lightweight rules that can be developed with simple patterns and
shallow linguistic analysis. The automatic application of rules is motivated by two potential use scenarios. In a technical forum, where users have varied linguistic knowledge and might not have particular interest in fixing linguistic issues, automatic rule application requiring no participation or effort is clearly valuable. In a case where forum posts were to attain a better quality and a manual post-editing phase performed by bilinguals was necessary, automatically applying our rules beforehand could reduce both effort and time involved in this task.

We have developed 27 monolingual post-editing rules for French. The rules treat two types of phenomena: (1) spelling and grammar errors and (2) system-specific errors. We have used different resources to develop and infer the rules: manual analysis of previously post-edited data, bilingual terminology extraction on source and raw translation, and spell-checking of the raw translation using Acrolinx\textsuperscript{10}.

Examples of errors and monolingual automatic rules for French can be found in Table 1.

<table>
<thead>
<tr>
<th>Incorrect negation</th>
<th>Wrong word order</th>
</tr>
</thead>
<tbody>
<tr>
<td>Je n'ai pas accès à distance.</td>
<td>Le Norton technicien Norton m'a conseillé de [...].</td>
</tr>
<tr>
<td>C'est ce n'est pas bloqué par le fichier.</td>
<td>Votre PC périphérique PC doit être [...].</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Incorrect punctuation and elision (comma, hyphen)</th>
<th>Reformulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Je comprends mais comprends, mais...</td>
<td>je suis en espérant &gt; j'espère</td>
</tr>
<tr>
<td>As-tu As-tu lu ça ?</td>
<td>Veuillez aider. &gt; Aidez-moi, s'il vous plaît.</td>
</tr>
<tr>
<td>Est-ce qui il qu'il s'agit de...</td>
<td>Ti Bonjour, merci pour le message.</td>
</tr>
<tr>
<td>Blocage des appels Pas de message &gt; appels. Pas de</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Incorrect verb form (imperative, infinitive, participe, subjonctif)</th>
<th>Agreement errors (subject-verb, determinant-noun, noun-adjective)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Il n'a pas faire fait ça.</td>
<td>Lorsque je faire fais une recherche [...].</td>
</tr>
<tr>
<td>J'ai dû fait faire ça.</td>
<td>commentaires apprécié- appréciés</td>
</tr>
<tr>
<td>Bien que je ne comprends comprenne ça, [...].</td>
<td>nouveau nouveau article</td>
</tr>
<tr>
<td>Regards Regarde en bas.</td>
<td>le les deux chose choses</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Casing error</th>
<th>Wrong term and anglicisms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Il a demandé à Si si je savais [...].</td>
<td>Les mises à jour norton Norton [...].</td>
</tr>
<tr>
<td>tout Tout en supposant que [...].</td>
<td>Veuillez la mettre à jour asap au plus tôt.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Missing or extra spaces</th>
<th>Doubled words</th>
</tr>
</thead>
<tbody>
<tr>
<td>4GB &gt; 4 GB</td>
<td>Je ne l'ai pas pas pas fait.</td>
</tr>
<tr>
<td></td>
<td>J'ai mis à jour mon les mes pilotes.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Avoid direct questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tu as As-tu lu le message ?</td>
</tr>
</tbody>
</table>

Table 1: Example of phenomena treated by French automatic post-editing rules

In this section, we describe the methodology followed to test our hypotheses. We introduce the tasks designed to this end, the data selected and the participants recruited for the study.

3.1. Method

In our study, two tasks were designed to evaluate automatic monolingual post-editing rules in terms of readability and usefulness (as discussed in Section 1): a comparative evaluation task
aimed at eliciting judgments on the impact of our rules on readability (the extent to which a translated segment reads naturally), and a post-editing task aimed at determining the usefulness of changes introduced by rules in an actual post-editing context (their beneficial and practical use). Results for readability and usefulness were cross-analysed.

3.2. Data

An original corpus of 5000 English sentences extracted from the Norton Community forum was pre-edited using the project’s pre-editing rules for English. The data was then translated into French using the project’s baseline system, which is a phrase-based Moses system, trained on translation memory data supplied by our partner, Symantec, and supplemented with Europarl and news-commentary data (ACCEPT D4.1).

We automatically applied our post-editing rules to the translated corpus and removed sentences with more than 40 words to avoid long sentences. We classified the resulting sentences according to the Levenshtein distance between the automatically post-edited (APE) output and the raw output, and then according to the number of rules that had been applied in each APE sentence. Our intention was to focus on sentences with the highest number of changes in order to cover a larger number of post-editing rules. We kept for this study a sample consisting of the first 200 sentences appearing at the top of the resulting classification. One sentence was duplicated and eliminated from the selection. The selected 199 sentences totalled about 3700 words.

3.3. Participants

For both tasks performed in this study, we recruited three translation students in the second year of the MA programme at the Faculty of Translation and Interpreting of the University of Geneva. They are native French speakers with English as their main working language. None of the participants had specific technical knowledge.

3.4. Comparative Evaluation Task

This task was meant to test our first hypothesis (see Section 1). We let annotators comparatively evaluate pairs of raw and APE sentences. They rated each pair on a 3-point scale: first better—equal—second better, according to which of the versions they considered to be more readable. For this task, annotators were not shown the corresponding source. The evaluation focussed on readability alone, with no consideration of adequacy. The two versions were shown to annotators in random order to avoid bias. In addition to evaluating the overall readability of the sentence, the annotators rated all individual edits (IE)\(^2\) automatically introduced by our rules using the same 3-point scale mentioned above. Annotators were provided with guidelines and evaluated 199 sentences and 391 IEs using Excel sheets.

3.5. Post-editing Task

To test the second hypothesis, we asked the same annotators to manually post-edit the APE output with access to the source text.

\(^2\) We understand by “individual edits” any sequence of adjacent words modified by the automatic application of our monolingual rules for French.
The post-editing task was performed using the post-editing environment of the ACCEPT portal (ACCEPT D5.6; Roturier et al., 2013) in bilingual mode, as shown in Figure 2. Participants were provided with post-editing guidelines and a glossary of the domain. They were asked to render a grammatically correct target sentence, which should convey the same meaning as the original, and to use as much of the raw MT output as possible. Style was not given priority. No time limit was given, and all participants were paid.

Figure 2: Interface of the ACCEPT post-editing environment

Once the task was completed, we compared the resulting sentences with the APE version to identify the actual differences between the versions. We used an in-house tool to automatically identify the IEs that our rules had introduced in the raw output and that human post-editors had kept during the subsequent manual processing. This allowed us to check the rate of preservation of the IEs. The results of the automatic extraction were checked manually to ensure that all IEs were detected, including insertion/deletion of spaces and use of capitals.

4. Results

This section presents the results obtained by applying the method described in Section 3. We proceed by presenting the findings related to the hypotheses put forward in Section 1.

4.1. Comparative Evaluation Task – Readibility

The number of sentences and individual edits (IEs) deemed as better was higher in the case of automatically post-edited (APE) sentences than in the case of raw sentences. A total of 199 sentences and 391 IE were evaluated. The results of the comparative evaluation task are shown in Table 2.

Ratings were similar both at the sentence level and at the IE level. On average, 74% of sentences (78%-80%-64%) and 75% of IE (74%-84%-68%) were considered better in terms of readability when automatic post-editing is applied. While in each evaluation a mean of 20% of annotated pairs were considered equal, the amount of raw sentences and IEs considered better
was negligible (1% to 6%). A one-sample chi-square test of goodness of fit was performed to test the difference in proportions. For all three annotators, the difference in proportions between the three categories is significant at sentence level ($\chi^2(2, N=199) = 186.5 / 211.4 / 114, p < 0.001$) and at IE level ($\chi^2(2, N=391) = 298.6 / 455.5 / 255.7, p < 0.001$).

<table>
<thead>
<tr>
<th></th>
<th>Sentence level</th>
<th></th>
<th>raw better</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotator 1</td>
<td>156 (78%)</td>
<td>34 (17%)</td>
<td>9 (5%)</td>
</tr>
<tr>
<td>Annotator 2</td>
<td>161 (80%)</td>
<td>36 (18%)</td>
<td>2 (1%)</td>
</tr>
<tr>
<td>Annotator 3</td>
<td>128 (64%)</td>
<td>66 (33%)</td>
<td>5 (3%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>IE level</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotator 1</td>
<td>288 (74%)</td>
<td>80 (20%)</td>
<td>23 (6%)</td>
</tr>
<tr>
<td>Annotator 2</td>
<td>328 (84%)</td>
<td>51 (13%)</td>
<td>12 (3%)</td>
</tr>
<tr>
<td>Annotator 3</td>
<td>268 (69%)</td>
<td>111 (28%)</td>
<td>12 (3%)</td>
</tr>
</tbody>
</table>

*Table 2: Comparative evaluation task – Results for readability*

The observed agreement for judgements at the IE level was of 56% (unanimous = 219/391) and it reached 58% at the sentence level (unanimous = 115/199).

We assessed inter-annotator agreement (IAA) to validate this observation. At the sentence level, we first calculated Cohen's kappa for each pair of annotators (Cohen, 1960). Although the observed agreement was relatively high, results only showed fair agreement (average $k = 0.277$), probably due to the effects of prevalence (Artstein&Poesio, 2008). Because $k$ may become unreliable when used on skewed data, we decided to assess IAA using a two-way intra-class correlation (ICC) (McGraw, 1996). The resulting ICC was in the good range, ICC = 0.64 (Cicchetti, 1994), indicating that annotators had a relatively high degree of agreement and a low amount of measurement error.

We did the same for the evaluation at the IE level. Cohen's kappa was equally low (average $k = 0.245$) and the two-way ICC (McGraw, 1996) was also in the good range, ICC = 0.62 (Cicchetti, 1994).

The results of this first experiment confirm our hypothesis that our automatic monolingual rules significantly improve readability.

### 4.2. Post-editing Task – Usefulness

For this task, the analysis focused on the IE level. We assessed the rate of preservation of the 391 IEs that had been introduced by our automatic rules.

Our analysis showed that a high percentage of IEs (70%) was kept during manual post-editing, suggesting that our rules perform useful modifications that reduce the number of changes post-editors have to perform to reach the final output (see Table 3). Some sentences were not edited at all (4%, 8%, 10%). A one-sample chi-square test of goodness of fit was performed to test the difference in proportions between the Found and Missing at IE level. For all three annotators, the difference is significant ($\chi^2(1, N=391) = 64.7 / 68 / 78.3, p < 0.001$). Results are shown in Table 3.
To assess agreement of the three post-editors on the IEs that were kept, we again computed Cohen’s kappa (Cohen, 1960), which showed moderate agreement, $k = 0.559$ (Landis&Koch, 1977). A two-way ICC assessment indicated excellent agreement, ICC = 0.79 (Cicchetti, 1994).

To quantify the share of work performed by the automatic post-editing rules, we chose to measure edit distance by means of TER (Snover et al., 2006). Our assumption was that the TER score would be lower for the automatically post-edited (APE) output than for the raw output.

We computed TER for the raw MT and APE output using the manually post-edited sentences as reference. The raw MT output achieved a TER score of 0.42, while the APE output dropped to 0.27. This suggests that, in terms of edits, our rules contribute to making the MT output more similar to the human output (lower values indicating higher similarity).

Since the manual post-editing was done by using the automatically post-edited as a basis, it might be argued that the final human output will be closer to the APE version than to the raw MT output because of this methodological choice. To obtain scores not subject to this bias, we computed TER scores against a human reference built from scratch. This reference was produced by a native French speaking professional translator with domain knowledge. The translator used the same guidelines as the post-editors. Against this second reference, the raw MT output achieved a TER score of 0.66 against 0.59 for the APE version. While the difference between scores is smaller, it is still in favour of the APE version. These results confirm that the changes introduced by the automatic rules bring the text closer to the final version and reduce the post-editing "technical effort" (as defined by Krings, 2001).

In view of the above, we can conclude that our second hypothesis was also confirmed. We had assumed that most individual edits (IE) introduced by our rules would be kept in the final version of the selected sentences.

### 4.3. Readability vs Usefulness

Our third hypothesis was that the IEs preserved during manual post-editing would not necessarily be the same as those IEs judged as enhancing readability. We expected a low correlation between readability and usefulness. To test the hypothesis, we crossed the data obtained in the comparative evaluation task (readability) and the post-editing task (usefulness).

The cross-data analysis was very similar for all three annotators (see Table 4). Results show that, on average, 60% of the IEs introduced by our rules (58%-66%-54%) were considered better during the comparative evaluation task and also kept during the manual post-editing of the output. A lower percentage (16%-17%-16%) was discarded. The rate of preservation of IEs considered equal was of about 10%, while a similar percentage (11%) was discarded. Finally, only 1% to 3% of the raw versions were found better and either discarded or kept.
Table 4: Combined results for readability and usefulness at the IE level

<table>
<thead>
<tr>
<th></th>
<th>IE level</th>
<th>Annotator 1</th>
<th>Annotator 2</th>
<th>Annotator 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>APE better</strong></td>
<td>Found</td>
<td>227 (58%)</td>
<td>260 (66%)</td>
<td>213 (54%)</td>
</tr>
<tr>
<td><strong>APE better</strong></td>
<td>Missing</td>
<td>61 (16%)</td>
<td>68 (17%)</td>
<td>55 (16%)</td>
</tr>
<tr>
<td><strong>Equal</strong></td>
<td>Found</td>
<td>38 (10%)</td>
<td>15 (4%)</td>
<td>64 (16%)</td>
</tr>
<tr>
<td><strong>Equal</strong></td>
<td>Missing</td>
<td>42 (11%)</td>
<td>36 (9%)</td>
<td>47 (12%)</td>
</tr>
<tr>
<td><strong>Raw better</strong></td>
<td>Found</td>
<td>10 (3%)</td>
<td>2 (1%)</td>
<td>6 (2%)</td>
</tr>
<tr>
<td><strong>Raw better</strong></td>
<td>Missing</td>
<td>13 (3%)</td>
<td>10 (3%)</td>
<td>6 (2%)</td>
</tr>
</tbody>
</table>

Table 5 illustrates correlation cases. We provide one example for each case, but for space limitations, we will only comment on APE better-Found and APE better-Missing cases. A thorough study of all correlation cases is still needed to draw complete and definitive conclusions.

<table>
<thead>
<tr>
<th>English Source</th>
<th>Raw MT-Output</th>
<th>APE sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>R: APE better U: IE Found</strong></td>
<td>Also are there any programs you recommend doing the job?</td>
<td>Il y a également des programmes que vous recommande de faire ce travail?</td>
</tr>
<tr>
<td><strong>R: APE better U: IE Missing</strong></td>
<td>If you have already done this, but the space is not showing released [...]</td>
<td>Si vous avez déjà fait, mais l'espace n'est pas faire preuve de la sortie de message privé [...].</td>
</tr>
<tr>
<td><strong>R: Equal U: IE Found</strong></td>
<td>Are they still valid?</td>
<td>Ils sont toujours valables?</td>
</tr>
<tr>
<td><strong>R: Equal U: IE Missing</strong></td>
<td>I was looking for Ghost [...]</td>
<td>J'étais en train d'pour Ghost [...]</td>
</tr>
<tr>
<td><strong>R: Raw better U: IE Found</strong></td>
<td>Is the post below also posted by you?</td>
<td>C'est le post ci-dessous également publiés par vous?</td>
</tr>
</tbody>
</table>

Table 5: Sample cross-data for readability and usefulness

Combinations APE better-Found are the most common. They correspond mostly to corrections of shallow errors related to grammar and structure, and some specific reformulations. Row one of Table 5 illustrates this correlation.
Combinations APE better-Missing are the second most common. These can be explained by the characteristics of the rules themselves. Due to the chosen technology and to the fact that monolingual rules do not refer to the source text, our rules are incapable of correcting long distance dependencies, detecting incorrect lexical choices, producing perfect agreement between words or choosing the right verb tenses. They are developed to treat mainly local and highly recurring phenomena. As a consequence, a rule might correct a sequence of words and thereby locally improve the natural flow of the text (e.g., by inversing verb and subject in questions or correcting wrong verb forms), but when considering the entire sentence, these changes might not be relevant. Row two of Table 5 illustrates these cases. The APE version was considered better in the readability task, but the individual edits introduced were not kept. Taken out of context, the sequence "n'a pas fait preuve de" is better than "n'est pas faire preuve de". However, considering the entire sentence and the source text, both versions are wrong and the correction made by the IE is useless.

5. Conclusions

Our study has shown that lightweight automatic post-editing rules such as the ones developed in the ACCEPT project for French user-generated content are beneficial both in terms of readability and usefulness for subsequent manual post-editing. About 74% of the sentences and IEs evaluated were deemed better when automatic post-editing rules were applied, and 70% of the IEs that the rules had introduced were kept. The TER results confirm that an APE version can reduce post-editors' technical effort.

The cross-data analysis confirmed that certain rules induce changes that are more adequate for readability purposes than for the actual post-editing task. This analysis has allowed us to better understand and explain why our rules may produce a divergent effect, that is, they improve readability but do not help in the post-editing task, or vice versa. Although a high percentage of IEs improve readability and are useful for manual post-editing, a non-negligible percentage fell in other categories.

In future work, we plan to perform a more detailed analysis of the results obtained in this study. In particular, we want to look into the specific rules that produce the divergent effect mentioned above. This will allow us to classify and filter rules depending on the purpose they may serve the best. We also plan to perform the extrinsic evaluation of post-editing rules, in an actual forum context. Since the rules are tailored to social platforms and in particular to technical forums, we would like to perform evaluation using real users, in order to assess both the readability of rules and their contribution to solving users' problem at hand.

Acknowledgements

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 288769.

References

ACCEPT Deliverable 2.4 (2014). Available at: http://www.accept.unige.ch/Products/D-2-4-Definition-of-Post-editing-Rules.pdf
Translating and The Computer 36

ACCEPT Deliverable 4.1 (2012). Available at: http://www.accept.unige.ch/Products/D_4_1_Baseline_MT_systems.pdf

ACCEPT Deliverable 5.6 (2013). Available at: http://www.accept.unige.ch/Products/D_5_6_Browser-based_client_demonstrator_and_adapted_post-editing_environment_and_evaluation_portal_prototypes.pdf


Quality Assurance process in translation

Jerzy Czopik

ABSTRACT

Quality does not start, when the translation is finished. To deliver a high quality product a well designed process is necessary. In the best case it starts already before the translation is assigned to a translator.

The session will start with defining quality and looking at the measures to achieve it. After having done that, we have a good starting point to talk about checking quality.

Quality of a translation cannot be achieved by using tools like CAT or QA-tools. These tools can only provide some help, but cannot replace the human. Nevertheless good quality can be improved, if the tools are used properly. But only then – improper use will cause a lot of misunderstandings and problems.

We shall thus talk about quality checking, focused on the target language. Tools like SDL Trados Studio, MemoQ or Xbench allow you to configure the QA-checking modules, but in quite different ways. Here not only the knowledge of the tool, but also some understanding of the target language is necessary. Best case QA-checking should be done by people understanding both source and target language. Unfortunately very often this process is done by project managers, who typically cannot have command of as many languages as the languages of the projects they manage.

During the session I would like to show why understanding target is also necessary when doing QA-checking.

1. Introduction

First things first: this paper will not talk about linguistic quality. Measuring the quality of a language is not my aim here, as it is highly complicated and remains quite subjective – our perception on language may differ, even though a definition of proper language is of course possible. So instead I will concentrate on parameters, which can be easier checked and controlled.

This session and this paper are addressed mainly to freelance translators and small LSPs.
2. Defining quality

More or less everybody is talking about quality nowadays. And of course nearly everybody talks about quality management, quality assurance, quality control and all related processes. But what IS quality?

There are several definitions for quality. If you look at Wikipedia (http://en.wikipedia.org/wiki/Quality_%28business%29 – to take an example), you’ll find some interesting descriptions of how quality can be defined in business. And in the end what we do is business, so applying this definition to translation quality is surely not unreasonable.

From the definitions given there I would chose these:

- A combination of quantitative and qualitative perspectives for which each person has his or her own definition; examples of which include, “Meeting the requirements and expectations in service or product that were committed to” and “Pursuit of optimal solutions contributing to confirmed successes, fulfilling accountabilities”.
- Degree to which a set of inherent characteristics fulfils requirements. The standard defines requirement as need or expectation (ISO 9000).
- Fitness for use. Fitness is defined by the customer.
- The result of care.
- Conformance to requirements.

Personally, I very much like the last three with the last one being my absolute favourite.

But what does this “conformance to the requirements” mean for us?

3. Quality in translation

Translation is not just done for the sake of the translation itself. The purpose of the translation is to make information given in the source language to the audience, speaking the target language.

So to provide a “high quality translation” we must meet the needs of this audience, as this audience is the “customer” here. However, our customer is the one who gives us the job. He is the one who knows the audience, so he defines the expectations – and this is why I particularly like the last definition. If the customer just needs the text to be understandable without any special requirement regarding the language, a machine translation with some post editing may be fully sufficient. In this situation the “high quality” is defined not in the common sense, as we would understand it if applied to a product we would like to buy. Here “high quality” means to meet the expectation of the customer with the lowest possible effort. A very good example of this approach are articles in the knowledge base of Microsoft, which are machine translated and still very understandable – at least from what I can tell for their German and Polish language versions.

Using some common sense definitions, the quality of a translation shall include, but is not limited to:

- correct transmission of the meaning from the source to the target language
• no spelling and grammatical errors
• linguistic perfection
• nicely written sentences (good readable)
• and many more...

Certainly there are much more criteria we could define.

But how to achieve these?

4. Steps to quality

Being bilingual or having a so called “very good” knowledge of a foreign language is commonly seen as a good base for being a translator. For most people outside of our industry translating is more or less the same as taking a dictionary and retyping the source in target language. I think most of us have already heard similar statements from other people. But is an in-depth knowledge of another language really enough to be a translator? Certainly not – I think we all agree on that. So what is really needed to provide a translation, which can be considered a good one?

Again there are so many aspects of this, so I can list here just few of them, but I think the following are quite important:

• perfect language skills
• knowledge in the field of translation
• research
• continuous learning
• ...

But even though people may study languages and be an expert in several languages, knowing the important aspects of grammar, vocabulary and so on, such people are not necessarily good translators. Same applies to bilinguals. There must be something more, what we need to be a good translator. Something similar to what an artist needs to be a good one or an interpreter need to carry the audience away. A translator must “fit” to the purpose and must have that “something”, what makes him more or less an artist. This is quite an important aspect, which cannot be taught – either you have it or you have it not. All this may sound a bit philosophical, but if you look at your own career I think you’ll agree. Your have been successful because you are not “just a translator” – you are a kind of a virtuoso.

5. How to achieve quality

Now, as we already defined what kind of quality we would like to achieve and which steps can be taken, it is time to define a process for achieving quality.

Of course language cannot be pressed in a form and moulded like aluminium. Language is like fluid. Do you remember the definition of fluids from your school time? Here is a definition from Wikipedia: “In physics, a fluid is a substance that continually deforms (flows) under an applied shear stress.” (http://en.wikipedia.org/wiki/Fluid) This fits quite well I would say.
But even if the language is like fluid, we can catch it. A fluid will take the form of a vessel, so we just need a vessel:

(https://slodkotko.blogspot.de/2012/03/inspiracje.html)

Now let’s try to define a process of getting our translation there:

- The customer sends a request
- We check, if we’re able to fulfil the request
- The job is accepted and confirmed
- Materials provided by the customer are being translated
- A spell- and grammar-check will be performed
- The text shall be checked for omissions or number errors
- A proofreading is performed (better to call this step revision)
- Target text is being created
- The target text shall be formatted
- A final check is done
- The result is delivered to the customer

Of course this is a very simplified process and the steps above can be supplemented by many more, but this reflects the core of our work.

6. Quality Assurance (QA) vs. Quality Control (QC)

When we talk about quality and means to achieve it in our work, in most cases the whole process is being called “Quality Assurance”. Unfortunately, usually in our business this is not correct.
If we talk about quality assurance, the process MUST start before the text even comes to us. A QA process must already begin from the creation of the source, which will then be translated. But in nearly all situations no one of us – both freelancers and agencies – can take influence on the creation of source materials. So our QA starts from the receiving of the materials from the customer. So it starts before we start to translate. However, this fact is commonly being forgotten – and people talk about Quality Assurance, but start to perform it when the translation is already done. In this situation no influence can be taken on the input – so if the input is bad, there is I think no doubt that the translation cannot be as good as it would be with a good input.

So if we want to perform a “real” Quality Assurance, we MUST start before the job even begins. This is quite important – and even if we cannot influence the creation of source materials, we can of course check them before we start to translate.

Many errors (not only translation errors) are caused by problematic source. The source may be badly understandable, and so the translation will remain – it is at least not trivial to translate a non understandable source to an understandable target. However in such cases many translations really read better than the original. But the source may have much more traps than the linguistic one. And all these “small” traps are usually overseen in the process of translation – by all involved parties, starting from the customer and ending with the translator. I will return to this problem in another section of this paper.

What is usually performed by translators and translation agencies, is not Quality Assurance (even though many of us really call this process QA – just look at the tools we use), but Quality Control. The difference is quite simple. While QA starts from the very beginning of the process and follows it in every step, the QC is a task performed after certain process steps have already been done. QC is the proofing done after the translation has been finished. It is part of the Quality Assurance, but cannot replace the QA. The translation can be checked for several purposes – spelling, grammar, no omissions, correct numbers, comprehensibility and many other. All this checks are parts of the Quality Control.

7. Checking of the quality in CAT tools

All CAT-tools I know have a build-in module for checking quality. Very often this module is called “QA Checker” (SDL Trados Studio) or “Run QA” (MemoQ), even though no real quality assurance is done here.

Obviously also integrated is at least a spell-checking module and sometimes a grammar checking module.

Beside the integrated checking modules also external programs can be used. There are several programs on the market, which can perform quality checks. Some of those are even for free, like Xbench in version 2.9. The current version of Xbench is 3.0 and it is a paid one. Other program for quality checking is ErrorSpy or QA Distiller and many other.

The checks performed include, but are not limited to:

- numbers
- translation consistency
8. How to achieve quality

A good question I think. And a question with no general answer unfortunately.

There was no, is no and will not be any universal method to achieve quality, valid for all of us and for all languages and kinds of texts. Quality must be defined and checked individually.

However, some general steps can be defined, which shall help to achieve the desired quality. But quality does not start with the translation and does not only apply to the translated text. No, it starts even before we have got the job... Quality starts with the author of the text. A badly written text may not deliver as good target text as a source text, which has been written with the intention to be published in several languages and thus having the need of translation in mind.

Of course we cannot affect the author of the text. But still we can do something with the text we shall translate – we can try to tidy it up.

When for example translating from German into English you may sometimes encounter something like this:

\textit{Mit der Maschine können:}
\begin{itemize}
\item Dosen,
\item Flaschen
\end{itemize}
produziert werden.

It means more or less, that the offered machine is designed for producing cans and bottles. But of course it is not possible to keep this word order in English, so the last part "\textit{produziert werden}" must be integrated in the first part and the last segment will be empty or contain just a period:

\begin{align*}
\textit{Mit der Maschine können:} & \quad = \quad \textit{The aim of the machine is to produce:} \\
- \textit{Dosen,} & \quad = \quad - \textit{cans,} \\
- \textit{Flaschen} & \quad = \quad - \textit{bottles.} \\
\textit{produziert werden.} & \quad .
\end{align*}

If you now confirm all the segments, they can be used for a pretranslation of your next job. And in your next job your customer will send you this text:
Now this machine is not for producing, but for filling of cans and bottles. But you have already this segment in your translation memory:

Mit der Maschine können:  
- Dosen,
- Flaschen

abgefüll werden.

So you are now trapped – and this is why such text needs correction before you start translating or you need a different way of treating segments within your translation project to avoid such things happening. This ways may differ from tool to tool, but the aim is not to store such translations in the TM.

From my own experience it is very important to go through the source and not only to check for such examples like above, but also look for superfluous paragraph breaks, tabulators and other means of formatting, which are breaking the text flow into separate segments.

This is what you see in a PDF file or in printout:

**Adapter Vierkant** (1/4" Sechskant)  
Nutzung mit Schnellwechselfutter und Führungshülsenmodellen, soweit nicht anders vermerkt.

But when you open the file in the corresponding editing program and switch on the view of non-printable characters, you will see this:

**Adapter Vierkant** (1/4" Sechskant)
Nutzung mit Schnellwechselfutter und Führungshülsenmodellen, soweit nicht anders vermerkt.

The above text is splitted into four lines by paragraph breaks. Of course this will cause severe problems in your translation. So to make your job easier remove the paragraph breaks. If you perform these steps, you will move from quality checking to real quality assurance.

After having prepared your text for translation the core process may now start. During the process, beside the obvious things like correct grammar, perfect wording, usage of correct terminology, consistence and other factors some other much less important things do also...
matter. These small things are commonly considered not to be very important. These are mainly elements of typography, but also spaces and dashes and so on. Quite often I see linguistically good translations where double spaces between words may be found – of course this is a minor error, but such text goes to print without correction you will see the disturbing bigger space between words. These thins include quotation marks and also correct spacing between numbers and measurement units. As translators we must remember, that the person creating the final document in a DTP application for example may very well not know our language. So she or he will not know, if the correct typographic quotes are ” ” like in English or „ “ like in German or maybe « » for French or „ “ for Polish. She or he will not know that the German notation of percentage with a space like 5 % is not correct in Polish, where it should read 5%. The delivered translation needs to fit to the standards of the target language and make the job of the layouter easy. Checking for such errors is a part of the QC process.

9. Why even small things do matter

All the steps taken to achieve high quality translation may sometimes fail, because some small things have been overseen.

As you've already seen, translating badly prepared and thus badly segmented text may bear a danger of mistranslation in it because of the use of a translation memory system, which will pretranslate text, taking matches from the TM, which may not fit. But this kind of errors will usually be fixed with the proofreading of the text. However, some other kind of errors tend to be overseen quite often. One of the most common mistakes is the omission of spaces between the number and the measurement unit. For example instead of 40 kW, you see 40kW. Or in many American texts you can find 24VDC, even though it should state a voltage of 24 V direct current – 24 V DC. Many engineers do not pay attention to that, so many translators do simply use same way of writing. However, a mistake in the source does not really legitimate us to repeat it in target.

Please look at the following example. This small piece of text is a part of a description of a Polish airplane from the time of the Second World War.¹


¹See "Andrzej Glass, Samolot bombowy PZL. 37 ŁOŚ, TBIU 1/2012(5), Wydawnictwo ZP 2012"
It lists the capacities of fuel and oil tanks, given in litre. The unit symbol is “l”. But in the font used, the lower case letter “l” (L for litre) and the number “1” are very similar. This is the case for example if you use Times New Roman:

Here is the letter “L” in lower case in TNR: l

And here the number “1” in TNR: 1

They are really very similar. Now please look at this sentence from the above example:

This sentence does give the capacity of the main fuel tank. But it is NOT 5321 litres! The plane has had a total own weight of approx. 6250 kg and a load capacity of 4250 kg. Over 5000 litres of fuel would mean more than the load capacity of the plane. The correct number would be 532 l – 532 litres, but due to the missing space you really may get the impression the capacity were over 5000 litres.

Another nice example is the loss of the Mars Climate Orbiter. Due to the usage of wrong unit systems this NASA space probe was totally lost. This example is not directly correlated with linguistic translation, but shows how important a proper QA and QC is.
Intelligent Translation Memory Matching and Retrieval Metric Exploiting Linguistic Technology

Rohit Gupta       Hanna Bechara       Constantin Orasan
University of Wolverhampton       University of Wolverhampton       University of Wolverhampton

ABSTRACT
Translation Memories (TM) help translators in their task by retrieving previously translated sentences and editing fuzzy matches when no exact match is found by the system. Current TM systems use simple edit-distance or some variation of it, which largely relies on the surface form of the sentences and does not necessarily reflect the semantic similarity of segments as judged by humans. In this paper, we propose an intelligent metric to compute the fuzzy match score, which is inspired by similarity and entailment techniques developed in Natural Language Processing.

1. Introduction
Most of the Translation Memories research has been carried out in the industry. The focus of this research has been on providing a good graphical user interface to the translators, developing different filters to handle different file formats (e.g. pdf, xml, txt, html, word, xlf, subtitle etc), and project management features. Apart from this, current TMs are also equipped with some tools like terminology managers and plugins to support machine translation from MT service providers. One of the core features of a TM system is retrieving previously translated similar segments for post-editing in order to avoid translation from scratch when an exact match is not available. However, this retrieving process is still limited to edit-distance based measures. Although, these measures provide a strong baseline, they are not sufficient to capture the semantic similarity between the segments as judged by humans. This results in uneven post-editing time required for the same fuzzy match scored segments and non-retrieval of semantically similar segments. In this paper, we propose an intelligent metric to compute the fuzzy match score, which is based on SemEval Task 1 semantic similarity and textual entailment system (Gupta et al., 2014).

2. Our Approach
The system described in (Gupta et al., 2014) calculates the similarity and entailment between a pair of sentences. This system was adapted to measure the similarity between two TM segments. Given the amount of calculation involved in the task, we kept only those features which can be quickly calculated and proved the most useful for the original system. The system
uses features based on surface form, parts of speech information, lemma, typed dependency parsing, named entities, paraphrasing, machine translation evaluation, and corpus pattern analysis (Hanks, 2013). Stanford CoreNLP3 toolkit (Manning et al., 2014) provides lemma, parts of speech (POS), named entities, and dependencies relations of words in each sentence. We used the PPDB paraphrase database (Ganitkevitch et al., 2013) to identify paraphrases.

After extracting these features, we employed a support vector machine (SVM) in order to build a regression model to predict semantic similarity. The training dataset for the SVM is a set of 4934 parallel sentences of the SICK dataset (Marelli et al., 2014) annotated with similarity scores by humans. The SVM used an RBF kernel with $C = 8$ and $\gamma = 0.125$. More details about the method can be found in (Gupta et al., 2014).

The trained SVM system works as a similarity calculator between any pair of sentences provided that the same feature values are available for this pair of sentences.

3. Experiments and Results

We carried out evaluations on two different sets. The test sets were generated by a random selection of segments from DGT-TM corpora (Steinberger et al., 2012). We used English as source and French as target. The target side (French) of the input was considered as a reference for evaluation. We used the word based edit-distance measure implemented by OmegaT\(^1\) as a baseline. The statistics for our test sets is given in the Table 1 below:

<table>
<thead>
<tr>
<th></th>
<th>Test-1 (# segments)</th>
<th>Test-2 (# segments)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input</td>
<td>500</td>
<td>2500</td>
</tr>
<tr>
<td>TM</td>
<td>5000</td>
<td>10000</td>
</tr>
</tbody>
</table>

**Table 1: Test sets statistics**

We performed both a manual and automatic evaluation. For our automatic evaluation, we used the machine translation evaluation metrics METEOR (Denkowski and Lavie, 2014) and BLEU (Papineni et al., 2002). For each input segment, we retrieved the most similar sentence (and their proposed translation into French) as indicated by the baseline and our similarity metric. Table 2 presents the results of automatic evaluation when having a threshold of 70% over the edit-distance. BLEU-ED-70 represents BLEU score using edit distance, BLEU-SS- 70 represents BLEU score using our approach, METEOR-ED-70 represents METEOR score using edit distance, and METEOR-SS-70 represents METEOR score using our approach. The proposed method yields better results for Test-1 but not for Test-2.

<table>
<thead>
<tr>
<th></th>
<th>Test-1</th>
<th>Test-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLEU-ED-70</td>
<td>77.32</td>
<td>81.34</td>
</tr>
<tr>
<td>BLEU-SS-70</td>
<td>81.61</td>
<td>77.14</td>
</tr>
<tr>
<td>METEOR-ED-70</td>
<td>91.5</td>
<td>87.35</td>
</tr>
<tr>
<td>METEOR-SS-70</td>
<td>92.6</td>
<td>84.55</td>
</tr>
</tbody>
</table>

**Table 2: Results automatic evaluation**

\(^1\) http://www.omegat.org
To gain a deeper understanding of our system’s performance, we also performed a manual evaluation on Test-2. We considered the source side (English) of the segments for this evaluation. A native speaker of English performed the manual evaluation. Three different options were given to the evaluator: Semantic similarity is better; Edit-distance is better; or both are similar. When keeping the 70% threshold and ignoring exact matches, we retrieved 266 different fuzzy matched segments. In these 266 segments, 258 segments were tagged as similar, for 6 segments, edit-distance retrieved better and for 2, our semantic similarity approach retrieved better. Some of the examples from Test-2 are given in Table 3. Example 1 shows our approach (SS) performed better, while examples 2 and 3 show edit-distance (ED) performed better.

The initial results, as stated earlier, show comparable results. Although our approach does not perform better overall, there are several factors, which should be taken into consideration.

<table>
<thead>
<tr>
<th></th>
<th>Input</th>
<th>ED</th>
<th>SS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>For the purposes of this Regulation: For the purpose of this demonstration: For the purposes of this Regulation the following definitions shall apply:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>This Decision shall enter into force on the date of its publication in the Official Journal of the European Union. This Decision shall enter into force on the day of its publication in the Official Journal of the European Union. This Decision shall enter into force on the date of its adoption.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>The Commission sought and verified all information deemed necessary for the determination of dumping. The Commission sought and verified all the information deemed necessary for the purposes of the review. The Commission sought and verified all the information provided by interested parties and deemed necessary for the determination of dumping, resulting injury and Union interest.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Examples from Test-2

The genre of the training set and test set were very different. The SICK dataset consists of simple sentences extracted mostly from image captions while DGT-TM corpus has much larger and complex sentences from mainly legal domain. The average words per segment for TM is 27.9 and for input is 32.54 for test set, whereas for SICK training dataset average words per sentence is only 9.63.

4. Conclusion and Future Work

In this paper we suggested an initial approach to employ a semantic similarity system in a TM framework. Our initial experiment shows some positive indication in this direction. We are in a stage of improving and speeding up our system and extend our experiment to a similar training and test set. In the future, we would also like to develop a human annotated corpus of the same domain to get the better training model. Other similarity calculation techniques involving less computation will also be explored.

References


Improving fuzzy matching through syntactic knowledge

Tom Vanallemeersch  
University of Leuven

Vincent Vandeghinste  
University of Leuven

ABSTRACT

Fuzzy matching in translation memories (TM) is mostly string-based in current CAT tools. These tools look for TM sentences highly similar to an input sentence, using edit distance to detect the differences between sentences. Current CAT tools use limited or no linguistic knowledge in this procedure. In the recently started SCATE project, which aims at improving translators' efficiency, we apply syntactic fuzzy matching in order to detect abstract similarities and to increase the number of fuzzy matches. We parse TM sentences in order to create hierarchical structures identifying constituents and/or dependencies. We calculate TER (Translation Error Rate) between an existing human translation of an input sentence and the translation of its fuzzy match in TM. This allows us to assess the usefulness of syntactic matching with respect to string-based matching. First results hint at the potential of syntactic matching to lower TER rates for sentences with a low match score in a string-based setting.

Acknowledgments

This research is funded by the Flemish government agency IWT\(^1\) (project 130041, SCATE).

1. Introduction

Computer-aided translation (CAT) has become an essential aspect of translators' working environments. CAT tools speed up translation work, create more consistent translations, and reduce repetitiveness of the translation work. One of the core components of a CAT tool is the translation memory system (TMS). It contains a database of already translated fragments, called the translation memory (TM), which consists of translation units: segments of texts (sentences, titles, cell tables, etc.) together with their translation. Given a sentence to be translated (which we will call the query sentence), the CAT tool looks for source language sentences in a TM which are identical (exact matches) or highly similar (fuzzy matches), and, upon success, suggests the translation of the matching sentence to the translator.

In current CAT tools, techniques for retrieving fuzzy matches from a TM mainly consider sentences as simple sequences of words\(^2\) and contain very limited linguistic knowledge, for

\(^{1}\) Innovation by Science and Technology
\(^{2}\) Words are generally defined as sequences of characters surrounded by spaces.
instance in the form of stop word lists. Few tools use more elaborate syntactic knowledge. In the SCATE project (Smart Computer Aided Translation Environment), which primarily aims at improving translators' efficiency and consistency, we study the use of syntactic information for detecting TM sentences which are not only similar when comparing words but also when comparing the syntactic information associated with the sentences (such information consists for instance of parse trees with constituents and part-of-speech tags). We investigate whether such abstract, syntax-based matching is able to increase the number of matches (recall) and produce matches which lead to more useful translation suggestions. As the SCATE project is a collaboration between university teams and companies such as translation agencies, which provide feedback on techniques developed within the project and on their potential application in an industrial environment, we do not neglect the aspect of matching speed. This is for instance relevant in case of syntax-based matching methods like comparison of parse trees.

The remainder of this paper is structured as follows. In the next section, we provide background on matching methods which are based on words or syntactic information, and on how to evaluate their results. In the subsequent section, we detail our methodology, i.e. the matching and evaluation methods we applied. We then proceed to present the results from our first, exploratory tests, and finally provide conclusions and a lookout to future research.

2. Background

There is a great variety of methods for matching flat sequences of elements and for matching complex structures like trees. Flat sequences may consists of various elements, like words or characters (one may compare a pair of words based on their characters). Trees are structures with hierarchically linked nodes. For instance, parse trees structure a sentence into its syntactic constituents, assign part-of-speech tags to words, and assign dependency relations to constituents (for instance subject). An example of a parse tree is shown in Figure 1.

![Parse tree with syntactic constituents and part-of-speech tags](image)

Figure 1: Parse tree with syntactic constituents and part-of-speech tags

2.1. Strings

Methods acting upon flat sequences are called string-based. They include methods like Levenshtein distance (Levenshtein 1966) and percent match (Bloodgood and Strauss 2014). Levenshtein distance is one of the most commonly applied matching methods. It looks for the shortest path to transform one sequence into another sequence through deletions, insertions and substitutions. This shortest path is expressed as a distance. The distance can be converted

---

3 One example is the tool Similis ([http://www.similis.org](http://www.similis.org)), which determines constituents such as noun phrases in sentences and allows for retrieving TM sentences which share constituents with the query sentence.

into a score by normalizing on the sentence length. Although commercial TMS developers do not tend to provide details on their matching methods, Levenshtein distance on sequences of words is widely believed to be the major approach (Bloodgood and Strauss 2014), and is also referred to in publicly available system descriptions (Eisele and Lavecchia 2011, Koehn and Senellart 2010a). Table 1 shows an example of Levenshtein distance calculation.

<table>
<thead>
<tr>
<th>A</th>
<th>total of</th>
<th>470</th>
<th>workers</th>
<th>had</th>
<th>accepted</th>
<th>early retirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Many staff also took</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

→ 4 substitutions + 3 deletions = cost 7

Table 1: Levenshtein distance

2.2. Trees

Tree-based methods include tree edit distance (Bille 2005) and tree alignment distance (Jiang et al. 1995). These are two methods with a mathematical rather than linguistic origin. The first method looks for the shortest path to convert one tree to another one by deleting, inserting, and substituting nodes. The second method looks for the easiest way to combine the two trees into a single one by adding empty nodes. When applying these methods to parse trees, node comparison may involve several features on a node, such as word form, lemma (for instance, the infinitive of a verb, the singular form of a noun), part-of-speech tag (for instance, verb) or constituent (for instance, noun phrase). Tree-based methods have also been proposed in the literature on example-based machine translation, see for instance Cromieres and Kurohashi (2011). Figure 2 provides an example of a parse tree alignment based on the shortest tree alignment distance. In this example, the parse trees have a stronger match than the sentences in Table 1, thanks to the fact that the trees contain similar syntactic information: not only some words match, but also some constituent labels and part-of-speech tags.5

Figure 2: Aligned parse trees

5 The VBD node matches the VBN node as the tags both refer to verbs and the surrounding nodes also match.
2.3. Trees as strings

Linguistic knowledge can be exploited in both string-based and tree-based methods. The most obvious matching configuration in this context is applying tree edit distance or tree alignment distance to parse trees. However, it is also possible to use linguistic knowledge in string-based methods. For instance, flat sequences need not consist of word forms, but may also contain word lemmas, part-of-speech tags, etc. Lexical knowledge may be exploited by matching words in flat sequences through lexical similarity. For instance, Gupta (2014) uses a paraphrase database.

Considering tree-based matching methods, the aspect of speed should not be neglected. While methods like Levenshtein distance are highly efficient, even for long sentences, methods like tree alignment distance perform a high number of calculations on a pair of trees. One way to diminish the problem of speed is to convert trees to flat sequences in order to be able to apply string-based methods to them. One such method (Li et al. 2008) is based on work by Prüfer (1918). It converts a tree to a flat Prüfer sequence by visiting the nodes in the tree in a certain order. Prüfer sequences capture some of the hierarchical information in the tree. Figure 3 illustrates this. Matching on Prüfer sequences helps to select matches in an efficient way before applying a more finegrained and time-consuming method like tree alignment distance.

Another way to tackle the aspect of speed is the use of an index, which allows to reduce the set of source sentences in the TM which are candidate for fuzzy matching methods. One example of such an index is a suffix array; see Koehn and Senellart (2010b). For the sake of briefness, we will not go into further detail on the creation of indices.

![Figure 3: Conversion of parse tree to Prüfer sequence](image)

2.4. Metrics

While the usefulness of a translation suggestion should ultimately be determined by the user of a CAT tool, human evaluation is time-consuming. Automatic evaluation metrics may be used as a proxy during the development of fuzzy matching methods, similarly to the modus operandi in the development of machine translation (MT) systems. A wide range of metrics have been designed for MT since the advent of statistical MT, such as BLEU (Papineni et al. 2002), NIST (Doddington 2002) and TER (Snover et al. 2006). The latter stands for Translation Error Rate and is an edit distance which does not only take account of substitutions, deletions and insertions but also of shifts of blocks of words (which take little effort but change the word order drastically).

---

6 Originally called Translation Edit Rate.
The minimal TER rate is 0 (no editing required, i.e. sentences are equal). There is no predefined maximum, as TER results from dividing the number of edits by the number of words in the reference translation. MT metrics can also be used in the context of fuzzy matching, to compare the translation suggestion with the desired translation (Simard and Fujita 2012). This is especially true for TER, as it expresses the effort needed to correct a translation suggestion. TER can not only be calculated for a single sentence pair but also for a set of sentences. In the latter case, long sentences requiring much editing effort have a stronger influence on TER than short sentences requiring much effort.

3. Methodology

In order to measure the usefulness of syntax-based matching methods, we compare their translation suggestions to the desired translation using TER. Our baseline consists of translation suggestions produced by Levenshtein distance on words, as this is the standard matching method. In order for a syntax-based matching method to provide added value, we expect its TER to be lower than that of the baseline, as a lower TER indicates less editing effort. In this regard, we make a distinction between fuzzy match score ranges. CAT tools typically use a fuzzy matching threshold of 70% (Bloodgood and Strauss 2014), which means matches with a Levenshtein score below 0.7 (exact matches have score 1) are ignored, as their translations are considered not useful enough to be provided as suggestions. Therefore, we would like to know whether syntax-based matching methods can make a difference not so much in the high fuzzy range (70 to 100 %) but in lower ranges, in order to improve the recall of the TM, and ultimately improve translators’ efficiency.

Applying syntax-based matching methods on a TM requires preprocessing its source sentences. We use the Stanford parser (Klein and Manning 2003, de Marneffe et al. 2006) to parse English sentences, and derive Prüfer sequences from the parses. We mainly focus on the language pair English-Dutch. We match each sentence in the TM enriched with parses and Prüfer sequences to all other TM sentences using the baseline method (Levenshtein on words) and using syntax-based matching methods, more specifically Levenshtein on Prüfer sequences. While Levenshtein distance commonly involves an identical cost for each type of operation (deletion, insertion, substitution), we apply a more sophisticated weighting scheme to elements in Prüfer sequences: we compare features on nodes to one another and combine them in a weighted manner, according to the formula in Figure 4.

\[
COST_{\text{subt}} = \begin{cases} 
1 & \text{if Term → Non-Term} \\
\lambda_D \Delta_D + \lambda_S \Delta_S + \lambda_C L + \lambda_W W & \text{otherwise} 
\end{cases}
\]

where
\[
\Delta_D = \begin{cases} 
0 & \text{if Dependency relations are equal} \\
1 & \text{otherwise} 
\end{cases}
\]
\[
\Delta_S = \begin{cases} 
0 & \text{if Syntactic constituents are equal} \\
1 & \text{otherwise} 
\end{cases}
\]
\[
C_L = 1 - \frac{1}{1 + \Delta_{\text{Levenst}}} \\
C_W = 1 - \frac{1}{1 + \Delta_{\text{Levenst}}} \\
\Delta_{\text{Levenst}} \text{is the Levenshtein distance between the lemmas} \\
\Delta_{\text{Levenst}} \text{is the Levenshtein distance between the words}
\]
\[
\sum \lambda = 1
\]

**Figure 4: Substitution cost of node pair in Prüfer sequence**
The comparison of word forms and lemmas is based on a character-based Levenshtein comparison. If two terminal nodes have a different lemma or word form but these are very short (say, determiner the vs. a), a substitution costs less than if the lemma or word form is long, as the distance is longer in the latter case. The substitution cost also accounts for deletions and insertions. Some examples:

- Comparison of terminal node `hd|VB|fall|fall` with terminal node `hd|VBZ|fall|falls` leads to cost 0+0+0+\lambda(\text{word form})*(1-0.5)). The label `hd` stands for syntactic function "head", `VBZ` stands for part-of-speech tag "verb, third person singular present", and the last two features are lemma and word form.

- Deletion of non-terminal node `nsubj|NP|:` each feature value is compared to the empty string.

By increasing the two lambdas for syntactic information, we can emphasize the importance of syntactic structure during matching. By increasing the two lambdas for lexical information, we emphasize lexical similarity between sentences. In case of tree alignment distance, we apply a similar weighting scheme when comparing nodes.

In order to speed up matching and avoid having to compare a query sentence to all source sentences in the TM, we build a suffix array from flat sequences using the SALM toolkit and exploit it in a way similar to Koehn and Senellart (2010b). This speeds up the matching process by a factor of 6.

4. Results

We applied the methodology in the previous section to a TM which was provided to the SCATE consortium by one of the companies participating in the project. The TM contains about 2,800 sentences (110,000 words). We parsed the English source sentences using the Stanford parser, and derived Prüfer sequences. We used each source sentence in the linguistically enriched TM as query sentence and looked for the highest-scoring match in the remainder of the TM using the baseline matching method, Levenshtein distance on words. The fuzzy threshold was set to 0.2. Figure 5 shows the distribution of the match scores. All matches below 0.2 are considered useless and get score 0. As expected, low-scoring matches are more frequent than higher-scoring ones. After retrieving the translation of the highest-scoring match of a query sentence from the TM, we calculated its TER with the desired translation (the translation of the query sentence in the TM). We also calculated a single TER for the set of best matches in each fuzzy match score range (see previous section on combining TER rates of multiple sentences). Figure 6 shows the TER rate per range (for instance, the leftmost dot refers to range 0.2-0.3). This figure also complies with expectations: TER diminishes as match scores increase.

---

7 Characters which only differ in case get a cost of 0.5.
8 http://projectile.sv.cmu.edu/research/public/tools/salm/salm.htm
9 Interestingly enough, though, there are far more matches between 0.2 and 0.3 (their frequency is higher than the maximal frequency shown on the y scale) than matches with score 0. Hence, sentences are less likely to be dissimilar to all other sentences than to be slightly similar to some of them.
In order to find out the added value of syntax-based matching for specific fuzzy match score ranges, we applied Levenshtein matching to the Prüfer sequences of the sentences instead of the sentences themselves (we refer to this method as Prüfer matching). We set the lambdas of the weighting scheme described in the previous section higher for lexical features than for syntactic features in order to emphasize lexical similarity. For a given fuzzy match score range, we took all query sentences whose best baseline match had a match score in that range. We then applied Prüfer matching to the same query sentences and calculated TER for the translation of the best matches. Table 2 lists the Prüfer and baseline TER rates per range; rows where rates differ across matching methods are marked in bold.

<table>
<thead>
<tr>
<th>range</th>
<th>baseline TER</th>
<th>Prüfer TER</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2 – 0.3</td>
<td>0.94</td>
<td>0.93</td>
</tr>
<tr>
<td>0.3 – 0.4</td>
<td>0.89</td>
<td>0.88</td>
</tr>
<tr>
<td>0.4 – 0.5</td>
<td>0.80</td>
<td>0.79</td>
</tr>
<tr>
<td>0.5 – 0.6</td>
<td>0.64</td>
<td>0.64</td>
</tr>
<tr>
<td>0.6 – 0.7</td>
<td>0.42</td>
<td>0.42</td>
</tr>
<tr>
<td>0.7 – 0.8</td>
<td>0.33</td>
<td>0.34</td>
</tr>
<tr>
<td>0.8 – 0.9</td>
<td>0.16</td>
<td>0.16</td>
</tr>
<tr>
<td>0.9 – 1</td>
<td>0.09</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Table 2: Baseline and Prüfer TER per score range of baseline matches

Although the differences in TER rates in Table 2 are very limited, they indicate that TER tends to be lower for Prüfer matching in lower fuzzy match score ranges. In the upper ranges, both matching methods tend to have the same best matches, leading the TER rates for the range to coincide. Hence, there is a potential for making matches in lower ranges more useful: replacing baseline matches with Prüfer matches in these ranges leads to a decreased effort when correcting translation suggestions.\(^{10}\) This also holds a potential for improving recall for

\(^{10}\) It should be said, though, that translators using CAT tools may opt for a setting in which they are offered multiple translation suggestions. In such a setting, there should be sufficiently large differences in ranking for a matching method to improve over other ones.
Levenshtein matching: if lower range matches can be made more useful, one may envisage to lower the minimal fuzzy match score, now typically set at 0.7. However, the small TER rate differences show that there is still a clear need to exploit Prüfer sequences in a better way, for instance by varying more on the weights of lexical and syntactic features and testing on larger translation memories. The tests we are currently performing are still in the exploratory phase.

An example of a query sentence leading to differently ranked match scores for two specific TM sentences is shown in Table 3. The second TM sentence is both syntactically and lexically more similar to the query sentence than the first TM sentence, but the baseline method can only detect part of this similarity because it merely matches word forms. Therefore, it considers the first TM sentence as the best match. Table 4 shows the TER rates of the corresponding translation suggestions; the translation of the second TM sentence is very close to the reference (it only differs in the period).

<table>
<thead>
<tr>
<th>Query sentence</th>
<th>Match score</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>TM sentence 1, baseline match</td>
<td>Merger falls within scope of Regulation but raises no serious concerns</td>
<td>0.56</td>
</tr>
<tr>
<td>TM sentence 2, baseline match</td>
<td>The merger does fall within the scope of the Regulation and does raise serious doubts as to its compatibility with the common market</td>
<td>0.48</td>
</tr>
<tr>
<td>TM sentence 1, Prüfer match</td>
<td>Sentence: Merger falls within scope of Regulation but raises no serious concerns</td>
<td>0.81</td>
</tr>
<tr>
<td>TM sentence 2, Prüfer match</td>
<td>Sentence: The merger does fall within the scope of the Regulation and does raise serious doubts as to its compatibility with the common market</td>
<td>0.82</td>
</tr>
</tbody>
</table>

Table 3: Two TM sentences with different ranking according to matching method

<table>
<thead>
<tr>
<th>Query sentence, translation</th>
<th>TER</th>
</tr>
</thead>
<tbody>
<tr>
<td>TM sentence 1, translation</td>
<td>De fusie valt binnen het toepassingsgebied van de verordening en er bestaat ernstige twijfel over de verenigbaarheid ervan met de gemeenschappelijke markt</td>
</tr>
<tr>
<td>TM sentence 2, translation</td>
<td>De fusie valt weliswaar binnen het toepassingsgebied van de verordening, maar er bestaat geen ernstige twijfel</td>
</tr>
</tbody>
</table>

Table 4: Editing effort needed to correct the translation of the two TM sentences in Table 3

Besides our tests with Prüfer matching, we also performed match tests with tree alignment distance. As the former type of matching is much faster than the latter, the former may act as a filter before tree alignment distance is applied. However, we noted tree alignment distance is prohibitively slow even when only a small amount of parses is involved. Therefore, we need to optimize our implementation of tree alignment distance, which currently performs an exhaustive
search for the optimal alignment. A drastic pruning of possible alignment paths will have to take place in order to make application of the method viable, especially when two parses with a large number of nodes are compared. Apart from the problem of speed, we also noted that low parse quality can hamper syntax-based fuzzy matching. Similar sentences may happen to be parsed in very different ways. This problem may be diminished through the use of parse forests, which describe a multitude of parses for a single sentence.

5. Conclusions and future research

We have presented a methodology for testing whether the exploitation of translation memories using syntax-based fuzzy matching can improve the recall (by increasing the number of matches) and the usability of TM matches. We compare a sentence to translate, the query sentence, with TM sentences through standard Levenshtein distance on the one hand (the baseline) and through the parse trees of the sentences on the other. The aspect of matching speed is approached through the use of an index (suffix array) and through the derivation of flat sequences from parse trees, Prüfer sequences, which allow to apply string-based methods. Our tests are still in an exploratory phase. First results indicate that replacing baseline matches by matches based on Prüfer sequences slightly improve the TER of translation suggestions for low fuzzy match score ranges. This shows a potential for improving usability of matches and increasing recall of standard Levenshtein matching. We also performed tests with tree alignment distance, which acts directly on trees, but our current implementation of this matching method is still prohibitively slow.

We will pursue tests by applying matching methods using different settings and optimizing them, by making use of additional, larger translation memories in order to increase the likelihood of finding useful fuzzy matches, and by investigating the use of parse forests (Mi et al. 2008) to overcome parsing errors during matching. We will also create an additional type of translation suggestion, which does not consist of the full translation of a sentence in TM, but from the translation of the matching parts only; this will be realized through the use of word alignment from a statistical MT system.

References


Integrating Machine Translation (MT) in the Higher Education of Translators and Technical Writers

Marion Wittkowsky
Flensburg University of Applied Sciences

ABSTRACT

This paper describes how MT is integrated into a course project for translation and technical writing students. The course project is based on the idea of combining controlled language and a pre-editing step in order to achieve an effective way to prepare contemporary technical documentation for rule-based machine translation (RBMT). I will explain what I mean by “contemporary” within the context of technical documentation and why this attribute plays an important role within the decision-making process to integrate CL, pre-editing and MT in the course project, which also includes practical exercises for the students. In addition, the reason why RBMT is the MT method chosen within the context of multilingual text production is explained.

1. Introduction

Over the last decade the knowledge and skills required for a career as a professional translator or a technical writer working within the environment of multilingual technical communication have changed drastically. Translators are not only working as translators; they have to manage projects, define the language and terminology to be used, and they have to have knowledge of language, translation and project management software and systems. Today, technical writers not only have to be at least bilingual experts in writing well-structured technical texts, they also have to create style guides and need to know how to optimize information distribution and knowledge transfer. The skills and knowledge that today's students of technical writing and translation acquire, can be seen as wide-ranging knowledge within the field of international technical communication, i.e. they become international technical communication experts. In this context, I am referring to those students who have been educated in this occupational area over the last 10 years. “Contemporary” technical documentation, no matter whether it is of informative, instructive or descriptive content, is produced to a great extent by technical authors who did not study international technical communication with all its various
aspects. A very high number\(^1\) of persons writing or generating single-language technical contents today are not trained technical writers, but rather engineers or experts in a certain technical area, so-called “career changers”. Why should this be an important factor when it comes to translation processes? Writing technical texts without considering the fact that a subsequent translation process only works properly if the source text follows certain rules, in most cases results in texts that are of course comprehensible and readable yet not necessarily easy to translate, neither for humans with or without for CAT nor for MT systems. As this probably affects more than 70% of the technical documentation that is currently being written, there has to be an approach on how to interact within the translation process so that the source texts meet the language requirements of the different translation resources to be used afterwards.

2. Content of the course project

Starting with the properties of natural languages (NL), the students are given an introduction into the RBMT-system to understand how the system works, so that they will be able to experiment with the use of controlled language (CL) as a pre-edit step prior to RBMT.

2.1. Natural languages

Natural languages have certain properties that may cause misunderstandings among the recipients regarding the intended message sent by the sender. One of the main problems with natural languages is ambiguities. Lexical ambiguities, for example, can easily be solved by defining unique terms to be used in a specific subject area. Other types of ambiguities such as syntactical, semantic or contextual ambiguities need to be revised and changed. In addition to language variants that should be avoided, various language styles may lead to a lack of understanding for the target audience.

2.2. The MT system

This section discusses why RBMT is used and why this RBMT system is Lucy LT.

Statistical machine translation (SMT) systems work with huge databases to look up whether the desired contents can be found anywhere in the databases. Nowadays, because they have been fed for a long time, SMT systems do find translations for sentences like: German “Die nächste Schraube an vorderster Stelle einsetzen.” You get a translation that reads: “Insert the screw next to the forefront”.

Why did I choose this example? I didn’t choose it, but simply entered a meaningless, yet grammatically correct sentence in Google translate\(^2\). The system took all the words, looked them up and the result is the above.

What does “nächste” in German mean? In different context it can have the meaning of 1. next; 2. next to or 3. very near.

\(^1\) More than 70% of the employees working in technical documentation departments have no formal educational training in technical writing. ([http://www.tekom.de/fileadmin/Dokumente/de/2013-08_Branchenkennzahlen_2013_DE.pdf](http://www.tekom.de/fileadmin/Dokumente/de/2013-08_Branchenkennzahlen_2013_DE.pdf))

\(^2\) Google translate is a service by Google Inc. providing translations for written text in many language combinations.
However, the system translated “nächste” with “next to” (the forefront). From a grammatical point of view, it is not even possible to translate it this way, because “nächste” in German with the meaning “next to” requires an indirect object, which does not exist in the sentence.

This is only one example that shows why SMT is not useful to obtain proper translation results. There are no linguistic rules programmed in SMT systems – that is the reason why we have chosen to use RBMT for our course project.

The MT system used in this course project is Lucy LT. Lucy LT is a rule-based MT system with a long history. The system’s origin is METAL, a system that had been developed by the Linguistic Research Center of the University of Texas, Austin. “It was based on the interlingua philosophy and was also influenced by the then current theories of Transformational Grammar. A system for German-English translation based on these ideas was implemented during the years 1972-1975.” (Whitelock, Kilby 1995: 171)

For our purpose to work with the language pair German – English the METAL system delivered the best working base; it was used until the end of the millennium at the Flensburg University of Applied Sciences. Lucy LT is the successor of METAL and the decision was made some years ago to continue using this system for course work with the students. A very helpful feature within the program is the so-called Scratchpad, where one can translate e.g. single sentences and display the analysis and transfer tree of the source and target sentence to understand how the translation process is performed by the system with regard to syntax rules, etc.

2.3. Controlled language

Technical documentation today often requires that authors need to be questioned regarding the original intended meaning of their texts. In order to find a way to obtain a certain quality level, efforts have been made to define rules on how to use language, avoid ambiguities and achieve a high quality level both in the source and in the target texts. (cf. Babych, Hartley, & Sharoff, 2009) There are guidelines on how to write rule-based texts in German that have been published by associations like tekom or one can use software like Acrolinx to check the correct application of controlled language.

Technical writers are asked to apply CL to avoid ambiguities, inconsistencies, redundancies, and other problems that may cause problems in understanding. However, we are currently very far away from the point that the majority of technical writers apply these rules, as mentioned above. The future experts of international technical communication need to know what results can be achieved by applying CL, not only for the original texts but also for succeeding translations.

3. Course project scenario

Let me first introduce the target audience, i.e. the students who take part in this course. These students who are studying in this master programme International Technical
Communication have quite varying backgrounds regarding their general knowledge of multilingual text production and translation processes as well as in their professional experience level – we have translators as well as technical writers, and the so-called career changers as the programme can also be studied part-time by people who are already working, either as employees or freelancers. What the students are supposed to gain from this course is not only the knowledge about how MT systems function but also to gain a deeper knowledge of why certain properties of NL can be a problem in the source and in the target texts resp. for the source and target text audience, and how the use of CL can help solve those problems.

Methodology

The course guides the students through the following steps:

a. After an introduction into MT systems in general and a short briefing about NL and CL theories, the students receive instructive and descriptive excerpts from a contemporary technical text\(^5\) (service manual or similar) in order to execute a first machine translation using Lucy LT. The subject area the technical text is taken from is not of special interest because the dictionaries provide a broad variety of technical subject areas.

b. Then the first actual work done by the students is entering entries into the dictionary either by replacing or adding wrong or missing terminology within the monolingual, bilingual and transfer lexicons. Here the students enter the canonical forms, make simple lexical additions to the new terminology, e.g. by entering the word stem, affixes, etc. or specifying the lexical category as for example AST, NST or VST for adjective, noun or verb stem. Because of the fact that we do not educate computer linguists, the work with the dictionaries should not superimpose the actual subject matters of this course.

c. After having finished the dictionary work, the students are instructed to perform another translation and to view and analyze the results. In many cases the resulting texts at this point are not comprehensible for persons who do not have the same context as the students. Some sentences are perfect and some are completely incomprehensible. The Scratchpad in Lucy allows the students to translate small segments of the source text and to display the analysis tree and the transfer tree where the system behavior can be followed.

d. Now, the main experimenting part begins. As mentioned above, the quality level of the translated contemporary technical texts reached so far is often poor so that the students now start to apply CL rules combining style and grammar to the source texts in order to achieve higher quality target texts. They shorten sentences, simplify the syntax, find solutions for various kinds of syntactical ambiguities, try out the system behavior when using transitive and intransitive verbs, avoid complex phrases and tenses, eliminate redundancies, change the grammatical gender, etc. (tekom: 2011). This experimenting step is very important for the following reasons:

\(^5\) In the first part of the course, it is a German text, in the second part the source text is English.
(1) The students reduce their distance to the software in a positive way. Observations over the last years working with students have shown that they normally have great respect for computers and software.

(2) They learn that they are able to make changes and to interact with a computer.

(3) The students see that changing things may bring about better results.

4. The course project in practice

In this section I would like to show examples of the practical work performed by the students. The language combination relevant in this course project is German - English. As mentioned before, the students focus on modifying a given source text using CL rules in order to improve the translatability of source texts. The students work with the RBMT system Lucy LT, however this work could also be done using any other RBMT system.

Let us start with step 1, the introduction to MT systems, a quick overview of traps of natural language and the start guide for using CL. This introduction is important to ensure that all students have the same knowledge level when starting the actual project.

The students get a copy of the original technical text file, we look at the contents and answer possible questions regarding e.g. the subject matter. After having prepared the MT system for the translation such as specifying the subject area, or making some basic language pair settings, or define display options, the students then let the MT system translate the original source text without modifying it in advance. The sentence used in the following section has been taken from an automotive manual but from a linguistic point of view it could be part of any contemporary technical documentation and serves as an example only.

Example source sentence (DE):

*Der Gangwechsel wird mit den Schaltwippen oder dem Wählhebel vollzogen.*

(Meaning: You have two possibilities to change gears: use the paddles or the gear selector.)

1st translation (EN) by Lucy LT:

*The gear trimmer is carried out with the switching rocker plates or the choosing lever.*

We find that we have three apparently problematic terminology components within this text:

- Gangwechsel ≠ gear trimmer
- Schaltwippen ≠ switching rocker plates
- Wählhebel ≠ choosing lever

The three terms (combinations of two words each) are new to the system, which proposes the usage of two independent target terms as a combination of single words that have been stored in the system.

The result of the first translation of the correct and understandable source sentence makes no sense.
Step 2 involves entering the terminology into the three dictionaries, two monolingual lexicons and one transfer lexicon. In these dictionaries the students enter the canonical forms as well as basic grammatical information like the lexical category (in this case NST for noun stem) for every term. For our sentence, in detail this means the following:

Schaltwippen is a plural word so that the singular word entered is Schaltwippe with paddle as target.

Wählhebel can be entered into the dictionary as it is with gear selector as target.

The third term that has not been translated properly by the system is Gangwechsel, however it should not be entered into the system, because this word as compared to the two words above is not a concrete but an abstract compound noun made up of Gang and wechseln. In English the corresponding words are gear and shift. This problem has to be solved in a later step.

At the end of this step, the students have added or replaced two terms in all three dictionaries.

In the 3\textsuperscript{rd} step the students translate the same source text again:

Der Gangwechsel wird mit den Schaltwippen oder dem Wählhebel vollzogen.

The system applies the newly entered terminology. The output is as follows:

The gear trimmer is carried out with the paddles or the gear selector.

We now have the correct terminology for the two changed words, yet we still have the problem with the word gear trimmer in English.

At this point the experimenting step starts. The students perform the following substeps as often as they find possibilities to make amendments to the source text without changing its intended meaning:

1) They enter the changed source text sentence by sentence again and again into the source text window of Scratchpad.

2) They have Lucy LT translate the text.

3) When the sentence\textsuperscript{6} has been translated, the analysis tree and the transfer tree may be used as support, when the students analyze the system behavior and their source and target texts. The students can write down all important observations, copy useful and less useful examples of the source and the target texts as well as the trees into their working files so that they have something to work with and to refer to in their course reports.

The following section shows how our example sentence could be changed, i.e. pre-edited and what the results of the translations are. I speak about pre-editing here and not about applying CL, because the students are not asked to apply language rules that follow CL only but

\textsuperscript{6} It is also possible to enter more than one sentence into Scratchpad, yet the number of characters is restricted.
they may also change the source to that extent that the grammar is still correct and the original meaning is retained. For every source and target sentence listed below you can read some comments referring to the meaning, language and quality that have been made during the translation procedure.

Since it is the third change + translation, I name it correspondingly:

3rd source

_Vollziehen Sie den Wechsel der Gänge, indem Sie die Schaltwippen oder den Wählhebel verwenden._

The German term _Gangwechsel_ needs to be separated into two words as mentioned above, yet we still have two nouns to be able to keep the original verb _vollzogen_ (canonical form: vollziehen). In addition, the passive voice has been changed to active in such a way that the user is addressed directly using _Sie_ (EN you). Regarding the quality, the sentence does not follow the rules of CL, it is very complicated.

3rd target

_Carry out the trimmer of the gears when you use the paddles or the gear selector._

The separation of _Gangwechsel_ into _Wechseln der Gänge_, did not improve the translation at all. Lucy still translates this part of the sentence with “trimmer of the gears”. This is a mistake.

4th source

_Zum Wechseln der Gänge verwenden Sie die Schaltwippen oder den Wählhebel._

Here, the problematic word _Gangwechsel_ has again been separated, yet we have a nominalized verb _Wechseln_ and the plural form of _Gang->Gänge_, which results in _Wechseln der Gänge_, but which means the same thing. We avoid the verb _vollziehen_ because it is not the correct style in this context and use the verb _verwenden_ instead.

4th target

_You use the paddles or the gear selector to the change of the gears._

The resulting sentence is not correct regarding the grammar, and the expression _change_ could be misunderstood.

5th source

_Sie verwenden die Schaltwippen oder den Wählhebel zum Wechseln der Gänge._

We have the same content and the same words as in example 4, however, the sentence structure has been changed. The sentence does not follow the rules of CL, i.e. one should be informed about the result before taking an action (tekom: 2011, p. 55).
5th target

You use the paddles or the gear selector to the change of the gears.

Very interesting here, the target sentence is the same as in example 4 and could still be misunderstood.

6th source

Wechseln Sie die Gänge, indem Sie die Schaltwippen oder den Wählhebel verwenden

The sentence in German is correct and written by applying CL rules. The noun Wechseln has been changed to Wechseln Sie (verb+addressing the user).

6th target

Change the gears when you use the paddles or the gear selector.

The system response is correct by changing the noun change to a verb, yet the conjunction when is wrong – it should be by + gerund.

7th source

Sie können mit den Schaltwippen oder mit dem Wählhebel die Gänge wechseln.

In this example we again do not have the right order regarding action and objective and we have a modal verb that should be avoided (tekom: 2011, p.62).

7th target

You can change the gears with the paddles or with the gear selector.

Apart from the modal verb, the fact that with should be replaced with using and the second appearance could be deleted; the sentence is of an acceptable quality.

8th source

Um die Gänge zu wechseln, die Schaltwippen oder den Wählhebel verwenden.

We have a correct German sentence following CL rules.

8th target

In order to change the gears, use the paddles or the gear selector.

The resulting target sentence is perfect.

9th source

Gänge wechseln mit Schaltwippen oder Wählhebel.
Last but not least a very short sentence that summarizes the information and could be used as a title for purely informative purposes.

9th target

*Change gears with paddle or gear selector.*

The translation depicts the same as the German source.

There are at least 10 other possibilities to write the intended content in other words. These examples represent an excerpt only.

5. Conclusion

The course project presented is intended to make the students aware of other translation scenarios than human translation or computer-aided translation CAT. It is supposed to introduce the students into machine translation, let them form their own informed opinion about MT by showing them how the system behaves and how an advanced human interaction may increase the quality of the outcome. Since the students are to be regarded as experts in international technical communication, they should not only have an interest in the translation process but also in the process of generating source texts of a certain quality level – depending on various factors like target audience, text type etc.

From my point of view it is a valuable side-effect that the students generate different quality levels when experimenting with the texts, so that future research in this area could probably bring up an approach to define different quality levels and to find better ways to generate multilingual text units from existing documentation on demand.

References


tekom (2011): *Leitlinie für regelbasiertes Schreiben - Deutsch für die Technische Kommunikation*, Gesellschaft f. technische Kommunikation e.V., tekom

Top-down or bottom-up: what do industry approaches to translation quality mean for effective integration of standards and tools?

Joanna Drugan
University of East Anglia

ABSTRACT

The diverse approaches to translation quality in the industry can be grouped in two broad camps: top-down and bottom-up. The author has recently published a decade-long study of the language services (Quality in Professional Translation, Bloomsbury, 2013). Research for the study covered translation providers from individual freelance translators working at home, to large-scale institutions including the European Union Directorate-General for Translation, commercial translation companies and divisions, and not-for-profit translation groups.

Within the two broad 'top-down' and 'bottom-up' camps, a range of further sub-models was identified and catalogued (e.g. ‘minimalist’ or ‘experience-dependent’). The shared distinctive features of each sub-group were described, with a particular focus on their use of technologies.

These different approaches have significant implications for, first, the integration of industry standards on quality, and, second, the efficient harnessing of technology throughout the translation workflow.

This contribution explains the range of industry approaches to translation quality then asks how these map on to successful integration of standards, and features of the leading tools which are designed to support or enhance quality.

Are standards and technologies inevitably experienced as an imposition by translators and others involved in the translation process? Significantly, no straightforward link was found between a ‘top-down’ or ‘bottom-up’ approach to assessing or improving translation quality and effective use of tools or standards. Instead, positive practice was identified across a range of approaches.

The discussion outlines some painless ways these developments are being channelled to improve quality, or more frequently, to maintain it while meeting tighter deadlines. Some models existed beyond, or were partially integrated in, ‘professional’ translation (e.g. pro bono translators, and volunteer Open Source localizers).
What lessons can we learn from enthusiasts in such communities, who sometimes adopt or create approaches voluntarily?

1. Introduction

Translation quality matters in the industry, and for different reasons than in translation studies. Providers have to measure, compare and guarantee quality throughout the translation process. Before winning contracts, they must convince clients they can deliver translations more reliably or efficiently than rivals. During translation, feedback on aspects of quality might be expected. Post-project, decisions affecting quality must be justified or repaired at no additional cost. At the strategic level, quality is important when planning, allocating resources, designing training and support, ensuring return on investment (ROI), or measuring the impact of change. The industry is driven to maintain quality while reducing costs or deadlines, or to improve quality, usually without increasing costs or extending the time needed for translation.

None of this is new, but there has been a fresh turn to translation quality, due to the combined effects of the Information Age, changes in the types of translation needed, rising demand, and growth in opportunities for international trade when home markets in many regions are stagnating or declining. Translation happens faster today for more - and more diverse - clients, into an increasing number of target languages, across more technical formats, using more complex tools, and is increasingly subject to international standards. Source texts are also more complex, perhaps co-authored by teams working in a shared language which is not their mother tongue, and regularly updated.

2. Research methods and questions

Research on translation quality in such real-world contexts has barely begun. Researchers have devised models to assess quality then tested them on short - often literary - texts in a single language pair/direction (Al-Qinai, 2000; House, 1997); or measured the effects of a single intervention, such as using a translation memory (TM), in artificial settings - usually a small group of student subjects translating a short text in a single language pair/direction in a university lab (Bowker, 2005; Teixeira, 2011); or focused on post-editing machine translation (MT) (Fiederer & O'Brien, 2009); or assessing quality in student translations (Delizée, 2011). In contrast, the research reported here observed language service providers (LSPs) of all sizes over a decade. Challenges for this approach are significant. How can we study translation quality across dozens of language pairs, in different specialisations, for diverse clients, produced by thousands of translators to varying deadlines, using a range of tools and resources? Methods were tested and revised during the research, including the use of questionnaires, workshadowing, interviews, and Think-Aloud Protocols (TAPs), drawing on a range of disciplines. A modified form of Grounded Theory (Glaser & Strauss, 1967) was then used to describe common features of real-world approaches. Initial conclusions were tested with some providers before arriving at the published classification (Drugan, 2013).

Industry approaches to translation quality can be grouped in two broad camps: top-down and bottom-up. I identified eight further sub-models within these two camps, and described their distinctive features. The research initially focused on classifying the range of industry approaches, and mapping their key features. This article digs deeper on two questions. First,
what does a top-down or bottom-up approach to translation quality mean for use of tools? Second, what does a top-down or bottom-up approach to translation quality mean for integration of standards? These questions are related to one another, and to translation quality. Tools and standards are designed and adopted to guarantee, measure or improve quality; or, at least, to maintain quality levels while producing translations more efficiently. Their likelihood of success may be linked to how translation quality is understood in the different models.

3. Top-down and bottom-up models

What is meant by ‘top-down’ and ‘bottom-up’ in practice? Top-down approaches are hierarchical, driven from the top. They harness translation expertise and aim to manage or control quality levels. Associations with this group of approaches are traditional, conservative, authoritarian, or paternalistic. Bottom-up approaches, in contrast, are led by users or suppliers themselves. They draw on target-language ability and/or product expertise, combined with end-user feedback loops, rather than translation competence. Associations with this group of approaches are novel, radical, egalitarian, or democratic. In the top-down category, I identified five sub-models: Maximalist, Client-driven, Experience-dependent, Content-dependent, and Purpose-dependent. In the bottom-up category, I found three sub-models: Minimalist, Crowdsourced, and User-driven. For each of these eight sub-models, a relatively ‘pure’ form was outlined in detail (Ibid.: 127-173), based on a real provider which hosted one or more research visits, sometimes over several years. I describe the main features of each approach, including details of how suppliers are recruited and assigned to projects, any pre-translation tasks, tools and resources used, quality checks during the project lifetime, post-translation checks, return of work, post-project review and ongoing planning. As well as the ‘pure’ forms of sub-model, a given project or provider might combine aspects from more than one sub-model in a hybrid approach. The Organisation for Economic Co-operation and Development (OECD) uses a model combining aspects of Content-, Experience- and Purpose-dependent sub-models, for example (Prioux & Rochard, 2007). Similarly, providers operated various models or hybrids for different translation projects.

This discussion is based on broad definitions of translation quality and professional translation. The inclusion of some approaches, such as user-generated translation, might be questioned, as these generally do not involve professionals. A broad understanding of LSPs was chosen to include emerging bottom-up approaches which are increasingly filling gaps in professional provision. Demand for translation is not met by the industry, so it seemed important to capture what was happening in these contexts too.

4. Top-down and bottom-up models and tools

Is there a relationship between top-down or bottom-up approaches and integration of tools? Do top-down models impose use of certain tools, while bottom-up models leave users to decide, for instance? A review of a representative range of providers within each of the eight sub-models demonstrated significant diversity. As Cronin points out (2003), translation is tools. Without tool use, we would be discussing interpreting. Translation is based on a long history of harnessing tools, whether parchment, quill and early dictionaries, or current combinations of terminology management, MT, TM, localisation tools and add-ons, or potential ‘personalized’ MT
environments integrating predictive text and adaptation to individual users’ styles and preferences (Green et al., 2014). Given this rich history and today's diverse industry, a range of approaches to tools was predictable.

Top-down sub-models were first reviewed, concentrating on requirements regarding the use of tools. Some Maximalist approaches mandated use of given tools and resources (e.g. imposing locked segments in TM content). Even in the most extreme Maximalist settings, however, tool use was not required for all jobs. In most top-down settings, users decided whether and when to use tools, though they were offered ‘hidden’ or unprompted resources and support, with investment at the design stage (creating and supporting highly customised versions of tools). Translators were encouraged to harness useful features by default suggestions, via concordance features in editing interfaces or colour coding of source texts to highlight potential matches in previous translations, even without TMs.

The principal tools used in top-down models were for terminology, TM (including localisation), corpora\(^1\), bespoke MT, and automated quality assurance (QA). These tools were used in heterogeneous combinations, alongside personal resources such as specialised glossaries. A common feature of top-down models was the quality ‘gatekeeper’: where translators suggested new content, gatekeeper authorisation was needed before incorporation in databases. In-house and external suppliers were separated, so content was only approved for databases if authored by in-house translators or freelance suppliers who met imposed quality ‘standards’ (e.g. translation for the organisation for several years). This aspect of top-down approaches had perverse effects for consistency and quality, because excellent content was excluded. Top-down approaches to tools usually accorded significance to training in ‘appropriate’ tool use.

Bottom-up sub-models were next reviewed. Use of tools was occasionally mandated for specific formats (e.g. in Free and Open Source Software (FOSS) localisation projects, users had to select one of four interfaces). Because many such contexts depend on tools to exist (e.g. free online machine translation (FOMT) to generate website translations), tool use was effectively imposed, but selection of any particular tool or workflow remained user-driven. Bottom-up sub-models shared an emphasis on offering tools, resources and support then letting translators decide whether to adopt them. The community approach meant additional support for novice users and informal training resources. Instead of emphasising initial training and mentoring, bottom-up models archived records of previous translation issues then made these easily searchable by users. Discussion boards, wikis, YouTube videos and blogs provided peer support for collaborative working. Unlike top-down models, where providers must reassure clients as to translation quality, bottom-up participants were encouraged to admit weaknesses so others could help. Feedback from motivated end-users meant translation was never viewed as complete. Different tools were used: terminology and TM tools were widely available, but Open Source (OS) or customised collaborative platforms and editors were the norm, rather than proprietary tools. Informal corpora were widely integrated, particularly through quick search features to identify similar previously translated content. FOMT was harnessed as a matter of

\(^1\) Usually informal collections of texts, with search buttons giving quick access to previous translations, alignments or related source documents.
course. No bespoke automated QA tools were observed, though translation environments included QA features such as terminology verification.

Integration of tools relied on some top-down approaches. User behaviour was directed via ‘rules for translators’, style guides, or getting started guides stipulating tools and workflow. Localisation leads, super-users or experienced contributors performed management roles akin to those of quality gatekeepers in top-down models: ranking translations, deciding when translators disagreed, or signing off approved versions of translations before release. Contributors might be unaware of these features.

How might common top-down attitudes to tools affect translation quality? One strength lies in their ability to impose aspects which prove useful for quality, while maintaining scalability. Best practice can be observed and disseminated efficiently. Understanding of clients and the industry has benefits for quality and continuity, as do staff retention, effective training, and high levels of experience. Top-down approaches dominated in large providers with extensive customisation: dedicated IT support staff managed bespoke solutions, interoperability and transfer of resources across tools and formats, and fixed bugs, so translators could concentrate on content. Top-down approaches mean problems can be caught rather than delaying delivery - the ultimate quality failure being to miss a deadline.

During the research, larger providers began recruiting to new roles in translation quality management. This enables and emphasises ongoing review of quality processes. Gatekeepers for translation content can make similarly beneficial contributions: checks by experienced staff meant only high quality material was re-used. Large clients with ongoing translation needs preferred top-down models because problems can be addressed then prevented for future projects. They also appreciated guaranteed levels of quality across multiple languages. For multilingual contexts, language leads could discuss shared projects and share knowledge or innovation. Clients saw quality as linked to other aspects of translation provision, notably confidentiality and ownership of resources. Top-down providers were able to guarantee such aspects.

Some of these features are absent in bottom-up models. Imposing processes or tools was rare or impossible. How then did common bottom-up approaches to tools affect quality? Rather than imposing certain features, the bottom-up approaches could instead harness an unusual degree of provider enthusiasm and commitment, both to design and use tools. Community support and provider knowledge of products for translation were key. Bottom-up models integrated end-user feedback to improve translation quality via buttons, voting features or discussion boards. The chief drawback for quality in this method lay in the high turnover of volunteers, making ongoing quality planning and review challenging. Most bottom-up providers had some limited management behind the scenes for this reason.

Ethical aspects of translation quality were more apparent in bottom-up models. Some translated content with high quality requirements is not, and is unlikely to be, funded. For instance, Wikipedia articles on medical topics are the most used online healthcare resource\(^2\) in

---

the world and are widely accessed in translation, but translation quality depends on volunteers. In contrast, some amply funded translation content is ephemeral, or commissioned but never used. Allocation of high quality translation resources is not linked to where the need for high quality is greatest. Conversely, bottom-up approaches enabled translation that would otherwise never happen, sometimes with wider societal and ethical benefits. An illustration here is the localisation of OS resources for Lao, with integration of translation tools and resources enabling important progress in standardisation of the complex script and the language itself.

5. Top-down and bottom-up models and standards

Standards are linked to tools and translation quality. The term refers both to technical standards to enable translation (e.g. Unicode) and to quality management standards applied to translation processes (e.g. the ISO 9001:2008 standard). The translation industry is unexceptional in its embrace of the latter sort of standards, as the ISO series relates to process quality management across industries. Did top-down and bottom-up approaches predict a different attitude to such standards? The sub-models in each camp were reviewed with this question in mind.

For top-down approaches, adoption of standards was observed mainly in the Maximalist, Client-Driven and Content-Dependent models; among Multi-language vendors (MLVs) rather than smaller providers, with some exceptions in regulated sectors such as automotive translation; and where ROI on substantial long-term contracts justified the effort and investment to devise and refine the requisite policies and procedures. A ‘bandwagon’ effect was observed: once standards were adopted by some MLVs in a sector or region, there was a need for others to demonstrate compliance or ability to meet the same standards. Suppliers who qualified for certification to quality standards emphasised this prominently in sales and marketing materials, making clear links between certification and a commitment to translation quality in general. Certification to quality management standards does not mean that all jobs are certified, however. Certification is for companies, rather than projects, and the additional resources implicated in complying with the standards make them prohibitive for most projects. This may be misunderstood by clients, with associated potential for misplaced confidence in such apparent badges of quality. Beyond ISO or related national standards for translation quality, a more common feature across the top-down models was integration of informal industry or sector norms relating to translation quality processes. These were poorly understood by both providers and clients. Freelance translators were often unclear whether their translations would be revised in full or via sampling after return to agencies, for instance, and clients were clearly surprised to learn during interviews that their texts might not be edited and proofread from start to finish. Some top-down models did place greater emphasis on their own internal quality standards and processes, notably in the Maximalist models.

Standards for translation quality processes were conspicuous by their absence from bottom-up approaches. Given the emergence of such approaches for technical or ICT settings, technical standards dominated rather than those such as the ISO or CEN standards. These latter standards

---

were not found in any in the bottom-up models. Instead, informal quality ‘standards’ were widespread, relying on style or quick-start guides, ‘rules for translation’, archives, community guidelines and norms. This absence of formal quality standards does not straightforwardly predict low levels of eventual translation quality, however. Alternative methods filled the gap in bottom-up approaches, bringing some novel benefits for quality.

Nor was there a straightforward positive relationship between top-down adoption of standards and effects for translation quality. Because standards for quality management focus on processes, compliance and certification might not actually guarantee high quality translation products. Indeed, the costly and lengthy certification process itself, and subsequent administration of the standard, tend to divert resources to process management instead of translation. Unless clients value the standard and pay extra for certified providers, the proportion of project funds spent on translation can be cut to support the processes. The nature of the standard protocols and processes make compliance inefficient and inflexible, unless it is sensitively managed. Once processes are defined and agreed, any subsequent revision is costly, cumbersome and again diverts resources from translation. This means that standards affect providers’ ability to respond nimbly, for instance to the rapid pace of change in society or technology.

A risk for quality of such standards is that they encourage providers and clients to focus on what can be measured and recorded, rather than on important aspects of quality which are challenging to define or guarantee. The National Standard of the People's Republic of China GB/T 19363.1-2003 stipulates a dress code for receptionists working in the industry, which is unlikely to have any direct bearing on the quality of the translation products. A more important absence from quality standards is the end-user of translations. Standards are based on processes agreed between clients and providers, so cut a significant stakeholder - the eventual beneficiary or user of translation - from the discussion, unless providers choose to integrate user feedback as a standard workflow process. This was not observed in any visit to certified providers or in their documentation relating to standards.

In contrast, bottom-up approaches emphasised user input to assess and improve translation quality. Bottom-up approaches operated outside formal translation quality management standards, but (perhaps as a result?) relied on an unusual degree of user feedback to refine translations. End-users were integrated widely in evaluating translation quality and improving it, via buttons, wikis, voting, discussion boards, blogs and other interactive features designed into the tools. This process was also viewed in bottom-up models as an evolving one. In this, it was in evident contrast to the sort of in-country review (ICR) sometimes integrated in top-down models near the end of the translation workflow. In the bottom-up models, translations continued to evolve as contexts for their use evolved or new users brought different quality expectations.

6. Conclusion: Findings and lessons

No straightforward pattern was found between top-down or bottom-up approaches to translation quality and respective attitudes to tools or standards. Instead, positive practice was identified across the range of approaches. Top-down approaches were more able to impose good practice; harness tools effectively through training, customisation and support; invest
Translating and The Computer 36

strongly in resources; retain and reward staff expertise; ensure client confidentiality and confidence; and scale processes to suit different contexts, including certification to internationally recognised standards, perhaps with benefits for the status of the profession. Bottom-up approaches were more able to draw on emerging technological features to enhance translation quality in imaginative ways; adapt quickly to changing contexts for translation, without having to confront institutional barriers or standards; harness positive features of the top-down models behind the scenes; and draw on providers’ enthusiasm and technical skills, and end-user feedback.

Negative impacts for translation quality were found in both models too. The top-down approaches’ reliance on gatekeepers and mandated uses of some tools or workflows meant high quality content was excluded from resources and lost for future re-use. Compliance with standards had perverse effects for translation quality. One conclusion of this study is that standards and tools interact with each other in the real world, with as yet unmeasured effects for translation quality. Top-down models’ integration of standards for quality management can actually impede nimble reactions to evolving translation contexts, lead to a ‘one-size-fits-all’ approach, and block early adoption of new technical features. Standards can in this view be seen as potentially in opposition to effective use of tools, and indeed translation quality itself.

Further work is needed. Significant sectors and providers remain unmapped, notably some language pairs and regions (South America, Eastern Europe), specialisations (literary translation, game localisation) and translation providers working with direct clients at the high end of the profession. Problems remain in identifying methodologies to research the industry in the workplace rather than the lab. Encouragingly, increasing attention is being paid to this sort of approach across the piece: in translation studies (e.g. by the PACTE group, Barcelona, Spain; and the EXPERTRANS research group, Oslo, Norway); in the translation industry (e.g. by the W3C World Wide Web consortium); and via industry-academic cooperation (e.g. under the auspices of TAUS).

References


Getting the best out of a mixed bag

Terence Lewis
MITI

ABSTRACT

This paper discusses the development and implementation of an approach to the combination of Rule Based Machine Translation, Statistical Machine Translation and Translation Memory technologies. The machine translation system itself draws upon translation memories and both syntactically and statistically generated phrase tables, unresolved sentences being fed to a Rules Engine. The output of the process is a TMX file containing a varying mixture of TM-generated and MT-generated sentences. The author has designed this workflow using his own language engineering tools written in Java.

1. Introduction

There is broad agreement today that improvements in the fluency of machine translation output can be achieved by the use of approaches that harness human translations. This paper discusses the development and implementation of an approach to the combination of Rule Based Machine Translation, Statistical Machine Translation and Translation Memory technologies. This “multi-faceted approach” can be applied in a vendor and platform independent environment.

Early methods for the combination of machine translation and translation memory tools involved the use of an analysis made by translation memory software to produce an export file containing unknown segments which were then fed into a machine translation system. The results were subsequently imported into the translation memory software where they received an “MT” penalty. This technique has been superseded in practice by the introduction of MT plug-ins which are now available in the major commercial translation memory applications. Some professional translators use these plug-ins, in a “pre-translate” stage in preference to accepting fuzzy matches to produce draft translations which they then revise. In the automated translation work flows referred to above, the process is controlled through the translation memory application, and in applications like SDL Studio 2014 and memoQ the user can set which, if any, machine translation services are to be consulted.

The author has implemented an approach whereby the machine translation system itself is able to consult and draw upon translation memories and a statistical translation model as part of an automated translation process. This paper does not claim to describe a novel approach as the
literature contains many useful accounts of attempts to combine translation memory and machine translation, such as the paper by Koen & Senellart¹, and the very detailed account by Kanavos and Kartsaklis²of attempts to combine a variety of third-party translation tools on real translation projects. It describes a practical way of harnessing a number of different data resources which the author has found useful for handling major projects such as the one described further on in this paper. The author acknowledges that the methods discussed in this paper have been broadly applied in the recently launched MateCat project³.

The author is both a translator and a self-taught software developer – not a computational linguist – and this contribution is intended to be a personal account of experiences rather than a scientific paper. The approach to translation automation described is a practical one. For nearly two decades the author has worked as an independent provider of translation automation services and a language technology consultant, for Siemens Nederland and other Dutch companies and institutions. The translation memories utilised to deliver these services have been built up in decades of work as a translator and language service provider. They include sentences translated by other professional translators and post-edited machine translations. All these data have been combined into one large TMX file, which is a record of the author's knowledge and experience. These services are supplied within a private network and are not delivered "in the Cloud", something that is important to many of the author's industrial and government customers.

At TC21, the author reported on his attempts to combine the use of his growing translation memories with his early Dutch-English Machine Translation system.⁴ His clients in those days mostly wanted to receive fully formatted MS Word files. His initial efforts relied on the use of the Trados Translator's Workbench to create export files of unknown words which were machine-translated, with the translations being imported into the translation memory. Nowadays, however, corporate clients for his automatic translation services have themselves acquired commercial translation environments – commonly but not exclusively SDL Studio and memoQ, and want to receive the output of the MT system in a format such as TMX which can be imported directly into their translation memory software for post-editing. Significantly, the supply of a TMX file is billed at a lower rate than a fully formatted MS Word file!

Having worked extensively as a translator for companies in the chemical, transportation and telecommunication industries, the author has built up a wide-ranging “master translation memory”. As a developer of a machine translation application he has investigated various ways of exploiting these translation memories on the assumption that output derived from human translations will generally be more fit for purpose than that generated solely by the application of syntactic rules.

³See www.matecat.com
⁴The Best of Both Worlds – or will two mongrels ever make a pedigree?, T. Lewis, TC21, 1999.
2. Background

The approach described in this paper was a practical response to the challenges posed by a large translation automation project. The author was asked to translate 250,000 words in hundreds of small files for the HAN University of Applied Sciences in the Netherlands. The translations were needed quickly and costs had to be kept down. After comparing the results of various MT services offering Dutch-English translation, the university decided to avail itself of the author’s machine translation software. At the time of placing the order, the university had not even decided exactly how the machine translation output was going to be processed further. Working with veteran Localization Consultant, Lou Cremers, the author decided on a workflow which involved machine translating a TMX file in such a way that the client would receive a raw translation memory. The HAN translation office wanted to post-edit this machine translation output in a translation memory environment so that it could be brought to a standard fit for publication on the university’s website. The institution eventually decided to use memoQ and arranged for a post-editing team to be trained in its use. After terminological preparation of the project, the MT software produced a series of TMX files (TMX 1.1). The post-editors were given an opportunity to provide feedback which was incorporated into the translation memory being built as the project progressed and even led to the improvement of some of the MT rules.

When a second project for the HAN came along, the author knew that many of the sentences in this project already had translations in the translation memory built from the first project. The client wanted to handle the project in the same way as the first one and receive a TMX file. Wishing to carry out the new project completely via his own machine translation software rather than use third-party translation memory software, the author wrote the code to enable his machine translation engine to search this translation memory directly in order to enjoy the benefit of the translations contained in that memory within the automatic translation process.

3. The process

In terms of file handling, the process is a simple one. A TMX file is prepared in which the target elements, in this case `<tuv lang="EN-GB">`, initially contain the source text. The MT engine reads the input TMX file line by line. Only the content of the target element is of interest, everything else being written straight to the output buffer. The engine first sends off a query to the TranslationMemoryConsultor class. If the selected translation memory contains a 99% or 100% match, the corresponding translation is entered in the target element in place of the “source text”. After making any required minor editorial adjustments to the target segment, the engine moves on to the next segment. If the selected translation memory does not contain a suitable match, the segment is sent off to the internal translation server for further processing. The translated content is returned from the server segment by segment and also replaces the source text in the target element.

The output of the process is a TMX file containing a mixture of directly TM-generated and MT-generated sentences. The client’s translators can review this file for “sanity checking”, post-editing or full-blown revision, depending on the intended purpose of the automatic translation.

---

6This can be done using the Okapi Tools: [http://www.opentag.com/okapi](http://www.opentag.com/okapi)
The advantage of delivering a TMX file is that the post-editing work can be done in any commercial (or non-commercial) translation environment, in a dedicated TMX editor such as Olifant or even in a simple text editor on any platform. The translators at the author's main client for language technology services – Siemens Nederland N.V. - import the MT-generated TMX file directly into their translation memory in SDL Studio 2011. Other clients use different CAT tools.

As stated above, the MT engine goes through the submitted TMX file on a segment by segment basis, and if a translation memory contains a 100% or 99% match, the target language segment is inserted in the output TMX file and the engine then moves onto the next segment in the input file. The MT engine also recognises segments entirely in English (often the case in Dutch documents) and inserts these directly into the output TMX file.

In practice, parts of sentences in the input file will frequently match the content of the translation memory database at subsegment level. In translation memory terminology, these are the “fuzzy matches” for which users can determine an acceptability percentage (many translators set this threshold at 75-80%) in a “Pretranslate” run. The problem with these “fuzzy matches” is that a sentence in the translation memory can be displayed to the user as an 80% match or higher, even though it means the exact opposite of the source sentence. Figure 1 gives a simple example of this.

![Figure 1: The problem with fuzzy matches](image)

From the earliest translation memory environments attempts have been made to use colours or other devices to alert the user to the fact the displayed target segment is not a complete translation of the source segment. However, the author found that users of his Dutch-English automated translation service, who were paying for “unrevised machine translation”, were not prepared to receive translations potentially containing glaring inaccuracies. For this reason, it was decided to set a very high threshold (99-100%) for transferring segments automatically from the translation memory into the output file via the MT engine. On the other hand, his translation memories contained millions of potentially useful segments.

This awareness of possessing translation memories that didn't always tell the whole truth has led the author to investigate ways of storing potential subsegment matches in phrase tables which the machine translation engine can consult. The advantage of storing data at subsegment level is that the translations retrieved by the MT engine are NOT fuzzy matches but 100% matches for the part of the sentence to which they correspond. In practice, the author's MT engine consults two phrase tables to search for matches at subsegment level: one is created by the application of syntactic rules; the other is statistically derived.

The **Clean Data Repository** is created by decomposing segments in translation memories into meaningful fragments or subsegments. The term “clean” refers to the fact that the data have been checked by a human reviewer. They have been assembled by automatically aligning
meaningful subsegments of the segments contained in translation memories. This is done by “looping through” source and target sentences from complete sentence down to bigram level. Through the application of a series of syntactic rules, source and target segments are divided into noun phrases, prepositional phrases and verbal phrases, and short sentences are also retained as sentences. Many of these entries will correspond to Multiword Expressions\(^7\) (MWE’s). The human checking goes beyond grammar checks; it is made sure that terms are in-domain and the subsegment is in the right register. Entries in the Clean Data Repository in the author’s Dutch-English machine translation system look like these:

```xml
<trans-unit>
<source xml:lang="nl-NL">bij het uitvoeren van een beveiligingsfunctie</source>
<target xml:lang="en-GB"><mf>when a security function is performed</mf></target>
</trans-unit>

<trans-unit>
<source xml:lang="nl-NL">virtueel diagnostisch systeem</source>
<target xml:lang="en-GB"><n1>Virtual Diagnostic System</n1></target>
</trans-unit>

<trans-unit>
<source xml:lang="nl-NL">wens over te gaan tot</source>
<target xml:lang="en-GB"><vts>wishes to proceed to</vts></target>
</trans-unit>
```

*Figure 2: Examples of entries in Clean Data repository*

These data are stored on-disk in the form of an XLIFF file and are loaded into a Java data structure at run-time. No translation scores are involved as it is assumed that any entry in the Clean Data Repository, having been checked, is 100% correct, or has a probability of 1. The user can add project-specific data to this repository on the fly before a translation run. This is done by breaking down the source document into ngrams, which are presented with their frequency of occurrence as shown in Figure 3. The user can add the translations to the source segments in a text file and the program then converts the entries into the XLIFF format and adds them to the repository.

```
1 : zijn de volgende afspraken gemaakt welke verder in het document uitgebreider zullen worden toegelicht
1 : welke verder in het document uitgebreider zullen worden toegelicht
1 : in het kort zijn de volgende afspraken gemaakt

*Figure 3: Phrases to be translated and added to Repository*
```

The author’s program also allows post-edited TMX files to be “decomposed” so that the subsegments are added to the repository on the fly – this is particularly useful on large projects with multiple files as improvements made by post-editors/revisers can be incorporated into the data repository immediately, literally by the click of a mouse button. Figure 4 shows subsegments extracted from a TMX file ready to be added to the Clean Data Repository.

\(^7\)For a simple explanation of Multiword Expressions, see [http://en.wikipedia.org/wiki/Multiword_expression](http://en.wikipedia.org/wiki/Multiword_expression)
One of the problems of using a static phrase table is that in real language phrases are not set in stone but come to life in an engagement with other words. The Dutch verbal phrase “brengen op de hoogte”, literally translated as “bring someone on the height”, means “to inform”. We inform somebody about something so the software has to enable us to link “brengen” to “op de hoogte” while taking into account an intervening object. The author has written code to deal with these “gappy phrases”, so that a sentence like “hij bracht de raad op de hoogte” will be correctly translated as “he informed the board”. A special repository of such gappy phrases is built as a subset of the Clean Repository Data, using segments contained in the main repository, which has in turn been derived from the main translation memory.

If the MT engine fails to find matches for every subsegment in the Clean Data repository it may proceed to consult a “dirty data” repository. As the name suggests, the entries in this phrase table will not have been individually reviewed. Their accuracy is reliant upon the successful building of a statistical translation model using the tools provided in the Moses Statistical Translation Toolkit. The repository is built using the train-model.perl script. Existing translation memories are saved as TMX files and then divided into source and target language text files, which are subsequently tokenized and cleaned in the manner known to SMT practitioners. Train-model.perl is employed in the same way as for building the statistical translation model to be used by the Moses decoder in a statistical machine translation environment. The resulting phrase tables will be as general or specific as the TMX files from which they are derived.

The content of the file phrase-table.gz generated by the “Moses” training process is stored as byte code in a serialised Java data structure, or HashMap. The translation model is created during a training stage, i.e. in advance of its deployment. The data to be stored in this Java data structure are selected during the training operation on the basis of scores produced by train-model.perl, which means that the entire phrase table created by train-model.perl will not be stored in memory at run-time. At present, the generated scores are utilised in a fairly naive way - low-scoring phrase pairs are excluded from the phrase table loaded into memory. There is certainly scope for a more sophisticated use of these scores.

At run-time, the MT engine “decides”, sentence by sentence, on the basis of a number of criteria whether the Statistical Model should be used to translate that particular sentence. These
criteria include the length of the sentence, the complexity of the sentence and the percentage of
subsegments that have already been translated with subsegments from the Clean Data
Repository. For example, the Dutch sentence “Het product wordt getest” (The product is tested)
will not be sent to the SMT model because the Rules Engine can perfectly well handle it, and it is
probably also in the Clean Data Repository. One the other hand, in the case of the more complex
sentence “In het ideale geval wordt er niet getest” (In the ideal situation no testing is done) a
search will be made in the statistically derived phrase table. This is consulted, not by the open
source Moses decoder, but by the author’s own decoder.

As already stated, the Clean Data Repository mostly comprises noun phrases and verb
phrases and short whole sentences. Unlike the results of the translation memory search, the
target subsegments retrieved from the Clean Data Repository are POS tagged and may even
contain some semantic information. The “clean data” are therefore useful input to the Rules
Engine which resolves any untranslated parts of the sentence and composes the subsegments
from these different sources into the final English sentence. On the other hand, the “dirty” or
unvetted, statistically derived data do not currently include any syntactic or semantic
information. Their relevance and usefulness is dependent upon the domain relevance of the
TMX files from which they are derived. Figure 5 shows a translation unit derived from the
Statistical Machine Translation model. In most translation memory applications, the source of a
translation is displayed in the editing panel so it is easy for the post-editor to identify whether the
translation comes from a translation memory, via a Statistical Translation Model or from the
Rules Engine.

```
<tu creationdate="2014/10/15 09:41" creationid="SMT">
<tuv lang="NL-NL">
<seg>Actielijsten moeten worden aangevuld cq worden beheerd.</seg>
</tuv>
<tuv lang="EN-GB">
<seg>Action lists must be supplemented or, as the case may be, managed.</seg>
</tuv>
</tu>
```

**Figure 5: Translation unit supplied via the Statistical Machine Translation model**

On a large project it is possible to gauge the usefulness of particular resources by running
test files and then consulting the job log. Figure 6 shows a job log from a project on which both a
translation memory and a statistical machine translation model supplied translations. The Clean
Data Repository is consulted by default and its supplied translations are always piped into the
Rules Engine.

```
Starting new logfile
Number of errors logged: 0
Segments translated from Translation Memory= 35
Segments/subsegments supplied by Clean Data Repository= 195
Segments translated by Rule-Based MT Engine= 661
Segments produced from SMT Data= 45
```

**Figure 6: Example of a job logfile showing translation sources**
Any segments (or subsegments) not resolved by the above three approaches will then be tackled by the Rules Engine, which has its own ways of dealing with multiword expressions. The whole process – from input to output – is summarised in Figure 7.

![Figure 7: Automatic translation process from input to output](image)

### 4. Practical experience

This three-pronged approach is useful for large-scale translation automation projects in domains for which a relevant translation memory is available. Typically, a project memory, which is a subset of the main memory, is created. That translation memory is then used to generate new entries in the Clean Data Repository. A statistical translation model may then be trained from the same translation memory. It has to be remembered that the resulting statistically derived phrase table is not used as the sole supplier of the translation of the sentence but only to offer translations of phrases at subsegment level.

In the “HAN project” described above, the preparatory work and the feedback of the translation team was the key to its success, which led the university to consider the adoption of this form of automated translation for future projects. The translators had an opportunity to provide translations of a system-generated list of unknown words and phrases in advance of the first translation run. The unigrams (single words) were entered in the Core Dictionary, while the phrases were added to the Clean Data Repository. The project involved three translation runs of a document of some 250,000 words. After run_1 and run_2 respectively the data were corrected as some of the target sentences were considered unacceptable for submission to the post-editors. The author fed the corrections into the system by adjusting entries in the translation memory and entering phrases and complete sentences into the Clean Data repository. The results achieved in run_3 were judged acceptable as they met the basic undertaking that the sentences would be grammatically correct and the output would not contain any garbage or
jumbled phrases. At the end of the project, the translation team was handed a large translation memory.

5. Judgement on usefulness

The SMT Research Survey Wiki lists more than 60 publications (since 2005) dealing with domain restriction. The literature suggests that the restriction of the domain in which training is done delivers results that require less correction of technical terms in post-editing than general-purpose resources. The author's practical experience of how different types of output from his software are received by his customers bear out this claim.

In day-to-day practice the author implements domain restriction by using a project translation memory and entering relevant subsegments in the Clean Data Repository. This has broadly been found to deliver a more fluent – though sometimes disjointed - output than use of the Rules Engine alone. The disjointedness is frequently one of style, reflecting the multiplicity of translation originators, and the disconnect is between sentences rather than within a sentence. In particular, it has been found that long-standing translation memory entries that have never been updated are sometimes less accurate in terms of terminology than the “pure MT” output.

This methodology has been used to provide English translations of Dutch technical texts to members of international teams on large-scale projects in the fields of transport, highway engineering and healthcare. In nearly two decades of providing machine translation services, the author has only received two serious complaints about the quality of the output. The key to user acceptance has undoubtedly been the Clean Data Repository, which is continually enlarged and kept up to date by deconstructing “approved” translation memories, something which is done at the end of every project. In fact, many of its short sentences, such as technical instructions, could equally have been included in a translation memory, and a script allows them to be converted from the XLIFF to the TMX format so that they can be used in any third-party translation memory such as Studio 2014.

The incorporation of unsupervised or “dirty data” is recent and the extent to which the use of the statistically derived model improves the output of the MT run has not yet been fully investigated. Figure 8 shows the English output from the machine translation of a Dutch sentence. The first version is generated by piping the results from the Clean Data Repository straight into the Rules Engine. The second version is built partly by finding matches in the statistically derived phrase table.

| DUTCH: De uitvoer van het rapport dient geëxporteerd te kunnen worden (als excel-, csv- en pdf-file). |
| RBMT: It must be possible to export the output of the report (as Excel, csv and pdf-file). |
| SMT MODEL: The output with the report should be capable of being exported (such as excel, csv and pdf-file). |

Figure 8: Output from Rules Engine and from Statistical Translation model

3 http://www.statmt.org/survey/Topic/DomainAdaptation
In the above example the Rules Engine achieves a more fluent output than that produced if the SMT model is accessed (Google Translate provides a similar translation to our in-system SMT model). The conversion of the awkward passive “should be capable of being exported” to “it must be possible to export” is the result of the application of a specific rule. However, the translation provided by the SMT model is not POS-tagged so the Rules Engine has less to work on. The author is therefore considering POS-tagging of the English phrases translated from the SMT model “on the fly”. This POS-tagging can be accomplished by the BNC_Frequency_List_Investigator, a Java class written by the author to retrieve the parts of speech of items contained in the British National Corpus.

Nevertheless, at subsegment level, the author’s general-purpose Statistical Model (trained from Europarl\(^\text{10}\)) can provide reasonably credible translations as shown below\(^\text{11}\):

<table>
<thead>
<tr>
<th>Dutch</th>
<th>With SMT</th>
<th>Without SMT</th>
</tr>
</thead>
<tbody>
<tr>
<td>beschikbare personeel</td>
<td>human resources available</td>
<td>available personnel</td>
</tr>
<tr>
<td>gewenste middelen</td>
<td>desired equipment</td>
<td>desired means</td>
</tr>
<tr>
<td>onderdeel van de planning</td>
<td>component of the timetable</td>
<td>component of the planning</td>
</tr>
<tr>
<td>ingreep</td>
<td>surgical procedure</td>
<td>intervention</td>
</tr>
<tr>
<td>bijzondere aandacht voor</td>
<td>particular attention to</td>
<td>particular attention for</td>
</tr>
<tr>
<td>er kan ingesteld worden</td>
<td>you can set</td>
<td>it can be set</td>
</tr>
<tr>
<td>bezetting van de OK</td>
<td>workforce of the OR</td>
<td>occupation of the OR</td>
</tr>
<tr>
<td>medische gegevens</td>
<td>medical statistics</td>
<td>medical data</td>
</tr>
<tr>
<td>wordt er gekeken naar</td>
<td>we look at</td>
<td>it is looked to</td>
</tr>
<tr>
<td>in de volgende functies</td>
<td>within the next functions</td>
<td>in the following functions</td>
</tr>
<tr>
<td>actuele omstandigheden</td>
<td>topical situations</td>
<td>current circumstances</td>
</tr>
<tr>
<td>registratie van het OK-team</td>
<td>registration of surgical team</td>
<td>record of the surgical team</td>
</tr>
<tr>
<td>indien een patiënt overlijdt</td>
<td>where a patient dies</td>
<td>if a patient dies</td>
</tr>
<tr>
<td>standby functie</td>
<td>standby task</td>
<td>standby function</td>
</tr>
</tbody>
</table>

The above phrases are taken from a healthcare specification which was translated both with and without the data provided by the SMT model. Given that the SMT model was trained from

\(^{10}\)http://www.statmt.org/europarl/

\(^{11}\)Phrases extracted from machine translation of healthcare technology specification – with and without use of SMT
the broad-ranging Europarl data without any tuning, it seems reasonable to assume that models trained on in-domain data will yield at least comparable results in terms of the usability of subsegments. The author does not intend to throw out his Rules Engine but rather to use the SMT model as an additional source of translations at subsegment level.

6. Conclusions

This flexible approach to automatic translation has been designed for handling large-scale projects involving professional translators at the beginning and end of the production line. The author prepares these projects in close collaboration with the translators, who will ultimately import the “machine output” into their respective translation memory tools. The full process sees the machine translation engine first consulting any provided translation memory and then, if there is no TM match, consulting a Clean Data Repository of subsegment data and (if the engine so decides) a statistically derived phrase table, before piping the “mixed bag” of phrases into a Rules Engine which generates the final English sentence.

Based on his results and customer acceptance achieved from using a large general-purpose translation memory and a statistical translation model based on the Europarl corpus, the author plans to build a series of in-domain translation memories and use them to train in-domain statistical translation models.
Terminology finding in the Sketch Engine: an Evaluation

Adam Kilgarriff
Lexical Computing Ltd.

ABSTRACT

The Sketch Engine is a leading corpus query tool, in use for lexicography at OUP, CUP, Collins and Le Robert, and at national language institutes of eight countries, and for teaching and research in many universities. Its distinctive feature is the ‘word sketch’ a one page, automatic, corpus, derived summary of a word’s grammatical and collocational behaviour.

Very large corpora and word sketches are available for sixty languages.

A number of tools and resources have recently been added with translators and terminologists in mind. The resources are parallel corpora: EUROPARL-7 and the various datasets available in the OPUS collection. The tools are bilingual word sketches and the term finder. These have been reported on in previous Asling/Aslib conferences.

One remaining task is to make the Sketch Engine functions conveniently accessible to translators and terminologists. We have recently done this via IntelliWebSearch, a tool which lets the user highlight text in the environment they are working in, which could be a CAT tool or Microsoft Word, and, with a key sequence, query their preferred database. So now the key sequence can take the translator or terminologist to a browser window showing the word sketch, or parallel concordance, or any of a number of other reports, for the expression they are working on.

1. The term-finding functionality

The term-finder starts from a domain corpus, and a reference corpus. First it finds all the noun phrases, and their frequencies, in both corpora. It then takes the ratio, and the items with highest ratios will be terms, as in Figure 1 (data supplied by lead users, the World Intellectual Property Organisation).
In some cases, as with WIPO, the user will have domain corpora, but in others they will not. In that case they may use the BootCaT procedure (Baroni and Bernardini 2004). The user, typically a translator working in a domain where they are not an expert, inputs a few domain-specific ‘seed words’; these are sent to a search engine, and the hits identified by the search engine are gathered, cleaned, de-duplicated and processed to give a domain-specific corpus. This functionality has been found to support translators well (Bernardini et al. 2013). For some time, the Sketch Engine has incorporated a BootCaT tool, allowing users to create an instant corpus for a domain, which means they can then compare this corpus with a reference corpus to find the keywords of the domain. The functionality has recently been extended so the user can find the terms alongside key words. Thus, where the user has Bootcatted an English environment corpus, the Sketch Engine provides the "key words and terms" report shown in Figure 2.

The requirements for the term-finding functionality are:

- a processing chain, comprising tokeniser, lemmatiser and part-of-speech tagger, installed and ready to apply to the user’s domain corpus
- a reference corpus processed with the processing chain
- a term grammar.

At time of writing, these are in place for Chinese, English, French, German, Japanese, Korean, Russian, Spanish and Portuguese.

<table>
<thead>
<tr>
<th>Term</th>
<th>Frequency</th>
<th>Freq/mill</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>station de base</td>
<td>28612</td>
<td>3292.2</td>
<td>3293.2</td>
</tr>
<tr>
<td>station mobile</td>
<td>12514</td>
<td>1439.9</td>
<td>1440.9</td>
</tr>
<tr>
<td>communication sans fil</td>
<td>8189</td>
<td>942.3</td>
<td>943.3</td>
</tr>
<tr>
<td>liaison montante</td>
<td>6561</td>
<td>754.9</td>
<td>737.5</td>
</tr>
<tr>
<td>terminal mobile</td>
<td>7406</td>
<td>852.2</td>
<td>709.8</td>
</tr>
<tr>
<td>liaison descendante</td>
<td>5434</td>
<td>625.3</td>
<td>626.3</td>
</tr>
<tr>
<td>stations de base</td>
<td>5010</td>
<td>576.5</td>
<td>577.5</td>
</tr>
<tr>
<td>réseau de communication</td>
<td>4255</td>
<td>489.6</td>
<td>490.6</td>
</tr>
<tr>
<td>communication mobile</td>
<td>4722</td>
<td>543.3</td>
<td>462.5</td>
</tr>
<tr>
<td>point d’ accès</td>
<td>3907</td>
<td>449.6</td>
<td>450.6</td>
</tr>
<tr>
<td>modes de réalisation</td>
<td>3486</td>
<td>401.1</td>
<td>402.1</td>
</tr>
<tr>
<td>réseau d’ accès</td>
<td>3241</td>
<td>372.9</td>
<td>373.9</td>
</tr>
<tr>
<td>réseau sans fil</td>
<td>2903</td>
<td>334.0</td>
<td>335.0</td>
</tr>
<tr>
<td>accès radio</td>
<td>2412</td>
<td>277.5</td>
<td>278.5</td>
</tr>
<tr>
<td>transfert intercellulaire</td>
<td>2408</td>
<td>277.1</td>
<td>278.1</td>
</tr>
</tbody>
</table>
2. Evaluation

To evaluate a term-finder for a language and a domain, a list of all the ‘true terms’ is required. Then we can compute precision and recall.

One problem: how to define the domain? The straightforward answer: provide a corpus of it. Then we have the more constrained task of assessing recall and precision, from a given corpus, when the terms in that corpus are known.

Another problem: won’t two different terminologists inevitably deliver two different lists?

We approached the task by hunting for research datasets comprising domain corpora and term-lists derived, by human experts, from them. In most cases, this had been done as part of a term-finding task, so there were also published papers, with term-finding results, over these datasets, so we had a reference result to compare our results with. This addressed the second problem, as, whatever the lists, we were confronted with the same challenge as the resource developers. We found datasets for seven languages and six domains. In each case we entered the corpus into the Sketch Engine, ran the term-finder, and computed precision and recall (which we could then compare with the performance figures of the group who developed the dataset.) The paper will present these results.
3. In sum

The Sketch Engine has for some years been a leading tool for lexicography and corpus linguistics. Its terminology function is now a year old. We present a thorough evaluation.

References


Machine Translation Quality Estimation Adapted to the Translation Workflow

Sabine Hunsicker  
euroscript Deutschland GmbH

Alexandru Ceausu  
euroscript Luxembourg S.à r.l.

ABSTRACT

The varying quality of machine translation (MT) poses a problem for language service providers (LSPs) which want to use MT to make the translation production process more efficient. In this user study we describe the MT confidence score we developed. It predicts the quality of a segment translated by MT and it is fully integrated into the translation workflow.

1. Introduction

The varying quality of MT poses a problem in the translation workflow as it requires different levels of post-editing effort. As one sentence may be translated well, the next one may turn out to be completely unusable. There might be cases where discarding the MT suggestions and translating from scratch is going to be faster. This decision time also increases the post-editing time.

In order to be able to exploit the full potential of MT suggestions, they should be annotated with a score that is indicative for the translation quality. The translators are already familiar with such scores from the usage of translation memory (TM).

MT does not assign a score to its output. The decoder calculates internal scores to find the best hypothesis from the translation options, but these decoder scores cannot be used to estimate a level of quality.

For an LSP, a predictive score for MT quality would be very useful, as this would be in line with the way TMs are used in the workflow. Another important advantage is that it provides an upfront estimation of the cost for a given translation.

2. Confidence Score

The translation workflow at euroscript involves several MT-related stages. One of these stages contains the quality estimation component which we call confidence score, e.g. a component that would answer the question on ‘how confident is the MT that a particular sentence was well-translated?’
In order to reduce the annotation effort, we developed this score starting from the automatic scores. The approach can be easily automated so that it can be run immediately after training a new MT system. Another advantage is that there is no time lost in finding human annotators and for data creation before the new MT system is deployed in production.

The prediction model makes use of a combination of system-independent and system-dependent features. For example, the sentence lengths of the source and the MT candidate are taken into account. The system-dependent features vary on the MT system that should be evaluated. SMT systems usually provide different scores calculated during decoding.

Each training instance includes the source sentence, the target sentence, the MT candidate, the feature vector and the automatic score.

The training algorithm automatically chooses a well-distributed sample of training instances to train the prediction model. As the confidence score is integrated into the MT workflow, each MT request is automatically annotated with the confidence score.

The confidence score is optimized to predict which of the following levels of quality the current translation belongs to:

- good (no or little editing needed)
- usable (some editing is required)
- useless (discard)

3. Experiments

For exemplification, we present our confidence score experiments on the language direction English→Danish.

The texts used in our experiments come from the public domain. The training data was created by translating these texts with MT and then post-editing the results. As such, the translations are usually very close to the MT candidates, except where MT was of such a bad quality that it was discarded. The test set contains 1074 sentences.

During translation with MT, the automatic scores for the classifier were collected and a confidence model was trained on the resulting data. After integration into the translation workflow, the model was evaluated on unseen texts from the same domain.

We trained prediction models for the following three automatic scores:

- BLEU (Papineni et al., 2002)
- normalised Editing Distance
- Fuzzy Match

The editing distance (ED), based on the Levenshtein distance, allows us to draw conclusions concerning the post-editing effort—the lower the editing distance, the lower the effort for post-editing. In the original version a low score means that the MT candidate was close to the reference translation, a high score respectively that the MT candidate varied a lot from the reference. In contrast to BLEU, this distance is not contained in a closed interval, therefor we use
a normalised version that transposes the scores to the interval \([0,1]\). Additionally we reverse the score, so that 1 is the best score, analogous to the other scores used.

Fuzzy matching (FM) is another indicator of how close the reference and MT translation are. FM is usually used when evaluating a new text against a translation memory and works on the source sentence. In our experiments, we used the fuzzy matching algorithm implemented in the Okapi Framework\(^1\). The reference translation is set as the original translation (that would be saved in the TM) and the MT candidate is set as the new text.

Neither FM nor ED take into account the source sentence.

Each confidence score model is evaluated compared to the three scores: BLEU, editing distance (ED) and fuzzy matching (FM).

To compare the different metrics, we calculate three types of measures: the mean absolute error (MAE), the root mean squared error (RMSE) and Pearson's correlation coefficient \((r)\).

<table>
<thead>
<tr>
<th>Confidence Score</th>
<th>Score</th>
<th>RMSE</th>
<th>MAE</th>
<th>(r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{conf}_{\text{BLEU}})</td>
<td>BLEU</td>
<td>0.4577</td>
<td>0.2894</td>
<td>-0.1615</td>
</tr>
<tr>
<td></td>
<td>ED</td>
<td>0.6765</td>
<td>0.6469</td>
<td>-0.2414</td>
</tr>
<tr>
<td></td>
<td>FM</td>
<td>0.5305</td>
<td>0.4618</td>
<td>0.2358*</td>
</tr>
<tr>
<td>(\text{conf}_{\text{ED}})</td>
<td>BLEU</td>
<td>0.5743</td>
<td>0.5422</td>
<td>0.0063</td>
</tr>
<tr>
<td></td>
<td>ED</td>
<td>0.1878</td>
<td>0.1537</td>
<td>0.1980</td>
</tr>
<tr>
<td></td>
<td>FM</td>
<td>0.3611</td>
<td>0.2658</td>
<td>-0.1414</td>
</tr>
<tr>
<td>(\text{conf}_{\text{FM}})</td>
<td>BLEU</td>
<td>0.3861</td>
<td>0.3257</td>
<td>0.4063*</td>
</tr>
<tr>
<td></td>
<td>ED</td>
<td>0.4707</td>
<td>0.4348</td>
<td>0.2743</td>
</tr>
<tr>
<td></td>
<td>FM</td>
<td>0.3858</td>
<td>0.3483</td>
<td>-0.1044</td>
</tr>
</tbody>
</table>

Table 1: Evaluation statistics for EN \(\rightarrow\) DA confidence score correlation

Table 1 shows the evaluation results for all three prediction models. We see that the error rates differ considerably between the evaluation metrics and the error rates. The prediction model based on the editing distances performs quite well: it achieves the lowest error rates and correlates moderately with the score it tries to predict.

As predicting the full range of scores is a very complex task, we decided to scale down and only predict the three quality levels described in Section 2.

To determine the thresholds of these levels, we ran two experiments, one with a very high level (95% for good, 75% for usable) and the other with a moderate level (75% for good, 50% for usable).

From these experiments, we can tell that the moderate quality levels are easier to predict, as we achieve higher correlation values with them. The editing distance model performs well here as well: choosing the minimum thresholds, we achieve a correlation of 0.2588 of the confidence score to the actual editing distance score.

\(^1\) http://okapi.opentag.com/
In a human evaluation, we used a random sample of the evaluation data to be judged by professional translators. Of 221 sentences, 102 scores were judged to be appropriate and 119 to be inappropriate, indicating that the scoring mechanism needs more fine-tuning, but that it is still usable.

4. Conclusion

We presented a practical example on how to incorporate the confidence score into the traditional translation workflow. The first prototype of our confidence score works well in our production set-up. To fine-tuning the predictions require more evaluation and data, both of which are created automatically during the translation post-editing in production.

Fine-tuning the predictions requires more evaluation and data, both of which are created automatically during the translation production. Our on-going work is to introduce more features for the prediction model, such as linguistic analysis. Another constraint for assessing MT, however, is the time required to do so, as the MT needs to be provided quickly. Our current model provides judgments in less than half a second, and further improvements need to scale accordingly.

References

The Dos and Don’ts of XML document localization

Andrzej Zyroń
CTO XTM International

ABSTRACT

XML is now ubiquitous: from Microsoft Office to XHTML and Web Services it is at the core of electronic data communications. The separation of form and content, which is inherent within the concept of XML, makes XML documents easier to localize than those created with traditional proprietary text processing or composition systems.

Nevertheless, decisions made during the creation of the XML structure and authoring of documents can have a significant effect on the ease with which the source language text can be localized. For example, the inappropriate use of syntactical tools can have a profound effect on translatability and cost. It may even require complete re-authoring of documents in order to make them translatable.

This presentation highlights the potential pitfalls in XML document design regarding ease of translation and provides concrete guidance on how to avoid them.

1. Introduction

The adoption of XML as a standard for the storage, retrieval and delivery of information has meant that many enterprises have large corpora in this format. Very often information components in these corpora require translation. Normally, such enterprises have enjoyed all of the benefits of XML on the information creation side, but very often, fail to maximize all the benefits that XML based translation can provide.

The separation of form and content which is inherent within the concept of XML makes XML document easier to localize than traditional proprietary text processing or composition systems. Nevertheless decisions made during the creation of the XML structure and authoring of documents can have a significant effect on the ease with which the source language text can be localized into other languages. The difficulties introduced into XML documents through inappropriate use of syntactical tools can have a profound effect on translatability and cost. It may even require complete re-authoring of documents in order to make them translatable. This is worth noting as a very high proportion of XML documents are candidates for translation into other languages.
A key concept in the treatment of translatable text within XML documents is that of the "text unit". A text unit is defined as being the content of an XML element, or the subdivision thereof into recognizable sentences that are linguistically complete as far as translation is concerned.

2. Designing XML documents for translation

It is very important to consider the implications for localization when designing an XML document. Wrong decisions can cause considerable problems for the translation process thus increasing costs. All of the following examples assume that the text to be translated is to be extracted into an intermediate form such as XLIFF (XML Localization Interchange File Format). Anyone planning to provide an XML document directly to translators will soon be disabused of this idea after the first attempt. The intermediate format protects the original file format and guarantees that you get back an equivalent target language document to that of the original source. An additional concept which is important regarding the localization of XML documents is that of the 'inline' element. Inline elements are those that can exist within normal text (PCDATA - Parsable Character DATA). They do not cause a linguistic or structural break in the text being extracted, but are part of the PCDATA content.

The following is a list of guidelines based on (often bitter) experience. Most of the problems are caused by not following the fundamental principles of XML and good XML practice. It is nevertheless surprising how often you can come across instances of the following type of problem. Please note that this is not a prescriptive list, there may be special circumstances where the proposed rules may have to be broken:

2.1. Avoid the use of specially defined entity references

Although entity references can look like a 'slick' technique for substituting variable text such as a model name or feature in a publication, they can cause more problems than they resolve.

Example 1: Incorrect use of Entity References

Entities can cause the following problems:

- **Grammatical difficulties**

  If the entity represents a noun or noun phrase this will potentially cause serious problems for languages in which nouns are strongly inflected, such as many Slavonic and Germanic languages. What appears fine as an entity substitution in English can cause insurmountable problems in inflected languages.

  The solution is to resolve all entities in the serialized version of the XML document prior to translation.

- **Parsing difficulties**

  During the translation process the text will typically be transformed into different XML based translation formats such as XLIFF where the entity will cause a parsing error.
• **Problems with leveraged memories**

The use of specially defined entity references can also cause problems with leveraged memories. The leveraged memory may contain entities not declared in the current document.

It is generally better to use alternative techniques rather than entity references:

<para>Use a <tool id="a1098">claw hammer</tool> to release the CPU retention catch.</para>

**Example 2: Proposed solution**

One area where entities CAN be used to great effect is that of boilerplate text. The technique here is to use parameter entities to store the text. The text must always be linguistically complete in that it cannot rely on positional dependencies with regard to other entities etc. Boiler plate text is used solely within a DTD. There need to be parallel target language versions of the DTD for this technique to be used which can add to the maintenance cost, although judicious use of INCLUDE directives and DTD design can mitigate this.

2.2. **Avoid translatable attributes**

Translatable attributes can also look like a smart way of embedding variable information in an element.

<para>Use a <tool id="a1098" name="claw hammer"> to release the CPU retention catch.</para>

**Example 3: Incorrect use of translatable attributes**

Unfortunately, they present the translation process with the following difficulties:

• **Grammatical difficulties**
  The same problems can arise as with entity references. If you want to use the text for indexing etc., then you cannot rely on the contents of translatable attributes to be consistent for inflected languages.

• **Flow of text difficulties:**
  With translatable attributes there are two possibilities regarding the flow of text:
  o The text is part of the logical text flow.
  o The text should be treated outside of the text flow.

If the text is to be part of the text flow then the translatable attribute causes the insertion of extra inline elements in translatable format (typically XLIFF format) of the file. If it is to be translated separately, then the translatable attribute forms a new text unit. The translator then needs to know if it is to be translated within the context of the original text unit or in isolation. With extra inline elements the burden is on the translator to preserve the encapsulating encoding, baring in mind that there may be significant changes in the sequence of such attribute text in the target language. Translation may often require that the position of the various components of a text unit are significantly rearranged.
Use a <tool id="a1098">claw hammer</tool> to release the CPU retention catch.

Example 4: Proposed solution

There is a good rough rule of thumb that if text has more than one word then it should not be used in attributes. As a syntactical instrument attributes are much more limited than elements. For a start you can only have one attribute of a given name. The use of attributes should be reserved for single "word" values that qualify in a meaningful way an aspect of their element.

2.3. Avoid using CDATA sections that may contain translatable text

CDATA sections are typically used as a means of escaping multiple '<' and '&apos;' characters. Unfortunately they pose particular problems for tools that are extracting such text. The problem is not one of the escaped characters, but how to treat the CDATA text.

Example 5: CDATA section problems

The problem is a similar one to that posed by translatable attributes. Is the text to be treated as 'inline' to the surrounding text? What of the escaped characters. Are they to be replaced on translation with the appropriate characters that were originally escaped, or are they to be left in their escaped form. How is the software to know?

I have come across whole XML documents being embedded as CDATA within an encompassing XML document. This poses significant problems regarding the treatment of the CDATA text. It must first be extracted and then re-parsed before it can be extracted for translation.

Unless the text within CDATA sections is specifically never to be translated, please avoid using CDATA sections and use the standard built in character references to escape the text.

Example 6: Proposed solution

Example 7: Or alternatively us a link to an external resources

2.4. Avoid the use of infinite naming schemes

Do not use the following type of element elm001, elm002, elm003 in well formed documents.
Example 8: Example of infinite naming scheme usage

This presents problems for extraction programs and is not regarded as good XML practice. A much better way of doing this is to use the ID and IDREF attribute mechanisms to link elements together.

Example 9: Proposed solution

2.5. Avoid placing translatable text in Processing Instructions (Pls)

Processing instructions are a very 'weak' syntactical instrument in XML. There is no built in mechanism in XML to assist syntactically in the preservation of Processing Instructions. Similarly any extraction tools will need to have special knowledge of the structure of the data in processing instructions.

Example 10: Incorrect use of translatable text in Pls

Example 11: Proposed solutions
It is generally not a good idea to have any processing instructions present within translatable text. The main reason is that there is no guarantee that they will survive the translation process. It is better to strip out any PIs prior to translation.

2.6. Avoid the use of text in bitmap graphics

There should be no excuse with the existence of SVGs to use bitmapped graphics. They pose particular problems in that the original bitmap will need to recreated for the target language with the translated text. This is usually a very costly and error prone process and requires appropriate target language knowledge of the person that is editing the graphics.

2.7. Never make any assumptions about text length sizes in your design

Always allow for the fact that the target language text may be significantly longer than the source. For example "Welcome" becomes "шчыра запрашаем" in Belarusian and "maligayang pugdating" in Tagalog. Design your output with flexibility in mind.

2.8. Always use UTF-8 (or alternatively UTF-16) encoding

With English source we can often get tempted to use 7 bit ASCII or ISO 8859/1 encoding. As soon as you find that you are required to translate into a language that is not covered by ISO 8859/1 you will find that trying to maintain documents in different encoding schemes a real problem. Always use UTF-8 from the start. It gives you immediate access to commonly used punctuation characters such as 'm-dash' and 'n-dash' etc. It also significantly simplifies your document processing. All XML parsing tools have to be able to cope with UTF-8 and UTF-16. UTF-8 is more economical in terms of space usage for most European Languages whose scripts are based on the Latin alphabet.

2.9. Never break a linguistically complete text unit

Never start a sentence in one non-inline element and continue it in another. You cannot rely on the translated text being in the same word sequence in terms of the sentence content as the target. It also makes the job of translation much more difficult as the translator does not see the whole sentence.

Example 12: Example of a sentence broken over more than one element

2.10. Avoid the use of “typographical” elements

Use logical elements instead that encompass the text.

Example 13: Example of typographical element usage
Use *emph* instead of **bold**. Encompass any text that requires to be on a line with line elements.

```xml
<para>
  <emph>Do not use</emph> 'br' type elements.
</para>
```

**Example 14: Suggested correct usage**

Avoid at all cost introducing any line breaks into the text stream. You can unconditionally guarantee that this will cause problems in some if not all of the target languages.

2.11. **Do not mix translatable and non-translatable text in the same elements**

Keep non-translatable PCDATA in different elements than translatable PCDATA.

```xml
<data-items>
  <data id="class">com.xmlint1.data.dataDefinition</item>
  <data id="text">Replace generic data definitions with specific instances.</item>
</data-items>
```

**Example 15: Example of mixed PCDATA**

Most XML translation tools will have problems with this type of construct. It is only when inspecting the 'id' attribute that a decision can be made as to whether the PCDATA should be extracted or not.

```xml
<data-items>
  <class id="com.xmlint1.data.dataDefinition">
    <text>Replace generic data definitions with specific instances.</text>
  </class>
</data-items>
```

**Example 16: Suggested solution**

2.12. **Avoid holding Source and target PCDATA in the same document**

This can cause all manner of problems for processing and extraction tools.

```xml
<para>
  <text xml:lang="en">My hovercraft is full of eels.</text>
  <text xml:lang="fr">Mon aéroglisseur est plein d'anguilles.</text>
  <text xml:lang="hu">Légpárnás hajóm tele van angolnákkal.</text>
</para>
```
Example 17: Example of mixed source and target PCDATA

Unless your document requires mixed language content use a separate document instance to store each target language version. If you store both source and target data in the same document it will become unwieldy, overly large and cumbersome to process.

2.13. Clearly define text that requires translation

Keep any PCDATA that requires translation in different elements from PCDATA that does not require translation. Use special elements for text within PCDATA that is specifically not to be translated.

The following part of this sentence should not be translated at all.
ABSTRACT

AutoLearn<word> extracts new translation relations for words and multiword expressions (MWE) of any category from bilingual texts of any size in high quality and prepares the information found as a conventional dictionary entry - with morpho-syntactic and semantic classifications and contextual use conditions.

The function uses Lingenio's MT-system and analysis components as knowledge source, integrates its results into these and, by this bootstrapping approach, adapts dictionary and MT to the needs of the customer. Manual intervention is restricted to a very reduced number of difficult cases and can be carried out easily in an ergonomic graphical user interface, without need of effortful training. This is enabled by the underlying MT-architecture with rule-based core and additional statistical features.

The use conditions connected to the new dictionary entries are derived from the local representation the considered word or expression is part of in the considered reference(s). They restrict the corresponding translation to similar cases so that interferences with other translations in the dictionary are avoided.

A basic version of the function is already available in the current version of Lingenio's translate.

1. Motivation

Unknown words and lack of information about translations of words and expressions are notorious problems of Machine Translation (MT). The problems differ however depending on the type of MT system considered and the coverage of the lexicon aimed at: In a statistical machine translation system (SMT), 'unknown' means that the form is not listed in the frequency lexicon of the source language and the translation model doesn't show corresponding elements in the phrase tables. In a so-called rule based Machine Translation system (RBMT), 'unknown' means that for a form detected in the source text information about lemmatization, morphological classification and about syntactic and semantic properties is missing. With respect to translation, lack of information may mean that besides the pure (weighted) transfer relations use conditions are missing or unsufficient, such that a contextually justified selection of the target form cannot
be made. For hybrid MT systems the problems may vary depending on the type of system considered.

Vocabulary and translation of expressions may change significantly when considering different text topics and domains. No system can provide constant translation quality for any text. Therefore, terminology extraction from (monolingual and bilingual) texts and corpora is a very important prerequisite of customization and effective professional translation. Corresponding research and development have a rather long tradition.

As the purposes of extraction are many and the environments an extraction tool shall be part of different (consisting of, or integrating, different CAT-tools and MT-systems respectively), the research area covers a wide range of subtasks, questions and approaches to solutions. Main topics and distinctions are: extraction of references for search terms in monolingual versus bilingual texts, corresponding monolingual versus bilingual search of new terms, where 'term' may refer to single words or, in contrast or additionally, to multiword expressions and to specific categories (mostly noun) or alternatively to expressions of any category. Search may be carried out on the basis of 'knowledge-rich' or 'knowledge-poor systems' where 'knowledge' may be statistical or 'rule-based' or hybrid. Well known commercial tools in the area of bilingual extraction for CAT and/or MT respectively are, or are included in: Systran Business Translator, ProMT Professional, Across Personal Edition (crossTerm), SDL Trados (MultiTerm). Examples of (free) web-based monolingual term recognition and extraction tools are FiveFilters and TermMine. 'Concordancer' like AntConc find reference sentence pairs for glossary items.

An older but still valid overview of the architectural attempts made in the field is (Thurmair 2003). A good selection of research papers can be found at the homepage of the EU FP7 project Terminology Extraction, Translation Tools and Comparable Corpora (TTC) which terminated recently.

Some of the systems available combine linguistic analysis and statistical filtering, in particular if they don't only extract references, but suggest dictionary entries. However, if so they typically confine to subtypes of NPs only (in English typically: adjective-noun, noun-noun, noun-PP-combinations). Suggestion of words and MWEs for any category and adding transfer conditions is seldom. AutoLearn<word> is such a system.

2. Background

The background of AutoLearn<word> is a hybrid MT architecture with rule-based core and statistical features (cf. Babych et al. 2012). The rule-based core traces back to the longtime IBM R&D project Logic Based Machine Translation (LMT, McCord 1989b) which had been initiated in the late 80s at IBM research. The systems resulting from this have been developed further and extended by Lingenio GmbH, which is an IBM spin-off from 1999.

The architecture follows the transfer approach: It segments the text, assigns analyses to the segments (sentences), translates these and generates the target from the target structures. The analysis system is built on slot grammar (McCord 1989a), a unification-based dependency grammar. Slot grammar is also used in deep linguistic processing of IBM Watson (McCord et al 2012).
From the slot grammar analysis of a sentence the system constructs a shallow semantic representation, called a dependency tree. These trees are compact encodings of so-called flat underspecified discourse representation structures (FUDRSs). FUDRSs are the representation items of flat underspecified discourse representation theory (FUDRT; cf. Eberle 2002, 2004), which is an extension of (underspecified) Discourse Representation Theory (DRT / UDRT; cf. Kamp 1981, Kamp/Reyle 1993, Reyle 1993). The trees represent the predicate argument structure(s) of the sentence as can be derived from the syntactic structure and from ontological knowledge about the words. They are compact as they don't spell out the semantics of the lexical elements, but only point to them, i.e. they aren't the FUDRS of a sentence but determine its construction. FUDRSs allow to include additional information, in particular pragmatic information from the discourse history and from intersentential links, and to use this for stepwise disambiguation of lexical and structural ambiguities, if needed. The tree encoding can account for this.

The dependency trees are translated recursively into corresponding structures of the target language using the knowledge of the bilingual dictionary. From these 'deep structures' of the target side slot grammar structures are generated and from these, finally, the target sentence. (In the older LMT systems the shallow semantic level was missing and transfer happened at the level of slot grammar analysis).

The advantage of the representation for the purpose of AutoLearn is that it abstracts from the syntactic details of the surface structure and concentrates on the representation of the events and states of the sentence and the corresponding roles without going too much into detail with semantic specification. As on this level source representation and target representation are typically more similar to each other than on the syntactic level, it is more appropriate for the extraction of lexical source-target correspondences than this.

Note however that the details of the more 'lower' level analyses are not lost; they can be used by transfer and generation as they are connected to the 'higher' levels through corresponding links (so that, for instance, calculation of the surface order of the target sentence may use the shaping of the source sentence as an additional knowledge resource).

3. Algorithm

After aligning the sentences of a bilingual text, AutoLearn carries out structural analyses of the corresponding source-target sentence pairs and then uses the translation knowledge of the MT system to relate words and expressions of the respective source structure and the target structure to each other.

The example that we will use is from the solvency consulting knowledge series (cf. Munich Re), which is about European solvency law.
3.1. Dependency tree representations

Fig 2. shows the dependency trees for source sentence and target sentence of the first pair of sentences in fig. 1 marked in red. The picture rotates the dependency trees by 90 degrees to the left, such that the top is on left and the leaf information on right, and the different information types are structured in a column-like form where the first column represents the dependency relations between the lexical items and the second and the third the corresponding semantic and morpho-syntactic reading of these items. The trees neglect the order of the lexical items in the sentence. (As said before, this information can be accessed, if necessary). Expressions of the form \( s(\text{stem}, \text{ID}) \) in the second column are called senses. The first argument represents the homonym stem and the second the lexical identifier of the corresponding reading via which the respective semantic interpretation in the sense of DRT or predicate logic can be constructed, if needed (for instance in order to represent temporal and referential links of the discourse history). For the purposes of AutoLearn<word>, mostly the shallow ontological classifications represented in the dependency tree will do: \( s(\text{artikel},46752) \) in the second column of the second row of the first tree in fig. 2 stands for the reading of the word \text{Artikel} as an \textit{article of a document} like a \textit{newspaper article} - in contrast to the reading as an \textit{item of a product portfolio} - and consequently it is classified as a \textit{doc}, i.e. a document, in the third \textit{column}. This column combines the morpho-syntactic features - assigned to the item in the sentence representation selected and the ontological classification available for it (if any) in the form \textit{Feas:Types}, where 'Types' arranges the available semantic supertypes of the word reading in disjunctive normal form as a list of lists. (Typically, a word reading is classified by one disjunct only).

Though the dependency trees of fig. 2 resemble very much slot grammar representations of the syntactic level because the compact representation format chosen 'hides' the proper semantic contribution of the lexical items, they nevertheless significantly abstract from syntactic details.
Figure 2: Dependency tree analyses of a pair of sentences from the bilingual text

For instance, consider the core sentence of the (first conjunct of the) English sentence of the example, (1.a) below. (1.b) - (1.d) are variants which develop from (1.a) by using other mood and/or other tense forms respectively, including forms with auxiliary constructions.

(1.a) *This article presents the principles of the total balance sheet approach.*

(1.b) *This article has presented the principles of the total balance sheet approach.*

(1.c) *Does this article present the principles of the total balance sheet approach?*

(1.d) *Have the principles of the total balance sheet approach been presented by this article?*

Though syntactically different, all sentences are represented by structurally identical trees; the differences are represented by different features assigned to the verb node only. There is no other difference with respect to dependency structure.
Fig. 3 gives the representation of (1.b) and screenshots of the verb rows of the representations of (1.c) and (1.d). The features of (1.a) are as in fig. 2.

The Features of verb nodes in dependency trees represent mood, tense and voice (mtv).

*ind*:*dcl:*nwh means that the node is not dependent on something else (*ind*) i.e. it stems from the main clause, is introduced by a declarative sentence (*dcl*) which doesn’t come with a *wh*-element (as relative clauses and *wh*-questions do). In contrast: *q* indicates question (*q:*nwh yes/no-questions, *q:*wh wh-question). The representation of the meaning of tenses used here follows the three-dimensional analysis suggested in (Kamp/Rohrer 1985), also used in (Kamp/Reyle 1993) and in many contributions to tense meaning since then. According to this, the tense forms are analyzed into tense level (*pres*, *past*, *fut*), perfectivity (yes/no (*1/0*)), and progressivity (yes/no (*1/0*)). For the voice we use: active (*a*), passive (*p*) and resultative passive (*rp*).

This is sufficient to describe the differences between (1.a)-(1.d) on the semantic level. When constructing a DRS or FUDRS, this information would be expanded to different temporal relations between the event, the speech time and contextual perspective time and reference time.

Dependency trees don’t always avoid ‘unfolding’ of the representation as they do here. Parts of the representation may be ‘multiplied out’, if this helps to make the source tree and the target tree more similar and supports structure-preserving translation therefore. An example is a sentence pair like ‘the experiment was difficult and unsuccessful in the end – das Experiment war schwierig und missglückte am Ende’. Here, ‘be’ in the English sentence is used as a copula in a predicative structure and obtains different translations depending on the respective conjunct of the coordination which fills the predicative argument of the structure. The factored form with the coordination raised to VP eases translation a lot, because now there are two occurrences of predicate structures with ‘be’ (… was difficult and was unsuccessful …) that can easily be translated separately. Other types of coordination and (some types of) ellipsis present further examples. As, nevertheless, these phenomena are not very frequent, ‘unfolding’ is carried out only if triggered by needs in transfer. (For a more detailed description of this compare Eberle 2003). Another feature of dependency trees is that they can ‘underspecify’ dependency relations, if the context doesn’t provide enough information about how some expression shall be related to the context. Frequently this is the case with prepositional phrases: In the example of fig. 2, *for a property*-
casualty insurer may relate to the second conjunct of the coordination only or, in contrast, may relate to the entire coordination (i.e. to both conjuncts). Depending on the options selected in the analysis system, the representation may disambiguate such ambiguities (as is the case in fig. 2) or not. We don't go into detail with this too (see Eberle et al. 2009 for a more elaborate description).

3.2. Relations between dependency trees

AutoLearn uses different types of relational knowledge of the system in order to find correspondences between nodes of the source dependency tree and nodes of the target tree. The strategy is as follows: First, correspondences are searched that provide 'anchors' of the transfer relation, i.e. correspondences that are safe knowledge; then, on the basis of this, additional correspondences are searched that represent likely correspondences. This is done using several levels of reliability. On the basis of the resulting bi-structural knowledge, the system extracts those relations found which are not yet covered by the dictionary of the system. The set of these is suggested to the user for (automatic) integration in user dictionaries.

The relational knowledge used by the system in the first step consists of:

a) the translation relations between the lexical elements of the system dictionary (and already existing user dictionaries respectively) and
b) formal similarity between source and target expressions if the corresponding source or target item is unknown to the system (names for instance).

As words (and translations) may be used several times in the pair of segments considered, the anchors of the first step must satisfy to a number of uniqueness constraints to be sure that they represent fixed points of a translation.

Starting out from the fixed points the subsequent steps of the algorithm use similarity considerations about the dependency relations connected to the anchor node pairs of the trees in order to recursively apply the search procedure of the first step to the remaining nodes. As the search spaces shrink step by step in this procedure, uniqueness constraints may be satisfied in subsequent steps that are not in preceding steps. At the same time the reliability conditions may be reduced step by step. The algorithm uses a number of parameters for regulating the behavior of the algorithm in this respect.

The options include to trigger computation of factored representations for parts of the dependency tree or to underspecify the trees further, if seems reasonable. Another option is to integrate information from statistical word alignment as an additional knowledge resource of the computation of the bi-structural relations.

3.3. Candidates for new dictionary items

From the cascaded procedure in 3.2 we obtain node relations and with this a coverage of source- and target sentence by pairs of partial structures whose quality reflects the information as available to the system. As, through this, the pairing doesn't relate to nodes only but is extended to the local substructures connected to the nodes, the suggestion of new lexical relations can go beyond single words and can specify multiword relations also and even larger phrasal relations if desired. The new relations include appropriate linguistic annotations as they
can reuse the morpho-syntactic and semantic properties of the senses they refer to in the trees. In addition, through the cross-language links morpho-syntactic and even semantic information can be inferred for words which, before, have been unknown to the system.

On the basis of the system dictionary used in the Lingenio product *translate pro* version 12.1 (the current version) the system suggests 6 new relations for the sentence pair of figure 4:

<table>
<thead>
<tr>
<th>German</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>to present</td>
</tr>
<tr>
<td>S2</td>
<td>total balance sheet approach</td>
</tr>
<tr>
<td>S3</td>
<td>solvency</td>
</tr>
<tr>
<td>S4</td>
<td>II</td>
</tr>
<tr>
<td>S5</td>
<td>property-casuality</td>
</tr>
<tr>
<td>S6</td>
<td>property-casuality insurer</td>
</tr>
</tbody>
</table>

S1 relates single words to each other where both words are known to the system, but not the relation between them (a gap in the system dictionary). S2, S5 and S6 relate MWEs where neither the source expression nor the target expression is known to the system, but only (the) leaves of the corresponding structures. Note that the respective internal structures are not pairwise homomorphic (see fig. 2) and therefore illustrate different subcases of the multiword suggestion type of AutoLearn<word>. S3 and S4 again relate single words to each other, but here the source word is unknown to the system in the one case and the target in the same category in the other (cf. fig. 2). When including the suggestions into the dictionary the derivable information is added: DE-Solvency is treated as a common noun like EN-solvency and the MWEs obtain standardized representations concerning the parts they consist of. Of course, the user can edit the entries and correct or add supplementary information.

### 4. Extensions

S1 is specific as it relates expressions to each other which are familiar to the system and therefore already obtain translations in the dictionary. This means that the new relation is unsatisfactory as, until now, it doesn't define under which circumstances the new relation shall be used in contrast to those already known by the system. The availability of local context helps to define corresponding conditions: The system knows the respective subcategorization frames and can infer from the dependency trees how the slots have been filled. We can obtain the more complete suggestion of fig. 4 for S1 therefore:
In fig. 4 the translation to present is restricted to the case where the subject is headed by Artikel and the direct object is filled by a conjunction made up by NPs headed by Prinzip and Auswirkung, as in the dependency tree of fig.2.

The restriction of figure 4 omits details of the complements in the dependency tree: neither does the determiner condition, det(Var1,Var2), specify whether the NPs should be sg, pl, definite or indefinite, nor are there included representations of modifiers (of the complements or the verb). It depends on the parameter setting whether a more restrictive extraction of conditions is selected or a more general one or a mixture of both. Typically, the system assigns conditions that relate to the ontological classifications of the words instead of the words themselves. In the example of fig. 2 this means to weaken the conditions to stipulating that the subject be of type doc (document) and the direct object of type regel0 (rule) or event.

Currently, the system is trained on large corpora in this respect, in order to optimize the conditions of the dictionary. This is joint work of the university of Leeds and Lingenio funded by the EU project HyghTra. Quantitative results are on the way, but couldn't made available yet. As the method is of type 'knowledge-rich' its reliability is very high, if validity and coverage of the underlying resources are, as is the case for the Lingenio analysis systems.

5. Conclusion

We described an algorithm for extracting new dictionary entries from bilingual corpora. It considers words and multiword expressions of any category, where 'multiword' is any linguistically closed structure whose consideration is justified by the administrator of the system - who may include phrasal templates too as well as statistical collocation information, if desired.

The new relations are annotated by morpho-syntactic and semantic classifications and by use conditions as can be derived from the system knowledge and the context in the sentence. The use conditions specify the circumstances under which the translation is triggered.

A basic version of AutoLearn<word> is already available in the current version of Lingenio’s MT-series translate. A more comprehensive version with features as described here will be integrated in the next product version, which is currently worked out.

References


Kurt Eberle (2004): Flat underspecified representation and its meaning for a fragment of German (Habilitation) Universität Stuttgart.


TTC: http://www.ttc-project.eu/
Quality Evaluation Today: the Dynamic Quality Framework

A. Görög
TAUS

ABSTRACT
Translation quality is one of the key topics in the translation industry today. In 2011, TAUS developed the Dynamic Quality Framework (DQF) in an attempt to standardize translation quality evaluation. In this paper, we will describe common approaches to translation quality and introduce the TAUS framework for QE. We will show that the development of this framework, initiated by the industry, was necessary to fill the gap between theory and practice. In We will give a short summary of the survey on quality evaluation and DQF that was conducted in the summer of 2014 among users of the DQF tools. Finally, we will suggest some ways academia and industry could and should collaborate with each other in the field of quality evaluation in the future.

1. Introduction
Translation quality is one of the key concepts in the translation industry. Measuring and tracking translation quality is essential for all players of the industry: more and more translation vendors offer different types and levels of quality resulting in dynamic pricing; translation buyers are seeking to know whether their customized Machine Translation (MT) engine is improving and would like to compare different MT providers; finally, translators need to set the threshold of TM/MT matches at the most optimal levels. And these are just a few examples where translation quality becomes central and increasingly tuned to user satisfaction.

This said, translations are evaluated using one arbitrary model (usually error-typology) while ignoring the fact that several models are available for this purpose. In 2011, TAUS developed the Dynamic Quality Framework (DQF) in an attempt to standardize translation quality evaluation. Quality in DQF is considered dynamic since today's translation quality requirements change depending on content type, purpose and audience. DQF contains a rich knowledge base, resources on quality evaluation and a number of tools to profile and evaluate translated content. The framework is based on the assumption that the evaluation type selected should always match the content type, purpose, and communicative context of the given translation in a flexible, dynamic way. There is no one-size-fits-all approach to translation quality evaluation (QE).

In this paper, we will describe common approaches to translation quality (section 2) and introduce the TAUS framework for QE (DQF). We will show that the development of this
framework, initiated by the industry, was necessary to fill the gap between theory and practice (section 3). In section 4, we will give a short summary of the survey on quality evaluation and DQF that was conducted in the summer of 2014. Finally, we will suggest some ways academia and industry could and should collaborate with each other in the field of quality evaluation in the future (section 5).

2. What is translation quality?

Quality is when the user or customer is satisfied. A longer and more scientific definition of quality is as follows:

“A quality translation demonstrates required accuracy and fluency for the audience and purpose and complies with all other specifications negotiated between the requester and provider, taking into account end-user needs.” (Melby, 2014, forthcoming)

Unfortunately, quality measurement in the translation industry is still not always linked to customer satisfaction and specifications. Very often, quality evaluation is the task of quality managers on the supply and demand side who have one specific evaluation model. This model is often based on error-typologies that assign different weights to different error types. Input from customers is usually missing or ignored and every translation receives the same treatment.

Despite very detailed and strict error-based evaluation models, it seems that satisfaction levels with both translation quality and the evaluation process itself are low. The major problem is that models and metrics used are not always measuring the right thing. Little consideration is given to multiple variables such as content type, communicative function, end user requirements, context, perishability, or mode of translation generation (whether the translation is created by a qualified human translator, unqualified volunteer, machine translation system or a combination of these). Traditional one-size-fits-all approaches to quality do not satisfy buyers and vendors of translation services anymore. QE models such as the LISA (Localisation Industry Standards Association) QE model, the J2450 or the EN15038 do not seem to take into account the different varying user requirements, communicative goals and content types. Are existing (ISO, LISA, ASTM, etc.) standards and certificates to ensure quality then useless. Standards that certify the translation vendor and/or the quality management process that lie beneath are certainly useful but cannot give a 100% guarantee that the product itself meets the required quality level. The only way to ensure that is by evaluating and doing that exactly the same way each time preferably across the whole industry. Today, there is an increasing appetite for such an approach to quality within the industry, an approach that measures the right quality level with the right method.

To offer such an approach and to standardize human evaluation of translated content, TAUS created the Dynamic Quality Framework (DQF). The DQF platform consists of a rich knowledge base on Quality Evaluation with best practices, reports, templates and a number of tools to evaluate translations made both by human translators and MT engines. The tools enable evaluators to compare translations, assess their accuracy and fluency, to measure post-editing productivity and to score translated segments based on an error typology. The Content profiling wizard enables users to select best-fit evaluation methods.
3. The Dynamic Quality Framework

3.1. Aim

Quality in DQF is considered dynamic as translation quality requirements change depending on the content type, the purpose of the content and its audience. The Framework provides a commonly agreed approach to select the most appropriate translation quality evaluation model(s) and metrics depending on specific quality requirements. The underlying process, technology and resources affect the choice of the quality evaluation model.

The Framework is underpinned by the recognition that quality is when the customer is satisfied. It is used when creating or refining a quality assurance program. DQF provides shared language, guidance on process and standardized metrics to help users execute quality programs more consistently and effectively. Improving efficiency within organizations and through supply chains. The result is increased customer satisfaction and a more credible quality assurance function in the translation industry.

3.2. Development

The development of DQF started in January 2011 by over fifty companies and organizations. Contributors include translation buyers and vendors as well as academic institutions. Users continue to define requirements and best practices as they participate in regular (online) meetings and events. Since the end of 2014, DQF is part of the TAUS Evaluate platform.

In the first phase of DQF development, TAUS carried out a benchmarking exercise to review evaluation models and this showed that existing QE models are relatively rigid\(^1\). For the majority, the error categories, penalties applied, pass/fail thresholds etc. are the same no matter what communication parameters were involved. The models are also of such a detailed nature that applying them is time-consuming and evaluation can only be done for a small sample of words. No standard tool was used for sampling neither for quality evaluation at the time. What's more, QE models are predicated on a static and serial model of translation production, which doesn't match 21\(^{st}\) century expectations of dynamic pricing.

DQF offers a more flexible approach to the common static quality evaluation models since it is based on the three parameters of utility, time and sentiment (UTS). This model considers the communication channel – Regulatory, Internal, or External (B2C, B2B, C2C). It is informed by the results from the content profiling exercise performed by TAUS enterprise members collaborating in this project, which shows that it is possible to map content profiles to the evaluation parameters utility, time and sentiment.

\(^1\) [https://www.taus.net/reports/translation-quality-evaluation-is-catching-up-with-the-times](https://www.taus.net/reports/translation-quality-evaluation-is-catching-up-with-the-times)
Figure 1: Importance of Utility, Time and Sentiment attributes distributed according to content types

The results of the content-profiling exercise also suggest that there are clear content differentiators for utility and sentiment while the parameter of time is much fuzzier. Reason is that most companies require a quick turnaround time for translations. Some examples of the mapping between content types and UTS rating are as follows:

- User Interface text and website content are rated highest for utility while audio/video content is rated lowest
- Marketing material and social media content are rated highest for time while user documentation, training material, online help and audio/video content are rated of medium importance for time.
- Marketing material and website content are given highest importance for sentiment while training material and online help are rated lowest for this parameter.

The Content Profiling wizard available on the TAUS Evaluate platform is one of the results of the TAUS benchmarking exercise described above. The DQF Content Profiling feature is used to help select the most appropriate quality evaluation model for specific requirements. This leads to the Knowledge base where you find best practices, metrics, step-by-step guides, reference templates, and use cases.

3.4. DQF tools

The DQF tools provide a vendor independent environment for the human evaluation of translation quality. Users gather vital data to help establish return-on-investment, measure productivity enhancements, and benchmark performance, helping to ensure that informed decisions are made. One of the aims of DQF tools is to standardize the evaluation process and make it more objective and transparent. The benchmarking and reporting functions provide users with a wealth of information on quality problems related to certain language pairs, text types, industries or domains.
DQF tools are created with the non-technical user in mind. The interface is extremely user friendly, which also makes it into an excellent teaching aid (more on this in section 5). The project manager creates a project, defines the evaluation task and uploads the translation file(s). The evaluators receive an email and begin the task. When the task is completed, the project manager receives an email asking to review the results. After clicking through, automatically generated reports are provided. Data can also be downloaded to create customized reports. The project manager can discuss the findings with the evaluators or compare the results to previous findings.

3.4.1. MT ranking and comparison

The Comparison Task helps users select MT engines or human translators based on the quality of the output. DQF limits the number of sources you can compare to three. Shared experience at TAUS member companies has shown that an evaluator's ability to make robust judgments is impaired if he or she has to score more than 3 options segment-by-segment. After the translation files are uploaded, evaluators are invited to compare the translated segments and to give a ranking. The tool randomizes the order in which the target segments are presented. This means the evaluator(s) do not get conditioned into giving anticipated rankings.

At the end of the task, the project manager can see which engine or translator yields better results for a certain language combination on a given text-type. Users can also gain insight into common errors.
3.4.2. Productivity testing

Post-editing productivity testing is becoming one of the most practical ways of generating evaluation scores. This evaluation type enables you to assess the difference in speed between MT post-editing and translating from scratch. This DQF tool removes half the target side (MT output) segments from your uploaded file(s). Users therefore have to translate half the segments from scratch and edit the other half. The system measures the time taken to complete these tasks. When assigning the task to users, you need to specify which of the two types of post-editing is required (i.e. light or full).

![Figure 4: Productivity results in the DQF reporting tool](image)

The results provide insight into the difference in time and effort between light and full post-editing. Users will also learn about the impact of certain errors on translation quality, the variance across languages and content types, the correlation with certain metrics and scores or the influence of the translator's profile (age, gender, experience, etc.) on post-editing. This test can also be used to compare different MT engines in a more indirect way. DQF also offers the possibility to post-edit the whole text offering time-measurements and edit-distance information at the end of the evaluation project making DQF a simple, user-friendly post-editing environment with productivity reporting.

3.4.3. Error typology

Error typology is the standard approach to quality evaluation currently. There is some consistency in its application across the industry, but there is also variability in categories, granularity, penalties and so on. The DQF error typology tool offers a standardized way to categorize and count translation errors using commonly used industry criteria for accuracy, language, terminology, style and country standards. The DQF error typology was developed by considering existing error-count metrics (such as the LISA QA Model). Another example of an error typology is the Multidimensional Quality Metrics (MQM) developed in the European QT Launchpad project and owned by DFKI.

Tracking and comparing the errors found in computer-generated translations offers insights into the weaknesses of MT engines and MT in general. Besides, a comparison of SMT with RBMT based on an error typology can be an interesting exercise that makes the differences between the two types engines more tangible for users. TAUS has published best practice guidelines on the error-typology approach. These guidelines enable users to adopt standard approaches to error typology evaluation, ensuring a shared language and understanding between translation buyers, suppliers and evaluators.
As of September 2014, TAUS and DFKI have started the harmonization of DQF and MQM with the aim of bridging the gap between the definitions and specifications of the two models. TAUS acts as the industry outreach platform for the harmonized model for all stakeholders: translators, language service providers, translation buyers, government institutions and NGOs. TAUS will continue offering access to the harmonized models through direct contacts with its membership as well as through partnerships with other associations and members. Both TAUS and DFKI has decided to reach out to and work with standards organizations like ISO and ASTM to share the harmonized model and offer the agreed specifications in the standardization process.

3.4.4. Adequacy/Fluency

This evaluation type is in use in machine translation evaluation and can be equally adopted for human translation quality evaluation. It involves measuring two text quality attributes:

- **Adequacy:** “How much of the meaning expressed in the gold-standard translation or the source is also expressed in the target translation.”
- **Fluency:** To what extent the translation is “one that is well-formed grammatically, contains correct spellings, adheres to common use of terms, titles and names, is intuitively acceptable and can be sensibly interpreted by a native speaker”

4. Quality evaluation today

In the summer of 2014, a survey was carried out by TAUS. The aim was to collect user feedback on TAUS DQF and on Translation QE in general. By that time, more than 500 regular and irregular users had been using the DQF tools to evaluate translation quality. The full survey will be published for public consultation on the TAUS website by mid-December 2014.

A third of the participants in this survey were translators themselves or academic staff. They were using DQF as an evaluation tool to evaluate MT output or human translations. In-house staff (the largest percentage of respondents – 38%), freelance translators or linguists using the DQF tools, on the other hand, were doing that commissioned by LSPs or translation service buyers. They were conducting MT comparison/ranking to help decide which MT solution is best.
fit for the purpose. A smaller percentage (5%) of the users at the buyer side outsourced translation QE to another company, usually to another LSP or a specialized linguistic consultancy firm while another 14% chose to outsource evaluation to freelance translators.

Respondents noted that evaluators of translation quality do not always have the necessary skills. Very often untrained translators, interns or bilingual staff members are asked for the job. Training and providing ample information prior to an evaluation project are essential. A video might serve as a first step to provide information and training. The downside of this is that a video might also be difficult to consult as a quick reference on an ongoing basis. Generally, some written instructions are needed. They should be brief and positive (Do's rather than Don'ts). Additional instructions could be provided depending on the translation evaluation purpose. The advantage of written pointers is that they can easily be consulted at any stage. Providing examples can be also very useful.

On the methodology side, the difficulty of differentiating between preferences and errors, the lack of consistency, objectivity and time were pointed out as major problems. Subjectivity is a major concern when it comes to evaluation even when the TAUS DQF tools are used. While DQF offers a transparent and standardized workflow for translation QE, increasing the number of evaluators as well as providing the necessary instructions beforehand, can improve the credibility of the evaluation results. The decision on how detailed evaluation should be is necessary before every evaluation process. One can choose for a monolingual fluency evaluation or a bilingual accuracy evaluation or a combination of these two topped with an error-typology evaluation. Different types of evaluations and different combinations require different amount of time and expertise to conduct.

Finally, participants would like to see a definition emerging for good quality and for different quality levels. They mentioned the lack of transparent evaluation criteria when conducting evaluations in existing tools and metrics. Finding the right metrics already causes problems to some. Others complained about the lack of standardized sampling algorithms. Though human evaluation remains a valuable method for assessing translation quality, it can be time-consuming and expensive. A recent model to obtain quicker and cheaper human evaluations is by means of sampling. Sampling can be used in several situations within the translation workflow; the two main use cases pointed out during the breakout session at the TAUS QE Summit Dublin 2014\(^2\) being Quality Assessment of human translation and machine translation evaluation. Sampling is appropriate in the most evaluation scenarios, but the scenario in which sampling takes place has a strong influence on how the sample will be designed and analyzed. Sampling can also be a useful technique in translator training and continuing development.

It was interesting to see the variety of tools and methods applied by the respondents before starting to use TAUS DQF including automated metrics (WER, BLEU), Excel forms, open source tools (e.g. Appraise), the LISA QA model and in-house tools. TAUS's approach to QE was justified by the recommendation of some respondents that urged the industry to use an agreed set of standards and metrics (such as DQF) since quality is something every Language Service Provider (LSP) offers to clients but without defining or measuring it in a proper way.

\(^2\) [https://www.taus.net/taus-quality-evaluation-summit-2014](https://www.taus.net/taus-quality-evaluation-summit-2014)
5. DQF in training and research

Although translation QE has always been an essential part of the translation process in the industry, it is only now that it's gaining importance in academic research. Translation quality evaluation data enables researchers to answer several of the following questions: what exactly are the key features of good content and how can we measure them? What are the general problems in enabling machines to 'understand' language? Which text types are most amenable to MT? What are the advantages and disadvantages of different MT approaches (RbMT vs SMT)? How can we compare two translations of the same source text in a consistent way? How can a user improve the performance of an MT system? What are the requirements of effective post-editing?

Since DQF is freely available for academic research and education purposes, an increasing number of universities have been using the tools. DQF tools and reports enable researchers to investigate the achievements and limitations of (commercially available) MT systems such as Google translate, Bing etc. They can also assess which text types are suitable for processing with these technologies. And they can also evaluate human translations or compare post-editing to translation from scratch. Although DQF is free for research, large volumes of evaluation data are still missing. Work in the area has been hampered by the lack of availability of relevant data to train metrics. Companies are not keen on offering their data to research purposes even though this type of data is often abundant among providers and buyers of automatic translations, since they routinely need to assess translations for quality assurance. Research on better automatic evaluation metrics would therefore greatly benefit from a closer relationship between industry and academia.

Platforms and tools such as TAUS DQF (Dynamic Quality Framework) can facilitate such collaboration between industry and academia by providing systematic ways of collecting and storing quality assessments (according to specific requirements for a given content type, audience, purpose, etc.) that can be directly used to train metrics. Additionally, quality evaluation and quality estimation could be integrated into such platforms to support human evaluation. Academia needs to obtain more feedback, information and requirements from the industry to better focus research activities on solutions to the problems that the industry is actually facing. Industry also needs better software solutions from academia, both in terms of usability and performance, in order to test the techniques and solutions designed by the industry.

6. Conclusion

Since TAUS launched the Dynamic Quality Framework in 2011, we have learned to apply different methods of QE such as adequacy, fluency, productivity testing and MT ranking. We have also learnt to compare results to previous projects and to minimize subjectivity by using a standardized workflow. What's still missing is benchmarking to satisfy user needs and to provide the right level of quality for each user. In order to develop and improve translation quality, we need to measure quality constantly and consistently. But how can we achieve that when budgets and resources set-aside for this purpose are so tight. How to become efficient in QE? Using DQF tools, users can now research the achievements and limitations of their MT engines. They can assess which text types are most suitable for processing with these technologies. They can also
evaluate human translations or compare post-editing to translation from scratch. The final aim, of course, remains satisfying the customer.

References


Lena Marg, Sharon O'Brien, Attila Görög, Miguel Gonzalez: TAUS Best Practices on Community Evaluation

Luigi Muzii: Quality Assessment and Economic Sustainability of Translation


Sharon O'Brien: Towards a Dynamic Quality Evaluation Model for Translation

Sharon O'Brien: Translation Quality - It's time that we agree
Far from the Maddening Crowd: Integrating Collaborative Translation Technologies into Healthcare Services in the Developing World

Erin Lyons
University of Maryland

ABSTRACT

Crowdsourced and collaborative translation technologies have been at the centre of a heated debate in the translation industry in recent years, as questions have been raised regarding labour practices, the widespread integration of machine translation (MT) as well as concerns regarding quality and professional practices. However, despite the criticism of this emergent technology, the union of collaborative translation platforms and mobile communication technology has bridged a knowledge, resource and communication gap in the developing world, allowing healthcare and medical services to be re-imagined to reach a previously unimaginable community – often instantaneously. The rich data network supplied by mobile phones, when combined with automated data integration, can now be merged with translation services to contribute to initiatives, such as slowing the spread of malaria or stopping stock-outs of life-saving drugs at local clinics. We will take a closer look of the role of translation (machine translation, human translation and controlled language) in some of the leading crowdsourced translation applications, how translators bridge the gap between algorithm and on-the-ground communication and the implications for the development of “lite”, mobile-ready versions of CAT tools and TMs.

1. Introduction

Crowdsourced and collaborative translation technologies have been at the centre of a heated debate in the translation industry in recent years, as questions have been raised regarding labour practices, the widespread integration of machine translation (MT), as well as concerns regarding quality and professional practices. However, despite the criticism of this emergent technology, the union of collaborative translation platforms and mobile communication technology has bridged a knowledge, resource and communication gap in the developing world, allowing healthcare and medical services to be re-imagined to reach a previously unimaginable community – often instantaneously. The rich data network supplied by mobile phones, when combined with automated data integration, can now be merged with translation services to contribute to initiatives, such as slowing the spread of malaria or stopping
stock-outs of life-saving drugs at local clinics. We will take a closer look of the role of translation (machine translation, human translation and controlled language) in some of the leading crowdsourced translation applications, how translators bridge the gap between algorithm and on-the-ground communication and the implications for the development of “lite”, mobile-ready versions of CAT tools and TMs.

2. Computers are Ineffective, People are Inefficient

Crowdsourced and other collaborative translation methods occupy a unique position, regardless of the computational model used, populating the spectrum between machine translation (MT), professional human translation (HT) and open-source innovation, as illustrated in Figure 1. Scalable crowdsourced translation and information processing is still a fairly nascent technological application undergoing continual development, as the cutting edge of the data science industry trickles down and mobile services and capabilities continue to expand – in even the most remote regions of the world. This continual proliferation of mobile and Web-based technology means that speakers of more than 5,000 languages now have real-time access to data and voice communication (Crowley and Chan 2011).

![Figure 1: Computation models rely on a combination of machine translation, open source innovation and crowdsourcing to populate the spectrum between professional human translation and machine translation](image)

Yet, despite the myriad of revolutions in information technology, there has been a gap between technological innovation and the ability to process and understand large amounts of data to support linguistically diverse populations. The volume and velocity of data has overwhelmed the same technology that had been the driver behind this dynamic. Yet, with about 85.5% of the world subscribing to mobile services (versus about 33% using the Internet) (Lorentz 2004), the penetration of mobile phone services and Short Messaging Service (SMS), combined with crowdsourcing and micro-tasking have emerged as forms of collective intelligence in the field of translation, adding value through shared data communication, and transforming technology into a means of including the world’s many under-represented languages. Consequently, mobile technology provides robust architecture that is virtually ubiquitous, allowing crowdsourced translation to become the missing link in a globalised world where
linguistic variation is the norm, enabling global collaboration, humanitarian development and action.

3. On-the-Ground Applications of a Nascent Technology

The “crowd” has emerged as a buzzword in recent years, as virtually all problems – from start-up financing, to searching for missing persons, to voting for Coca-Cola World Cup promotions (Joseph 2014) – seem to be easily resolved via the input of voices from the collective crowd. Crowdsourced translation and the micro-outsourcing of translation- and language-related tasks has served as the momentum behind a wave of inclusion, wherein translation has made our global, multilingual and multicultural society just that much smaller and more closely intertwined.

The role of technology, when combined with the power of translation – whether outsourced to local crowdsourcing translators or combined with professional human translation – has the power to open the flood gates to new voices and strengthen humanitarian action and manage big data. Harnessing digital communication platforms through crowdsourcing translation is the missing link that will become the driver behind optimising and extending humanitarian and social aid in years to come. An examination of four ground-breaking applications of the technology provides insight into current capabilities and achievements when these innovations are applied to cross-lingual communication, as well as the areas of weakness requiring further development and integration. As better synthesised by Robert Munro, the computational linguist and data scientist behind Mission 4636, the first application of crowdsourced translation for humanitarian relief during the 2010 Haiti earthquake, “The future of how we talk to each other is changing and even crowdsourced translation looks nothing now like it did 12 months ago. Who knows how it will look in just a few more years?” (Junglelightspeed 2014).

4. Real-Time Intelligence for Healthcare Services in the Developing World

In recent crowdsourced translation campaigns for healthcare initiatives (Stop Stock-Outs, ProMED-mail, MalariaSpot, etc.), human-centric multilingual systems have been designed around major and minor language processing systems to engage users and non-professional translators to bridge socio-linguistic gaps in framing and contextualising reports from otherwise geographically and/or linguistically excluded populations. Collaborative human translation and interaction add precision and quality control that are otherwise lacking in machine translation, particularly in the context of low-density languages (Eidelman 2011).

5. Aggregating Multilingual Information to Stop Stock-Outs

The Stop Stock-Outs campaign is an initiative based in Kenya, Uganda, Malawi, Zambia and Zimbabwe to gather data from citizens on stock-outs of ten essential medications (first-line antimalarials, zinc, benzathine penicillin, first-line antiretrovirals, metronidazole, ciprofloxacin, amoxicillin, ceftriaxone, cotrimoxazole, ORS) at public health facilities. By gathering data via SMS and aggregating the information on interactive maps, the campaign has raised awareness about health rights and access to essential medicines and fostered a culture of greater community engagement and institutional accountability in the implementation of public health policy.
The Stop Stock-Outs campaign was born out of an epidemic sweeping across many regions of Africa, wherein pharmacies and health centres temporarily had no medicine on the shelf, often lasting several weeks. For example, in Uganda, prior to the 2009 launch of the initiative to combat these supply shortages, only 45.7% of public facilities had a basket of 28 essential medicines. In 2007-2008, stock-outs at public facilities averaged 72.9 days per year, as a result of poor funding, ineffective coordination or drug procurement and distribution, as well as due to gaps in management and pilferage (Medicines 2010).

Armed with the data, activists collaborated with Ushahidi and FrontlineSMS to create a project to map the availability of essential medicines at public health facilities in several African nations in real-time, supplying public and private health centres with vital information, while holding them accountable. By gaining citizen support and participation in the project, through grass-roots campaigns and local media, the campaign was also able to fight the resistance and denial of local officials faced early-on regarding the very existence of the pervasive and life-threatening issue.

In fact, the Kenyan Ministry of Medical Services released a press statement stating, “There are no stock-outs of essential drugs experienced in the country as reported by the media” (Ole Kiyiapi 2009). However, 5 months after the Stop Stock-Outs campaign was launched, including additional momentum gained during “Pill Check Week” (22 - 26 June 2009), garnering the attention of international media because of the innovative use of emerging technology in the developing world, the Minister of Medical Services, Prof. Peter Anyang' Nyong'o admitted on a Kiswahili radio station that this was indeed a potentially lethal problem threatening public health facilities in Kenya and, furthermore, he confirmed to lawmakers that additional funding would be made available to purchase essential medicines and pre-empt further stock-outs (Stopstockouts 2010).

In practice, Stop Stock-Outs was aimed at checking stock levels of essential medicines. These included first-line anti-malarials, zinc tablets, penicillin, and first-line anti-retrovirals (ARVs) to treat HIV/AIDS. All of these medicines are essential in varying degrees to fighting disease and illness, and widely used in the countries where the project was carried out. After visiting clinics and pharmacies, activist-participants used their mobile phones to report their results using structured, coded SMS messages: “x, y, z” – where x represented the country code (Kenya, Malawi, Uganda, Zambia or Zimbabwe), y the district or city and z the medicine that was found to be out of stock. The messages were then received by FrontlineSMS, a free open source software, which would then run an automatic script to validate the data before sending it over the Internet to a Ushahidi-powered website. From there, the results could be automatically aggregated and visually displayed on an interactive map, as seen in Figure 2.
In this project, while crowdsourcing was used to engage the participation of locals in the initiative, the encoded numerical messages served as a workaround to the problem of engaging a population speaking a variety of local languages. The project organisers determined that the availability of real-time information on medicine inventory superseded the need to gather more complex multilingual data from citizens. Instead, information on the initiative could be translated for local health clinics, explaining how to use the service and input data via a mobile network.

By implementing this translation strategy upstream from the crowdsourcing, crowdmapping and data analysis steps, the linguistic barrier was removed, along with the challenge of sourcing a network of on-call crowdsourcing translators. Although, this encoded system also removed the possibility for two-way communication, framing the data around extremely specific data points. Furthermore, since only numerical data was gathered according to a key, the Stop Stock-Outs programme was somewhat error-prone, as a typographical error, too few or too many input numbers or a failure to adhere to the requested code order could potentially generate erroneous and misleading data. Ultimately, the campaign, combining off-line translation and multi-lingual community outreach with crowdsourcing mobile data collection was effective in obtaining real-time reports from those on the ground, thereby including a cross-lingual population on the ground that is sometimes missed with more targeted online viral marketing campaigns.

6. ProMED-mail: An Adaptable Health Initiative with Far-Reaching Potential

Another crowdsourcing-translation initiative launched in the healthcare sector, serving to deliver real-time intelligence on a range of emerging infectious diseases and outbreaks is ProMED-mail. The completely human-based information collection programme, established in 1994, uses reports gathered from local media, experts and eyewitness reports, which are then processed and translated by humans for systems in English, French, Russian and Spanish. This
crowdsourced translation application for global public health differs from many related projects in its ability to organise multilingual open-source data and in its ability to offers users complete and unfiltered control of searching the aggregated data, rather than framing and constructing fixed analysis tools (Madoff 2004).

In the simplest of terms, ProMED-mail is an e-mail list used to gather data and provide early warnings of disease outbreaks around the world 7 days a week, 24 hours a day. Whereas World Health Organization surveillance reports require official clearance before posting, and are then only posted in French and English, ProMED-mail is able to pick up and disseminate reports over the Internet much sooner, through its network of global multilingual moderators (Lyon 2012). Moderators start by triaging reports in their own language(s), rejecting them or forwarding them to subject moderators. The subject moderators then edit the reports and add commentary, subsequently returning the report to the top moderator, who assigns a colour-coded level of urgency (green, yellow or red – the most urgent) (Cowen et al. 2006). Green reports are published within 24 hours, while red and yellow reports receive expedited review. On a typical day, 7 reports are published: 1 red, 1 yellow and 5 green (Madoff 2004).

Since disease and epidemics know no borders, ProMED-mail runs simultaneous language-specific regional networks in Portuguese, Spanish, Russian, English for Southeast Asia, French for Francophone Africa and English for Anglophone Africa. Local languages are translated by regional specialists in infectious diseases, epidemiology and public health into the official language of the local ProMED-mail network. Then, particularly emergent or cross-border reports are translated by a moderator-expert for the main English-language site. Based on the urgency of these reports and the availability of further emerging information, the main English-language page and the local network pages may report translations with slight variations, and may also have been edited to highlight particularly pertinent information for a more global or more local audience. One such example, regarding the recent outbreak of the Ebola virus in West Africa shows two similar, but somewhat divergent reports on the Ebola outbreak, as of 7th April 2014, as per the Francophone and Anglophone Africa pages:

A cumulative total of 151 clinically compatible cases of Ebola virus disease (EVD) including 95 deaths had officially been reported from Guinea. 65 samples were tested of which 34 tested positive for Ebola.

La Guinée avait notifié officiellement un total cumulé de 151 cas cliniquement compatibles, incluant 95 décès. Les investigations se poursuivent au laboratoire de l’Institut Pasteur de Dakar à Conakry (65 échantillons analysés, dont 34 ont été positifs à la PCR pour le virus Ebola) et au Laboratoire mobile de l’Union européenne (EMLab) qui a mis en place une équipe à Guékédou (36 échantillons analysés/20 positifs) [Guinea officially reported a cumulative total of 151 clinically compatible cases, including 95 deaths. Investigations are continuing in the laboratory of the Dakar Pasteur Institute in Conakry (65 tested samples, of which 34 are positive from PCR for the Ebola virus) and in the European Union Mobile Laboratory (EMLab), which has dispatched a team to Guékédou (36 tested samples/20 positive).] (ProMED-mail 2014).
Rather than providing perfect, linear translations of cross-national reports in its epidemiological surveillance, ProMED-mail focuses on rapid-fire, cross-border, multilingual communication, with mapping capabilities and logical information filters (Lyon et al. 2011). It is a sensitive, but not necessarily specific system, using local scientific and linguistic intelligence to pinpoint outbreak vulnerabilities, which may in turn be visually mapped and processed, as shown in Figure 3.

**Figure 3:** ProMED-mail delivers real-time intelligence on emerging diseases and outbreaks, based on reports from local media, experts and eyewitness reports, which are then translated by humans for systems available in multiple languages (Lyon et al. 2011)

Not constrained in its reporting by the need for official clearance, as with the World Health Organization (WHO), ProMED-mail allows a human-based multilingual crowdsourcing project to outperform more automated processing systems. For example, specialists in Ukraine, Russia and Uzbekistan post an average of 2-3 daily reports with expert comments 7 days a week. These selected reports are then translated from the Russian on ProMED-RUS post to the main ProMED-mail English page in an average of 0-4 days, following a translation and editing lag time (Rakhmanova et al. 2014). This far-reaching cross-lingual communication project outperforms WHO reporting, despite its human component, ensuring massive multilingual cover through its crowded resources. In 23 outbreaks on WHO’s Outbreak Verification List (OVL), ProMED-mail reports preceded the OVL listing, versus 5 outbreaks in which the OVL preceded ProMED-mail (Woodall 2001).

In both the Stop Stock-Outs and ProMED-mail crowdsourced translation campaigns for healthcare initiatives and reporting, human-centric multilingual systems designed around major and minor language processing systems were means of engaging users and non-professional translators to bridge socio-linguistic gaps in framing and contextualising reports from otherwise geographically and/or linguistically excluded populations. Such non-profit health campaigns transformed new technology and the buzzword of the new millennium into powerful transcendental tools to use crowdsourced translators as a vehicle for social impact and inclusive change.
7. Monitoring Outbreaks and Epidemics with a Click

Diseases and epidemics know no borders, making collaborative translation projects with integrated data components that deliver real-time intelligence on topics such as emerging infectious diseases and outbreaks invaluable. Crowdsourced translation applications for global public health can be used to organise multilingual open-source data and give users complete and unfiltered control of searching the aggregated data, rather than framing and constructing fixed analysis tools. Custom tools supply alternatives to perfect, linear translations of cross-national reports on epidemiological surveillance, instead focusing on rapid-fire, cross-border, multilingual communication, with mapping capabilities and logical information filters. It is a sensitive, but not necessarily specific system, using local scientific and linguistic intelligence to pinpoint outbreak vulnerabilities.

8. Conclusions

Crowdsourced translation initiatives in the healthcare field in the developing world, as with Stop Stock-Outs and ProMED-mail, have all contributed to pioneering and co-innovating crowdsourced translation as a tool for inclusion, access and assistance, quickly assembling the data, know-how and resources necessary to carry out a task or solve a problem by facilitating multi-channel communication among people and organisations, collaborating across disciplinary, geographic and linguistic boundaries. However, despite reductionist descriptions of crowdsourcing in the context of language processing, this type of translation is not a singular task, but a translation eco-system, combining human and automated work-flows and interfaces to harness professional, crowdsourced and machine translation that remain inclusive and flexible enough to motivate contributors and engage broad general participation in such relief, development and aid initiatives.

Crowdsourced translation will continue to have varied and cross-platform applications, some of which we can likely not even predict at present; however, what remains clear is that a connected world means translation is now a much broader and scalable possibility, regardless of the number of speakers, the far-flung location or the lack of or blocked social-media footprint. We are already speaking to one another cross-linguistically and this is changing very quickly, as are the means used to do so. The place for MT, HT and crowdsourced translation in humanitarian work and social initiatives will continue to be re-shuffled in coming years as technology gains precision and humans are better able to integrate automated technology into existing work-flows. Regardless, the outlook is optimistic for the harnessing of mass communication to serve as a mechanism of inclusion through translation and cross-lingual communication in a growing digital world that has made our geographic footprint that much smaller.

References


Medicines Transparency Alliance (2010) “Case Study: Stop Stock-outs Campaign A CSO initiative to increase access to medicines”.


Is Machine Translation Ready for Literature?

Antonio Toral  
Dublin City University

Andy Way  
Dublin City University

ABSTRACT

Given the current maturity of Machine Translation (MT), demonstrated by its growing adoption by industry (where it is mainly used to assist with the translation of technical documentation), we believe now is the time to assess the extent to which MT is useful to assist with translating literary text. Our empirical methodology relies on the fact that the applicability of MT to a given type of text can be assessed by analysing parallel corpora of that particular type and measuring (i) the degree of freedom of the translations (how literal are the translations) and (ii) the narrowness of the domain (how specific or general that text is). Hence, we tackle the problem of measuring the translatability of literary text by comparing the degree of freedom of translation and domain narrowness for such texts to texts in two other domains which have been widely studied in the area of MT: technical documentation and news. Moreover, we present a pilot study on MT for literary text where we translate a novel between two Romance languages. The automatic evaluation results (66.2 BLEU points and 23.2 TER points) would be considered, in an industrial setting, as extremely useful for assisting human translation.

1. Introduction

The field of Machine Translation (MT) has evolved very rapidly since the emergence of statistical approaches two decades ago (Brown et al., 1993). MT is nowadays a reality throughout the industry, which continues to adopting this technology as it results in improved translation productivity, at least for technical domains (Plitt and Masselot, 2010).

Having reached this level of maturity, we explore the viability of current state-of-the-art MT for literature, the last bastion of human translation. To what extent is MT useful for literature? At first glance, these two terms (MT and literature) might seem incompatible, but the truth is – to the best of our knowledge – that the applicability of MT to literature has not been studied rigorously from a empirical point of view.

2. Background

The first work on MT for literature we are aware of (Genzel et al., 2010) translates poetry by constraining a SMT system to produce translations that obey to particular length, meter and rhyming constraints. Form is preserved at the price of producing a worse translation. However, this work does not study the viability of MT to assist with the translation of poetry.
The only other work on MT for literature we are aware of (Besacier, 2014) presents a pilot study where MT followed by post-editing is used to translate a short story from English to French. Post-editing is performed by non-professional translators and the author concludes that this pipeline can be useful as a low cost alternative to translate literary works to a broad number of languages at the expense of sacrificing translation quality.

3. Methodology

The applicability of statistical MT (SMT) to translate a given type of text for a given pair of languages can be studied by analysing two properties of the relevant parallel data.

1. Degree of freedom of the translation. While literal translations can be learnt reasonably well by the word alignment component of SMT, free translations result in problematic alignments.
2. Narrowness of the domain. Constrained domains lead to good SMT results. This is due to the fact that in narrow domains lexical selection is not really an issue and relevant terms occur frequently, which allows the SMT model to learn their translations accurately.

We conclude that the narrower the domain and the smaller the degree of freedom of the translation, the more applicable SMT is. This is why SMT performs well on technical documentation while results are substantially worse for more open and unpredictable domains such as news (cf. WMT translation task series\(^1\)).

We suggest to study the applicability of SMT to literary text by comparing the degree of freedom and narrowness of parallel corpora for literature to other domains widely studied in the area of MT (technical documentation and news). Such a corpus study can be carried out by using a set of automatic measures. The degree of freedom of the translation can be approximated by the perplexity of the word alignment. The narrowness of the domain can be assessed by using measures such as repetition rate (Bertoldi et al., 2013) and perplexity with respect to a language model (Ruiz and Federico, 2014).

Therefore, in order to assess the translatability of literary text with MT, we put the problem in perspective by comparing it to the translatability of other widely studied types of text. Instead of considering the translatability of literature as a whole, we root the study along two axes:

1. Relatedness of the language pair: from pairs of languages that belong to the same family (e.g. Romance languages), through languages that belong to the same group (e.g. Romance and Germanic languages of the Indo-European group) to unrelated languages (e.g. Germanic and Sino-Tibetan languages).
2. Literary genre: from novels to poetry.

We hypothesise that the degree of applicability of SMT to literature depends on these two axes. Between related languages, translations should be more literal and complex phenomena (e.g. metaphors) might simply transfer to the target language, while they might have more

\(^1\)http://www.statmt.org/wmt14/translation-task.html
complex translations between unrelated languages. Regarding literary genres, in poetry the preservation of form might be considered relevant while in novels it may not.

As a preliminary study, we evaluated the translation of a recent best-selling novel for a related language pair (Spanish to Catalan). The scores obtained – 66.2 BLEU (Papineni et al., 2002) points and 23.2 TER (Snover et al., 2006) points – would be considered, in an industrial setting, as very useful for assisting human translation (e.g. by means of post-editing or interactive MT). We expect these scores to generalise to other related language pairs such as Spanish–Portuguese or Spanish–Italian. 

4. Conclusion

In summary, we have proposed a methodology to assess the applicability of MT to literature which aims to give an indication of how well SMT could be expected to perform on literary texts compared to the performance of this technology on technical documentation and news. While we may be far from having MT that is useful to assist with the translation of poetry between distant languages such as English and Chinese, we have provided evidence that state-of-the-art MT can already be useful to assist with the translation of novels between related languages.

References


The lexical similarity between Spanish and Catalan (0.85) is close to that between Spanish and Italian (0.82) and Spanish and Portuguese (0.89). http://en.wikipedia.org/wiki/Lexical_similarity
ABSTRACT

We describe a prototype platform for creating multilingual voice questionnaires. Content is defined using a simple form-based language with units for questions, question-groups and answers; questionnaire definitions are compiled into efficient speech recognition packages and tables, and the resulting applications can be deployed over the web on both desktop and mobile platforms. We sketch our initial questionnaire application, which is designed for gathering information related to availability of anti-malaria measures in sub-Saharan Africa. It contains 114 question-groups and 218 questions.

1. Introduction

There are many circumstances where it is potentially useful to be able to administer multilingual voice questionnaires. A familiar example in Western society is admission to the accident and emergency room of a hospital: the nurse on duty will most likely start by asking for personal details, the nature of the patient’s immediate problem, previous medical history, and so on. If the nurse and the patient do not share a common language, difficulties arise. Another example, which will occupy us more in this paper, is information gathering for demographic and health topics (DHS Program 2014).

We describe an easy-to-use architecture, inspired by the RAMP framework (Salihu 2013) which can be used to generate voice questionnaires of this type. Our questionnaires are deployed on mobile platforms — smartphones, tablets or laptops — linked over a 3G connection to a remote server. The person administering the questionnaire chooses the next question by speaking it in their own language. The application uses speech recognition, performed on the server, to identify it, speaks a pre-recorded translation in the respondent language, and displays a set of icons on the touch-screen. The respondent answers by pressing on the icons; each icon has an associated voice recording, in the respondent language, identifying its function.

In the rest of the paper, we briefly sketch the architecture, focussing on the formalism used to define questionnaires, and present an example questionnaire application.
2. Architecture of the system

The questionnaire designer specifies the questionnaire using a single file written in a simple formalism which supports three types of unit: Groups, Questions, and Answers. A Group specifies a top-level item on the questionnaire, a list of permitted Fillers, and a pointer to the next Group. A Question specifies one possible way to attempt to assign a Filler to a Group; it defines the Group which the Question belongs to, a list of surface realization of the question in the questionnaire administrator's language, a translation in each target language, and a list of permitted Answers, each one optionally associated with a Filler. An Answer defines an associated translation for each target language. The questionnaire description is compiled into an accurate limited-vocabulary speech recogniser and a set of tables. It is deployed over the web using methods developed at Geneva University and elsewhere in the context of various speech translation and CALL projects (Rayner et al 2006, 2014; Fuchs et al 2012).

To give a simple example of the formalism, the following lines specify a Group called Religion, which belongs to a questionnaire called toy_questionnaire.

Group
Questionnaire toy_questionnaire
Name Religion
PrintName What is your religion?
Code REL
Fillers christian muslim no_religion
Next WhichIslam If muslim
Next MotherTongue If christian no_religion skip
EndGroup

The other fields have the following meanings. PrintName is the title displayed at the top of the screen when the Group is reached, and Code is the identifier used to store the result when the filled questionnaire is printed out. There are three possible ways to fill the Religion slot: christian, muslim and no_religion. Once the slot has been filled, the questionnaire moves to the Group WhichIslam (what type of Muslim) if the selected filler was 'muslim’, and moves to the Group MotherTongue on other fillers.

A Group will normally have several Questions associated with it; indeed, the value of the tool, compared to an application which mechanically asks a set of questions in a fixed order, is that it allows human judgement to be used to select an appropriate questions. In the current case, one Question might involve directly asking the interviewee what their religion is:

Question
Group Religion
Variant what is ( your | your family’s ) religion
Variant what religion do ( you | your family ) belong to
Translation/French Quelle est votre religion ?
Translation/Spanish ¿Cuál es su religión?
Answers christian muslim no_religion
EndQuestion

Here, the Variant lines define different ways for the (English-speaking) interviewer to ask the question, the Translation lines define how it will be asked in the different target languages covered, and the Answers line lists possible responses; as shown, it is possible to use regular expressions in the Variant lines in order to list different forms more compactly.

Similarly, another question might instead ask the interviewee if they are Muslim:

Question
Group Religion
Translation/French Êtes-vous musulman?
Translation/Spanish ¿Es usted musulmán?
Variant are you ?a muslim
Answers yes=muslim no
EndQuestion

Here, the element yes=muslim in the Answers line indicates that the answer yes should be interpreted as indicating the filler muslim. Evidently, the situation may make some questions more appropriate than others; in certain cases, the difference may be very important.

The surface forms of the answers are defined by Answer units; a typical example is the following, which specifies how to realize the Answer yes.

Answer
Content yes
PrintName Yes
Translation/French Oui
Translation/Spanish Sí
EndAnswer

3. Example questionnaire application and user interface

Our initial case study has used a questionnaire consisting of 114 Groups and 218 Questions, designed for gathering information related to availability of anti-malaria measures in sub-Saharan Africa. The overall goal is to measure nine internationally recognized vector control indicators that can be generated through household surveys. Some of the questions are about preventive measures — namely, the use of mosquito nets and indoor residual spraying — and bear a more immediate relevance to such indicators; others are about the characteristics of the household’s dwelling unit and are important to establish a connection between the socioeconomic status of the population and public health issues. All questions were taken from two model questionnaires recommended by the cluster of humanitarian actors that developed the indicators, which include the World Health Organization and the Roll Back Malaria Partnership.

When using the application, the first step is to choose the specific questionnaire the user wants to administer and the language of the interviewee. At the moment, the available choices
are to administer malaria-relevant modules of the Multiple Indicator Cluster Survey (MICS) or the Malaria Indicator Survey (MIS) in French or Spanish. Adding other target languages more suitable for the intended context of application would only involve writing extra Translation lines in the relevant Question definitions.

Figure 1: User interface example, main tab. The current Group, “HC4: Main material of the roof” is shown at the top; the user has just pressed the Help button (bottom right) to get a list of example questions (bottom center). They can ask a question by holding down the Input button (top right) and speaking. The interface shows the recognition result. If this is correct, the questionnaire administrator switches to the PROVIDE ANSWER tab.

Figure 2: User interface example, PROVIDE ANSWER tab. The top panel shows the interviewee language translation of the question (here in French). The clickable buttons at the bottom are the available answers, and are also displayed in the interviewee language.
After choosing the questionnaire and the language, the user is directed to the operative tab of the application (Figure 1), where they can start administering the questionnaire. The title at the top of the screen corresponds to a question in the original paper-based questionnaire. Just as in any conventional situation, the user can read the prompt out loud or use their own judgement to rephrase the question or ask something else. The prototype allows for some degree of syntactic and lexical reformulation, as well as for a number of relevant alternatives, including all possible yes-no questions stemming from predefined answers. For instance, instead of asking “What is the floor made of?” the interviewer can opt for “Is the floor made of wood?” if he or she deems it more appropriate in view of the context. When the interviewer has spoken the question, it is displayed in the Provide Answer tab together with the available answers (Figure 2).

The application reproduces the sequence structure of the original questionnaire, so the next question is always determined by the answer of the question at hand. For instance, the application will automatically skip question HC12 in the MICS questionnaire (“How many hectares…?”) if the respondent gives a negative answer to HC11 (“Do you own any agricultural land?”). Similarly, the application will force the interviewer to probe for additional information whenever it is required. All questions and answers are registered and shown in the show history tab, and there is a final tab, instructions, that contains question-specific guidelines for the interviewer.

This application is intended for use in a context where household surveys are conducted by mixed teams of international and local interviewers, as it is often the case for this type of programme-monitoring surveys. A mobile device integrating speech recognition and automatic translation can help international organizations reduce the need of training ad hoc teams of local people or having to rely on unqualified interpreters.

References


Translating implicit elements in RBMT

Irina Burukina
ABBYY
RSUH, Computational Linguistics
department

ABSTRACT

The present paper addresses MT of asymmetrical linguistic markers, in particular zero possessives. English <-> Russian MT was chosen as an example; however, obtained results can be applied to other language pairs (English – German / Spanish/Norwegian etc.). Overt pronouns are required to mark possessive relations in English. On the contrary, in Russian implicit possessives are regularly used, thus making it very important to analyze them properly, not only for MT but also for other NLP tasks such as NER, Fact extraction, etc. However, concerning modern NLP systems the task remains practically unsolved. The paper examines how modern English <-> Russian MT systems process implicit possessives and explores main problems that exist concerning the issue. As no SB approach can process IP constructions properly, linguistic rules need to be developed for their analysis and synthesis; the main properties of IPs are analyzed to that end. Finally, several rules to apply to RB or model-based MT are introduced that help to increase translation accuracy.

The present research is based on ABBYY Compreno © multilanguage NLP technologies that include MT module.

1. Introduction

MT industry has been recently growing and delivering ever better results. However, several crucial problems remain that prevent us from saying The perfect MT is achieved. Among them there are inherent linguistic problems: bilingual lexical ambiguity, bilingual structural ambiguities, structural asymmetries etc. [Hutchins 2007]

One of the possible reasons for structural asymmetry is zero elements (zero subjects, determinants, etc.) allowed in one languages and prohibited in others. For example, Spanish allows zero subjects, while in English overt pronouns are needed. Spanish -> English translation thus requires reconstruction of appropriate overt pronouns and English -> Spanish translation should include deletion of explicit elements, mostly at the beginning of the sentences.

1. Marco calentó el agua del té. Ahora tiene miedo de quemarse.

Marco warmed water for tea. Now he is afraid to burn himself.
The present research examines how modern MT systems can deal with sentences with structural asymmetry. In particular, we focus on possessive markers in English - Russian language pair as one of the least studied problems.

Implicit possessives (IPs) are zero possessive pronouns, used with inalienable nouns (kinship terms and body-part nouns) in the positions that can be occupied by overt possessives (pronominals or reflexives). The high frequency of inalienable nouns (for example, 1200.6 IPM for the word *рука* 'hand', 484.1 IPM for the word *отец* 'father' [Lyashevskaya, Sharoff 2009]) increases frequency of IP constructions; therefore it becomes crucial for MT system efficiency to take them into account.

This research is based on Compreno multilanguage NLP technologies that include but are not limited to model-based MT. Compreno provides opportunities for NER, Fact extraction, ontology creation etc. [Zuev et al. 2013] It turns out that taking structural asymmetry and IP into account is also important for these tasks and in particular for:

- Text analysis and situation modeling (including interpretation of elliptical structures):
  2. Петя позвонил маме, и Маша тоже.
     Peter.NOM called mother.DAT and Masha.NOM too
     ‘Peter called his (Peter's) mother and Mary called her (Mary's) mother.’

- Anaphora resolution;

- (Co)reference resolution:
  3. Петя позвонил маме. Машина мама всё слышала.
     Peter.NOM called mother.DAT Mary.PossADJ mother.NOM everything heard
     ‘Peter called his mother {PERSON-1}. Mary's mother {PERSON-2} heard everything.’ The problem of IPs remains almost unsolved. To approach the problem, we have analyzed how the most well-known in Russia SB, RB and model-based MT systems process IPs.

2. **Automatic translation of implicit possessives in English - Russian**

Processing IPs correctly is essential for English <-> Russian MT systems. As Russian IPs are not unique, the present research can also help to improve MT of other language pairs such as Norwegian <-> English, Spanish <-> English, German <-> English, Russian <-> French etc.

The problem of IPs can be broken down into two primary tasks.

First, in English -> Russian translation overt pronouns should be deleted to get more accurate results. IPs in Russian in many contexts are not only allowed but preferred, especially with body-part nouns.
4. Петя сломал ногу.

Peter broke leg.ACC

‘Peter broke his leg.’

?Петя сломал свою ногу.

Peter broke hisSELF leg.ACC

‘Peter broke his leg.’

Second, in Russian -> English translation explicit possessives should be synthesized instead of implicit ones considering properties of their antecedents. Anaphora resolution for IPs in Russian is necessary to synthesize appropriate overt pronouns in English.

5. Девочка любит маму.

Girl loves mother.ACC

‘The girl loves her/*his/*their mother.’

We have analyzed the most well-known in Russia English <-> Russian machine translation systems (rule-based as well as statistics-based) and, unfortunately, all of them showed rather poor results analyzing constructions with IP.

6. Петя подошёл к маме.

Peter.NOM came up to mother.DAT

‘Peter came up to his mother.’

(Google, SBMT) Peter went to my mother.

(Yandex, SBMT) Petya went to her mother.

(SYSTRAN, Hybrid MT) Pete approached the mom.

(PROMT, RBMT) Petya approached to mother.

(Compreno, Hybrid MT) Petya walked over to the mother.

7. Peter will be happy if Masha talks to her mother.

(Google) Питер будем рады, если Маша разговаривает с материю.

‘Peter (we) will be glad if Masha talks to mother.’

(Yandex) Питер будет счастлив, если бы Маша говорит ей мать.

‘Peter will be happy if Masha talks to her (non-reflexive) mother.’

(SYSTRAN) Питер будет счастливо, если Masha говорит к ее матери.
Peter will it be happy if Masha talks towards her mother.’

(PROMT) Питер будет счастлив, если Маша будет говорить со своей матерью.

‘Peter will be happy if Masha is talking to her (reflexive) mother.’

(Compreno) Питер будет счастлив, если Маша будет говорить со своей матерью.

‘Peter will be happy if Masha is talking to her (reflexive) mother.’

As is evident from these examples, SBMT cannot process IP properly. It leaves no opportunity for anaphora resolution and appropriate overt pronouns deletion. Altogether we tested almost one hundred typical IP examples, and only most frequently used collocations (idioms) were processed correctly by Google and Yandex MT systems.

8. a. Петя засунул руку в коробку.

(Google) Peter put his hand into the box.

(Yandex) Peter put his hand in the box.

b. Петя надел на голову шляпу.

(Google) Peter put on his head a hat.

(Yandex) Petya put the hat on his head.

c. Peter put his hand into the box.

(Google) Петр положил руку в коробку.

‘Peter put his hand into box.’

(Yandex) Петр сунул руку в коробку.

‘Peter put his hand into box.’

Current research includes several experiments on bilingual English – Russian corpus attempting to automatically extract sentences with IP. Both recall and precision are insufficient; lots of “noise” sentences were found.

As for the modern RB or model-based MT systems, most of them usually contain no rules for IP or lack accuracy. The present paper provides a detailed description of ABBYY Compreno MT system as an example.

3. The Compreno system

ABBYY Compreno MT system is part of ABBYY Compreno multilanguage NLP technologies.

Compreno represents a hybrid model-based MT system. The core of the system is the Universal Semantic Hierarchy (USH), a thesaurus-like structure with universal semantic concepts
as its nodes, that provides semantic analysis of the text. Complete syntactic analysis represents the second part of the technology.

While translating the input text, Compreno analyzes syntactic structures of each sentence that are further converted into intermediate semantic structures. During synthesis semantic structures are converted into output sentences. All the structures created by the system are trees. The crucial idea for Compreno is that while syntactic (surface) structures are language dependent, at the semantic (deep) level nodes of the trees are concepts from the USH, and arcs are universal relations between these concepts. Apart from this, the system establishes non-tree links between nodes of the tree to represent anaphora, coordination, etc.

Compreno is claimed to be a system applicable to any natural language. Language-dependent rules have to be created anew; however, universal part of the system fits different language models.

Below we give examples of syntactic semantic structures for the sentence *The boy showed his file to the girl*. For more information on the project see [Anisimovich et al. 2012, Zuyev et al. 2013].

![Figure 1: Syntactic structure](image1)

![Figure 2: Semantic structure](image2)

Normally, special diathesis is used to convert language specific syntactic structures to universal semantic structures. Diathesis is the description of correspondence between elements of two structures, for example, between syntactic positions and semantic slots (syntactic subject = semantic agent etc.). However, sometimes it becomes very difficult or almost impossible to get a semantic structure that can be further converted to the output syntactic structure because of language asymmetries. In these cases at the analysis stage one or several syntactic positions can
be removed, replaced or added to get an appropriate result. This is done with transformation rules.

Transformation rules (TR) apply to a part of a syntactic tree. Usually TR consists of several productions, describing different possible operations on syntactic components. The system tries to implement each of these productions one by one. If attempt is successful the system passes onto the next TR. The productions can change properties of the components, replace them, delete or add new ones.

As for the possessive constructions, TRs mostly describe external possessors. [Bogdanov, Leontyev 2013] For the implicit possessives analysis there is only one TR. Its productions describe several particular contexts in which Russian IP most commonly appear. At the synthesis stage the system allows deletion of overt possessive pronouns only for body-part nouns.

During Russian -> English translation Compreno restores overt possessives in several cases listed below.

I. Body-part noun with IP is a semantic orientational locative (Locative_PartAsOrientation); normally agent is chosen as antecedent, however if the main verb describes movement of an object by someone, the object becomes 'candidate antecedent':
   9. a. Мальчик лежал головой к окну. -> The boy was lying his head to the window.
      Lit.: Boy was lying head.INSTR to window.
   b. Мальчик поставил стол ножками вверх. -> The boy set a table with his legs upward.
      Lit.: Boy set table.ACC legs.INSTR upwards.

II. External body-part noun (нога ‘leg’, зубы ‘teeth’, хвост ‘tail’) with IP is a semantic instrument, locative or initial point locative; subject is chosen as antecedent:
   10. a. Мальчик расколол орех зубами. -> The boy cracked a nut with his teeth.
      Lit.: Boy cracked nut.ACC teeth.INSTR
   b. Мальчик держал в руках мяч. -> The boy kept a ball in his hands.
      Lit.: Boy kept in hands ball.ACC

III. Body-part noun with IP is a semantic object for verbs like ломать, трогать (‘break’, ‘touch’) and there is no external possessor; subject is chosen as antecedent:
   11. Мальчик сломал ногу. -> The boy broke his leg.
      Lit.: Boy broke leg.ACC

IV. Kinship term with implicit possessive is coordinated with non-kinship term; the last one is chosen as antecedent:
12. Президент с супругой приехали в Москву. -> The president and his spouse arrived to Moscow.
Lit.: President with wife arrived in Moscow

During English -> Russian translation Compreno deletes overt possessives only for body-part nouns if a body-part noun is a semantic orientational locative, semantic instrument, semantic locative inside the noun group with *with*, semantic locative for possession verbs.

13. a. The boy lies with his head to the window -> Мальчик лежит головой в окно.
Lit.: The boy lies head.INSTR into window
b. The boy nodded his head. -> Мальчик кивнул головой.
Lit.: The boy nodded head.INSTR
c. She fed the baby with her breast -> Она накормила ребёнка грудью.
Lit.: She fed baby breast.INSTR
d. A boy with a ball in his hand. -> Мальчик с мячом в руке.
Lit.: Boy with ball in hand
e. She had pills in her hands -> У неё были таблетки в руках.
Lit.: Of her were pills in hands

Compreno's TRs for IP are insufficient and need to be specified and expanded. To do this we should first provide the profound linguistic study of the main properties of IPs and IPs constructions.

4. **Definition and main properties of implicit possessives**

As we have said before, implicit possessives (IP) are zero possessive pronouns, used with inalienable nouns (kinship terms and body-part nouns) in the positions that can be occupied by overt possessives (pronominals or reflexives). They can be used as either deictic elements (when they point out entities in non-linguistic ‘real’ context) or proper anaphors (when they are coreferential with entities mentioned in the same text).

14. a. Мамы нет в доме.
Mother.GEN not in house.PREP
‘My mother is not in the house.’
b. Что случилось? Рука болит?
What happened Hand.NOM aches
‘What happened? Is your hand aching?’
As was mentioned earlier, Russian IPs are not a unique phenomenon. Zero possessive markers appear rather frequently in different world languages, European and non-European alike. Several examples are listed below:

- Implicit possessives with body-part nouns in Norwegian:
  15. Han stakk hendene i lomma.
  
  He put hands.DEF in pocket.DEF
  
  ‘He put his hands in the pocket.’ [Loedrup 2010]

- External dative possessors with body-parts nouns in German:
  16. Sie wäscht sich das Gesicht.
  
  She washes self face.DEF
  
  ‘She washes her face.’ [Loedrup 2012]

- Zero possessive markers with kinship terms in Dogon:
  17. u ba
  
  you father

  ‘your father’ [Haspelmath 2008]

Current research studies the behavior of noun phrases (NP) with IPs in different contexts in Russian and introduces the main properties of IPs that are listed below.

- The antecedent of IP (i.e. possessor) should be human;
- IP can point out persons in non-linguistic “real” context. However, they can indicate only the speaker (used in assertive direct speech) or the hearer (used in imperatives or questions) (1a,b);
- Distance between the NP with IP and its potential antecedent in fact doesn't matter much. IP can be locally bound by its antecedent, located in the same minimal clause. It can also be coreferential with element in another clause. A local antecedent is preferable. However, if it's not a subject then the semantically appropriate subject of the matrix clause should be chosen:
  18. Маша хотела, чтобы мама позвонила Пету.

  Mary.NOM wanted that mother.NOM called Peter.DAT.

  ‘Mary wanted her mother to call Peter.’

- The preferable antecedent is the subject of the clause;
5. Development of analysis/synthesis rules for implicit possessives

5.1. How general can these rules be?

The initial idea was to propose a small number of the most general rules. However, it proved infeasible. There are several problems that hold us back from analyzing every sentence unambiguously without exceptions. Large-scale MT systems like Compreno cannot risk the precision of translation even to obtain higher recall, more literary results and economical set of TRs.

First, it is impossible to formalize pragmatics. Implicit possessives are essentially ambiguous. In many cases (especially in direct speech) they may be used in deictic function to denote the speaker himself.

20. Вчера студенты позвонили матери.

Yesterday students called mother.DAT

‘Yesterday students called my mother.’

Second, even though we have analyzed IPs in different contexts and discovered their main properties, these properties should be considered regularities, but not strict rules. There are always exceptions that can be crucial for the evaluation.

21. a. Маша попросила Петю позвонить маме.

Masha asked Peter to call mother.DAT

‘Masha asked Peter to call her mother.’

b. Учительница попросила Петю позвонить маме.
Teacher asked Peter to call mother.DAT

‘The teacher asked Peter to call his (Peter’s) mother.’

Third, semantics of predicates should also be taken into account.

5.2. What rules can be proposed to improve translation

Nevertheless, we propose several rules that help improve the translation.

The present paper shows several rules that should be used during analysis stage to improve the results of Russian -> English MT. What is even more important is that these rules can be applied to the Russian IP analysis in general.

- In direct speech first person possessive pronoun should be inserted with kinship terms and body-part nouns; in imperative sentences and questions second person possessive pronoun should be inserted;
- If a body-part noun with IP is the subject of subordinate clause, the subject of the main clause should be considered as preferable antecedent if there is no external possessor construction), the appropriate overt pronoun should be inserted;
- If a kinship term with IP plays the ContrAgent semantic role of “social interaction” predicate, the subject of that predicate should be considered preferable antecedent, and the appropriate overt pronoun should be inserted:

22. Петя спорил с мамой.

Peter.NOM argued with mother.INSTR

‘Peter was arguing with his mother.’

As for English -> Russian MT we discovered a phenomenon that allows us to propose a more general rule. In several cases the deletion of pronoun helps not only to get more literary results, but also to preserve the initial sense of sentences. We argue that this approach can be applied as well to MT of other language pairs. In English there are no possessive reflexives and only pronouns like his, her are used. The sentences with such pronouns can be ambiguous.

23. Bill asked John to call his father.

However, while translating such sentences to Russian, the system should unambiguously identify the antecedent; either reflexive svoi or a pronominal is used. One of interpretations is missed and incorrect translation can be got.

The rule proposed below should be used at the analysis stage during English -> Russian MT.

In simple sentences if the subject (A) and the direct object (B) share the same grammatical and semantic characteristics, the kinship term with overt possessive is used in the same sentence and A or B is chosen as antecedent to that possessive, the pronoun should be deleted.

In composite sentences if the subject of the main clause (A) and the subject or the direct object of the subordinate clause (B) share the same grammatical and semantic characteristics,
the kinship term with overt possessive is used in subordinate clause and A or B is chosen as antecedent to that possessive, the pronoun should be deleted.

24. a. Bill introduced John to his father.

b. Mary asked Jane to call her mother.

c. The boys wanted the girls to show the books to their mothers.

The rule has been successfully applied to ABBYY Compreno MT system and verified.

6. Conclusion

This research is focused on formerly unsolved problem of English <-> Russian MT of implicit possessives however the proposed rules can be applied to other language pairs and can help as well to solve text analysis tasks in general.

The research is carried out on the basis of ABBYY Compreno NLP technologies and the obtained results allow to increase MT system efficiency.

As it turns out to be almost impossible to process IP statistically we had to start with theoretical linguistic research. Investigating main properties of IPs in Russian we managed to formulate several rules for automatic analysis and synthesis of IP constructions. Among them the pronoun deletion rule that helps to get more precise translation of English sentences into Russian or another language allowing zero possessives.

A more detailed study of the IP phenomenon is being carried out. Below we describe the main tasks of our work in progress:

- Investigation of IPs in other languages (Norwegian, German, Spanish, etc.); verification of the “deletion” rule to other language pairs;
- Implementation of proposed rules for other but MT NLP tasks;
- Further analysis of Russian IP including examination of semantics of predicates in sentences with IP.

Acknowledgements

We are particularly grateful to Vladimir Selegey (ABBYY) and Alexey Leontyev (ABBYY) for their help and support. A special thank goes to prof. Yakov Testelets (RSUH) for his help and advices on theoretical part of research.

References


Lødrup, H. 2012. *Inalienables in Norwegian and binding theory*.

http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.50.7636


Losses and Gains in Computer-Assisted Translation: Some Remarks on Online Translation of English to Malay

Mozhgan Ghassemiazghandi
Tengku Sepora Tengku Mahadi

University Sains Malaysia

ABSTRACT

The article begins with a concise study of the significance of the translation technology in modern life as well as machine and computer-assisted translation. It then describes the technology accessible to translators and examines the losses and gains of the used tools in computer-assisted translation such as the electronic dictionaries, and specifically Google translate. The paper studies the influence of the online dictionaries on the professional translator with a view to determining the extent translation done using such dictionaries can be accurate. Loss in machine translation is inevitable due to the differences between English and Malay as they are entirely two different languages and not-related language pairs for translation. The online dictionary and translation software cannot replace the human translator and guarantee high-quality translations, despite their efficiency and outlooks. Online dictionaries and other translation means accelerate and facilitate the translation process only by minimizing the expected time for translation. Combination of electronic technologies with comprehensive knowledge of the translator and translation theory may result in a high-quality translation. Translation software and programs nonetheless, will not replace humans even in the future. As mentioned, the main aim of the paper is to investigate the new technologies in machine translation tools to discover the losses and gains in translation of English to Malay by using online dictionaries. Machine translations employing online dictionaries are compared with the translation done by a human translator to analyze the probable errors in machine-translated texts.

1. Introduction

Nowadays, machine translation (MT) is a significant technology that represents an essential part of natural language processing systems. The quality and the number of specialized dictionaries of the software define the efficiency of machine translation (González-Rubio & Casacuberta, 2014). According to European Commission, one of the world’s largest translation services, the EU contracts 1,750 translators on a full-time basis due to global multilingualism. The EU employs external translation providers that produce almost one-fourth of its output
translation to cope with demand variations. The EU translation services translated more than 1,800,000 pages, which cost about one billion Euros in 2008 (EC, 2009). Given this volume, human translation alone without computer assistance would logically not be practical.

For the purpose of this paper, a human translator assisted only by a computer for specific tasks such as typing and looking up specialized terms and expression in the dictionary is referred to as human translation. Based on the data given above, human translating will be a high-priced and time-consuming method. There are not adequate qualified translators, while the need for superior translation has undeviatingly increased. To minimize this problem, computer and computer technology need to be relied upon to facilitate the computerized translation of extensive numbers of documents. Professional translators should post-edit the automatic translation for an accurate translation (Alabau, Sanchis, & Casacuberta, 2014).

Modern machine translation systems are not capable of delivering high-quality translation without human translator intervention. The human translator must edit the output of machine translation and translation software to improve the quality of translation. Computer-assisted translation (CAT) is an interactive translation process between human and computer. Human translator uses computer software just in order to facilitate and accelerate the translation process (Barrachina et al., 2009).

2. Losses and Gains in Computer Translating

Even the most advanced software does not have the skills of a professional translator and fluency of native speakers. Machine translation is not a straightforward task, as each word might have different meanings according to context. An accurate translation needs a good understanding of context as well as language structure and rules. Computer needs some capabilities to deal with translation difficulties in the same way as human.

A computer can only translate non-ambiguous texts that exist in the computer's dictionary with the same meaning. The output translation qualifies for personal knowledge and not academic aims. Computers translate technical texts quite well, as technical documents have a restricted topic and monotonous style (Precup-Stiegelbauer & Laura-Rebeca, 2013). Météo, a translation system for weather forecasting from English into French is an example of a successful computer translation in the domain-specific and controlled environment (Hutchins, 2001). However, computers are not able to produce a high-quality translation for other texts, which are more interesting and appealing to readers. On the other hand, professional human translator can translate all ranges of text.

Because many words have multiple meanings, the process of translation is made more difficult. Computers translate words based on a one-to-one substitution without considering other possible meanings, whereas humans choose the proper words according to the contexts. A word with different meanings may have various translations based on how the word is being used in the context. For example, the word ‘book’ has different meanings, even though the spelling is the same. Computer just translates it to one word ‘buku’ in Malay and ignores other existing meanings like ‘reservation’. It demonstrates that sense should be taken into account in the translation. Human translators can readily distinguish which meaning best fits the context among the multiple uses of the word. Although, it may still be difficult for them to choose the
best translation, it is still possible with effort. A precise translation necessitates a perception of the text, situation, and various uses of the word in the language to determine the appropriate substitution.

Computers need to distinguish between general vocabulary and specialized words, whereby the former might be culturally influential. Overused words should be avoided in general language, as the variety is highly valued. On the other hand, overused words are allowed in technical translation and specialized terminology, as the consistency is highly valued. Unfortunately, computers are not able to distinguish between general and specialized words (Nitta, 1986; Precup-Stiegelbauer & Laura-Rebeca, 2013). Humans can distinguish between general and specialized use of the words. There are many different terminology databases for specialized terms that help the translator to choose an appropriate word. Computers have an amazing memory as compared to humans, but computers cannot decide and choose the best meaning of the words based on the situation. Computers would be desperately disorganized to deal with general and specialized domains while humans can easily differentiate these two types of texts (Precup-Stiegelbauer & Laura-Rebeca, 2013; Şahin, 2013).

Computers and even novice translators might overlook the differences in meaning and therefore produce poor translation by using inappropriate words. Translators must be familiar not only with the source language, but also with target language and culture in order to produce a useful and reliable translation and deal with any translation difficulties appropriately. An accurate translation also needs to consider the intended audience, regionalism and culture in the total context. Computer is not a native speaker of any language and presumably the cultural knowledge in computer is not comparable with humans. Accordingly, it can be concluded that computers cannot translate like humans, as they are not equipped with the learning potential like humans. The computer is, however, a lifesaver as time is scarce in modern life, and people want to be productive in the shortest period. Computers aid people to accelerate translation work faster than a human translator (Dimitriu & European, 2006; Precup-Stiegelbauer & Laura-Rebeca, 2013).

Google translate is a multilingual translation service to translate written source text to target text that supports 80 languages. Google Translate is the most famous and easily accessible machine translation. Nowadays, Internet increasingly develops over recent years, and Google translate, for example, quickly helps people to get an idea of the foreign language contexts. Google Translate is only applicable to standard and coherent texts, which exist in the computer’s dictionary with the same meanings. (Precup-Stiegelbauer & Laura-Rebeca, 2013).

3. Discussion

Several examples of machine translations, from English into Malay are selected to support the aforementioned ideas. The English texts are extracted from Doctor in the House, The Memoirs of Tun Dr. Mahathir Mohamad. Both the English (Mohamad, 2011) and Malay (Mohamad, 2012) versions of the book are used in this study. The original text is being compared with machine translation and human translation to evaluate the efficiency of translation. Then it will be concluded to what extent the translation is trustable and how much human intervention is necessitated to achieve an accurate translation. Here is the Google Malay translation of some
English paragraphs of Tun Dr. Mahathir Mohammad’s A Doctor in the House. Although the hard copy of the Malay translation was used for comparison, comments made in this paper are general, and the translation texts are not presented here due to space constraint.

<table>
<thead>
<tr>
<th>Original Text</th>
<th>Google Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Though my mother had warned me many times that a doctor’s life meant having to forgo sleep and work odd hours, she and my father seemed silently pleased that I was going to go to college. Moreover, when I was on contract with the office, I had been paid $80 a month. Now that I had become full-time and pensionable, I was being paid RM60 monthly.</td>
<td>Walaupun ibu saya telah memberi amaran kepada saya banyak kali bahawa kehidupan yang doktor bermakna perlu melepaskan tidur dan bekerja jam ganjil, dia dan bapa saya seolah-olah senyap gembira bahawa saya akan pergi ke kolej. Lebih-lebih lagi, apabila saya berada di atas kontrak dengan pejabat, saya telah dibayar $ 80 sebulan. Sekarang bahawa saya telah menjadi sepenuh masa dan berpencen, saya dibayar RM60 setiap bulan.</td>
</tr>
</tbody>
</table>

**Example 1: P. 128 Para. 3, A doctor in the house**

Reading both the original text and the translation, one would get an understanding about the meaning of the text, but the text structure is grammatically poor. A long process is required to edit the text and to achieve an excellent translation. The sense of a few words has not been captured. For example, the word ‘odd’ has been translated to ‘ganjil’ by computer, which is a literal translation, while the correct translation is ‘bekerja tidak tentu masa’. The literal translation, or one by one substitution of words does not seem enough to give an idea to the reader as more explanation is needed to clarify the text. The preposition used in the paragraph has been translated to ‘di atas kontrak’ by computer instead of ‘secara kontrak’. The linking words are missed in Google translation. As discussed earlier, context of culture affects the specific meaning of the language. Computers are not able to analyze the cultural context of Malay language. For example, the word ‘ibu’ has been translated to ‘dia’ that is correct in terms of general meaning but the cultural aspect of respect and politeness is missed in computer translation.

<table>
<thead>
<tr>
<th>Original Text</th>
<th>Google Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>I went to the lab where the second-year students were being introduced to anatomy and asked one of them to uncover one of the corpses. It was a Chinese man and he smelled strongly of formalin. I told myself it was dead and could do me no harm. Slowly, by facing my fears, I managed to reach a point where dead bodies no longer frightened me.</td>
<td>Saya pergi ke makmal di mana pelajar-pelajar tahun kedua sedang diperkenalkan kepada anatomi dan bertanya salah seorang daripada mereka untuk mendedahkan salah satu daripada mayat-mayat. Ia adalah seorang lelaki Cina dan dia berbau kuat formalin. Saya memberitahu diri saya itu telah mati dan saya boleh melakukannya tidak membahayakan. Perlahan-lahan, dengan menghadapi ketakutan saya, saya berjaya sampai ke titik di mana mayat tidak lagi menakutkan aku.</td>
</tr>
</tbody>
</table>

**Example 2: P. 292 Para. 5, A doctor in the house**

This translation is somehow senseless. It is just disordered words that are translated word-by-word by machine. The second sentence has a vague meaning. The structure and formation of the sentences are not natural, e.g. the phrase ‘dia berbau kuat formalin’ should be ‘berbau
formalin yang kuat’ (smelt of formalin). The computer cannot distinguish the different meanings of the words as some words occur in more than one meaning. For example, the verb ‘menyuruh’ has been translated to ‘bertanya’ by computer, while it is not about asking question but it is asking to instruct. The pronoun used in the paragraph has been translated to ‘itu’ (that) by computer instead of ‘ia’ (it). The literal translation of ‘boleh melakukannya’ does not seem enough in the last sentence. More explanation, ‘dan ia tidak mungkin membahayakan saya’ (it is not likely to harm me), is required for transferring the message appropriately. As can be seen, computer translation is not perfect without post-editing of human.

<table>
<thead>
<tr>
<th>Original Text</th>
<th>Google Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>I did not think we could face too many problems establishing these industries if we gave them enough careful thought. But I overestimated the Malaysian capacity to learn how to operate a major industry. At the time it was built, the cement factory was the best in the country. It was located by the sea on the island of Langkawi, off the northwestern coast of Peninsula, close to the limestone hills, which were to be quarried for the plant.</td>
<td>Saya tidak fikir kita boleh menghadapi terlalu banyak masalah mewujudkan industri-industri ini jika kita memberi mereka pemikiran yang cukup berhati-hati. Tetapi saya di luar tafsiran keupayaan Malaysia untuk belajar bagaimana untuk mengendalikan industri utama. Pada masa yang ia dibina, kilang simen adalah yang terbaik di negara ini. Ia terletak di tepi laut di Pulau Langkawi, di luar pantai barat laut Semenanjung, berhampiran dengan bukit-bukit batu kapur yang tidak akan dikuari untuk kilang.</td>
</tr>
</tbody>
</table>

**Example 3: P. 329 Para. 3, A doctor in the house**

The computer translation of the selected text did not translate most of the words appropriately. The computer cannot make a distinction between the functions of the words. Thus, the first sentence has been translated by ‘memberi mereka pemikiran yang cukup berhati-hati’ instead of ‘memikirkannya dengan cukup teliti’ and ‘di luar tafsiran’ instead of ‘terlebih menjangka’. Moreover, the translation even omitted words such as ‘orang’, which refer to Malaysian people whereas other words have not been translated correctly. A word can change the whole meaning, for example, ‘untuk mempelajari cara mengendalikan’ has been translated to ‘untuk belajar bagaimana untuk mengendalikan’ rendering it a stilted translation. Another obvious mistranslation is the affirmative sentence that has been translated with a contrary meaning. The phrase ‘yang akan dikuari untuk’ (will be quarried) has been translated to ‘tidak akan’ (will not) in which the original meaning is lost in computer translation.

<table>
<thead>
<tr>
<th>Original Text</th>
<th>Google Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Their armed forces are well organized with groups of increasing sizes under officers of different ranges wielding ascending orders of power and authority. Their huge armies move and act with precision.</td>
<td>Angkatan tentera mereka dengan baik dengan kumpulan-kumpulan yang semakin meningkat saiz di bawah pegawai julat yang berbeza memegang tampak perintah menaik kuasa. Tentera besar mereka bergerak dan bertindak dengan tepat.</td>
</tr>
</tbody>
</table>

**Example 4: P. 391 Para. 4, A doctor in the house**

The first sentence is a string of ineffective words with inappropriate associations. The phrase ‘pegawai julat yang berbeza’ has been translated by computer to ‘kumpulan menjadi
bertambah besor’, which is a mistranslation and meaningless. Moreover, the Malay translation even omitted verbs, for example, ‘diuruskan dengan baik dalam kumpulan’ has been translated to ‘dengan baik’. Other words have not been translated appropriately as computers cannot differentiate between the senses of the words. The word ‘julat’ is has been used wrongly from the point of syntax and vocabulary. It is an inappropriate translation for ‘pegawai-pegawai dari bahagian yang berbeza’ that emphasizes the power and authority of the army in the context.

malaysia is a developing country with a small domestic market, which means nothing that we produce enjoys economies of scale. To industrialise, we needed to protect our infant industries. Accordingly the national car was protected by a lower excise duty than import duty on foreign car. No one really complained because they were used to paying high prices for imported car. The government did not lose out either as higher sales of Proton and an expanding market earned it increased revenue. At the time it was built, the cement factory was the best in the country. It was located by the sea on the island of Langkawi, off the northwestern coast of Peninsular, close to the limestone hills which were to be quarried for the plant.

Example 5: P. 517 Para. 2

Reading the original text and Malay translation, the meaning of the text is dropped. The original text means the current activity did not benefit from economies of scale, while Google has been translated with the opposite meaning that every production benefits from economies of scale. It is an obvious mistranslation when the text includes ‘apa-apa’ (anything) for ‘nothing’ which should be ‘tiada yang kita hasilkan’. A long process is required to achieve an accurate translation. Some words such as ‘meningkatkan’ (to increase) are missing in translation. Moreover, the sense of many words has not been taken into account. Another example is a passive form ‘digunakan’ (used to) which describes an action that was normal for them, while it has been translated as ‘apply’ or ‘use’ in the Malay translation that is totally different in meaning. Computers are not able to put the words into the coherent and relevant order. The words ‘juga tidak rugi kerana jualan Proton lebih tinggi dan pasaran lebih luas’ has been translated by computer to ‘tidak ketinggalan sama...’ and thereby, reflected a mistranslation. Other near-mistranslations include ‘kita diperlukan’ for ‘we needed it’ which should mean ‘kita perlukan’. Some words are mistranslated in Malay translation as well; ‘kita’ (teller and listener both are involve in the activity) has been translated to ‘kami’ (while listener is not involved) by computer.

4. Result

From the small number of selected examples, we can see that there are many mistranslations and near-mistranslations and potentially ambiguous translations. Computer-
assisted translation or Google Translate in this case study is a useful tool restricted only to the aspect of memory and time. The product is not yet high-quality translation without the intervention of the human mind. Computers undoubtedly do not translate like humans despite the increasingly advanced specialized dictionaries added to the software. Sometimes, computers skip some words, as they cannot recognize them all. Computer software is unable to recognize the vagueness, complex sentence compositions, allegories, metaphors, and unknown sentences. While computer-assisted translation is time and effort saving, machine translations are of little value without human intervention. Machine translation cannot capture the hidden details and the mood of the source text because they do not resolve like the human mind does. To sum up, computer translation technology, specifically Google translate is not accurate. Imperfections of computer translation must be compensated by the intervention of intelligent humans with a computer-assisted translation.

References


EC. (2009). Translating for a multilingual community. European Commission,


Affective Impact of the use of Technology on Employed Language Specialists: An Exploratory Qualitative Study

AnneMarie Taravella
Université de Sherbrooke

ABSTRACT

A well-established fact in the information systems literature is the importance of human aspects of technology use. In our doctoral research, we look into the emotional effort that employed language specialists have to put in their daily work, in the light of an increased use of language technology tools (LTT) by language service providers. In 2011 and 2012, we conducted qualitative studies to understand how LTT were perceived by language specialists. We observed translators and other language specialists at work and conducted 12 in-depth interviews. We noticed that respondents often mentioned affective constructs, such as stress or anxiety, even when not prompted to describe their affective state. We then reanalyzed our transcripts and written notes in search for answers to the following specific question: “What affective variables do language specialists spontaneously mention when asked to describe their use of LTT?” Using content analysis, we found that respondents often mention some form of occupational stress, or relief of occupational stress, along with other affective variables, in relation with the use of LTT. We argue that emotional well-being and stress relief should be measured and serve as a guide for the design and implementation of language technology tools.

1. Background to the research

The DBA (Doctorate in Business Administration) at the Faculty of Administration of Université de Sherbrooke (Canada) is a research doctorate program with a focus on bringing theoretical and managerial solutions to real industry-based issues. When DBA students have completed their course load, they have to conduct an on-site research before they are allowed to present their research project. This on-site research is to make sure that the theoretical and practical contribution of the final research will be useful for the industry. This on-site research is called the residency phase in the DBA curriculum.

1 This research was supported by the Social Sciences and Humanities Research Council (Government of Canada). It is partly based on data that were first collected for a Mitacs-Accelerate internship fostered by AILIA, a Canadian language industry association.
In our case, we were interested in the translation industry. Initially, we aimed at understanding the success factors of implementing new language technology tools (LTTs) within translation service providers (TSPs). For our residency, we were awarded a 4-month Mitacs\textsuperscript{2} internship in 2012. During this residency, we observed language specialists at work in a medium-sized Montreal-based TSP, and we conducted in-depth interviews with employees of this TSP. We also conducted interviews with other language professionals and language technology vendors in Québec and Ontario.

The data analyzed below were collected during the residency phase of our doctoral research, in 2012. We conducted a first data analysis in 2012, to answer our residency research question, and we reported the results in our residency research report. The methodology and the results of data collection and first data analysis are presented in section 2.

However, even though the in-depth interviews were focusing on LTT use, we observed that the data also suggested that use of LTT induced several affective reactions. For the 36th Translating and the Computer conference, we conducted a second data analysis centered on affective reactions to LTT use. The results of this second data analysis are presented in section 3.

2. Residency — First data analysis

2.1. Theoretical background and research questions

We were interested in the use of language technology tools by Canadian translation service providers. Were the LTTs useful to TSPs?

To answer that question, our starting point was a classification proposed by Gurbaxani and Whang (1991). Those authors suggested that an information system can play five primary roles in an organization:

« a) it increases scale efficiencies of the firm’s operations (operations); b) it processes basic business transactions (transaction processing); c) it collects and provides information relevant to managerial decisions and even makes decisions (decision support); d) it monitors and records the performance of employees and functional units (monitoring and performance evaluation); and e) it maintains records of status and change in the fundamental business functions within the organization and maintains communication channels (documentation and communication). » (p. 66)

We wanted to know if language information systems (of which language technology tools are components) were actually playing these roles for TSPs.

Besides, we wanted to know whether language technology tools were supporting mostly production processes or support processes of TSPs. Along with Rivard and Talbot (2001), we define a process as an activity that transforms an input into an output, using resources available in the organization. A production process deals with manufacturing the product or providing the service itself, while all other processes support the production process and are called support processes.

\textsuperscript{2} MITACS is a Canadian organization funding internships that give postgraduate students an opportunity to apply their theoretical knowledge in real-life work settings. Companies that partner with students for those internships also benefit from current research approaches.
processes. For a translation service provider, the production process deals with transforming a document in a source language into another document in a target language. The support processes include technology procurement, training, human resources management, sales or financial management, to name only a few. Are language technology tools particularly supportive of production processes or support processes in TSPs?

Finally, we also wanted to know what translation industry players thought a « language technology tool » was. As far as we know, there is no general consensus about what a LTT does. In our view, a LTT was something that helped a translator translate. It could be anything from the electronic version of a dictionary, a concordance program, a translation memory tool or a machine translation engine to a voice recognition or project management software. On another note, from the user's point of view, we were wondering whether the use of a specific tool occurred on a voluntary or a mandatory basis.

Consequently, our global research question was : « What do translation service providers' employees, managers and vendors think of the usefulness of language technology tools for the TSP? ».

Specific research questions included : « How do TSP's employees, managers and vendors define a language technology tool? », « Are LTTs more useful to production processes or support processes? » and « What would you like LTTs to do better? ».

2.2. Methodology

Since the residency internship was to serve as an exploratory study for our doctoral research, we chose an exploratory, qualitative research design based on an inductive approach. An exploratory study called for an inductive approach, that is, one that let the respondents voice their main concerns and opinions, with as little guidance from the interviewer as possible.

We developed a brief interview guide, with only a couple of open-ended questions, including questions about how respondents would define a language technology tool, what LTTs they were using, how useful they felt those LTTs were, and what they thought a « perfect » LTT should do. We collected data within our main Mitacs internship partner (the « Mitacs partner »), a medium-sized Montreal-based TSP, through non-participating observation and in-depth individual and group interviews. We also conducted interviews with other players in the industry.

Non-participating observation took place during four three-day-stays at the Mitacs partner offices. We shadowed employees for 30-minute to 1-hour periods, taking notes and occasionally asking questions. Employees were performing translation, editing, and project management tasks.

Combining all respondents, 12 in-depth individual interviews and two group interviews were conducted; the interviews lasted 33 minutes on average. In total, 27 language specialists from 9 companies were either interviewed or observed, or both. All companies were Canada-based (provinces of Québec and Ontario).

The table below shows, for each of the 27 respondents, the respondent's role (manager, junior or senior translator, reviser, or a combination of those roles). Each respondent has also
been given a code, so that we could quote their answers without having to refer to their identity. GEST stands for « gestionnaire » (French for « manager »), while TRA stands for « traducteur » (or « translator »).

<table>
<thead>
<tr>
<th>GEST01</th>
<th>Manager and senior translator</th>
<th>TRA01</th>
<th>Translator and reviser</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEST02</td>
<td>Manager</td>
<td>TRA02</td>
<td>Junior translator</td>
</tr>
<tr>
<td>GEST03</td>
<td>Project manager</td>
<td>TRA03</td>
<td>Reviser</td>
</tr>
<tr>
<td>GEST04</td>
<td>Manager and senior translator</td>
<td>TRA04</td>
<td>Reviser</td>
</tr>
<tr>
<td>GEST05</td>
<td>Manager</td>
<td>TRA05</td>
<td>Junior translator</td>
</tr>
<tr>
<td>GEST06</td>
<td>Manager and senior translator</td>
<td>TRA06</td>
<td>Senior translator</td>
</tr>
<tr>
<td>GEST07</td>
<td>Manager and senior translator</td>
<td>TRA07</td>
<td>Reviser</td>
</tr>
<tr>
<td>GEST08</td>
<td>Manager and senior translator</td>
<td>TRA08</td>
<td>Junior translator</td>
</tr>
<tr>
<td>GEST09</td>
<td>Manager</td>
<td>TRA09</td>
<td>Junior translator</td>
</tr>
<tr>
<td>GEST10</td>
<td>Manager</td>
<td>TRA10</td>
<td>Translator and reviser</td>
</tr>
<tr>
<td>GEST11</td>
<td>Manager</td>
<td>TRA11</td>
<td>Reviser</td>
</tr>
<tr>
<td>GEST12</td>
<td>Project manager</td>
<td>TRA12</td>
<td>Junior translator</td>
</tr>
<tr>
<td>GEST13</td>
<td>Senior translator</td>
<td>TRA13</td>
<td>Junior translator</td>
</tr>
<tr>
<td>GEST14</td>
<td>Manager and senior translator</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Role(s) and codes of all 27 respondents

2.3. Results

The general consensus is that a language technology tool can be defined as « a tool that is useful for language specialists ». It could be either a software, a device or a database. Its main characteristic is that it can help language specialists do what they do, whether it has been developed specifically for language specialists or not. In that view, a translation memory is obviously a LTT, but an electronic version of the Merriam-Webster dictionary is also a LTT, as is Twitter or MSN when translators use those platforms to share translation knowledge.

We observed that the use of LTTs for day-to-day work is common standard for the TSPs surveyed. More than that, all respondents find it normal to use an active language technology tool to translate documents\(^3\). Many of them use several active language technology tools on a daily basis.

\(^3\) We call « passive language technology » a technology that users can refer to when they want to do translation; users must then transfer the knowledge themselves into the translated material. An electronic version of a dictionary or a concordance program are passive language technologies. We call « active language technology » a technology that enables users to create or modify translated material. A translation
Use of a specific passive language technology tool is typically voluntary, while use of a specific active LTT is typically required either by the client or by the TSP. The TSP generally works with the translation environment that their big clients want. For smaller or occasional clients, the TSP typically uses the off-the-shelf translation environment it has chosen to implement company-wide. No TSP in the study has a proprietary translation environment.

Globally, language technology tools are doing a good job in helping the TSP do its daily work. Among the five roles proposed by Gurbaxani and Whang (1991) (see above), LTT are quite useful with the four first roles, i.e. helping with operations, transaction processing, decision support and performance evaluation. However, a lot still has to be done with the fifth role (helping with documentation and communication). For many respondents, the main problem with integrating LTTs within the production chain is the lack of a smooth communication between different tools or different modules within a tool. The « perfect » LTT would be made of several independent modules that would cover all the TSP needs and would seamlessly share the same information.

[The perfect tool] would do everything. It would acknowledge receipt of the order. It would analyze the project. It would know the perfect translator for the job and ask him or her directly – it’s a perfect tool, remember : « Can you do the job? Is the deadline acceptable? » The translator would be able to translate directly within the tool; when the translation is ready, the tool would tell the reviser : « The translation is ready for revision. » It would automate not only the translation process, but also the management process. And it would not need two, three, four different softwares; it would be a one-stop, global tool. That would be nice. (GEST07)  

2.4. Conclusions of first research

We first conducted an exploratory study to confirm that translation companies needed help in implementing language technology tools. However, it appears from the study results that TSPs are doing a very good job in using the tools to perform their daily tasks. What is not so easy is evaluating whether users are happy to use the tools.

3. Second data analysis

3.1. Theoretical background and new research questions

A well-established fact in the information systems literature is the importance of human aspects of technology use. When Glass, Ramesh and Vessey (2004) compared Computer science (CS), Software engineering (SE) and Information systems as the main academic subdivisions of Computing discipline, they found that IS was the only research field conducting analysis at the memory or a MT engine are active language technologies. For more details about the difference between active or passive language technologies, or for a snapshot of language technology tools used in Canada in 2011, see Taravella (2011).  

Original quote : « [L’outil idéal] ferait tout. Il ferait la réception du travail, il l’analyserait, ferait le lien avec le traducteur le plus compétent pour le réaliser, communiquerait la question – dans un monde idéal, là - avec le traducteur en question : « Peux-tu le prendre, peux-tu placer ça? », permettrait aussi au traducteur de travailler directement dans l’outil en question pour faire sa traduction, ensuite, quand c’est prêt, communiquerait directement avec le réviseur pour dire : « C’est prêt à être révisé. ». Automatiser non seulement le processus traductionnel, mais le processus de gestion aussi. Mais pas dans deux, trois, quatre logiciels, mais dans un seul outil global. Ça, ça serait bien. »
behavioral level, while the other two were conducting analysis at the technical level. Using machine translation or refusing to work as a post-editor are examples of behaviors that can be analysed in the view of Information systems.

However, intention-to-behavior theories within the Information Systems field are often based on characteristics of the tools, not characteristics of the users. For instance, Davis (1989)'s Technology Acceptance Model (TAM) features two main tool characteristics: perceived ease of use and perceived usefulness. When a tool is both useful and easy to use, according to the user's perception, the user is likely to behave a certain way, namely use the tool. But this theory says nothing about how the user feels. It deals only with what the user thinks. As Davis (1989) himself wrote: « the role of affective attitudes is an open issue » (p. 335).

For language specialists within the translation industry (other than those who are self-employed), the use of a specific LTT is seldom a chosen behavior; it is usually decided by the client and/or the employer. Language specialists still have many behavioral choices: Will they use the tool well? Will they make useful changes to suggestions provided by the tool? Will they try to be creative in doing so? Will they remain engaged at work? Will they express dissatisfaction or lack of motivation? Will they eventually leave their job, or even the industry?

All those positive and negative behaviors are related to affective variables: attitudes, emotions, and moods, as well as dispositional variables (personality traits). According to Affective Events Theory (Weiss and Cropanzano, 1996), reactions to work events shape workers' affective states; in turn, affective reactions cause workers to adopt specific behaviors. Thus, a better understanding of language specialists behaviors at work starts with a better understanding of their affective reactions to new technology tools.

Consequently, we performed a second analysis of our residency data, with a new global research question: « What affective or dispositional constructs do translation service providers' employees, managers and vendors mention when speaking of language technology tools? ».

3.1. Methodology

Analyzing qualitative data is trying to find, among a huge amount of data, those elements that are relevant for our research question. Since interviews were conducted with as little rigidity as possible, in a non-controlled work environment, we collected a lot of data that were only remotely relevant for this question. To paraphrase Gavard-Perret et al. (2008), the main difficulty in analyzing the data is then to find « nuggets ». Information nuggets are useful pieces of information that we are able to use.

To find those nuggets, we chose to perform a horizontal thematic content analysis. We identified affective themes and sub-themes throughout all interviews, grouping answers from different interviews to spot affective or dispositional variables that were consistently mentioned by several respondents.

Results are summarized below.

---

5 A French version of this section has been submitted to AILIA as part of our Mitacs internship final report (Taravella, 2012)
### 3.3. Results and discussion

<table>
<thead>
<tr>
<th>Sub-theme</th>
<th>Theme</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Personal history</strong></td>
<td></td>
</tr>
<tr>
<td>Being a « real » language specialist</td>
<td>Professional identity</td>
</tr>
<tr>
<td>NOT being a TI specialist</td>
<td></td>
</tr>
<tr>
<td>Human work is still necessary</td>
<td></td>
</tr>
<tr>
<td>Quality is not valued in rates</td>
<td></td>
</tr>
<tr>
<td>Translators do not value their own work</td>
<td></td>
</tr>
<tr>
<td><strong>Fear of change</strong></td>
<td>Lack of control</td>
</tr>
<tr>
<td>Guilt related to non-imputability of technical errors</td>
<td></td>
</tr>
<tr>
<td>Fear of being deceived or conned by vendors</td>
<td></td>
</tr>
<tr>
<td>Being confused with processes</td>
<td></td>
</tr>
<tr>
<td><strong>Interchangeability of translators</strong></td>
<td>Loss of individual identity</td>
</tr>
<tr>
<td>Very aggressive vendor competition</td>
<td>Negativity of the environment</td>
</tr>
<tr>
<td>Translation is an economic burden</td>
<td></td>
</tr>
<tr>
<td><strong>Time-related stress</strong></td>
<td>Stress</td>
</tr>
<tr>
<td>Quality-related stress</td>
<td></td>
</tr>
<tr>
<td>Income-related stress</td>
<td></td>
</tr>
<tr>
<td><strong>Technology is taking over translation time</strong></td>
<td>Frustration</td>
</tr>
<tr>
<td>Silly, time-grudging technical decisions</td>
<td></td>
</tr>
<tr>
<td><strong>Clients lack of process consistency is infuriating</strong></td>
<td>Anger</td>
</tr>
<tr>
<td>Technical discussions are boring</td>
<td>Boredom</td>
</tr>
<tr>
<td>Work is spoon-fed</td>
<td>Stress relief</td>
</tr>
</tbody>
</table>

Respondents tend to mention mostly negative aspects of using LTTs. There seem to be a bias toward highlighting problems more than benefits. Indeed, the fact that affective impacts of using technology are almost never mentioned in the management’s discourse can be a good reason to feel undervalued and to be wanting to express one’s frustrations. However, most respondents recognized that the use of LTTs has made their work easier, from a technical point of view, and that they get some stress relief in using technology tools.

It was quite apparent in the data analysis process that mentions of negative effects were voluntarily brought up by respondents, as if they wanted to take the unique opportunity to voice their frustrations. On the other hand, they almost always mention stress relief benefits of using LTTs, but in a casual way, as if benefits were commonly known and accepted.

This makes us hypothesize that negative affective impacts of using LTTs are emphasized by the lack of taking those negative impacts into account. Maybe language specialists would find an easy, yet necessary relief in being allowed to express their frustration, fear and criticism toward technology tools.
4. Next steps and conclusion

Using content analysis, we found that respondents often mention some form of occupational stress, or relief of occupational stress, along with other affective variables, in relation with the use of LTT. This is an interesting result, for well-being of human resources that use LTT is never mentioned as a design criteria. Yet, as O'Brien (2012) reminds us very firmly: “[i]t is how the technology is created, or implemented, that has a dehumanising effect. Technology created without consideration for the task or end users removes those end users from the equation.”

In the view of our residency research results, we decided to focus our doctoral research on the emotional effort that employed language specialists (terminologists, translators, revisers) have to put in their daily work, in the light of an increased use of language technology tools by translation service providers. How does the affective states of language specialists evolve throughout the day, the task, or the change of environment? To answer that question, we intend to measure affective, dispositional and environmental variables in a longitudinal multiple-case study.

We argue that emotional well-being and positive or negative affective states should be measured and serve as a guide for the design and implementation of those tools. As a paraphrase of Desilets et al (2009)’s argument saying that « translators might better be served by the research community if it was better informed about their work practices » (p. 1), we argue that translators might better be served by the LTT community and the management community it those were better informed about their affective reactions to LTTs.

References


Taravella, AM. (2012). Rôles des technologies langagières dans les organisations fournisseurs de services de traduction (FST). Rapport de stage présenté à l’AILIA. Unpublished manuscript.

Solving Terminology Problems More Quickly with ‘IntelliWebSearch (Almost) Unlimited’

Michael Farrell
IULM University, Milan

ABSTRACT

Michael Farrell received several descriptions of university courses to translate from Italian into English in early 2005. The syllabuses boiled down to a list of topics and laws of mathematics and physics: not many complex sentences, but a great deal of terminology which needed translating and double checking with the utmost care and attention.

To do this, he found himself repeatedly copying terms to his PC clipboard, opening his browser, opening the most appropriate on-line resources, pasting terms into search boxes, setting search parameters, clicking search buttons, analysing results, copying the best solutions back to the clipboard, returning to the translation environment and pasting the terms found into the text.

He quickly realized that he needed to find a way to semi-automate the terminology search process in order to complete the translation in a reasonable time and for his own sanity. He immediately started looking around for a tool, but surprisingly there seemed to be nothing similar to what he needed on the market. Having already created some simple macros with a free scripting language called AutoHotkey, he set about writing something that would do the trick.

The first simple macro he knocked out gradually grew and developed until it became a fully fledged software tool: IntelliWebSearch. After speaking to several colleagues about it, he was persuaded to share his work and put together a small group of volunteer beta-testers. After a few weeks of testing on various Windows systems, he released the tool as freeware towards the end of 2005.

At the beginning of his workshop, Michael Farrell will explain what prompted him to create the tool and how he went about it. He will then go on to describe its use and its limitations, and show how it can save translators and terminologists a lot of time with a live demonstration, connectivity permitting.
The workshop will conclude with a presentation revealing for the first time in public some of the features of a new version which is currently being developed under the code name "IntelliWebSearch (Almost) Unlimited" (pre-alpha at the time of writing).

The workshop is aimed at professional translators, interpreters and terminologists in all fields, especially those interested in increasing efficiency through the use of technology without lowering quality standards.

1. How did IntelliWebSearch come about?

I wrote IntelliWebSearch for my own personal needs: I am a professional translator and not a professional software developer. I wrote it to solve a specific problem I had. I was translating a set of syllabuses for degree courses in physics and mathematics from Italian into English. The files boiled down to a few introductory paragraphs followed by a list of various laws and theories of physics and mathematics. It was obviously extremely important to use the most standard name for each law and theory. So I found myself doing literally hundreds of terminology checks per page, which is of course extremely time-consuming. I quickly realized that I needed to find a way to semi-automate the terminology search process in order to complete the translation in a reasonable time and for my own sanity. So I immediately started looking around for a tool, but surprisingly I could not find anything similar to what I needed on the market.

2. Terminology research before IntelliWebSearch

Without IntelliWebSearch, I had to select the term to look up in my translation environment and copy it to the PC clipboard. It was then necessary to launch the browser and open the search engine, online dictionary or online encyclopaedia I wanted to use, sometimes a different one for each term or even several for the same term, if I was cross-checking. After that I had to paste the term into the search box in the web page, edit the search string if necessary, set the search parameters, click on the “search button” and wait for the result page to load. Seeing all this, one word comes immediately to mind: macro. However I could not write a MS Word macro because I was using the translation environment tool Déjà Vu X. So I chose an application-independent macro language that runs on Windows called AutoHotkey. IntelliWebSearch is a sophisticated free-standing AutoHotkey macro compiled with Ahk2Exe. The macro script cannot therefore be altered by the user, but its behaviour may be customized by changing a wide range of settings.

3. Terminology research with IntelliWebSearch

With IntelliWebSearch, the term search process becomes much simpler. You need to start, as before, by selecting some text in your favourite translation environment. Then you have to press CTRL+ALT+B. You can of course change this shortcut key to anything you like. This calls up the IntelliWebSearch search window, where you can edit the search string if need be. For instance, you may want to add another keyword or change a plural to a singular. You then have to click the button corresponding to the search you want to perform. If the search you require is the one highlighted in blue, you can simply press the Enter key on the keyboard. If you do not want one of the buttons displayed, you can always change the group: the searches are organized into five different groups of ten searches corresponding to the ten buttons. In other words, you can reach up to fifty different searches through the IntelliWebSearch search window. The buttons may
correspond to search engines, online encyclopaedias and online dictionaries, or even local
dictionaries and encyclopaedias on CD-ROM or your hard disk. They may even correspond to the
same search engine configured fifty times with different advanced settings and parameters. In
reality, it is possible to save more than fifty search configurations in IntelliWebSearch, but a
maximum of fifty may be associated with buttons at any given time. After selecting an on-line
resource, your default browser is launched and your chosen search engine, dictionary or
encyclopaedia appears with your search results. If you choose a local resource instead, your local
dictionary or encyclopaedia result window opens.

4. Direct search shortcut keys

A particular search can also be done even more rapidly using a direct shortcut key. You start
in the same way by selecting some text in your translation environment. However you then press
a previously assigned shortcut key to skip the IntelliWebSearch search window entirely and
display the results directly in your favourite search engine, online dictionary, online
encyclopaedia or local resource. Obviously, to do this, each search configuration must be
associated with a different shortcut key.

5. Return shortcut key

Another convenient shortcut key combination (CTRL+ALT+C by default) copies any text you
select in your browser/local dictionary result window to the PC clipboard and returns you to the
translation environment, where you can choose to paste it in using the application's own paste
function (Edit menu>Paste or CTRL+V). In other words, IntelliWebSearch remembers which
application you launched the search from and returns to it with the results of your search, saving
you having to click about with the mouse.

6. PluriSearch

To look for the search string in several pre-selected resources almost simultaneously, you
can choose the PluriSearch button on the IntelliWebSearch search window. The searches are not
perfectly simultaneous: one is performed a fraction of a second after the other. In most
browsers, the various results are shown under different tabs. Again in this case, you can skip the
IntelliWebSearch search window and go straight to the results using a customizable shortcut key
(CTRL+SHIFT+ALT+B by default).

7. GroupSearch

To look for the search string in all the resources belonging to the group displayed on the
search window almost simultaneously, you can choose the GroupSearch button. Note that it is
not necessary to fill all ten buttons in a group with search configurations, so GroupSearch might
correspond to just two or three searches.

8. How does the tool know which searches I want to do?

IntelliWebSearch comes with a new user’s starter pack that includes the settings for over
eighty search engines, dictionaries and encyclopaedias. There is also a settings database on the
IntelliWebSearch website with an even greater number of search configurations and resources. If
you cannot find the settings you need, you have to write them yourself. It takes a few minutes, but only needs doing once.

9. IntelliWebSearch Wizard

IntelliWebSearch comes with a Wizard which guides you when setting up your own searches. The Wizard steps you through the procedure with simple on-screen instructions.

10. Where can you find IntelliWebSearch?

IntelliWebSearch has its own website (www.intelliwebsearch.com). The tool comes with user interfaces in English, Catalan, Czech, Dutch, French, German, Italian, Russian and Spanish, thanks to the help of volunteer localizers. You can also follow IntelliWebSearch on Facebook, Twitter and other social media, and there is even a peer-to-peer users’ group on Yahoo! Groups.

11. Online help

If you have any problems, online help is available in English, French and Italian: just press F1 when any IntelliWebSearch window is active, or choose help from the tray icon menu.

12. Freeware

IntelliWebSearch is freeware, but not open source. If you are happy with it, you can show your satisfaction by becoming a paid registered user. Registrants get to choose how much they want to pay.

13. Evolution since 2005

After a few weeks of testing, IntelliWebSearch v. 1 was launched at the end of November 2005. During December the user interface was localized into German, Italian and Russian, and on the 18th of the same month, IntelliWebSearch v. 2 was released. The tool was now UTF-8 compliant and it was possible to choose a specific browser for web term searches (not necessarily the OS default browser). The tool interface was localized into Dutch in May 2006 and French in December. The readme (help) file was translated into French and Italian in January 2007. A Spanish user interface was added in May and “Form Fill” mode was introduced so that the tool could access a few websites that had, up to then, eluded it (notably Le Grand Dictionnaire). IntelliWebSearch v. 3 went beta in December and was released in January 2008. The new features included PluriSearch, direct search shortcuts, term search in local resources (dictionaries on HD or CD-ROM), five groups of ten buttons (originally there was only one group) and the new user’s starter pack of search configurations. The IntelliWebSearch user interface was localized into Catalan in February 2008. In March the same year, the Wizard went beta and was released with IntelliWebSearch Version 3.1.0.3 in April. GroupSearch was introduced in October 2009 along with a Czech user interface. In November it was possible to become a paid registered user. In September 2010 IntelliWebSearch could be installed as a portable app, and in January 2013 the settings database was launched on the IntelliWebSearch website. Virtually all the changes listed were made thanks to feedback and requests received from users, not to mention the generosity of volunteer localizers.
14. Comparable tools that have appeared on the market since 2005

Around 2011 Rolf Keller, a German translator, launched Multifultor (freeware). It was described on a popular web forum for translators as the “new IntelliWebSearch”. The main difference is that, unlike IntelliWebSearch, it works with its own built-in browser. This has the advantage that Multifultor can access some sites directly which IntelliWebSearch can only access through the “last-resort” form fill mode. The downside is that you cannot use your favourite browser.

Some people use the freeware application Click-to, which I imagine is particularly appealing to “mouse lovers”. Web searches are however not the primary purpose of this tool and its search capabilities are comparatively limited (see comparative article listed in the webography).

Several translation environment tools have incorporated IntelliWebSearch-like capabilities, including Wordfast Classic, memoQ, Fluency Translation Suite and CafeTran. Wordfast Classic has actually had web search capabilities since before 2005 and CafeTran was first released at around the same time as IntelliWebSearch. CafeTran has an option to import IntelliWebSearch resource settings. All these TEnT implementations suffer from the same major drawback compared with IntelliWebSearch: it is only possible to launch searches from within the environment itself. IntelliWebSearch is instead a system-wide application.

Furthermore memoQ, Fluency and CafeTran do searches in a built-in browser window, thus consuming part of your “screen real estate”, although CafeTran also allows users to search with an external browser if they prefer. Wordfast Classic only allows searches with MS Internet Explorer. The much-respected tool guru Jost Zetzsche recently described memoQ’s web search feature as “comparable to (an anemic) IntelliWebSearch” [sic] (see webography). At the risk of sounding immodest, I believe that IntelliWebsearch is still today the most feature-rich tool of its kind on the market.

15. New version, new features and future developments

A new version is currently being developed under the code name “IntelliWebSearch (Almost) Unlimited” (pre-alpha at the time of writing). The aim, which is evident from the code name itself, is to remove most of the arbitrary limitations found in the previous version. Why only five groups of ten buttons? Why not an unlimited number of groups of an unlimited number of buttons? Another major change will be the introduction of direct shortcuts for GroupSearches, thus removing all differences between GroupSearch and PluriSearch. To put it another way, there will be an unlimited number of different PluriSearches. The Wizard will also be redesigned to make it easier to use and several other small changes are planned.

References


Twitter Crowd Translation — Design and Objectives

Eduard Šubert
Czech Technical University
in Prague

Ondřej Bojar
Czech Technical University
in Prague

ABSTRACT

This paper describes our project to support translation of streaming texts on social networks, in particular Twitter. Since machine translation of this type of content is still almost unusable, we rely on volunteers to provide and score the translations. The translations will serve as a testbed and development data for our MT systems tuned for this domain. The project thus serves multiple purposes: From the users' point of view, we would like to provide a smooth access to timely information in foreign languages. From our translators' point of view, we want to provide them with interesting content and material to improve their language skills. Finally, we admit that our project is still primarily a research exercise: As MT researchers, we are interested in learning to handle the specific challenges that this type of content brings. We hope to acquire an interesting collection of data for MT development and to gradually improve our MT processing pipeline for this type of text.

1. Introduction

Twitter Crowd Translation (TCT) is our project aimed at the development of an infrastructure for online translation of social media. Through this, we also want to gather relevant training data to support machine translation of such content. We focus on Twitter\(^1\) and the open-source machine translation toolkit Moses. Our project heavily relies on crowdsourcing.

The paper is structured as follows. We first briefly motivate our task and then provide an overview of the system. A section describing some technical details of the implementation follows. Having used the system for a few months in a dry run, we collected some interesting observations on human translation of tweets into Section “Translation Aspects of TCT”. Finally, we review the related work of machine translation of this content and add our preliminary experience and plans.

\(^1\) [http://www.twitter.com/](http://www.twitter.com/)
2. Motivation

Social networks have gained tremendous popularity and have successfully replaced many established means of communication. While geographical location of the users has little to no impact on communication, the obstacle of languages used remains.

For stable and long-lasting content, the problem is less severe: services such as Wikipedia have shown that volunteers are able to provide translations into many languages. Machine translation is easy to train on such content and it delivers moderately good results.

In contrast, social networks are used in a streaming fashion, Twitter being the most prominent example. Anybody can contribute a message, which is forwarded to a number of followers. They, in turn, are flooded with messages from sources they select. Given the constant flow of new information, nobody looks back at older messages. Therefore any potential translation needs to be instant.

Providing translation to “streaming networks” is much more challenging. The input is much noisier, significantly reducing MT output quality, and the community is less interested in providing manual translations.

The social motivation of our project is to break the language barrier for streaming social networks. The technological motivation is to advance MT quality by collecting more and better-fit data. What Wikipedia and online MT services manage for stable content, we would like to achieve for streaming networks and casual, unedited text.

Following the open-source culture, we will keep the data and code created in this project fully available to the community.²

3. Design of TCT

This section provides a high-level overview of the principles we followed when designing TCT.

3.1. Overview of Interactions

Our initial intention for TCT was to be as thin layer as possible, to cause minimal disruption. The majority of users would stay within their platform – Twitter in this case.

Therefore, we have designed following system: To select tweets interesting for translation, we manually identified a number of Twitter users and we collect all their tweets. For topicality and attractiveness for our beta-testers, we picked sources from Ukraine that cover the current Russian-Ukrainian conflict in English and Ukrainian.

Once tweets enter TCT, they are immediately sent out to our translators by e-mail. Later the system collects submitted translations and presents them to judges for evaluation. The best translation is tweeted back to Twitter to be published in the most approachable way.

² TCT source code is available at http://github.com/cifkao/tct
A few of our beta-testers however demanded the possibility to submit translation through the web. This is in fact in line with the study of Petrović et al. (2010) based on tweets collected in Dec 2009 and Jan 2010: more than 40% of them were created in the web interface, so many Twitter users don’t use any special Twitter application. Having added the option to submit through the web, our original workflow is slightly altered. Figure 1 depicts the current situation.

3.2. Benefits for Users

A crucial assumption we make in our project is that there exists a sufficiently large group of users who are happy to contribute for the sake of improving machine translation and therefore the world. This may sound like a ridiculous wish however the online encyclopedia Wikipedia stems from similar ideas and it evolved into a giant information source.

![Figure 1: The TCT workflow in a nutshell](image)

We speculate that the only incentive for users to contribute to Wikipedia is the prestige. Authors of successful articles may brag about their good work. We will try to deliver similar acknowledgement to our translators for being among the fastest and most accurate.

Therefore, we have implemented a leaderboard for translators. Translators are given an additional space to share information about themselves, including potential contact information, see Figure 2 for an example. This may serve as a registry for prospective customers looking for translation services.
Figure 2: Sample profile of a TCT translator. The system distinguishes between source and target languages the users is willing to help with

Aside from gaining reputation as a translator, TCT also offers a field for practice of language skills. We distinguish between languages that translators want to translate from and into. While still learning a language, people can opt for just reading tweets in it and offer to produce text only in languages they are skill enough at. For very early stages of language learning, users can also contribute by judging translation quality, e.g. from a language they understand only very vaguely.

4. Technical Aspects of TCT

This section describes the current TCT workflow in more detail. Technically, the back-end of the application is written in PHP and it is based on the CakePHP framework. The website front-end is built with the aid of Zurb Foundation 5.

In the following, we trace one tweet on its journey through our system from the entry to the publication of its translation.

4.1. Collection of Tweets

So far, tweets automatically enter the TCT system only when they are posted by one of the manually selected authors.

In the long term, we definitely plan to add other input ways. For example, we consider searching for tweets based on their originating location to provide translation for a live events. For “on-line events”, following a hashtag would be the most appropriate.

We collect tweets periodically through the Twitter REST API.

4.2. Translation of Tweets

Once a tweet enters our system, it is backed up in our database and immediately sent out to all available translators who indicated they can translate the required language pair. We also include our Moses system for some language pairs. Future versions of TCT will need a module for selecting a small subset of translators, mainly for balancing the number of requests sent to each translator based on his or her preferred involvement.
The source language of each tweet is obtained from tweet metadata as Twitter is running its own language detection system. The target language is set manually for each tweet source to reach appropriate audience.

As mentioned above, the initial intention was not to force translators to visit our website in order to submit translations. One reason was to avoid the need for switching into one more application in cell-phone access. Another reason is that we wanted to notify translators that a tweet needs to be translated. Providing the translation upon such a notification should be a one-click activity, to minimize all possible overhead. Therefore, we have based this part of TCT around e-mail as we find it global, simple and since the emergence of smart phones also instant. Translations are simply sent back as replies to our e-mail.

Below is an example of an e-mail sent to our translators. It clearly states the target language and presents the tweet. The URL is a genuine part of the tweet text and it links to an image, video or other supplementary material of the message. The last line labelled “ID” is our hash to identify the source tweet once the translation arrives.

Please translate the following post to language: Czech

My congratulations to Ms Dilma Rousseff for her re-election as President of Brazil. http://t.co/SKamkm5Uci

ID:51d46e56c720c0a20e98e7e628b5806a

To fulfill the wishes of our beta testers we have implemented possibility to submit translation through TCT website. As of writing this article, this functionality is still in an early alpha stage.

4.3. Translation Evaluation

Since our approach relies on volunteer translators and MT engines, some quality checks need to be part of the processing loop. We follow the common practice and solicit multiple translations. The next step is to select the best translation.

In future, we plan to use automatic quality estimation to either pre-select better translation or to actually perform the selection, see e.g. the system QuEst (Shah et al., 2013) or the shared task on quality estimation (Bojar et al., 2014). The current version of TCT is again limited to manual judging.

Judging via e-mail seemed somewhat cumbersome and we hope to be able to automate this process fully anyway, so we implemented a simple web-based interface.

To ease the work of judges we have first implemented relative pairwise comparison. Presented with the original tweet and a pair of translations, the judge was asked to select the better one or indicate that both are equally good or equally bad (inacceptable).

After some preliminary operation and a long discussion, we moved from relative comparison to absolute judgements. The relative comparison is perhaps easier for the judge but it does not help to decide if the better tweet is good enough to be published.
In the current version, the judge is thus presented with the original tweet, one of its translations and the options “hate it”, “dislike it”, “it’s OK”, “like it” and “love it” to indicate his or her attitude towards the translation. We are not entirely satisfied with this solution since we find it more demanding, nevertheless it is the only solution that we could come up with to obtain appropriate data.

With the use of absolute evaluation, it is relatively easy to select the best translation from database and publish it back to Twitter.

4.4. Publishing the Best Translation

Another obstacle we had to overcome was publishing the translation back to Twitter. We want to preserve the link between the source tweet and our translation at least informally. Moreover, we want our translations to reach a large audience but we certainly want to avoid being disruptive to users that do not want to see it.

We preserve the link between original tweet and our translation by using @mention, the standard Twitter way to reply to other users. It is technically done by starting the tweet with the other party’s username. This way though, our translations would be visible only to people who follow both us and the respective author of the original tweet. This would certainly limit our audience too much.

We therefore use a little technical trick and insert a dot in front of the original author’s name. Figure 2 provides an example where we translated the English tweet by Stefan Füle into Czech. The trick ‘detaches’ our tweet from being a direct reply and makes it visible also to people who follow only us and not the original author. Obviously, all our translations are also visible on our profile timeline on Twitter.

Figure 3: Timeline of our Twitter profile

5. Translation Aspects of TCT

The design of TCT as described above highlights some specifics of (human) translation of tweets.
5.1. Understanding Tweets

In contrast to common translation requests (books, news, legal documents, ...), tweets are too short to carry sufficient context information themselves. Arguably, the original context of a tweet is not completely unknown. At least the original author, the time and often also the location are recorded. On the other hand, the same author is likely to post tweets about diverse topics from diverse locations, so a more systematic examination of his or her posting history and circumstances would be desirable in order to provide a solid translation.

Our system so far does the exact opposite: we extract individual tweets and send them by e-mail, totally isolated from any available context. The translator is expected to make sense of the tweet without any supportive tools like web search. In a way, we are putting the translator to the position of a machine translation system, with little or no broader understanding of the message. That alone is a very interesting experience for MT researchers.

In the beta run, we picked a topical affair and set our system to follow some manually identified sources from Ukraine. This has already significantly constrained the context so that only very few unrelated items appeared in our collection and also the number of typos, ad-hoc abbreviations and other phenomena typical of social media was considerably smaller.

Some tweets are easy to understand on their own:

- EU-Ukraine Association Agreement provisionally comes into force tomorrow, 1 November. EU welcomes Ukraine! http://t.co/NNkLdh1nHR
- Some tweets use compressed wording that can be seen as vague or imprecise:
- Duma considers creating private military companies for "alternative settlement of armed conflicts outside Russia." http://t.co/RshtPyb06q

Translating this literally, we could cause the impression that the Duma (obviously the Russian parliament, not “male cheetah cub” in Swahili) is indeed planning to establish private companies.

Most tweets however need some rather specific or local world knowledge:

- Terrorists used #Russian-supplied Smerch MRLs against #Ukraine forces in th conflict zone http://t.co/s9ZKP2zsVk | EMPR http://t.co/9ElqFP7qlq
- Col. Lysenko @NSDC_ua spox: truce Memo signed yesterday in Minsk does not work http://t.co/MIYwWrKi9
- Mykolaiv armored plant handed over 10 APCS to the border guards http://t.co/ANOtAKcUdw via @HromadskeTV http://t.co/mUmfbYKuwf

A likely translator for these tweets in our pool of volunteers will not be an expert in military topics and not a native speaker of English. He or she would thus benefit from additional information looked up and extracted from open sources like Wikipedia, e.g.:
Translating and The Computer 36

- MRLs = multiple rocket launcher
- truce = ceasefire
- APC = armoured personnel carrier
- “Mykolaiv armored plant” is presumably “346th Mykolayiv Mechanical Armor Repair Plant” in Ukraine, not simply an armoured plant in the city of Mykolaiv.

We are currently working on the specification of such a module for the construction of similar on-demand translation dictionaries.

Sometimes, the picture clarifies the sentence:

- Russian SOF in Ukraine with “Polite People" patch and new gear #CrisisUcrania http://t.co/gqLvuehCof
  - Here, the “Polite People" patch is indeed a cloth badge on the uniform of a mercenary (Soldier of Fortune). (Note the typo in the hashtag.)
- German and soviet officers shaking hands in Brest, both celebrating their joint invasion of Poland, 22 Sept 1939. http://t.co/9EAsFe6cdD
  - Some target languages (e.g. Czech) need to specify if there was one officer per party or more.
- A massive column just outside Donetsk, we re firmly being asked not to film, take pictures.
  - With a picture, we would know if the “column” was of smoke, or if it was a traffic congestion.

And sometimes, the brevity leads to too vague statements:

- @OSCE Chair: so-called elections eastern Ukraine not in line wt Minsk Protocol. Calls for more dialogue, commitments http://t.co/po7W0IFKj1
  - Is it the OSCE chair who calls for more dialogue, or the situation?

Faced with such input sentences and insufficient details about the context, the translator is perhaps going to take a different strategy: to preserve as much ambiguity as possible. Short of any experience in translatology, we are not sure if preserving ambiguity is an established (or even promoted) technique for human translators. As MT researchers, we are not aware of any such maxim in MT whatsoever, but we find this direction of thoughts particularly intriguing and worth exploring.

5.2. Guidelines for Producing Translations

As of now, we don’t provide our translators with any requirements, hints or guidelines. During the dry run months, we observed that at least these issues deserve some centralized attention:

- URLs of pictures and detailed sources should be separated from the actual text for translation and reinserted mechanically to the translated tweet – if it fits the length limits.
- Some policy for hashtag translation has to be chosen. Some hashtags are standing away the sentence structure and as such, they can be preserved in the original language while others are part of the syntax and definitely need translation. The policy should say if the hashtag “#Russian” should be included even in the translated sentence where the word Russian has been already translated as ‘ruští’:
As #Russian diplomats increase usage of "Novorossiya" important to counter the fake term. The aim is to embed new reality. Old tricks.

Jak ruští diplomáté stále více používají termín Novorusko (Novorossiya), musíme se proti tomuto falešnému termínu postavit. Cíl je vnutit novou realitu. Staré triky.

Morphologically rich languages need some policy for hashtags in general. If we decide to translate hashtags to allow for exploration of related tweets in the target language, we may face the problem that hashtags should not be declinated. In this example, the translator decided to use the proper Czech version of the name Lugansk/Luhansk. To preserve the base form of the hashtag, a little trick with the character "|" was used to separate the necessary ending:

- Mass grave for the dead found during the ceasefire in #Lugansk http://t.co/Bmf8g0bMx6
- Během příměří byl v #Luhansk|u nalezen masový hrob http://t.co/Bmf8g0bMx6

Usernames like "@someone" may or may not be preceded with an introducing noun (e.g. "uživatel", user) which allows for declination rendering the syntax of the input sentence, the policy should prefer one of the two options.

- photo by @dondyuk http://t.co/WDj6IfNBDo
- Fotka od uživatele @dondyuk http://t.co/WDj6IfNBDo

Length limit has to be somehow considered by the translators since the translation is going to be posted as a tweet again. We are planning to cast no technical limit on the translations but to carry out automatic abbreviations if necessary.

6. NLP Aspects of TCT

The specifics of social media from the point of view of computational linguistics and natural language processing have been well studied in the past, see e.g. Hachey and Osborne, 2010. Most of this research so far has focused on input normalization and adaptation of NLP tools like taggers, parsers or named entity recognizers for this domain, see e.g. Baldwin et al. (2013) or Bontcheva et al. (2013) for a number of references.

In contrast to this, there seems to be much less research on translation of social media. Microblogs can certainly serve as an interesting source of parallel or comparable corpora, see e.g. Ling et al. (2013), Xing et al. (2013), Rajjem et al. (2013) and Jehl et al. (2012). Gerlach et al. (2013) combine pre-editing and post-editing for user-generated content in a tech forum.

Since the phrase-based approach to MT as implemented e.g. in the Moses toolkit (Koehn et al., 2007) has been shown to successfully circumvent the need for most of linguistic processing, we would like to jump-start MT for tweets in a similar fashion. Nevertheless, we are well aware that most of the mentioned pre-processing tools could bring us an improvement and we plan to gradually add them to future versions of TCT.

Some of such tools are already being developed in open source: Bertoldi et al. 2010 evaluate the utility of confusion networks (code available in Moses) for the recovery from spelling errors and there have been two related MT Marathon 2013 projects: MTSpell (spell checking for machine translation) and SMMTT (Social Media Machine Translation Toolkit, see http://ufal.mff.cuni.cz/mtm13/projects.html). We have already started adapting MTSpell for our needs and our languages of interest.
7. Conclusion

We presented TCT, an infrastructure for translation of tweets. As of now, the system is not much more than a playground for researchers in MT. The complete pipeline is nevertheless available for any prospective readers or users: our system follows certain tweet sources, manages registrations of volunteer translators, delivers requests and collects translations from them, operates a manual evaluation of the translations and finally publishes the best translation back to Twitter.

We have already started collecting interesting observations and parallel data necessary to improve MT and MT evaluation for this type of content. We will be gradually automating more and more from our translation pipeline.

Acknowledgement

The work on this project was supported by the grant FP7-ICT-2011-7-288487 (MosesCore). This work has been using language resources developed, stored and distributed by the LINDAT/CLARIN project of the Ministry of Education, Youth and Sports of the Czech Republic (project LM2010013).

References


Ben Hachey and Miles Osborne, editors. WSA’10: Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics in a World of Social Media, Los Angeles, California, 2010. ACL.


SMT for restricted sublanguage in CAT tool context at the European Parliament

Najeh Hajlaoui
European Parliament
Luxembourg

ABSTRACT

This paper shows that it is possible to efficiently develop Statistical Machine Translation (SMT) systems that are useful for a specific type of sublanguage in real context of use even when excluding the exact Translation Memory (TM) matches from the test set in order to be integrated in CAT “Computer Aided Translation” tools. It means that the included part is quite different from the existing translations and consequently harder to translate even for an SMT system trained on the same translation data.

Because we believe on the proximity of sublanguages even though it is still hard to practically define the sublanguage notion, we are proposing on the framework of the MT@EP project at the Directorate General for Translation (DG TRAD) of the European Parliament (EP) to develop SMT systems specific for each EP Parliamentary Committee optimised for restricted sublanguages and constrained by the EP’s particular translation requirements.

Sublanguage-specific systems provide better results than generic systems for EP domains showing a very significant quality improvement (5-25% of BLEU score), mainly due to the EP context specificity and to the proximity of sublanguages. This approach is also satisfactory for pairs of under-resourced languages, such as the Slavic families and German.

1. Previous work

In general, a sublanguage is a subset of the language (Harris, 1970) identified with a particular semantic domain or a linked family of domains (Kittredge, 1978), (Kittredge, 1982). In our previous research (Hajlaoui, 2008), we showed that, despite the great distance between mother languages (e.g. Arabic and French) (Hajlaoui, Daoud, & Boitet, 2008), the two correspondent sublanguages are very near one to another as shown in Figure 1. It was a new illustration of Kittredge's analysis (Kittredge, 1993), (Kittredge & Lehrberger, 1982).
Figure 1: Sublanguages are very near one to another

We also showed that SMT system works very well for small sublanguages with a very small training corpus (less than 10,000 words) (Hajlaoui & Boitet, 2008). This proves that, in the case of very small sublanguages, SMT may be of sufficient quality, starting from a corpus 100 to 500 smaller than for the general language. We are proving in this work the validity of this approach in real context of use clarifying and answering some related questions. We describe also the type of resources we need, mainly Thematic Translation Memory (TTM) presenting some promising results.

The issue consists to do further research on the type of sublanguages for which it is possible to develop efficient (useful) SMT systems in the context of CAT tools.

In the following section, we are testing our conjuncture which consists that SMT systems work very well for domain sublanguages using small training corpus.

2. SMT applied on sublanguages

Assuming that a sublanguage is a subset of the language identified with a particular semantic (family of) domain, in our EP context, health, environment, economy, etc. seems to constitute the restricted sublanguages we are looking for.

In the context of existing applications developed in the DG TRAD, one of the constraints to take into account concerning the use of MT in the EP workflows is to take the ad-hoc vocabulary to translate EP documents (amendments, laws, etc.). Our objective is to help EP translators by reusing an existing base of Translation Memory data to better translate unmatched sentences. Our technical choice involves automatic selection of data to resolve problems of context and quality.

The corpus must obviously reach a critical size to allow reliable statistical treatment. SMT approach works very well for restricted domains with little or no human revision, for example the rules-based TAUM-METEO system is purposely developed for the weather service in Canada to provide weather forecasts in French and English.

Based on some statistical information, we know that environment (ENVI), economy (ECON), and Control of budget (CONT) are ones of the main domains in terms of number of documents treated at the EP. Consequently, we built SMT systems for those domains using the Moses decoder (Koehn, et al., 2007) with the phrase-based factored translation models (Koehn & Hieu, 2007) to mainly translate from English to French. The language models for French were 3-gram ones over each training domain data using the IRSTLM toolkit (Federico, 2008). We used
Minimum Error Rate Training (MERT) (Och, 2003) to optimize the systems. The domain data is randomly split to three sets (training, tuning, and testing).

The following tables (Table 1, Table 2 and Table 3) show BLEU, METEOR and TER scores of our English-to-French SMT systems build specifically for three domains announced above comparing them to the existing "Generic" systems and to "Google" MT systems. The scores are computed on tokenized, truecased text, using the MultEval tool version 0.5.1 (Clark, 2011). The BLEU score show that the improvement can reach 25% of a "Specific" system over the "Generic" system and much more (about 29%) over "Google" system depending on the domain.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Specific</td>
<td>49003</td>
<td>1020</td>
<td>1020</td>
<td>65.7</td>
<td>75.7</td>
<td>29.9</td>
</tr>
<tr>
<td>Generic</td>
<td>NA</td>
<td>NA</td>
<td>1020</td>
<td>40.6</td>
<td>56.8</td>
<td>45.0</td>
</tr>
<tr>
<td>Google</td>
<td>NA</td>
<td>NA</td>
<td>1020</td>
<td>36.1</td>
<td>53.9</td>
<td>47.4</td>
</tr>
</tbody>
</table>

Table 1: BLEU, METEOR and TER English-French scores for the CONT (Control of budget) domain

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Specific</td>
<td>106736</td>
<td>2207</td>
<td>2207</td>
<td>60.4</td>
<td>72.6</td>
<td>32.6</td>
</tr>
<tr>
<td>Generic</td>
<td>NA</td>
<td>NA</td>
<td>2207</td>
<td>44.7</td>
<td>61.8</td>
<td>43.6</td>
</tr>
<tr>
<td>Google</td>
<td>NA</td>
<td>NA</td>
<td>2207</td>
<td>43.3</td>
<td>61.2</td>
<td>42.3</td>
</tr>
</tbody>
</table>

Table 2: BLEU, METEOR and TER English-French scores for the ENVI (Environment) domain

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Specific</td>
<td>101669</td>
<td>936</td>
<td>936</td>
<td>58.6</td>
<td>70.9</td>
<td>35.8</td>
</tr>
<tr>
<td>Generic</td>
<td>NA</td>
<td>NA</td>
<td>936</td>
<td>41.5</td>
<td>57.8</td>
<td>45</td>
</tr>
<tr>
<td>Google</td>
<td>NA</td>
<td>NA</td>
<td>936</td>
<td>34.2</td>
<td>52.1</td>
<td>49.7</td>
</tr>
</tbody>
</table>

Table 3: BLEU, METEOR and TER English-French scores for the ECON (Economy) domain

The first results are promising. They showed a general improvement of 5%-25% of BLEU score over generic MT systems depending on the domain and on the test set. They don't concern only English-to-French but the approach is also satisfactory for pairs of under-resourced languages, such as the Finno-Ugric or Slavic families and German (tested for English-to-German, English-to-Estonian and English-to-Bulgarian). It is mainly due to the lexical convergence which is the main characteristic of restricted sublanguage. It is also due to the EP context specificity and to the proximity of sublanguages.
Contrary to the huge volume of data used to develop generic SMT systems\(^1\), the training data used to develop specific systems are very small\(^2\). However the choice of the data sets is very important. In fact, a specific sublanguage training set avoids the introduction of out-domain vocabulary and a representative test set is more relevant than a single EP document. Based on a single EP document the translation result cannot be generalized since it depends on the matching chance of that document with the training data. Consequently, it is very important to take a representative test set of the domain data.

In order to integrate the specific SMT service in CAT tool, we would like in the next section to see whether this approach is still working when we exclude the part of the test set which can be translate with Translation Memory (full matching).

3. SMT in CAT tool context

In order to combine SMT with Translation Memories (TM), we would like to exclude the part having exact TM matches from the test set keeping as possible the fact that the way to select a test set for a specific system is a bit different from the generic system case. The test set should be in-domain and it should be representative of the domain to be able to provide a general conclusion. We called this part of sentences to be excluded from the test set "natural overlap" in order to make difference between it and the "artificial overlap": the "natural overlap" is the basic function in the SMT approach, which might be important in the case of a restricted sublanguage (small domain) due to the lexical convergence and the limitation of the vocabulary; it is one of the main features of a given sublanguage. While an "artificial overlap" consists to include the test set or a part of it in the training set which is of course forbidden.

As defined, the "natural overlap" is the part of the test set\(^3\) which have an exact TM matches. Consequently, in order to detect the "natural overlap" called also "lexical convergence", four cases can be distinguished.

- Same source but different target
- Same source and same target
- Same target but different source
- Different source and different target

In our actual experiments, we defined it as having the same source and the same target because it happens that even with exact TM matches, users need to post-edit the translation as shows the following French-to-English example for research domain.

- Source: il fait de la recherche.
- TM source: il fait de la recherche.
- TM target: he is doing search.
- Reference: he is doing research.

---

\(^1\) For instance, it is around 20 millions of sentences: 380 millions of words for English and 415 millions of words for French.

\(^2\) In general, it is less than 100 000 sentences.

\(^3\) It might happen even though the domain data are randomly divided into three data sets.
By updating CONT domain data to train the specific system and taking a bigger test set, we obtained a 23.4% improvement of BLEU scores of the EP Specific system over the Generic system for the full test set (3678 sentences\(^4\)). The BLEU score went from 46.5% to 69.9% with the specific system as shown in Table 4. In this experiment, we used a larger test set to restrict it to only 2017 sentences by excluding the part having the same source and the same target.

By excluding sentences having exact matches with translation memories, we might generally reduce the representativity of the domain in the test set but we still have an improvement over a generic system 16.5% of BLEU score for the CONT domain (English-to-French). The BLEU score went from 47.2% to 63.7% with the specific system.

<table>
<thead>
<tr>
<th>MT system</th>
<th>Training set (Nb. sent)</th>
<th>Tuning set (Nb. sent)</th>
<th>Testing set (Nb. sent)</th>
<th>BLEU</th>
<th>METEOR</th>
<th>TER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specific</td>
<td>61185</td>
<td>1226</td>
<td>3678</td>
<td>69.9</td>
<td>79.6</td>
<td>24.8</td>
</tr>
<tr>
<td>Generic</td>
<td>NA</td>
<td>NA</td>
<td>3678</td>
<td>46.5</td>
<td>62.6</td>
<td>41.3</td>
</tr>
<tr>
<td>Specific</td>
<td>61185</td>
<td>1226</td>
<td>2017</td>
<td>63.7</td>
<td>75.3</td>
<td>30</td>
</tr>
<tr>
<td>Generic</td>
<td>NA</td>
<td>NA</td>
<td>2017</td>
<td>47.2</td>
<td>63.3</td>
<td>40.6</td>
</tr>
</tbody>
</table>

Table 4: Third comparison between specific and generic system results

This improvement on MT quality can be converted to monetary benefit. It is important to demonstrate by extrapolation that a huge number of words will be better translated showing at the end an important benefit by reducing the post-editing time.

4. Conclusion and perspectives

We showed that it is possible to efficiently build SMT systems that are useful for specific sublanguages. In the actual context of use, we excluded the exact TM matches from the test set since TM is prioritised to MT to keep the same translation terminology. Because our specific system is built based a local data selected for a given specific domain, we think that it might properly respect also the same terminology as the TM. It is one of the challenges that might be tested for the future work.

The performance of a specific system is proportional to the coverage of the domain. The coverage is usually reached after a certain size of training data. In the next experiments, we will try to automatically detect the domain coverage using some previous experiences as well.

We are developing in a first step a limited number of specific domain systems for some language pairs such as EN-FR, EN-DE, etc. Then, they will be hosted to allow human evaluation involving EP users. We will mainly measure the post-editing time spent by users, which is the

\(^4\) Average length equal to 27 words.
main indicator of MT quality improvement. We should note that human factors like correlation, instability over time should be taken into account during the human evaluation campaign.

The sublanguage specific systems will be integrated in the EP translation workflows to improve Translation Memory results offering in priority previous human translations. To more prioritize Translation Memory, the development of an algorithm to translate only unmatched segments with MT is in progress. It include all the sentences that have a higher match score (e.g. between 82% and 99%).

References


Task-based Teaching of Computer-aided Translation in a Progressive Manner

Jessica Xiangyu Liu
The Chinese University of Hong Kong

ABSTRACT

The teaching of computer-aided translation is commonplace in academic institutions in recent years. More research has been done and more works have been published in this area. While much has been written on the theoretical and conceptual aspects of computer-aided translation and the contents of the course, little has been done on its practical aspect.

This workshop will present the classroom practice modules of the Introduction of Computer-aided Translation, an MA course at the Department of Translation of The Chinese University of Hong Kong. The author will discuss how to teach the use of computer-aided translation systems in a progressive manner through demonstrations and classroom practice, and from basic functions to advanced operations.

This workshop will also present some pedagogical reflections on the teaching of computer-aided translation systems.

It is hoped that it will lead to a rethinking of the way of computer-aided translation systems should be taught.

What it intends to propose is to bring the learning of translation technology closer to the real world through systematic training, thus responding to the changing professional requirements that translators face in their workplace.

1. Introduction

Computer-aided translation (CAT) systems, a.k.a. translation memory systems or software, refer to the translation environment integrated a selection of tools for alignment, data analysis, project management, quality control, terminology management, and translation memory management, to increase the productivity of professional translators.

The massive use CAT systems of has spread from localization companies to traditional translation agencies, and freelance translators in the recent years in the Mainland China and Hong Kong. Knowledge of computer-aided translation has become more common seen in the job
requirements of translator recruitment. Interests on learning CAT systems increase accordingly among translation students (Zhu, 2010).

In China the knowledge can be acquired from the software developers, translation agencies, and forums of translation and localization, and academic institutions. The first three usually provide ad-hoc instructions and task-oriented and function-specific tutorials (e.g. aligning translated files, creating a translation project). Therefore, the training provided by academic institutions should be conducted in a more systematic way.

In the last decade, Chinese academic institutions began to open courses relating to computer-aided translation in for translation majors or foreign language-related majors at undergraduate and postgraduate level. For example, Practice of Translation Technology is an elective course in the curriculum of master of translation and interpreting (MTI) program in Peking University, Tsinghua University, Nanjing University and similar courses are offered more than 10 other universities in the Mainland China. In Hong Kong and Macau, similar courses are also found in The Chinese University of Hong Kong, The City University of Hong Kong, The Hong Kong Polytechnic University and University of Macau.

After a preliminary research, the author finds that most of the courses offered inadequate practice of CAT systems (Wang, 2013) and how they should be taught has been un- or under-explored. The obstacles are probably similar with the those hinder the integration of technologies in translation programme discussed by Bowker and Marshman (2009), involving training the trainers, accessing relevant resources for use with technologies, addressing the needs of a wide range of student learners.

The author has been provided training assistance of various CAT systems at a course named Introduction to Computer-aided Translation for three consecutive years at Department of Translation, The Chinese University of Hong Kong. The course outline, the CAT systems taught in the course and the source texts used as the translation tasks are adjusted every year according to the teaching experience of the previous year. A cumulative hierarchical learning process is gradually developed and a task-based teaching approach is adopted in the teaching CAT systems, which is presented in this paper. It also provides some pedagogical reflections in the light of some specific problems encountered in the training of computer-aided translation systems.

2. Teaching CAT systems

Introduction to Computer-aided Translation is a 3-unit required course for MA in Computer-aided Translation Programme and an elective course for MA in Translation Programme at the Department. It is offered at the first term for both full-time and part-time students who may have some translation experience but little experience of using CAT systems. The aim of the course is to introduce the use of translation technology in translation practice. The course is carried in the multimedia laboratory. Every student is ensured to have one desktop computer for practice in the class. The hands-on experience happens with a range of CAT systems taught or introduced in the course. The course normally consists of 13 classes, each of which lasts 2 hours and 15 minutes.
The course introduces various translation technologies yet the concentration is on the CAT systems. Indeed, in its broadest sense, the concept of computer-aided translation can be understood as “the use of a number of computerized tools by the human translators to increase their productivity.” (Chan, 2004) In this sense, the course can introduce tools such as using electronic dictionaries, bilingual corpus, grammar checkers, and optical character recognition (OCR) software that used by general users and assist translator’s work as well. But this paper will only focus on teaching CAT systems because all kinds of CAT tools have provided by the systems.

2.1. A cumulative hierarchical learning process

Technophobia is true for some students in the class, even though, presumably, students at this course should understand they are about to use a lot of new software and have interest on taking advantage of technology. It is recommended to start from something easy and gradually increase the complexity of the knowledge. Therefore, A cumulative hierarchical learning process is introduced to class according to the Knowledge and Cognitive Process dimensions of the revision of Bloom's Taxonomy (Krathwohl, 2002).

The original Taxonomy of Educational Objectives provides an organizational structure and is often used as a scheme for classifying educational objectives. The four categories in Knowledge dimension and six categories in Cognitive Process dimension are arranged in hierarchical framework: from simple to complex and from concrete to abstract. “The mastery of each simpler category was the prerequisite to the mastery of the next more complex one” (ibid.)

The core idea of teaching of computer-aided translation is to train students in a progressive manner. That is to say, build the foundation first and the achievement of the next and more complex knowledge or ability is based on the acquisition of the prior one. The following is a brief overview of what the knowledge in computer-aided translation should be subcategorized and included.

A. Basic concepts in computer-aided translation that students must know to use the software, e.g. translation memory, alignment, terminology database, quality control and project management.
B. Relation of the basic concepts and classifications of CAT systems, e.g. how aligned parallel texts create translation memory, and difference between standalone and networked systems.
C. Procedure and criteria of computer-aided translation, e.g. how to translate a document and what document format is supported.
D. Strategic knowledge, e.g. the selection of suitable computer-aided translation for a translation task.

The course contents in the 13 classes is arranged according to the above category. Each of the first 12 classes only focuses on only one concept and the skills related the concept. For example, translation memory is without doubt the first concept introduced to students. In the last class students are required to finish a task in groups using any software to practice the skills taught in the course.
Also, in the beginning of the course students learn the systems as freelance translators who often work independently and gradually students are asked to work together and experience collaborative translation as a team at the end of the course.

2.2. Task-based teaching

Since most students enrolled in the course have very little experience of using CAT systems, the teacher normally plays a central role—explaining everything and giving instructions at each step in the learning process. At first the author adopted this teacher-centred approach and found that it did have some advantages, especially in the control over the class. Thus, it was easier to carry out intensive practice within a short period of time. However, the students became passive learners and it seemed that most of them were unwilling to try new functions on their own. Eventually this approach was abandoned.

A task-based teaching of computer-aided translation is proposed in the paper. The idea is originated from a revision of task-based teaching of business translation proposed by Defeng Li (2013). Task-based teaching (TBT) is a method initiated and widely used in the filed of second language acquisition. Li made adjustments and adapted the model to the teaching of translation in general and business translation and his revised model consists of the following six stages:

Pre-task, task, reporting, analysis, revision and reflection. (ibid.)

The above order of the cycle is basically retained in the proposed task-based teaching of business translation but with important changes. One stage is renamed and two are merged together.

Figure 1: A cycle of task-based teaching computer-aided translation, a revision of Li’s TBT

The task-related aspects are kept. Reporting is renamed as Feedback to emphasize the teacher-student interactions. Since the Analysis and Revision stages are intended to assess and improve the textual translation, they are merged into one stage named Discussion that better reflects the process. For the sake of brevity, it is organized in the form of a table.

<table>
<thead>
<tr>
<th>A. Pre-task – Preparation of relevant knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>• The teacher introduces the new concept and helps students recall relevant knowledge / skills in the previous lessons;</td>
</tr>
</tbody>
</table>
The teacher provides necessary resources such as source text and translation memory;
The students conduct extensive readings of the concept and the CAT system to be used in the task;
The students fully understand the task;
The students discuss their procedure of doing the task in groups when necessary.

B. Task – Carry outing task

- The teacher provides students with guidance of the procedure of using specific functions in the computer-aided translation system;
- The teacher monitors students completing the task and executes time control;
- The students carry out the task following the procedure provided by the teacher or implementing the procedure agreed by their groups.

C. Feedback – Sharing knowledge acquired during the task

- The teacher encourages students' sharing of any knowledge relating the usage of the systems;
- The students summarize their process, report their problems or difficulties encountered during the task;

D. Discussion – Analysis of the problems and finding solutions

- The teacher coordinates the discussion session;
- The students contribute their solutions and discuss them with peers;
- The teacher may give feedbacks of the solutions proposed by the students and provide other possible solutions for students to consider;

E. Reflection – Evaluation of the task process

- The teacher and students evaluate the usage of the new concept and skills for translation practice;
- The teacher rethinks the entire task cycle and makes improvement for future classes;
- The students reflect on the execution of the task, the problems and solutions in the translation process.

<table>
<thead>
<tr>
<th>Table 1: Task-based model of teaching computer-aided translation systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Necessary preparation and guidance are highlighted in the pre-task and task stages to make up the students’ inadequate knowledge of CAT systems, particularly at the beginning of the course. Good preparation can facilitate the task cycle. The guided task is essential for the students to understand the abstract knowledge such as a concept or principle. For example, explanation of the definition is useful but hands-on experience of using translation memory will be much easier for the students to remember the concept.</td>
</tr>
</tbody>
</table>

The model is not totally in line with a student-centred approach. However, it does shift from a complete teacher-centred tradition towards a model with more teacher-student interactions and student initiative.

2.3. The achievement of the course objectives in the TBT stages

Compared with the traditional approach, the task-based model has several advantages in teaching CAT systems. First, it encourages students to engage more actively in the learning process. Some are eager to explore the complex functions of CAT systems after a couple of classes. Secondly, task-based teaching develops reflective thinking of students, which is very important for them to develop further skills in translation technology. The most important of all, it effectively develops the students' competence of using CAT systems.
The course objectives can be achieved in the task-based teaching stages as summarized in the Table 2.

<table>
<thead>
<tr>
<th>Course Objectives</th>
<th>TBT Stages in which the objective is accomplished</th>
</tr>
</thead>
<tbody>
<tr>
<td>Understand the common terminology in translation technology</td>
<td>pre-task, task</td>
</tr>
<tr>
<td>Apply the main functions effectively in CAT systems</td>
<td>pre-task, task, feedbacks, discussion, reflection</td>
</tr>
<tr>
<td>Receive a grounding in the principles of project management in the translation workflow of agency</td>
<td>pre-task, task, reflection</td>
</tr>
<tr>
<td>Use self-study methods to acquire knowledge of a new CAT systems</td>
<td>task, feedbacks, discussion, reflection</td>
</tr>
<tr>
<td>Provide practical skills necessary to develop further skills in translation technology</td>
<td>task, feedbacks, discussion, reflection</td>
</tr>
</tbody>
</table>

Table 2: The course objectives achieved in different TBT stages

3. Some reflections

During the course, some problems are encountered related to training of CAT systems, including the selection of source texts and preparation of related resources, and selection of the CAT systems.

3.1. Selection of Source texts

The selection of source texts (in either English or Chinese) is always the top priority in the to-do list when the author makes preparation for course. Texts picked randomly may not show the usefulness of CAT functions and students may not fully understand the concepts and skills introduced in the course.

The following six factors of the source texts require considering:

- Textual repetitions – It is a known fact that CAT systems work best with source texts with has to contain some inter-textual and intra-textual repetitions.
- Terminology – This is especially important when teaching the concept “terminology management” as CAT systems provide rigorous consistency of terminology. The number of terms and their frequency also require some thinking so that students are able to manage their own terminology databases or perform term extraction, owing to the importance of terminology management in preparing translators to deal with terminology in their day-to-day work.
- Difficulty of translation – The author recommend using source texts at elementary and intermediate level (for most students enrolled in the course) so that the texts are not too difficult to understand and translate within the time limit in the class. As the course does not concentrate on translation techniques, but on how to use CAT systems, it is not recommended to select something involving a lot of thinking and revision such as classical literary works. However, it doesn’t mean that the texts cannot be challenging. Difficult texts may provide a good practice for advanced users.
- Total word count – The texts have to be in the right length that most students manage to finish the translation in the time limit in the class.
• Authenticity – The author once created a text with a few simple sentences in a similar syntactical structure used at the beginning of the course but abandoned it soon. Even though it is useful in presenting the function of translation memory, it is tedious and students get bored easily. They prefer source text with meaningful content and closer to the real world. A possible explanation is that CAT systems are very practical in the real world. It seems that the authentic translation tasks can stimulate intrinsic motivation in students. A possible explanation is that the tasks fulfill the students’ need for realism.

• Text format – In the beginning of the course, the text should be simple. Then its complexity can be increased by various elements such as tables, charts, and graphics when the students master the editing functions in the systems.

The six factors indicate that it is easier to prepare source texts on the basis of genuine translation assignments. According to author’s experience, the excerpt from user manuals of IT products, financial statements and weather forecasts can be easily modified into a source text. In addition, the texts have to be altered in keeping with the main concept taught in each class. That is to say, when introducing the concepts and skills relating terminology management, the source text should have some important terms in different forms (one-word terms, multi-word terms and etc.) to be recorded in the user’s terminology database.

3.2. Preparation of related resources

Apart from suitable source texts, related resources, including translation memory and terminology databases, parallel texts for alignment, are crucial due to the interactive nature of the editing environment of CAT systems where translators consult all the resources available. Related resources also require in preparation in accordance with the topic and skills involved in the class. For example, when preparing the translation memory (by manually aligning parallel texts), the variety of different similarity rate is essential so that students can identify the difference among matches (such as context match, 100% match and fuzzy match). For different resources, the important factors vary a lot. However, the overall preparation is less complicated than that for the selection of source text.

3.3. Selection of the CAT systems and their order in the training

The selection of the CAT systems taught in the course is a very interesting process. Factors to consider include:

• Concepts and skills introduced in the class – The most important one of all, the selected systems have to fulfill the need of finishing the tasks.

• Competence of the students in mastering the skills of the CAT systems – The selection of the CAT systems has to be in accordance with the students’ understanding and mastery of the functions and skills of the CAT systems. Too much difficulty only frustrate the students.

• Popularity of the CAT systems – The students are usually very curious about the widely used CAT systems on the market. Thus the popular CAT systems such as SDL Trados, Wordfast, MemoQ (Proz, 2013) have a higher chance taught in the course.
---

**Work mode** – An easy-to-use standalone system may fulfill the needs of students at the beginning of the course. However, when the collaborative translation is introduced, a sever-based or cloud-based system with features like project management, and group-shared translation memory will be a more suitable selection.

Factors to consider also include the administrative factors such as budget provided by the University and hardware and software limitations such as hardware requirements and the operation system of the computers.

The order of CAT systems taught in the class is in line with the principle of the cumulative hierarchical learning process. When students master an easy-to-use system, a shift is made to a system with more comprehensive CAT tools.

## 4. Concluding remarks

The author thinks that the training of CAT systems in academic institutions should be conducted in a more systematic way. It is believed that the mastery of each simpler knowledge will become a good foundation for the mastery of the next more complex one, which is why a cumulative hierarchical learning process is discussed in the paper. At the practical level, a task-based teaching approach is proposed to organize learning activities carried out in each class, whereby the focus is more on students rather than the teacher. It is hoped that the course objectives can be achieved within the task-based teaching stages and the teaching pedagogy of CAT systems can be brought closer to the real world.

**References**


Index

alignment, 12, 13, 19, 20, 22, 58, 92, 93, 95, 97, 98, 151, 175, 235, 237, 241.
bilingual lexicons, 37, 39, 40, 41.
bottom-up models, 11, 112, 113, 114, 115.
CAT tools, 12, 13, 14, 16, 17, 18, 19, 20, 22, 56, 57, 58, 81, 90, 93, 94, 96, 117, 121, 129, 130, 146, 165, 166, 228, 229, 231, 237, 242.
cognitive process, 237.
corresponding corpora, 51, 52, 53, 54, 55, 57, 59, 61, 62, 64, 65, 132, 146, 225, 226.
conditional random fields, 44.
confidence score, 133, 134, 135, 136.
controlled language, 100, 101, 102, 165, 166.
correlation, 58, 59, 60, 61, 62, 63, 64, 65, 71, 72, 73, 75, 135, 160, 233.
cross-language information retrieval, 36, 38, 43, 44, 51, 53.
crowd, 27, 28, 165, 167, 217.
crowdsourcing quality, 29, 32.
Dynamic Quality Framework (DQF), 155, 156, 157, 158, 159, 160, 161, 162, 163.
exemplar based machine translation, 36, 37, 38, 40, 42, 43, 92.
fuzzy matching, 12, 90, 93, 94, 98, 135.
higher education, 100.
imPLICIT POSSESSIVES, 182, 183, 187, 188, 189, 190, 192, 193.
IntelliWebSearch, 129, 211, 212, 213, 214, 215, 216.
language communities, 66.
language technology tools (LTT), 202, 203, 204, 205, 206, 209.
LEXICAL KNOWLEDGE, 93.
lexicon, 36, 37, 39, 40, 41, 42, 43, 47, 50, 103, 105, 145.
machine translation evaluation, 87, 161, 162.
malaria, 165, 167, 168, 177, 179, 180.
multiword expressions (MWE), 122, 125, 145, 146, 153.
occupational stress, 202, 209.
parallel corpora, 36, 52, 56, 59, 62, 64, 65, 129, 174, 175, 227.

translating and the computer 36

- **parsing**: 87, 98, 99, 138, 142, 154.
- **quality assessment**: 116, 117, 162, 163.
- **quality assurance**: 6, 9, 77, 78, 80, 81, 83, 112, 157, 163.
- **quality checking**: 77, 81, 83.
- **quality estimation**: 7, 10, 133, 163, 221, 227.
- **questionnaires**: 7, 8, 10, 110, 177, 179.
- **rule-based machine translation**: 25, 76, 100.
- **search tool**: 57, 250.
- **semantic similarity**: 86, 87, 88, 89.
- **Sketch Engine**: 7, 10, 129, 130, 131, 132, 254.
- **social network**: 32, 34, 75, 217, 218, 226.
- **speech recognition**: 46, 177, 181, 245.
- **sublanguage**: 8, 11, 228, 229, 230, 231, 232, 233.
- **subsegment**: 5, 11, 19, 12, 13, 14, 15, 16, 17, 18, 22, 121, 122, 123, 124, 125, 126, 127, 128.
- **subsegment recall**: 12, 13, 14, 15, 16, 17, 19, 22, 23.
- **terminological knowledge base**: 45, 47.
- **terminological variation**: 5, 6, 9, 45, 46, 49, 253.
- **translating and the computer**: 36.
- **translation relation**: 145, 151.
- **translation resource**: 5, 6, 9, 45, 46, 47, 48, 49, 101, 114.
- **translation service providers (TSP)**: 203, 204, 207, 209.
- **translation technologies**: 27, 165, 237.
- **translator training**: 162.
- **Twitter**: 205, 214, 217, 218, 219, 220, 221, 222, 226, 227.
- **user-generated content**: 66, 74, 75, 76, 225, 226.
- **volunteer motivation**: 27, 28, 29, 30.
- **word sketches**: 129.
Speakers

Alejandro Armando

Alejandro Armando is a certified translator graduated from the School of Languages, Universidad Nacional de Córdoba, Argentina. He works as a freelance translator for a number of international organizations and private sector clients. He is currently finishing a Master of Arts in Translation at the University of Geneva and conducting research on automatic translation and speech recognition.

Hanna Bechara

Hanna Bechara previously studied at Eberhard Karls University of Tubingen, where she earned her Bachelor's degree in Computational Linguistics and Computer Science in 2010. After completing an internship at Dublin City University, she started research on statistical post-editing for machine translation and was awarded her Masters of Science (by Research) from Dublin City University in 2013. She joined the EXPERT project as ESR 12 on January, 6th 2014, at the University of Wolverhampton. Her research centres around evaluation methods of different machine translation systems, specifically the hybrid systems proposed by the EXPERT projects. This work is under the supervision of Dr. Constantin Orasan and Prof. Ruslan Mitkov.
Ondřej Bojar

Ondřej Bojar graduated in computer science in 2003 and received his Ph.D. in computational linguistics in 2008 at the Faculty of Mathematics and Physics, Charles University in Prague. He now works as an assistant professor at the faculty. His main research interest is machine translation. He participated in the Johns Hopkins University Summer Engineering Workshop in 2006 as a member of the Moses team. Since then, he is regularly taking part in WMT shared translation tasks mainly with systems based on Moses and adapted for English-to-Czech translation. He was the main local organizer of MT Marathons 2009 and 2013 held in Prague.

Irina Burukina

Irina Burukina completed her Diploma with Honours (BA equivalent) in June 2012 at the St. Petersburg State University, Department of Linguistics where she majored in Mathematical Linguistics; her final qualifying paper: "Integration of multiword expressions in the RussNet thesaurus structure". Then, in June 2014, she graduated from the Russian State University for the Humanities, Institute for Linguistics; with a Diploma with Honours (MA equivalent). She majored in Computational Linguistics; her final qualifying paper: "Syntax of implicit possessives in Russian. Implicit possessives recognition and translation". Irina gained other research experience working 2010 and 2012 as a Research assistant in the RussNet project: automatic extraction of collocations in Russian (semantic and statistical approaches), and 2013 as a Research assistant in the General Internet Corpus of Russian project. Her working experience:

June 2012 – present. Linguist, Lexical Semantics Group, Technology Development Department; ABBYY Headquarters (Linguistic Software Company), Moscow, Russia.
June 2011 – May 2012. Linguist, Tree Syntax Group, Technology Development Department. ABBYY Headquarters (Linguistic Software Company), Moscow, Russia.
Russian is Irina’s native language. She has proficiency in English, advanced reading skills in German and basic communication skills in Spanish.
Her Computer skills: proficient user; Java, Python, R.
**Hernani Costa**

**Hernani Costa** is currently a Marie Curie Early Stage Researcher in the Department of Translation and Interpreting at the Faculty of Philosophy and Humanities, University of Malaga, Spain. His main research interest lies in the Computational Linguistics and Artificial Intelligence areas, especially its practical application in the fields of Translation Technologies, Natural Language Processing, Information Extraction and Information Retrieval. He is also interested in (or has worked on) a number of other topics such as Recommender Systems, Multiagent Systems, Affective Computing, amongst others.

Hernani completed his BSc and MSc on Informatics Engineering in the Bologna model at the Department of Informatics Engineering of the University of Coimbra (UC) in 2010. It was during his Master degree that he started his research activities. In particular, he developed a system capable of acquiring semantic knowledge from any kind of Portuguese text. Besides that, he also analysed the benefits of applying similarity distributional metrics, based on the occurrence of words in documents, on the system outputs. In the same academic year, he applied for a research grant to work in the "Automatic Construction of Ontologies in Portuguese" project, where he explored popular distributional similarity measures with the purpose of quantify relational triples in an automatic fashion. Furthermore, in September 2011, he was invited by LAP LAMBERT Academic Publishing to publish his MSc thesis. The book was published on October 2011 named "Automatic Extraction and Validation of Lexical Ontologies from Text: Creating Lexical Ontologies from text". In a total of six years, he studied at the aforementioned institution, developing skills on the field of Computer Science, except in the academic year of 2007/2008 where he integrated the Erasmus program for a year at the University of Vigo, Spain. During this period, besides starting to acquire skills in the Natural Language Processing area, he developed his skills in interpersonal relations (meeting Erasmus students from other cultures and languages), teamwork, research, organisation and autonomy which enabled him to develop writing and speaking skills in Spanish, as well as in English. In October 2010, he applied for a scholarship and, between December 2010 and August 2013, he worked on the project "Forms of Selective Attention in Intelligent Transportation Systems", at the Cognitive and Media Systems (CMS) group, at the Department of Informatics Engineering of the University of Coimbra. In this project, a two-parted agent architecture was implemented, with an agent responsible for gathering Points of Interest (POIs) from a location-based service, and a set of Personal Assistant Agents (PAAs) collecting information about the context and the intentions of its respective user. In each PAA were embedded a set of Machine
Learning algorithms, with the purpose of ascertaining how well suited these classifiers are for filtering irrelevant POIs, in a completely automatic fashion. During the Autumn 2011, he also developed an online service for browsing Portuguese semantic relations for the Linguateca project. It is also important to mention that he has three years of experience on teaching area. In the first two years he taught at Lousã Professional School (2010/2011 and 2011/2012) and in the academic year 2012/2013 he taught at Coimbra Institute of Engineering (ISEC). As always, he is highly motivated to find new challenges that defy his competences and skills in computer science field. That is why he enrolled the doctoral program in September 2013 at the Department of Translation and Interpreting, at the Faculty of Philosophy and Humanities of the University of Malaga, Spain.

Jerzy Czopik

Jerzy Czopik was born and grew up in Cracow, where he started to study mechanical engineering. In 1986 he moved with his wife to Dortmund/Germany. Here he finished the mechanical engineering study and started a translators and interpreters career in 1990. Jerzy is approved trainer for SDL Trados Studio and MultiTerm. Together with his wife he is certified by LICS according to EN 15038. He also acts as LICS auditor for this standard. In 2011 he published a manual on SDL Trados Studio in Polish. He is also very active on various lists and forums, helping with SDL Studio problems.

Joanna Drugan

Dr Joanna Drugan is Senior Lecturer in Applied Translation Studies at the University of East Anglia, UK. Her main research interests include translation quality, translation ethics and translation technologies. Her most recent book is Quality in Professional Translation (Bloomsbury, 2013). She is currently researching real-world ethical challenges when professional translators and interpreters are not available, particularly in healthcare and social work, and ways in which training and technology might support professionals and service users faced with such challenges.
Jo holds an MA (Hons) and PhD in French from the University of Glasgow, Scotland. She previously worked at Reading University and Leeds University, where she was a founder member of the Centre for Translation Studies and ran the MA Applied Translation Studies for over a decade. She was awarded a National Teaching Fellowship and became a member of the Higher Education Academy in 2008. She has served as a member of the Peer Review Council for the Arts and Humanities Research Council since 2012 and was selected as a founding member of the Publication Integrity and Ethics Council in 2013. Since joining UEA in 2012, Jo has led specialist Masters modules in translation technologies, translation as a profession, and research methods, and an undergraduate module on translation and globalisation. She is Director of Graduate Studies for the School.

Kurt Eberle

Kurt Eberle received his dissertation and habilitation in linguistics and computational linguistics from the university of Stuttgart in 1991 and 2004. He holds master degrees in Romance Languages and Mathematics received from the universities of Freiburg and Heidelberg in 1983 and 1987. From 1987 until 1997 he was involved in various NLP projects at the university of Stuttgart and at IBM research in Heidelberg. In 1997 he joined the MT group there where he was responsible for the development of German-French. In 1999 he was one of the co-founders of Lingenio GmbH (named Linguatec Entwicklung & Services at that time). Since 2007 Kurt Eberle is associate professor at the University of Heidelberg and since 2009 general manager of the company. He has published approx. 50 articles and monographs in the fields of MT, syntax and semantics and designed and managed a number of innovative products in the fields of MT and dictionaries at IBM and at linguatec Development&Services/Lingenio.
Michael Farrell

Michael Farrell is primarily a freelance technical translator, but is also an untenured lecturer in computer tools for translation and interpreting at the IULM University (Milan, Italy). He is an Atril Certified Training Partner and the author of "A Tinkerer's Guide to Structured Query Language in Déjà Vu X". He is also the developer of the freeware terminology search tool IntelliWebSearch and a qualified member of the Italian Association of Translators and Interpreters (AITI).

Kevin Flanagan

Kevin Flanagan is pursuing a full-time PhD at Swansea University. He spent many years as a software developer, using language skills with French clients, prior to starting work as a freelance technical translator. Having used a number of translation memory (TM) systems in that capacity, he developed a prototype TM system providing more effective sub-sentential recall. His PhD research focusses on extending and refining the system, formalising the theoretical principles and delivering a production-ready implementation.

Johanna Gerlach

Johanna Gerlach started working as a Research and Teaching Assistant at the Translation Technologies Department of the University of Geneva in 2008. She contributed to the MedSLT and CALL-SLT projects, developing linguistic resources for German. Currently, she is involved in the ACCEPT European project, investigating pre- and post-editing technologies for user generated content. In 2012, she began working on her PhD thesis, which focuses on the development and evaluation of pre-editing rules for French forum content.
**Nizar Ghoula**

*Nizar Ghoula* is a PhD candidate in Information Systems at the University of Geneva. His fields of interest include the semantic web, knowledge representation and ontology alignment. He joined the OLANTO team in order to collaborate on the development and enhancement of many CAT tools. He is also a project manager within the Executive department of the University of Geneva to design a solution for executive programs management.

**Attila Görög**

*Attila Görög* has been involved in various national and international projects on language technology in the past 10 years. He has a solid background in Quality Evaluation, Post-Editing and Terminology Management. Attila is interested in globalization issues and projects involving CAT tools. His webinars and workshops discuss hot topics in the translation industry with aim of making participants future proof. As a product manager at TAUS, he is responsible for the TAUS Evaluation platform also referred to as the Dynamic Quality Framework or DQF.

**Jacques Guyot**

*Jacques Guyot* is a senior computer scientist with over 20 years of experience in turning break-through technologies into professional solutions. He is also a researcher at the University of Geneva. He holds a PhD in Computer Sciences from the University of Geneva.
Najeh Hajlaoui

Najeh Hajlaoui received his PhD in computer science from Joseph Fourier University (Grenoble, France) in 2008 on Multilingualization of ecommerce systems handling spontaneous utterances in natural language.
In 2002 he received his MS in information systems at Joseph Fourier University, and his Joint European Diploma MATIS (Management and Technology of Information Systems).
He is currently Senior Researcher and Project Manager for Machine Translation at the European Parliament in Luxembourg (since August 2013).
Before joining the Idiap Research Institute in December 2011, he has been a Research Fellow at the University of Wolverhampton (UK) in 2011, a Postdoctoral Researcher at Orange Labs (Lannion, France) in 2010, and an Associate Lecturer at Jean Monnet University (Saint-Étienne, France) from 2007 to 2009.

Sabine Hunsicker

Sabine Hunsicker is a computational linguist with experience in statistical natural language processing with a focus on machine translation.
Sabine completed her M.Sc. in computational linguistics at Saarland University in 2009. Her thesis concerned example-based machine translation enhanced with statistical methods.
Before euroscript, she worked at the German Research Center for Artificial Intelligence (DFKI) in Saarbrücken. Her research topics were hybrid machine translation and she was involved with the Euromatrix Plus and ACCURAT research projects with a strong focus on integrating linguistic analysis into the SMT workflow.
Her areas of expertise include linguistic analysis, data mining as well as information retrieval.
Miguel A. Jiménez Crespo

Miguel A. Jiménez-Crespo is an Assistant Professor in the Department of Spanish and Portuguese at Rutgers University, where he directs the MA program in Spanish Translation and Interpreting. He holds a PhD in Translation and Interpreting Studies from the University of Granada, Spain. His research focuses on the intersection of translation theory, translation technology, digital technologies, corpus-based translation studies and translation training. He is the author of *Translation and Web Localization* published by Routledge in 2013, and has published extensively in the top international journals in the discipline of Translation Studies.

Koen Kerremans

Koen Kerremans obtained his Master’s degree in Germanic Philology (Dutch-English) at Universiteit Antwerpen in 2001, his Master’s degree in Language Sciences - with a major in computational linguistics - at Universiteit Gent in 2002 and his PhD degree in Applied Linguistics at Vrije Universiteit Brussel in 2014 (Title of his dissertation: ‘Terminological variation in multilingual Europe. The case of English environmental terminology translated into Dutch and French’). His research interests pertain to applied linguistics, language technologies, ontologies, terminology (variation) and translation studies. He currently holds a position as post-doctoral researcher and teaching assistant at the department of Applied Linguistics (Faculty of Arts and Philosophy) of Vrije Universiteit Brussel (VUB) where he teaches applied linguistics, terminology and several Dutch language courses. He is a member of VUB’s research group ‘Centrum voor Vaktaal en Communicatie’ (Centre for Special Language Studies and Communication).
Adam Kilgarriff

Adam Kilgarriff is Director of Lexical Computing Ltd. He has led the development of the Sketch Engine, a leading tool for corpus research used for dictionary-making at Oxford University Press, Cambridge University Press, HarperCollins, Le Robert and elsewhere. His scientific interests lie at the intersection of computational linguistics, corpus linguistics, and dictionary-making. Following a PhD on “Polysemy” from Sussex University, he worked at Longman Dictionaries, Oxford University Press, and the University of Brighton prior to starting the company in 2003. He is a Visiting Research Fellow at the University of Leeds. He has been an Expert Witness in a number of legal cases involving trademarks. He is active in moves to make the web available as a linguists’ corpus and was the founding chair of ACL-SIGWAC (Association for Computational Linguistics Special Interest Group on Web as Corpus). He has been chair of the ACL-SIG on the lexicon and Board member of EURALEX (European Association for Lexicography).

Terence Lewis

Terence Lewis, MITI, entered the world of translation as a young brother in an Italian religious order, when he was entrusted with the task of translating some of the founder's speeches into English. His religious studies also called for a knowledge of Latin, Greek and Hebrew. After some years in South Africa and Brazil, he severed his ties with the Catholic Church and returned to the UK where he worked as a translator, lexicographer (Harrap’s English-Brazilian Portuguese Business Dictionary) and playwright. As an external translator for Unesco he translated texts ranging from Mongolian cultural legislation to a book by a minor French existentialist. At the age of 50 he taught himself to program and wrote a Dutch-English machine translation application which has been used to translate documentation for some of the largest engineering projects in Dutch history. For the past 15 years he has devoted himself to the study and development of translation technology.
Jessica Xiangyu Liu

**Jessica Xiangyu Liu** is a research postgraduate at Department of Translation, The Chinese University of Hong Kong, Hong Kong. She is in her first year of the MPhil programme in Translation. She has a strong interest in the teaching computer-assisted translation systems, and the hybrid of machine translation and computer-assisted translation. Before conducting her study, she worked as a research assistant at Centre for Translation Technology, CUHK (2010-2013) where she was engaged in the research and training of computer-assisted translation software.

Erin Lyons

**Erin M. Lyons** is a full-time French to English and Italian to English translator, medical writer and consultant and the Owner and President of BiomedNouvelle. She is also an Adjunct Professor of Translation in the Master’s Degree programme at the University of Maryland (USA). Her primary areas of focus include clinical research, pharmaceuticals, medical devices, and cosmetic products. Furthermore, she continues to work on the development of BabelNouvelle®, a mobile-based translation technology employing crowdsourcing and machine translation to facilitate medical services in the developing world. Ms Lyons has split her professional career between Europe and the U.S. She has a BA in Romance Languages and Literature from the University of Chicago and an MA in Italian and French Translation from the Monterey Institute of International Studies.

Ms Lyons frequently presents her work to private groups, universities and at international conferences, including at ATA Annual Conferences, the World Congresses of the International Federation of Translators and the International and France ProZ.com Conferences.
Tengku Sepora Tengku Mahadi

*Tengku Sepora Tengku Mahadi* (Associate Professor Dr.) is Dean of the school of Languages, Literacies and Translation at the University Sains, Malaysia. She lectures in translation theories and practice and supervises research in Translation Studies at MA and PhD levels. She is the author of *Text-wise in Translation* (2006), and co-author of *Corpora in Translation: A Practical Guide* (2010).

Victoria Porro

*Victoria Porro* holds a bachelor and a master in Translation Studies and Translation Technologies. She joined the Translation Technologies Department of the University of Geneva as a Research and Teaching Assistant in June 2012. Since then, she devotes most of her time to the EU-funded ACCEPT project she participates in and she is currently designing a PhD project in post-editing and machine translation. She is most interested in opening new lines of research in post-editing and advocates for the recognition of post-editing as a highly skilled activity.

Gábor Prószéky

*Gábor Prószéky* is a professor and vice dean of the Faculty of Information Technology and Bionics at the Pázmány Péter Catholic University (Budapest) and CEO of MorphoLogic, a leading Hungarian language technology company. He is in charge of the MTA-PPKE Hungarian Language Technology Group co-financed by the Hungarian Academy of Sciences and the Pázmány University. He graduated at the ELTE University both in software engineering and in general & applied linguistics. He holds a PhD (1994) in computational linguistics. In 2005 he received the title of Doctor of the Hungarian Academy of Sciences. Since his university years, he has been involved in more than
thirty R&D projects in human language technologies (HLT), and computational and theoretical aspects of humanities.

His research interest covers various aspects of computational analysis of highly inflectional languages, intelligent dictionaries and machine translation.

Aside of more than 140 scientific publications mainly on HLT, he is the author of three comprehensive books on human language technologies.

Among others, he was a Board Member of the European Language Resources Association and the President of the Lexicographical Committee of the Hungarian Academy of Sciences (2006-2012). Since 2013 he has been the president of the Association of Hungarian Applied Linguists, and in 2014 he became the president of the Council of Social Sciences and Humanities of the Hungarian Scientific Research Fund.

In 1991, with software engineer colleagues working on human language technology applications, he founded MorphoLogic, the first language industry company in Hungary. Since then, MorphoLogic’s various applications have been licensed by Microsoft, IBM, Xerox, among others. In 1999, MorphoLogic won the IST Prize of the European Commission.

Gábor Prószyéky received Hungary’s highest award, the Széchenyi Prize, for his activities in 2000. Among others, he also received the Kalmár Award of the John von Neumann Computer Society (1995), IT Manager of the Year (2002), Award for the Hungarian IT (2005), Special Prize to the IT Lecturer of the Year (2009) and Dennis Gabor Award (2010).

Emmanuel (Manny) Rayner

Emmanuel (Manny) Rayner is a Collaborateur Scientifique (senior non-teaching research post) at the University of Geneva’s Multilingual Processing Group, and has previously held positions at SRI International (1991-1999) and NASA Ames Research Center (1999-2005).

Over the last 20 years, his research has focused on construction of speech-enabled dialogue systems.

He has played key roles in several major projects in this area, including SRI’s Spoken Language Translator, NASA’s Clarissa and the Open Source Regulus platform.

At the University of Geneva, he has worked primarily with speech translation and CALL.

He has more than 100 refereed publications in speech technology, computational linguistics, machine translation and artificial intelligence.
Nasradine Semmar

Nasredine Semmar obtained his Ph.D. in 1995 at University of Paris-Sud (France) in 1995 on Multimedia software localization. He worked from 1996 to 2000 as an R&D engineer in Lionbridge Technologies – Bowne Global Solutions, he designed and implemented tools for Computer Aided Translation and he participated in delivering the multilingual version of MS Windows 2000. He then joined SAP - Business Objects where he worked from 2000 to 2002 as an expert in software internationalization and localization. Since 2002, he has been working as a research scientist at the Vision and Content Engineering Laboratory (LVIC) where he has implemented the treatment of Arabic in the CEA LIST NLP platform for a cross-language search engine and he has developed several tools for sentence and word alignment from parallel corpora. He is the convenor of the work group “Multilingual information representation” of the ISO/TC37/SC4 and he participates as an expert in promoting the MLIF standard (Multilingual information framework).

Eduard Šubert

Eduard Šubert studies informatics and mathematics at Czech Technical University in Prague. He was introduced to computational linguistics in course at the university. Among other interests, Eduard is responsible for the creation of science popularizing video content of his faculty’s YouTube channel. Additionally, he works on a computer simulation of lens polishing process for the Academy of Sciences of the Czech Republic.
AnneMarie Taravella

AnneMarie Taravella, cert. tr. (OTTIAQ) is a doctoral student and part-time faculty at Université de Sherbrooke, in Québec (Canada), as well as a member of Ordre des traducteurs, terminologues et interprètes du Québec (OTTIAQ).
She holds a BA in Translation and a MA in Translation Studies, both from Montreal-based Concordia University. She is also a graduate from Université de Paris-IX Dauphine, France, in Management Science.
AnneMarie is now pursuing a Doctorate in Business Administration (DBA) at the Faculty of Administration of Université de Sherbrooke, under the supervision of Alain O. Villeneuve, DBA.
Her research interests are translation work organization, adoption of information technologies in organizations, workplace well-being and positive organizational scholarship.
Her doctoral research focuses on the variation of language specialists' affective states in the workplace.
Her research is supported by the Canadian Social Sciences and Humanities Research Council.

Antonio Toral

Antonio Toral (Dr.), Research Fellow at Dublin City University (DCU).
Obtained his MSc in Computer Science in 2004 and PhD in Computational Linguistics in 2009 from the Universitat d’Alacant (Spain).
He worked as a researcher in CNR-ILC (Italy) from 2007 to 2009, involved in the EU-FP7 projects KYOTO and FLaReNet.
He joined DCU in 2010 where he has been working as a postdoctoral researcher to date, in the EU-FP7 projects Abu-MaTran (coordinator), QTLaunchPad, PANACEA and CoSyne.
He has published more than 70 peer-reviewed papers, has served in the scientific committee of international conferences and workshops and has reviewed papers for three indexed journals of the field.
He has also organised evaluation tasks at the SemEval and EVALITA forums.
Tom Vanallemeersch

Tom Vanallemeersch is a researcher at KU Leuven, Centre for Computational Linguistics. He has been working in the language technology sector for twenty years, both in academia and industry. His activities mainly involve translation memories, machine translation and alignment of bilingual resources. He performed work in these fields at Xplanation, LNE International, Lessius University College, Systran (project in collaboration with Centre for Computational Linguistics) and the European Commission. Other types of language technology he dealt with are multilingual dictionary processing (University of Liège), text-to-speech (Lernout and Hauspie), text mining (Temis) and terminology extraction (coordinator of project at Dutch Language Union).

Marion Wittkowsky

Marion Wittkowsky is a lecturer in the Department of International Technical Communication at the Flensburg University of Applied Sciences in Germany since 2007. She teaches courses in technical writing, technical translation, and applied computer linguistics. Prior to her position at the University she worked for ten years as a technical translator, project manager and finally as a business unit manager at a language service provider. A major focus of her translation work was post-editing the machine translation of SAP release notes.
Angelika Zerfass, a consultant, trainer and provider of technical support for users of translation tools is located in Bonn, Germany. After her studies (degree in translation for Chinese and Japanese plus Computational Linguistics) she came to the translation industry first as a training specialist for a tools provider in 1997 and has been working freelance since 2000. Her main focus being the translation technologies, she is traveling the world on a mission to make translation tools understandable.

Andrzej Zydroń is one of the leading IT experts on Localization and related Open Standards. Zydroń sits/has sat on, the following Open Standard Technical Committees:

- LISA OSCAR GMX
- LISA OSCAR xml:tm
- LISA OSCAR TBX
- W3C ITS
- OASIS XLIFF
- OASIS Translation Web Services
- OASIS DITA Translation
- OASIS OAXAL
- ETSI LIS
- DITA Localization
- Interoperability Now!
- Linport

Andrzej has been responsible for the architecture of the essential word and character count GMX-V (Global Information Management Metrics eXchange) standard, as well as the revolutionary xml:tm (XML based text memory) standard which will change the way in which we view and use translation memory. Andrzej is also chair of the OASIS OAXAL (Open Architecture for XML Authoring and Localization) reference architecture technical committee which provides an automated environment for authoring and localization based on Open Standards.

He has worked in IT since 1976 and has been responsible for major successful projects at Xerox, SDL, Oxford University Press, Ford of Europe, DocZone and Lingo24 in the fields of document imaging, dictionary systems and localization.
Andrzej is currently working on new advances in localization technology based on XML and linguistic methodology.

Highlights of his career include:
2. Writing a system for the automated optimal typographical formatting of generically encoded tables (1989).
3. The design and architecture of the Xerox Language Services XTM translation memory system.
4. Writing the XML and SGML filters for SDL International's SDLX Translation Suite.
5. Assisting the Oxford University Press, the British Council and Oxford University in work on the New Dictionary of the National Biography.
6. Design and architecture of Ford's revolutionary CMS Localization system and workflow.
7. Technical Architect of XTM International's revolutionary Cloud based CAT and translation workflow system: XTM.

Specific areas of specialization:
1. Advanced automated localization workflow
2. Author memory
3. Controlled authoring
4. Advanced Translation memory systems
5. Terminology extraction
6. Terminology Management
7. Translation Related Web Services
8. XML based systems
9. Web 2.0 Translation related technology
Gold Sponsor

Matecat

More Matches for your Translations

Free and open source enterprise-level translation software
From 10% to 20% more matches than any other CAT tool
Increased privacy, no more files via email

A professional tool for language service providers and MT specialists

Collect data to set a fair rate for post-editor and improve MT quality
Real-time progress report and quality control for your translations
Online adaptation and quality estimation for MT systems based on Moses

Start translating
www.matecat.com
What I like about SDL TRADOS Studio 2014

- **The ribbon** – I really like it and I also think that this will be helpful for new or basic users of Studio in particular. In previous training sessions, I found that trainees could not always find what they were looking for, whereas the new ribbon makes it much easier.
- **Automatic concordance search** – a very useful feature.
- **“One dick” batch tasks** – for example, this makes analysis of a single file and single language combination much faster.
- **Performance speed** – it is noticeably quicker when creating projects.
- **Larger font in Termbase Viewer** – much better for my eyes!
- **MS Word 2007-2013 comments options** – I’m pleased to see that we can now choose whether these are extracted as translatable or Studio comments.

Gemma Cooper
Translations and Training Manager, Alexika | www.alexika.com

Easier, smarter, faster… enjoy translation

www sdlcom/Studio2014
www.translationzonecom/studio2014

Purchase or upgrade to SDL Trados Studio 2014 today

Take it further, share projects with SDL Studio GroupShare
www.translationzonecom/groupshare2014
Silver Sponsor


I like memoQ because the tool reflects the people behind it: straightforward yet ingeniously creative, and truly committed to making their users’ lives easier.

Stefan Starling
German-English translator
 Sarasota, FL
For information on next year’s 37th Translating and the Computer conference, please check http://translatingandthecomputer.com and http://asling.org where information on calls for abstracts and posters, along with other information will be posted as it becomes available.