HAL open science

The Impact of Machine Translation on a Masters Course in Web Translation: From Disrupted Practice to a Qualitative Translation/Revision Workflow

Christopher Gledhill, Maria Zimina-Poirot

To cite this version:

Christopher Gledhill, Maria Zimina-Poirot. The Impact of Machine Translation on a Masters Course in Web Translation: From Disrupted Practice to a Qualitative Translation/Revision Workflow. Translating and the Computer 41, AsLing, The International Association for Advancement in Language Technology, Nov 2019, London, United Kingdom. pp.60-73. hal-02522883

HAL Id: hal-02522883

https://u-paris.hal.science/hal-02522883v1

Submitted on 28 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Translating and the Computer 41

LANGUAGE ASLING TECHNOLOGY

21-22 November 2019

One Birdcage Walk, London

Proceedings

ISBN 978-2970-10957-0

9 782970 109570

November 2019. Editions Tradulex, Geneva

© AsLing, The International Association for Advancement in Language Technology This document is downloadable from www.tradulex.com and www.asling.org

Acknowledgements

AsLing wishes to thank and acknowledge the support of the sponsors of TC41:

Gold Sponsors

Silver Sponsors

Preface

At a time when translation and technology are more closely than ever intertwined, the forty-first edition of Translating and the Computer (TC) could not be timelier. The TC conferences, organised by The International Association for the Advancement in Language Technology (AsLing), continue to serve as a unique forum for academics, developers, users, and vendors of computer aids for translators, of other translation technology tools, and this increasingly, for interpreters and others performing new roles in our industry. This distinctive event brings together freelance and in-house translators, interpreters, researchers and business people from translation companies, international organisations, universities and research centres, offering them an opportunity to exchange ideas, and to learn about and discuss the latest developments in translation technologies.

As translation memory, corpora and terminology tools evolved and were incorporated to various degrees into MT systems and suites over the last four decades translators have found their work, their workflows and even key aspects of their profession awash with change. The replacement of first-, second- and even third-generation tools by neural versions, increasingly augmented by artificial intelligence, has accelerated the rate of change and has resulted in more radical changes. The latest tools increasingly offer real benefits for interpreters. Serious professionals and astute business managers are best served by advanced and in-depth knowledge of the latest developments, tempered by well-founded observations based on solid research and advice from leaders in the professions and industry. TC conferences feature speakers from academia, industry and the professions that provide such insights.

AsLing is delighted to present this year's keynote speakers: Jean Senellart, CEO and Director for Research & Development at Systran, who discusses the neural revolution, whereas Jochen Hummel, co-founder and CEO of Coreon, founder of TRADOS and of Metaversum, focuses on unlocking machine translation. We are confident that these keynotes along with the presentations, posters, panels and workshops provide participants valuable insights and opportunities for inspiring discussions. We are also confident that the e-proceedings with the conference contributions, accepted after a competitive reviewing process, will be an important reference and stimulus for future work of all those active in AsLing fields of competence.

One session examines the resurging awareness of the importance of terminology. A presentation about sign language translation highlights the conference's commitment to tools for inclusion. Other presentations examine post-editing methods and tools in the world of neural MT, de-demonizing AI, applying AI to MT, training for and with regard to MT and post-editing as well as aspects of error evaluation and annotation.

We thank all who submitted proposals to the conference and those authors who produced full versions of their papers for the proceedings, as well as all whose slides are available on the AsLing website. A special thank-you goes to all the delegates wherever they have come from, who by taking part and interacting with presenters and fellow participants broaden this conference and give living acknowledgement of this special event.

We are grateful to the members of the Programme Committee who carefully reviewed the submissions: Juan José Arevalillo, Sheila Castilho, David Chambers, Caroline Champsaur, Gloria Corpas Pastor, Joanna Drugan, David Filip, Camelia Ignat, Raisa McNab, Vilelmini Sosoni, Paola Valli, Nelson Verástegui and Michal Ziemski. A big thanks to Sandra Chambers, Joss Moorkens and Maria Recort Ruiz who as fellow members of the Organising committee played key roles to make this conference happen, and to our Social Media Officers Maria Recort Ruiz and Nelson Verástegui, and David Chambers and Joss Moorkens who also provided ongoing editorial support. The biggest thanks this year, though, goes indisputably to Jean-Marie Vande Walle who served as conference coordinator. Last but not least, we thank our sponsors and all those who lent their support, helping to make both the conference and these proceedings possible.

Conference Chairs

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

London, November 2019

The Executive Committee of AsLing establishes several bodies each year, to organise and carry out the annual conference. Membership in these bodies overlap. The tables below show membership in these bodies for TC41.

Conference and Session Chairs:

João Esteves-Ferreira, AsLing President
Juliet Margaret Macan, AsLing Vice-President
Ruslan Mitkov, AsLing Vice-President
Olaf-Michael Stefanov, AsLing Vice-President

Conference Organising Committee:

Sandra Chambers, Organisation for Economic Co-operation and Development (OECD)
João Esteves-Ferreira, Tradulex - International Association for Quality Translation
Juliet Margaret Macan, independent translation technology consultant
Joss Moorkens, Dublin City University
Ruslan Mitkov, University of Wolverhampton
Maria Recort Ruiz, International Labour Organization
Olaf-Michael Stefanov, JIAMCATT, United Nations (ret.), Coordinator
Jean-Marie Vande Walle, Coordinator

Editors of the Proceedings:

David Chambers, World Intellectual Property Organization (ret.)
João Esteves-Ferreira, Tradulex - International Association for Quality Translation
Juliet Margaret Macan, independent translation technology consultant
Ruslan Mitkov, University of Wolverhampton
Olaf-Michael Stefanov, JIAMCATT, United Nations (ret.)
Jean-Marie Vande Walle, Publication Chair

Other Officers:

María Recort Ruiz, Social Media Officer
Nelson Verástegui, Social Media Officer

Programme Committee:

Juan José Arevalillo, Hermes Traducciones
Sheila Castilho, Dublin City University
David Chambers, AsLing Honorary Member
Caroline Champsaur, Organisation for Economic Co-operation and Development
Gloria Corpas Pastor, University of Málaga
Joanna Drugan, University of East Anglia
David Filip, CNGL / ADAPT
Camelia Ignat, Joint Research Centre of the European Commission
Raïsa McNab, CEO of the Association of Translation Companies
Bruno Pouliquen, World Intellectual Property Organization
Vilelmi Sosoni, Ionian University
Paola Valli, University of Trieste
Nelson Verástegui, International Telecommunications Union (ret.)

Contents

Innovation in the International Organizations: Can We Do Better? 5

Josep Bonet

Terminology: Towards a Systematic Integration of Semantics and Metadata 10

Denis Dechandon, Anikó Gerencsér, Maria Recort Ruiz

ACCOLÉ: A Collaborative Platform of Error Annotation for Aligned Corpora 27

Emmanuelle Esperança-Rodier, Francis Brunet-Manquat, Sophia Eady

Time is Everything: A Comparative Study of Human Evaluation of SMT vs. NMT 36

Emmanuelle Esperança-Rodier, Caroline Rossi

Preferences of end-users for raw and post-edited NMT in a business environment 47

Sabrina Girletti, Pierrette Bouillon , Martina Bellodi, Philipp Ursprung

The Impact of Machine Translation on a Masters Course in Web Translation: From Disrupted Practice to a Qualitative Translation/Revision Workflow 60

Christopher Gledhill, Maria Zimina

MT Evaluation in Translation Quotations at CPSL with ContentQuo 74

Lucía Guerrero, Kirill Soloviev

Assisting Sign Language Translation: What Interface Given the Lack of Written Form and Spatial Grammar? 83

Marion Kaczmarek, Michael Filhol

New Audiences in EU Terminology 93

Rodolfo Maslias

Terminology Extraction as a Tool for MT Output Assessment and Improvement 95

Jean-François Richard

Pilot Study on Medical Translations in Lay Language: Post-Editing by Language Specialists, Domain Specialists or Both? 101

Ayla Rigouts Terryn, Lieve Macken, Els Lefever, Robert Vander Stichele, Koen Vanneste, and Joost Buysschaert

Domain Adaptation for Machine Translation Involving a Low-Resource Language: Google AutoML vs. from-scratch NMT Systems 113

Margita Šoštarić, Nataša Pavlović, Filip Boltužić

Undergraduate Translation Students' Performance and Attitude vis-à-vis Machine Translation and Post-editing: Does Training Play a Role? 125

Maria Stasimioti, Vilelmini Sosoni

De-Demonizing AI Regarding Localization 137

Andrzej Zydroń

Poster: Usefulness of Translation Technology Training from Mexican Universities 145

Argelia Peña Aguilar

The Impact of Machine Translation on a Masters Course in Web Translation: From Disrupted Practice to a Qualitative Translation/Revision Workflow

Christopher Gledhill

Université de Paris UFR EILA (Etudes Interculturelles et Langues Appliquées)

cgl@eila.univ-paris-diderot.fr

Maria Zimina

Université de Paris UFR EILA (Etudes Interculturelles et Langues Appliquées)

mzimina@eila.univ-paris-diderot.fr

Abstract

The introduction of technology into translation curricula is a complex task in terms of translation competences and their acquisition. Computer tools and MT directly affect trainee translators. This study investigates the impact of technology on students on a Master's in Specialised Translation and Language Industries at Université Paris Diderot. We present the results of a teaching project "Website translation into English" which places strong emphasis on hands-on applications of MT. The aim of the project is to provide students with a semi-professional work experience in which they face real-life website translation problems. Students are expected to translate and revise webpages from French into English using a professional platform SystranLINKS. The first results of our study show that a more equipped translator's workstation results in assisted but also disrupted translation practice, and requires additional learning/teaching time. Intensive practice of MT raises students' awareness of the importance of a revision workflow, and gives students a broader understanding of translation quality.

1 Introduction

In this contribution, we describe the impact of Machine Translation (MT) on a Masters course on webpage translation. Our claim is that MT has two principal forms of impact on our students (and on ourselves, as translator trainers, as well as on the wider community). The first is 'disrupted practice', that is to say the initial challenge encountered by students in the context of what is – for many – a completely new learning experience. The second impact involves not only the development of 'survival strategies' when faced with a highly complex project simulation, but also more profoundly a qualitative change in attitudes among all participants, not only towards technology (the translation platform, neural MT output, corpora and other resources) but also our conception of the translation process itself. We characterise this emerging view of the translation project as a 'qualitative translation/revision workflow'.

This paper has six sections: 1) the web translation project and pedagogical workflow, 2) the specificities of webpage translation, 3) the tools used for website translation, 4) the role of neural MT in the project, 5) the impact of MT-enhanced technologies on the translation process, and 6) the role of metalanguage in students' reporting forms.

In the final sections of this paper, we suggest that an efficient and perhaps relatively underexploited means of examining the impact of technology on translation trainees is to analyse the metalanguage that our students actually use. As mentioned below, we use a 'Reporting Form' to assess our students' participation and progress on the course. Here, we use a hybrid approach to explore these texts (corpus analysis, textometrics, but also critical discourse analysis). The analysis of student reporting forms provides us with data on the representations that students form about the project: the students' professional worldview comes across in the metalanguage they use (terms about language, technical terms relating to the project workflow etc.), but also in the choices they

make regarding examples and in longer stretches of text in which they justify their evaluation of the course itself.

2 The Website Translation Project

In this section we set out the context for Masters-level university course entitled *Traduction de site web vers l'anglais* ('Website translation towards English') for which we use the acronym TSA. TSA is a 30 hour-long course which is itself part of a 2nd year programme known as M2 ILTS (*Industrie de la langue et traduction spécialisée,* 'Specialised Translation and Language Industry'), based at the Université de Paris, France[3]. M2 ILTS includes practical modules on technical translation in various specialised domains, translation tools, project management, as well more theoretical courses on terminology, corpus linguistics, etc. The particularity of M2 ILTS is that students must find an alternating work-study placement (one week in a company, the next at the university), a feature which exposes them to different translation tools and environments in the workplace, as well as intensive tuition at university.

The aim of the TSA course is to provide a simulated environment in which teams of students work on a single website translation project. Typically, this involves translation from French into English of webpages taken from a French university or other institution (an average of 1-2 pages per student, given that there are approximately 30-35 participants). Over the years we have worked on websites such as the *Institut de physique du globe de Paris* (IPGP): http://www.ipgp.fr. We generally only select pages on the website that have sufficient content: texts are at least 200 words long, and the genres involved range from scientific reports (descriptions of geological instruments, volcanic activity reports, etc.) to administrative and procedural announcements (job adverts, instructions for PhD enrolment, calls for research seminars, etc.). As discussed below, each individual student has one or more roles:

- Project Manager (responsible for communication between the teachers, translation team and website owners)
- Translator (responsible for translation/post-editing of a particular page or pages, one half of the Translator/Reviser pair)
- Reviser (responsible for post-editing/revision of a particular webpage, the other half of a Translator/Reviser pair)
- Auditor (responsible for an audit of the website and for dividing the website into manageable tasks)
- Contract writer (responsible for drawing up an agreement with the website owners or 'clients')
- Terminologist (responsible for drawing up a glossary of recurring terms, menu items, etc.)
- Task manager (responsible for managing tasks on the translation platform).

This division of labour is set out at the beginning of the course, and provides a tangible framework into which each student can see where she/he fits in the overall project.

While we are aware of many other simulated translation projects at this level (O'Brien, 2002; Blagodarna, 2018; Guerberof *et al.*, 2019), we believe that TSA presents a unique set of challenges:

[3] For more information: http://www.eila.univ-paris-diderot.fr/formations-pro/index (Coordinators: Geneviève Bordet and Nicolas Froeliger).

a) **Transversality**. The course is intended to be 'transversal' (i.e. to cover a wide variety of specific competencies that are taught as separate modules elsewhere on the ILTS programme). A key feature of transversality is that students are expected to become familiar with semi-automated translation and to undertake a process of 'transcreative intervention' (a critical approach to MT, a feature that we examine below). Most if not all of the modules in M2 ILTS cover the skills identified by the EMT Competence Framework (2017). The TSA course places particular emphasis on the skills set out below [the emphasis in bold is ours]):

- "[Competency No.] 4 Acquire, develop and use thematic and domain-specific knowledge relevant to translation needs (mastering […] **presentation standards**, **terminology** and **phraseology**, specialised sources etc.)"
- "10 Analyse and justify their translation solutions and choices, **using the appropriate metalanguage** and applying appropriate theoretical approaches"
- "11 Check, **review and/or revise their own work and that of others** according to standard or work-specific quality objectives"
- "18 Master the **basics of MT** and its impact on the translation process"
- "19 Assess the relevance of **MT systems in a translation workflow** and implement the appropriate MT system where relevant"
- "23 **Work** in a **team**, including, where appropriate, in virtual, multicultural and multilingual environments, using current communication technologies"
- "29 Clarify **the requirements, objectives and purposes of the client**, recipients of the language service and other stakeholders and offer the appropriate services to meet those requirements."

We discuss how each of these skills fit into the TSA project in later sections.

b) **Web tectonics**. The translation of web content requires flexibility and sensitivity to many different genres. We therefore expect our students to acknowledge that a website is usually made up of more than one register, and this typically involves sometimes quite distinctive strategies. For example, a translation solution adopted on a presentational page for casual visitors ('Who are we?') may not be the same as for a page aimed at members of the institution ('Jobs and Vacancies'), etc. In addition, some parts of the website are more or less permanent ('Legal Information') whereas other pages are more ephemeral ('Workshops and Events'). A particularity of the IPGP website is that pages on volcanic activity are constantly updated, a fact that reflects the constant activity of volcanos and the on-going work of IPGP observatories. Thus, ironically, the content of the IPGP website mirrors the conceptual world of the Earth Sciences: it is constantly shifting, a feature that is often overlooked when clients or translators talk about 'the translation product' or 'final delivery'.

c) **Non-linear task management**. The translation platform SYSTRANLinks we use for the course can constantly update the target text (TT) according to changes made in the source text (ST). The upshot is that translators have both an opportunity (there is less need to constantly review correspondences between ST and TT) and a constraint (the translation of many parts of the website can never be considered definitive). This raises a further issue to do with project management: since the website translation project is never completely finished, there is a constant need for feedback and dialogue between the translation team and the website owners/managers. Thus, although the TSA course has a natural academic cycle (with the students submitting their reporting

forms towards the end), the project itself is usually far from finished. This means that the Project Managers are usually not realistically involved in delivery or 'after sales service'. This is clearly one of the pedagogical aspects of the TSA course which deviates considerably from professional practice.

d) **Process-oriented assessment**. The evaluation of the TSA course is performed quite differently to the way this is managed traditionally in the French system (which usually focuses on the quality of the end-product, i.e. 'product-oriented assessment'). We argue that a product-based approach would be unrealistic for our purposes. The main explanation for this is that the TSA course is not about language proficiency alone. Furthermore, the students do not receive comparable pages to translate/revise, so ultimately the final quality of their translation work is just one of many factors we wish to assess. After some experimentation (student exposés, etc.) we have settled on the 'Reporting Form' as an ideal way of assessing students (see Appendix A). This presents several advantages. First, it obliges students to report on the basic facts about their assigned task (the volume of the ST and TT in words/signs, number of edits, revision cycles, typology and distribution of linguistic problems encountered, etc.). We find that such basic data are often overlooked by trainee translators unless they are specifically prompted. Another advantage is that the Reporting Form gives students an open forum so that they demonstrate explicit knowledge about certain features of the project workflow, as well as commenting in an open-ended way on aspects of the project that they found challenging. We acknowledge that the Reporting Form is not a 'protocol' (examples of more formal studies which use self-reporting forms of this type can be found in Piecychna, 2016). However, we claim that it is still interesting to see how our students express their understanding of the metalanguage of the translation process in the various ways in which the forms have been completed. We discuss this in further detail in section 6, below.

3 The Website Translation Platform SYSTRANLinks/Features

SYSTRANLinks is a multilingual translation platform that offers website translation and localisation features. In addition to on-demand Machine Translation with custom settings (MT engine configuration table), its Content Management System (CMS) also provides the tools to edit and manage all translated content collaboratively (from a centralised base) using a flexible review and publication workflow. When a new project is launched, the platform creates a copy of a specific website (which can be automatically updated according to changes made online).[4] Project members (having different roles[5] and access privileges) can view all pages of the website within a structured menu tree. The status of each page indicates whether the page has been crawled, or is in the 'to be crawled' list. It is possible to quickly retrieve all 'sentences' (or text segments) of any crawled page with corresponding MT translations, search particular sentences through filters, edit any translated sentence or select a set of sentences and create a translation 'Task' with it (for both 'post-editing' and 'full review': to 'edit', 'validate' or 'reject' any text segment). It is also possible to retrieve all non-textual elements parsed by the system (such as images, files, or external links). They can be changed in translated versions of the website using specific search and replace features.

Translation workflow with 'Tasks' management enables 'on-the-fly' translation and validation, directly from the 'sentences' list or from the website. Edited and reviewed sentences have a visibility status to track changes in the system. A sentence is visible on a translated website only

[4] Once a website translation has been fully translated, revised and fine-tuned, there are several ways to make it accessible to visitors using one of the "URLs navigation scenarios": http://support.systran.net/systranlinks/tutorials/website-settings/#translationsettings

[5] As mentioned above, the specific structure of the web translation project is framed by the assignment of different roles with corresponding access privileges on the translation platform.

when it is published (i.e. when an editor / reviser changes its status from 'validated' to 'published'). Whatever its validation status, this setting lets contributors control publishing progress in accordance with project management guidelines. Moreover, a revision history tool (tracing) gives an idea of how many revisions of a given text segment have been made: it gives users a way to make changes that are easy to spot and to review.

Along with this translation workflow, the platform has three features that we consider essential for professional website translation. First, it provides a specific 'context' view called 'Browse & Edit' mode that allows translators to edit any target text segment having access to the source page context and its translation. Second, all repeated text segments coming from different webpages (titles, tooltips, menu items, etc.) are automatically tracked in the 'Sentences' table to make relevant choices that create a coherent website translation. Third, translations can be enhanced and customised to a specific domain and project settings:

- With Specialised Neural MT engine: we use a connection with SYSTRAN Pure Neural Server (PNS) API account[6]
- With Translation Memory (TM): any sentence that has been reviewed/translated once will be identified and a TM can be uploaded or downloaded at any stage of the project
- With User Dictionary (UD): project terminology and glossary can be defined through Dictionary Manager: we use PNS UD for neural MT.

4 The Role of Neural MT in the Project

Our website translation course covers many specific competencies that are taught as separate modules, seminars or conferences on the M2 ILTS programme. As part of this intense training programme, students are expected to become familiar with recent advances in MT technologies and to develop a constructive-critical approach to the changing translator's workplace and its technologies. The partnership project between University of Paris and Systran allows our students to benefit from the latest MT technologies developed by Systran, such as neural MT specialisation with custom resources.

As mentioned above, prior to receiving human review or translations, SYSTRANKLinks can provide automated translations for a given website. If the default setting is set to 'Machine Translation', it becomes possible to set up an MT provider for a specific language pair (such as French into English). In our case, the translation platform is interconnected with a neural MT engine offered by PNS. This technical solution gives us an opportunity to adapt each website translation project to a new context of communication (or a new domain) using custom resources (standard dictionaries, user dictionaries, translation memories, language models). This adaptation is one of the core features of Systran technologies and services.[7] It provides an opportunity for our students to practise linguistic resource management, terminology management and corpus tools in computer-aided translation.

5 The Impact of MT Technologies on The Translation Process

Along with competencies that enhance project management and technological skills, such as customised MT, we focus on the impact that the latest translation technologies have on translation

[6] See the official product website: http://www.systransoft.com/systran/translation-technology/pure-neural-machine-translation. An open source neural machine translation system OpenNMT developed by the Harvard NLP group and Systran is also available online: http://opennmt.net

[7] For more information: http://www.systransoft.com/download/white-papers/systran-white-paper-PNMT-12-2016_2.pdf

practice, translation quality and revision workflow. In this approach, using MT technologies and custom resources is essential to demonstrate how 'a paper & pencil paradigm' (as though the translation is done on a blank sheet of paper) is replaced by quality revision of post-edited machine translation. The following example shows how a single segment has been edited and revised by one of our students (edits are in bold; the student's comments are in italics):

- SOURCE TEXT FR (see Figure 1): Malgré la magnitude du séisme meurtrier du 12 janvier 2010, le contexte sismotectonique et les failles actives dans la zone épicentrale étaient encore mal connus. *Followed by details:* 'Le sud de l'île d'Haïti est traversé d'est en ouest par…'[8]
- Machine Translation EN1: 'Despite the magnitude of the deadly earthquake of 12 January 2010, the seismotectonic context and active faults in the epicentral zone were still poorly understood.'
- Post-edited EN2: 'Despite the magnitude of the deadly earthquake of 12 January 2010, **its** seismotectonic **environment** and **the** active faults in the epicentral zone **were poorly understood until now**.' – *better but key info still not to the fore.*
- Revision EN2: *Better to turn sentence round and start with* **'Until recently…'** *so we insist on the fact there's a new development.*
- Revision EN3: *BUT we can make a more radical change. Rather than just switching sentence on its head we can switch whole text on its head. We wait until paragraph 3 to get the key information, which is that two authors have published a new study (in December 2015) which sheds light on this earthquake and exactly what happened. So, the ideal solution would be:* **'A new study has been published that helps to explain the deadly earthquake of 12 January 2010. Until recently the seismotectonics and active faults in the epicentral zone of this upheaval were poorly understood.'** *Structure but not just of the first sentence (as we first thought) but of the first three paragraphs. So, this is a question of context and structure (syntax). Readers are passing through. Put key information first. I compared the IPGP page with the 'Latest News' section of the CNRS site. Key info always comes first, often using present perfect, as in radio news bulletins – insists on the immediacy of the news item…*

This example shows that our trainee translators are aware of how information is not in free distribution across the webpage: as with traditional genres, the titles and introductory sections of webpages are not framed in the same way as other parts of the text, indeed they can have their own 'grammar'. The translator/reviser's comments also show how students ultimately become key decision-makers in the translation output, as they learn to understand the key cultural, institutional and lexico-grammatical features of website discourse.

As mentioned by the Directorate-General for Translation (2013), website language involves: high visibility, varieties of discourse and text types, short text paragraphs, clear wording, presence of keywords for search engines, hyperlinks, labels and navigation menus (self-service), always updated (never-ending process). Our students are expected to rely on computer-aided translation tools to address these issues.

[8] See: http://www.ipgp.fr/fr/contexte-sismotectonique-sud-dhaiti-un-nouveau-modele-seisme-de-magnitude-mw70-12-janvier-2010

One particular problem encountered by our students involves the fact that SYSTRANLinks both breaks up the text to translate into segments and makes an exact copy of the text that can be navigated using the 'Browse and Edit' mode. This is crucial to understanding why the overall structure of translated texts becomes extremely rigid. A generated bi-text is a table of translation correspondences that can be further edited or validated in the translation/revision workflow. As shown above, MT-enhanced translation can be followed by several revision cycles bringing major syntactic and structural changes to the translated text.

Figure 1: Example of a source text webpage from the IPGP translation project

In MT-enhanced computer-aided translation, TM management software divides texts into segments and uses the metadata each of these segments to trace back to a specific edit, change in status ('edited', 'validated', 'rejected'), date and time (Moorkens, 2013). This allows a translator/reviser to create translation tasks using specific sets of text segments, create and manage view filters, leverage segments that may contain specific terminology or phraseology, etc. The segment table is also an excellent tool to manage software interoperability using formats such as TMX (Translation Memory eXchange). However, this structural rigidity and lack of support for dynamic XML/HTML content may result in inappropriate translated text output being suggested in the TM segment table.

To work around this issue, a textual revision can either be implemented 'on the fly' (as in the suggested example of student revision) or by using one of the most interesting features of SYSTRANLinks: 'Rules'. A 'Rule' can affect all webpages or a given target language page. However, to enter a rule, the system user needs to master the format that the rules system understands[9]. It is then possible to edit the content of the 'Rules' pane, click 'Update' to tell the system to analyse and process it. However, this process is complex and time-consuming and it can be implemented only for some webpage elements which can be "found" or "selected" using CSS selectors[10]. Besides, it is hard to trace these changes using revision history as the generated text element displayed on the translated webpage is no longer registered in the bi-text segment table.

[9] See: http://support.systran.net/systranlinks/tutorials/manage-content/working-with-rules-in-systranlinks/
[10] https://www.w3schools.com/cssref/css_selectors.asp

We have noticed that these more or less ergonomic aspects of computer interfaces for translated text revision challenge students' perceptions of MT-enhanced CAT technology. Students progressively learn to navigate from the segment table and to the context view ('Browse & Edit'). However, multiple translation windows and over-segmentation result in a more assisted but also disrupted translation/revision practice, blurring comprehension processes, contextual interpretation of meaning, text coherence and communicative purpose.

6 The role of metalanguage in students' reporting forms

Here we explore the basic features of the Reporting Form by looking at data provided by students working on different website translation projects during the academic years 2017-18, 2018-19.

As can be seen in Appendix A, the first half of the form asks students to provide factual information about themselves (fields 1-2) and their assigned tasks (fields 3-7). This includes details on the genre/register of the source text, the style guide used, the number of pages assigned, the volume of words in the source text, the number of revision cycles and the number of revisions made. The second half of the form requires more subjective information, including examples of linguistic issues encountered (field 10), references to outside sources (field 11), the students' assessment of the project workflow (field 12), their feedback on SYSTRANLinks features (field 13), and on the role of tools used for corpus creation and analysis in computer-aided translation (field 14).

Regarding 'genre/register', the aim here is to ensure that all students acknowledge that the text they have been working on represents a specific variety rather than 'the general language' (a debatable concept, as discussed in Gledhill and Kübler, 2015). Of the students who do provide clear responses, we obtain a range of metalanguage labels, including register terms relating to a general type of discourse ('public understanding of science', 'academic' language, 'administrative' register) or judgements about level of formality ('formal', 'mostly formal', etc). Despite problems in classification for some students, this range of comments suggests that most students have a sufficiently clear approach to this type of classification.

Over all projects, students reported an average task size of approximately 177 segments (each segment corresponding either to a title, a website link, or a whole sentence) with an average task volume of 1500 words (this may involve one or several pages).[11]

Regarding 'revision cycles', the students reported an average of 3.2 cycles across multiple projects. A complete revision cycle should include: a) the generation of a machine translation and post-editing of all the task segments by the Translator, b) the revision of the same segments by the Reviser, and c) the validation of the same segments by the Editor (i.e. the teacher). Although the notion of 'revision cycle' seems clear to us (the teachers), it was not always that clear for a minority of students.

Regarding 'revision edits', the students reported an average of 70 edits per task. These figures are approximate. While many edits involve typos and single-character corrections which in theory count as just one edit, others involve large-scale changes in syntax or even the re-ordering of information above the level of the sentence (as seen above). In an attempt to simplify matters, in 2018-19 we suggested that students should just count the number of segments in which an edit had been made.

Field 10 of the Reporting Form requires students to cite four 'Key Examples of Translation Issues'. Students are also asked to associate each example with a linguistic 'stratum' (using M.A.K.

[11] These figures are provided by SYSTRANLinks. To obtain more reliable and comparable data about word counts, lexical density, etc., we intend to ask students to analyse the main body of their assigned text using external textometric tools.

Halliday's terminology). The highest category 'Context of situation' includes issues relating to editorial policy, such as 'Do not Translate' i.e. re-introduction of an original French segment, as the following example:

- FR: **Les Crédits d'Impôt Recherche** (CIR)
- EN1: **The Tax credits Searches** (CIR)
- EN2: **Les Crédits d'Impôt Recherche** (CIR) (Research Tax Credits)

Although our students are often able to find examples involving 'Context of situation', they are less certain about 'Discourse-semantics' and 'Lexico-grammatical structure'. Thus, while most students are comfortable with identifying simple lexical problems, as soon as an edit extends beyond the syntactic rank of 'group' they often have difficulties. In many ways this is understandable, because many examples involve both categories at the same time:

- FR: Le 8 juillet 1976 à 8h55, les premières **manifestations de surface** apparaissent.
- EN1: On July 8th, 1976 at 8:55, the first **demonstrations of surface** appear.
- EN2: On 8 July 1976 at 8.55 a.m, the first **demonstrations of surface** appeared.
- EN3: On 8 July 1976 at 8.55 a.m, the first **surface manifestations** appeared.

This example also allows us to mention the lowest stratum ('Phono-graphic substance'), which involves among other issues the harmonisation of date and time formats in order to conform to the DGT English Style Guide[12]. Almost all students were able to identify problems at this level. We note that many of these errors are essentially trivial, but also hard to predict (unless a custom training set is used for MT specialisation)[13].

We now examine the link between linguistic analysis and the second half of the reporting form dealing with "Project workflow issues", "SYSTRANLinks" issues and "corpus creation and analysis tools". At this point, it is interesting to examine our students' metalanguage not as individual lexical choices, but as elements of more extended patterns of expression. A first step in this approach is to treat the students' reporting forms as a corpus. Using textometric analysis tools, such as *iTrameur*[14] (Fleury and Zimina, 2014), it is possible to identify statistically salient examples of commonly shared lexical items: non characteristic elements of 0.0 *specificity* (Lebart *et al*., 1998). This gives the following items in decreasing order of frequency: **'main'** ('issue', 'problem', 'tool', etc.), **'search'**, **'information'**, **'meaning'**, **'collocations'**. For each of these items, it is then possible to analyse their habitual contexts using collocations and concordancing:

[12] Available online: https://ec.europa.eu/info/sites/info/files/styleguide_english_dgt_en.pdf

[13] Neural MT engine training and tuning are part of our future pedagogical challenges to prepare students for the evolving market. For example: https://translate.systran.net/translationTools/text

[14] The size of the corpus (composed of 40 entirely completed reporting forms) was measured online by *iTrameur* (http://www.tal.univ-paris3.fr/trameur/iTrameur): 38593 occurrences / 5824 words.

N°	Contexte
3	is to use the quote marks functionality in the google search engine, which allows users to search verbatim and to
4	in the google search engine, which allows users to search verbatim and to thus find other pages or articles where
5	different texts). I also used to corpus to search collocations. For instance, I typed "fault" to see
7	." 3. The system list resources were used to search for occurences of terms in other documents/places on the
8	uses a lot of specialized terminology. The ability to search the entire website through SystranLinks can be particularly helpful for
12	) I used the translation memory (in the concordance search) to translate but had to double check some names
13	played a crucial role. It was very useful to search the texts already translated on the website to check the
15	translate "Obsera". 3. I use the search bar in the Sentence window to look for the translations

Figure 2: Collocations of 'search' taken from student reporting forms

For example, the word **'search'** (as can be seen in Figure 2) is used either as a part of a technical noun ('**search** + bar, engine', 'concordance + **search**') or as a verb introduced by various words relating to 'use' ('it was very useful to', 'allows users to', 'I + use/used the <N> + to + **search**...', 'the <N> was/were + used + to + **search**...').

We suggest that such patterns of expression demonstrate that our students have internalised a very regular pattern of analysis which can be summarised by the formula: *<using (a tool) to search (examples of language)>*. Interestingly, this pattern of analysis is not only encouraged on the TSA course, but is explicitly taught throughout the M2 ILTS programme (in courses on corpus linguistics, phraseology, terminology, etc.).

The goal-oriented patterns of wording that can be observed around a word such as 'search' can be contrasted with the semi-technical use of a word such as **'meaning'** (see Figure 3), which is often used in contexts where the students are evaluating how the MT was generated (examples 2, 3, 5, 6, 8, 9), or explaining their own translation/revision choices (examples 1, 4, 7, 11). In many cases, the term is used in close proximity to an explicit evaluation of the translation of a segment as a whole ('devoid of all **meaning**', 'true **meaning** of the sentence', 'ambiguous **meaning**', 'good translation', 'perfect segmentation' 'unnatural word order', etc.). Thus, as an item of metalanguage, **'meaning'** often turns up as a marker of explicit personal evaluation of translation/revision quality:

N°	Partie	Contexte Gauche	Pôle	Contexte Droit
6	24	and those wich seemed to be a good translation had another	meaning	), I looked at the original article. It didn't have any
10	36	by a native English speaker. Therefore, I focused on the true	meaning	of the sentence which is that this meeting is the first
9	35	Issue #1 FR Quelques liens pour mieux comprendre la visioconférence : Ambiguous	meaning	: EN 1 - MT Some links to better understand videoconferencing: EN 2
4	13	it was a group or such, I managed to explicit the	meaning	(in brackets, or adding a few words...). 2. The term "rupture
8	34	long paragraph of scientific detail. This leads to a problem of	meaning	and begs the question 'encore mal connus' until what? What has
7	27	no equivalent structure in English. My solution was to translate the	meaning	and not the structure: "for medical purposes". 4. Typography "l'Institut
1	4	seismic tremor. I looked for the French term to understand its	meaning	, and found that the EN equivalent for trémor sismique is actually
2	9	the machine translation produced results that were effectively devoid of all	meaning	, often due to individual words that were translated without taking the
3	9	the west of it. 3) In some cases, all units of	meaning	were conserved, but with an unnatural word order that had been
11	36	translate by zombie nouns. To avoid this, I focused on the	meaning	of the sentence (see also « univers » → « processes »). 3. l'univers
5	22	would spend twice more time trying to get the perfect segmentation (	meaning	being able to find each of/ our segments in our task

Figure 3: Concordance of 'meaning' taken from student reporting forms

Thus far we have seen examples of lexical patterns of expression. The results of *Correspondence Analysis* (Lebart *et al.*, 1998) of a partitioned corpus of student records (reporting forms) suggest that grammatical items are also involved in regular phraseology, and they often characterise distinctive 'respondents' profiles'. To demonstrate this we look at two salient items **'I'** and **'This'** that are *characteristic elements* of typical student profiles. The pronoun **'I'** is involved in two patterns of expression in which students express vagueness or uncertainty (**I** + modal adverb of uncertainty, or negative polarity):

- **I mostly** agreed with the changes made by...
- **I** am **not** familiar with the tools

In the alternative pattern (the majority of cases) students express much clearer learning outcomes as '**I could** + (cognitive process)'; in which they either positively appraise translation workflow, or express an ability to appraise linguistic terminology:

- **I realized** I **could** get a more accurate translation.
- ...once I **had the topic in mind**, I **could** focus more on style during the...
- **I could recognize** the collocate "plane" which...

As Bordet (2018) has shown, the item 'this' is a key feature of formal academic writing, in which the author moves away from personal standpoint (as with 'I') towards a more sophisticated type of impersonal discourse. Our analyses suggest that students with advanced analytical skills use **'This'** in exactly this way. In one pattern **'This'** is a full pronoun introducing an explicit evaluation about the workflow or translation decision process:

- **This helped** with my translation choice.
- **This was useful** for verifying terms…

- **This will help** other translators in the future.

In the second pattern, **'This'** is a determiner relating a tool (or a process) to a positive learning outcome:

- **This tool gives us** a broad view of...
- **This corpus allowed me** to prove that the...
- **This feature allowed us** to search the whole...

Such examples show that in a representative corpus of texts it usually possible to find regular patterns of expression that are symptomatic of a particular type of discourse or 'register'. We suggest that many aspects of this discourse can be characterised as 'trainee academic writing', and that our students have assimilated characteristic elements of phraseology from various sources on the M2 ILTS course, including, of course, ourselves (their teachers). We would not wish to assess students' learning outcomes purely on the basis of their rhetorical skills. However, we take these results to suggest that those students who are able to express analytical skills have internalised at least some of the methodologies which the TSA course aims to transmit.

7 Conclusion

In this paper we have argued that the TSA course challenges its students by establishing 'disrupted practice' in each of these four areas: a) **Transversality** – an attempt to cover a wide variety of training outcomes as defined by EMT, b) **Web-tectonics** – an attempt to build a translation training model within a permanently shifting multimodal context, c) **Non-linear Task Management** – an attempt to manage a complex pedagogical workflow using a professional translation platform, and d) **Process-oriented Assessment** – an attempt to disrupt students' expectations by shifting their focus from a product-based 'paper and pencil' model to a qualitative self-assessment of the translation/revision workflow.

In addition, we have argued here that a crucial way in which we can identify learning outcomes is to explore the metalanguage of our students in their reporting forms. One surprising observation is that when students encounter difficulties with technology, this comes across in the very way they express themselves (vague expression, first-person narrative). On the basis of this, we may in the future be able to help these students to develop more constructive learning paradigms, perhaps conceptualising MT as a 'useful tool to explore language' (to paraphrase one of the recurrent patterns to be found in the corpus).

Finally, it is important to discuss the key notions of 'editing' and 'revision' within a 'qualitative translation/revision workflow'. We have seen that the revision process is facilitated by SYSTRANLinks as it generates a permanent editorial 'history' of changes made to each segment. However, this mechanism is not designed for teaching purposes, and does not produce any data which can be stored outside the system. In addition, although our students also follow courses on post-editing and revision elsewhere on the M2 ILTS programme, they generally see edits primarily as 'errors' and unless prompted they intuitively delete any 'initial' (non-revised) versions or MT that they might have generated. We suggest that this is because some students still see revision as a weakness rather than a strength, and in their reporting forms they often revert to the first person to report changes ('I mostly agreed with the changes made...'). This goes against the general philosophy of the course: TSA is a collaborative translation project, and revision should thus be the responsibility of the whole team.

Acknowledgments

We would like to thank previous generations of students on the TSA (M2 ILTS) course. We would also like to thank our collaborators and the website owners (the 'clients') who have generously provided feedback to our students.

References

Aziz, Wilker, Sheila Castillo Maria de Sousa, and Lucia Specia. 2012. PET: a Tool for Post-editing and Assessing Machine Translation. In *Proceedings of the 16th Annual Conference of the European Association for Machine Translation*, pages 3982-3987.

Blagodarna, Olena. 2018. Insights into post-editors' profiles and post-editing practices. *Revista Tradumàtica. Tecnologies de la Traducció*, 16, pages 35–51.

Bordet, Geneviève. 2018. "This dissonance": Bolstering credibility in academic abstracts. in Pilar Mur-Duenas and Jolanta Sinkuniene, editors. *Intercultural perspectives on research writing.* Amsterdam: John Benjamins, pages 83-104.

Directorate-General for Translation (European Commission), 2013. *Web translation as a genre*. Luxembourg: Publication Office of the European Union.

EMT. 2017. *European master's in translation competence framework 2017*. European Master's in Translation, European Commission. https://ec.europa.eu/info/sites/info/files/emt_competence_fwk_2017_en_web.pdf [last accessed October 20, 2019].

Fleury, Serge, and Maria Zimina. 2014. Trameur: A Framework for Annotated Text Corpora Exploration, In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: System Demonstrations*, pages 57–61.

Froeliger, Nicolas, Christian Balliu, and Lance Hewson, editors. 2018. *Equivalences* (Bruxelles), 45/1–2.

Gledhill, Christopher. 2011. A Lexicogrammar approach to checking quality: looking at one or two cases of comparative translation. In Ilse Depraetere, editor, *Perspectives on Translation Quality (Text, Translation, Computational Processing 9)*, Mouton de Gruyter, pages 71–98.

Gledhill, Christopher, and Natalie Kübler. 2015. How Trainee Translators Analyse Lexico-Grammatical Patterns. *Journal of Social Sciences – Special issue on Phraseology, Phraseodidactics and Construction Grammar(s)* (ed. Maria Isabel González-Rey.), 11/3:162–178.

Guerberof Arenas, Ana, and Joss Moorkens. 2019. Machine translation and post-editing training as part of a master's programme. *Jostrans: The Journal of Specialised Translation*, 31, pages 217–238.

Harris, Brian. 1988. Bi-Text, a new concept in translation theory. *Language Monthly*, 54, pages 8–10.

Lavault-Olléon, Elisabeth, and Maria Zimina, editors. forthcoming. *Traduction et technologie : regards croisés sur de nouvelles pratiques*. Des mots aux actes, 8. 2019. Classiques Garnier.

Lebart, Ludvic, André Salem and Lisette Barry. 1998. Exploring Textual Data. Kluwer Academic Publishers.

Moorkens, Joss. 2013. The Role of Metadata in Translation Memories. In Valerie Pellatt, editor, *Text, Extratext, Metatext and Paratext in Translation*, Cambridge Scholars Publishing Encyclopedia of Language & Linguistics. Elsevier, Oxford, pages 79–90.

O'Brien, Sharon. 2002. Teaching post-editing: a proposal for course content. In *Proceedings of 6th EAMT Workshop Teaching Machine Translation*, pages 99–106.

O'Brien, Sharon. 2012. Translation as human computer interaction. *Translation Spaces*, 1/1, Issue 1, 101–122.

Pauwels, Luc. 2012. A Multimodal Framework for Analyzing Websites as Cultural Expressions. *Journal of Computer-Mediated Communication*, 17, pages 247–265.

Piecychna, Beata. 2016. Integrated Problem and Decision Reporting in Translation Teaching – Advantages and Drawbacks from Translation Students' Point of View. *Białostockie Archiwum Językowe*, 16, pages 247–263.

Appendix A:

M2 ILTS

Teaching team: Chris Gledhill, Maria Zimina

Fiche d'évaluation

Traduction de site vers l'anglais (TSA)

PROJECT REPORTING FORM

Fill in all the cells in the right-hand column of this form in English

NAME, Given names	
Speciality (IL, TS)	
Task (Genre / Register)	
Title	
Source	
Style Guide	
Volume	
Revision cycles	
Revision edits	
Key examples of translation issues 1. Context of situation 2. Discourse-semantics 3. Lexico-grammatical structure 4. Phonic-graphic substance	
Remarks on bibliography, sources, reference material	
Project workflow issues 1. Editing workflow & workflow engine 2. Interacting with Revisers and Website Editor	
SystranLinks issues 1. Translation Memory (TM) 2. User Dictionary 3. Sentence List Resources 4. Rules (if applicable)	
Corpus creation & analysis tools Briefly explain the role of tools in website translation/revision. Give examples of specific problems that can be solved using CAT tools and corpora.	