Translating and the Computer 42

18-20 November 2020

On the Web

ISBN 978-2970-10958-7

October 2021. Editions Tradulex, Geneva

Acknowledgements

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

Gold Sponsors

Silver Sponsors

Preface

For over 40 years the TC conferences, organised by the International Association for the Advancement in Language Technology (AsLing), has served as a unique forum for academics, developers, users, and vendors of computer aids for translators, of other translation technology tools, and increasingly, for interpreters and others performing new roles in our industry.

In 2020, the 42nd edition of the Conference had to innovate, to meet the special situation caused by the Covid-19 pandemic, which prevented people from traveling to London, as for previous sessions, to meet in person and exchange face-to-face their experience, concerns, challenges and achievements. After weighing the pros and cons of a virtual conference during the early months of the pandemic the AsLing Executive Committee requested the Conference Coordinators to organize a fully online event. The result was a success and allowed a new, wider audience from all around the world to participate and join the circle of traditional in-person delegates for the 23 presentations and 14 workshops, covering a broad range of subjects and tools. The chill-out sessions also brought together freelance and in-house translators, interpreters, researchers and businesspeople from translation companies, international organisations, universities and research centres, and offered them opportunities to exchange ideas, and to learn about and discuss the latest developments in translation technologies.

TC42 featured speakers and workshop moderators from academia, industry and the professions that provided insights on the latest developments in the language industry. Recordings of most sessions, as well as many of the PowerPoint presentations presented during the Conference are available on the asling.org website. These Proceedings contain more elaborated texts of the main presentations during TC42 on the Web. The Conference chairs are delighted to thank this edition's prominent keynote and invited speakers: Martin Benjamin, the founder and director of Kamusi, an NGO dedicated to gathering linguistic data and setting that data to work within language technologies, with a major goal to include languages that are otherwise neglected in research and trade. He explained to the audience "How AI Cured Coronavirus and Delivered Universal Translation, and Other MT Myths and Magic". Joanna Drugan, Professor of Translation at East Anglia University and leader of a Transnational Organised Crime and Translation research project, captivated the attendees with her talk "The Dark Side of Translation: Crime, technologies, and translation today". She warned that transnational organised crime is often painted as the 'dark side' of globalisation. Like industry and commerce, it benefits from factors such as greater freedom of movement, global trade, and communication technologies. Cross-language communication is key to crimes such as human trafficking, online fraud and intellectual property theft, with the majority of organised crime groups working across more than two languages. However, she described how translation and language technologies are key to understanding, preventing and prosecuting transnational organised crime.

We thank all who submitted proposals to the conference and those authors who produced full versions of their papers for these Proceedings, as well as all whose slides and recordings are available on the AsLing website. A special thank-you goes to all the delegates wherever they were, who by taking part and interacting remotely with presenters and fellow participants, enriched this conference giving living acknowledgement to this special event. We are grateful to the members of the Programme Committee who carefully reviewed the submissions as well as all additional reviewers who helped assess some of the final papers and to our fellow members of the Organising Committee, who played key roles in ensuring that this year's special conference, AsLing's first ever online, took place and linked people from all continents. 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

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 TC42.

Conference and Session Chairs:
David Chambers, AsLing Honorary Member (Session Chair)
João Esteves-Ferreira, AsLing President (Conference and Session Chair)
Juliet Margaret Macan, AsLing Vice-President (Conference Chair)
Ruslan Mitkov, AsLing Vice-President (Conference and Session Chair)
Maria Recort Ruiz, TC42 Coordinator (Session Chair)
Olaf-Michael Stefanov, AsLing Vice-President (Conference and Session Chair)

Conference Organising Committee:
Sandra Chambers, Organisation for Economic Co-operation and Development
João Esteves-Ferreira, Tradulex, Coordinator
Juliet Margaret Macan, independent translation technology consultant
Ruslan Mitkov, University of Wolverhampton
Maria Recort Ruiz, International Labour Organization, Coordinator
Olaf-Michael Stefanov, JIAMCATT, United Nations (ret.)
Jean-Marie Vande Walle, AsLing Treasurer

Editors of the Proceedings:
David Chambers
João Esteves-Ferreira
Juliet Margaret Macan
Ruslan Mitkov
Olaf-Michael Stefanov

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
Eleanor Cornelius, University of Johannesburg
Gloria Corpas Pastor, University of Málaga
David Filip, CNGL / ADAPT
Camelia Ignat, Joint Research Centre of the European Commission
Raisa McNab, UK Association of Translation Companies
Joss Moorkens, Dublin City University
Vilelmini Sosoni, Ionian University
Antonio Toral, Rijksuniversiteit Groningen
Paola Valli, Project Manager, Tamedia
Nelson Verástegui, International Telecommunications Union (ret.)
David Verhofstadt, European Investment Bank
Michal Ziemski, World Intellectual Property Organization

Additional reviewers:
Rocío Caro Quintana, University of Wolverhampton
Souhila Djabri, University of Alicante
Marie Escribe, University of Wolverhampton
Viveta Gene, Intratranslations
Maria Kunilovskaya, University of Wolverhampton
Elena Murgolo, Aglatech14
Marine Ovesyan, University of Malaga
Tharindu D. Ranasinghe Hettiarachchige, University of Wolverhampton
Nikola Spasovski, University of Wolverhampton

Table of Contents

The Role and Perspective of the Post-Editor: What are the Training Challenges?

Viveta Gene

Intertranslations S.A., Athens, 17676, Greece

v.gene@intertranslations.com

ABSTRACT

Post-Editing is an increasingly in-demand language service that emerged along with Machine Translation (MT) technology to meet the rising demand for translation services, as a cost-effective and time-efficient solution. Post-Editing consists in applying necessary corrections to MT output to ensure linguistic quality. Although it has been largely adopted in the translation industry and is now considered as a service in its own right, there is still a fair amount of skepticism towards it, due, as this paper aims to show, to a lack of specialized, universally adopted guidelines and training. The present analysis explores the specific skills and competences required to perform Post-Editing tasks and links said skills to the concept of effort, as developed by Krings (2001), to define this nascent role. It goes on to present current attitudes and perspectives related to Post-Editing, in an attempt to reveal the source of skepticism within the community of language professionals. Furthermore, the present paper demonstrates the challenges involved in setting up a comprehensive Post-Editing training protocol and attempts to define its content, supported by empirical research results, as reflected in a digital survey among language industry professionals (Gene, 2020). The survey was conducted in the context of a Post-Editing Webinar involving representatives of LSPs (Language Service Providers). The results presented reflect the answers of 51 MLV (Multilanguage Vendor) and 12 SLV (Single Language Vendor) representatives.

Keywords: Post-Editing training protocol; Post-Editing competences; Post-Editing perspectives.

INTRODUCTION

This paper aims to provide a detailed analysis of the skills and competences required to perform Post-Editing tasks, in an attempt to set the tone for an effective and specialized Post-Editing Training Protocol. Post-Editing of Machine Translation (MT) is an increasingly in-demand linguistic service, already widely adopted within the language industry, as a cost-effective and time-saving solution with the potential to increase productivity. Initially conceived as a mere "human partner" to the MT engine (Vieira et al 2019), in recent years, the Post-Editor has emerged as a new role in its own right, characterized nonetheless by certain ambiguities regarding the scope and nature of his/her work. Previous analyses in the field have provided important insights into this new role, with the seminal works of Allen (2003) and Krings (2001) contributing to a theorization of the Post-Editing process, and that of Torrejon & Rico (2002) focusing on skills and training guidelines.

Drawing on these important contributions, this paper attempts to link Post-Editing-specific skills, following the typology developed by Torrejon & Rico (2012), to the concept of effort required to carry out mental processes during Post-Editing, as described by Krings (2001). To this end, we shall correlate linguistic skills, instrumental and core competences (Torrejon & Rico, 2012) to four types of effort, namely: temporal, technical and cognitive (Krings, 2001), and add the concept of psychological effort to Krings' traditional categorization. Recent research (O' Brien & Moorkens, 2014) as well as empirical

evidence suggest that Post-Editing is widely viewed as a disruptive force operating within the translation profession, with many linguists expressing concerns and skepticism towards Post-Editing as a service and some of them mounting stiff resistance to what they perceive as a threat to their future employability. We argue that, in this case, much of this negative criticism actually stems from a lack of comprehensive and widely available training that would provide specific guidelines and dispel ambiguities concerning speed and productivity standards, as well as compensation models. We further suggest that, on the – largely ignored – positive side, Post-Editing, combined with other activities, could enable linguists to expand the scope of their work and increase their overall job satisfaction. Post-Editing should not be deemed a replacement for traditional translation roles, but rather a shift in skills and common practices enabled by language technology.

A. What is Post-Editing?

The term "Post-Editing" first appeared in literature in the field of translation and language technology studies several years before it became a common practice adopted by industry professionals. A concise definition of the Post-Editing process is given in the seminal work of Allen (2003), which is held as one of the most significant contributions to the field and a decisive step towards the standardisation of a Post-Editing framework. He defines Post-Editing as "the correction of texts that have been translated from a source language into a target language by a machine translation system" (Allen, 2003).

In the early days of MT development, Post-Editing was merely considered a supportive task of secondary importance, the ultimate goal being the development of unaided Machine Translation (Nunes Vieira et al, 2019), ideally requiring no human involvement. The Post-Editor was viewed more as a "human partner" to the machine (Nunes Vieira et al, 2019), occupying an incidental yet necessary role in the process of automatically transferring meaning from one language to another. Although this rather dystopian vision has not materialized and human Post-Editors are established as legitimate agents in MT workflows, this notion has left lasting impressions on the way some language professionals regard Post-Editing, becoming one of the sources of negative attitudes towards the service (see chapter 3).

Within today's technological landscape, with language technology advancements significantly enhancing MT performance, Post-Editing continues to reflect Allen's (2003) definition, concerning the review and correction of MT output according to a pre-established quality level. A more "technical" definition of Post-Editing would be "the minimal edits required to transform a hypothesis into a fluent sentence with the same meaning as the source sentence". This description takes into account a significant factor at play in the Post-Editing process, namely the concept of effort, which will be discussed in detail in the following chapter: the "amount of edits" corresponds to technical effort while the aspect of "[transforming] into a fluent sentence" evokes the cognitive effort required to perform Post-Editing tasks.

This last element, i.e., transforming meaning into a fluent sentence in another language, bears close resemblance to traditional translation and/or review tasks, although we need to acknowledge that, at least in terms of effort, Post-Editing is a fundamentally different process. It consists in a combination of organizing the raw MT output, correcting errors, revising as well as retranslating entire sections when needed, which, most importantly, involves knowledge of the specific MT system used to produce said output, that is, the type and frequency of errors that the system produces. Following this definition of the Post-Editing process, the next section will delve into the role of the Post-Editor and the skills required to properly fulfill it.

B. The role of Post-Editors

To grasp the challenges involved in the training process and understand where they stem from, it is important that we first examine the different aspects of the Post-Editor's role, in terms of skills and competences as well as effort required to perform Post-Editing-specific tasks.

Typical tasks include both lexico-grammatical and syntactic corrections, as well as stylistic, linguistic and formatting adjustments, namely: correcting grammatical agreements (number and gender values etc.), punctuation and capitalization, sentence-level structural changes and changes in formatting and, in cases of more or less poor MT output, re-translation of certain words or even entire phrases and expressions which may include removing and/or adding words.

It is worth noting that, although the Post-Editor's role comprises all these different operations, the guidelines as to when and how they should be executed appropriately are still rather blurry. We shall return to this question towards the end of this section, discussing a common framework of best practices for Post-Editors.

The work of Krings (2001) is of particular importance here, as it has provided significant contributions to understanding the cognitive and practical requirements of Post-Editing. Krings (2001) distinguishes three types of effort at play in the Post-Editing process, describing the "amount of processing capacity" utilized in performing these tasks (Tyler et al., 1979): temporal, technical and cognitive effort. Technical effort refers to the Post-Editor's physical interaction with both hardware and software to manipulate the MT output, while temporal effort refers to the time needed to post-edit a segment (Post-Editing time). Cognitive effort is the actual effort required to identify errors in the MT output and apply any necessary corrections during Post-Editing. Unlike temporal and technical effort which are easily quantifiable, these corrections and/or amendments as described above (correcting word forms, word order and lexical choices, substituting and/or inserting missing terms or deleting unnecessary additions) are calculated based on the notion of *edit distance* (Zaretskaya et al., 2016). This metric is used to *"*assess the difference between two sentences which reflects the minimal number of deletions, insertions and substitutions needed*"* (Zaretskaya et al., 2016) to transform a sentence into its final version.

To this list we shall add the concept of psychological effort, which is linked to specific attitudes towards Post-Editing as a service in general. The negative connotations often attached to Post-Editing, concerning the gradual obsolescence of traditional translation tasks and their substitution by technologically driven processes, may affect the linguists' stance, who may view their new role as now limited to merely reviewing trivial errors repeatedly appearing in machine-generated texts. Indeed, many linguists have voiced their discontent with Post-Editing tasks, denouncing the repetitive and tedious nature or expressing fears for their future employability. It has also been suggested that these tasks are in fact more mentally demanding than traditional Computer-Assisted human translation.

However, despite these concerns and any obvious overlaps between the roles of Post-Editors and traditional reviewers, the two roles are, and should be considered as, distinct; not only does Post-Editing require a different skill set than traditional reviewing, but utilizing these Post-Editing-specific skills also requires a different type of effort on the part of the linguist.

Rico & Torrejon (2012) have described the necessary Post-Editing skills dividing them into three categories: a) linguistic skills, b) instrumental skills and c) core competences. Using this typology and drawing on the work of Krings (2001) we could attempt to match specific skills with a certain type of effort. Drawing such parallels could prove useful for setting up an effective Post-Editing training protocol that takes into account technical, cognitive as well as temporal demands.

According to Rico & Torrejon's categorisation (2012), linguistic skills include textual, cultural and subject matter competences, similar to those required from a typical translator. By this we refer to an excellent command of the source and target languages, communicative and textual competence in at least two languages and cultures, as well as cultural and intercultural competence, and subject area expertise.

Instrumental competences can be thought of as roughly equivalent to technical competences. The skills that fall under this category include terminology management and MT dictionary maintenance,

corpus quality assessment skills (for example-based and statistical engines), programming skills for creating macros or scripts for automated correction, controlled language pre-editing skills as well as excellent editing skills and word-processing experience, meaning full-key proficiency, efficiency in cursor positioning and effective use of search and replace functions. This last set of skills (Rico & Torrejon, 2012), requiring the ability to work and make corrections directly on screen, can be characterized as technically demanding according to Krings' typology of effort. In addition to this, instrumental competences also involve a fair amount of cognitive effort when it comes to manipulating MT output. Post-Editing entails a specific knowledge of MT systems and their affordances (whether rule-based, example-based, statistical or hybrid systems), as well as of their weak points to be able to expect potential errors in the final output. This point further supports the claim that human translation-review and Post-Editing of MT are two fundamentally different processes.

Core competences have been described as "psycho-physiological" and "strategic": the former concern the Post-Editor's ability to "cope with subjectivity" in the process of applying Post-Editing specifications, while the latter refer to those competences which allow them to "reach informed decisions" and "[choose] among different Post-Editing alternatives [...] showing no stylistic concerns" (Rico & Torrejon, 2012). Linking these competences to the concept of effort, we can clearly see a temporal dimension which is prominent in utilizing them, that is, the ability to balance Post-Editing speed to required quality while "choosing among Post-Editing alternatives".

Finally, psychological effort can be traced, not so much in the field of skills or competences, but to the linguist's attitude towards Post-Editing tasks and MT in general. We shall return to this point in the next chapter of this paper, discussing Post-Editors' perspectives.

The absence of a specific Post-Editing protocol can often lead to errors of varying severity that can significantly affect the overall quality of the final product. The Post-Editor's role lies exactly in utilizing all of the aforementioned skills to effectively manage the linguistic material and, most importantly, avoid errors which are intrinsic to the Post-Editing process. This requires the ability to detect and act upon MT issues based on a thorough understanding of how a machine-translated text came to be and, as described earlier, on an understanding of the engine's specific flaws.

C. Attitudes towards Post-Editing: positive and negative perspectives

Let us now turn our focus to language professionals themselves and discuss the attitudes associated with Post-Editing as a nascent profession at best met with skepticism or, as indicated by important research, downright resistance (O' Brien & Moorkens, 2014; de Almeida, 2013). This attitude is not unlike the resistance and discontent towards computer-assisted translation tools back in the 1990s on the part of linguists; in fact, the introduction of MT in regular workflows and the consequent need for human revision can be thought of in terms of the impact of a disruptive technology (O'Brien & Moorkens, 2014).

The arguments generally associated with this negative perspective are related to a) workload and payments, b) a lack of specific guidelines and c) employability and linguistic concerns.

- *Workload and payments*

Post-Editing tasks are perceived as poorly remunerated while involving the expectation of a higher productivity rate, which may not be sustainable. This situation is further exacerbated by the lack of a solid ground for Post-Editing metrics, which makes payment terms highly ambiguous. The same is true for quality standards, which may vary from project to project. Linguists also report that working with MT output leaves little room for creativity and that MT in general is only appropriate in specific situations.

- *Lack of specific guidelines*

The need to invest time and energy in Post-Editing training is often a source of frustration for both linguists and MLVs. On top of this, in many cases there are no specific guidelines in place, which makes the prospect of undertaking a Post-Editing project rather daunting. The instructions provided to linguists are relatively limited and cannot make up for the lack of comprehensive training or internally adopted guidelines. This situation of uncertainty adds to the generally negative perception of Post-Editing (de Almeida, 2013).

- *Employability and linguistic concerns*

Concerns over Post-Editing's overall impact on the translation profession may be the most important source of negativity and resistance. Many translators fear their role might be replaced by MT engines or see the adoption of MT as an attempt on the part of MLVs to force them into transitioning to Post-Editing. Another issue for them is the depreciation of their linguistic capacities. On the one hand, it has been suggested that being constantly exposed to texts of poor linguistic quality may affect the Post-Editor's linguistic judgement, leading to confusion. On the other hand, Post-Editing has been associated with the revision of trivial and repetitive errors due to MT engine flaws, as described in the previous section – a task that is not favorably viewed by professional translators.

Even though these concerns are valid, they seem to stem from an overall lack of experience with Post-Editing projects and, most importantly, they can be traced back to inconsistent training and inadequate guidelines, which is a legitimate objection on the part of linguists and should constitute the point of departure for the successful integration of Post-Editing into regular workflows. In addition to this critical step, we would like to suggest a change of perspective regarding Post-Editing, viewing it in a more positive light.

It is true that MT combined with Post-Editing can significantly speed up the translation process enabling linguists to deliver higher volumes of work. However, this should not be perceived as a necessary evil but as a productivity tool, offering the possibility to avoid easily automated, repetitive tasks. We should bear in mind that MT engines can be trained over time, based on linguist feedback, to automatically correct or predict error patterns, thus leading to better quality. The use of MT also allows for a standarisation of terminology, which has often been a source of frustration for translators, leading to fewer inconsistencies in the final versions of texts. Spending less time on automatically translated segments offers linguists the possibility to combine Post-Editing with other activities to increase their overall productivity and job satisfaction.

Furthermore, becoming familiar with such tasks and acquiring the new set of skills required to perform them may offer linguists a competitive advantage, meaning more opportunities for new projects. As was found with CAT tools back in the 1990s, it looks like MT is here to stay; the acquisition of knowledge and experience with MT in general is important as technology-driven processes are increasingly integrated into traditional workflows.

Regarding MT's compatibility with only a limited number of linguistic contexts, we could say that, in most situations, using MT is not so much a question of suitability, but rather estimating how much Post-Editing work would be required to reach the desired quality levels. With the appropriate process in place, MT can be used to respond to most translation requirements. That being said, we must acknowledge that there are still cases where MT might be less than ideal, such as in situations where the translation needs to be highly adapted to carry a specific cultural message, like marketing materials.

Finally, the best way to resolve employability concerns and establish a positive attitude towards Post-Editing within the translation community would be through investing in comprehensive, specialized and widely available training. Gaining a better understanding of Post-Editing tasks would definitely help make linguists more positively inclined towards them. In this context, the transition to

Post-Editing should be considered as a shift in roles and a multiplication of skills rather than a "replacement" strategy. The goal of any successful training protocol should be to establish Post-Editing not as a mere aid for translation, but as a new linguistic role in its own right.

D. Training challenges – GALA Webinar survey results

Despite its great significance for shifting common attitudes, setting up a comprehensive Post-Editing training framework is arguably no easy task; it requires that we first identify the main challenges and begin to define the objectives and scope of such training. This section attempts to do so, supported by the results of a survey conducted among language industry professionals regarding the management and training challenges of Post-Editing (Gene, 2020). The results presented are based on the answers given to 63 polls by 51 MLVs and 12 SLVs.

According to the results, it seems that the majority of the participants (60%) did receive some type of Post-Editing training [Image 1]. However, when asked if they considered this training to be comprehensive and concise enough to meet their expectations in terms of speed and quality, 28.57% of those replied in the negative. This reveals that in some cases, even when training is provided, it does not always successfully prepare linguists for Post-Editing tasks and further highlights the need to set up a common training protocol [Image 2].

Any successful training should, of course, cover the necessary skills and competences discussed in the second section of this paper, paying particular attention to the type of effort required to utilize them and making sure that any ambiguities in the Post-Editing process are clarified. In other words, it should meet two of the most important challenges faced by a Post-Editor in general, according to the survey's results, namely: ambiguous guidelines and the lack of Post-Editing skills. It is also clear, as mentioned before in this paper, that training should become widely available and set the tone for a cohesive and transparent compensation strategy for Post-Editors [Image 3].

Regarding the actual content of this training, in terms of temporal and cognitive effort, Post-Editing guidelines should focus on questions of managing errors and applying changes to the MT output. This includes what Krings (2001) refers to as "Target text evaluation processes" and "Machine translation-related processes". The former involve "making positive or negative evaluations of the MT output and [...] comparing it with the source text", while the latter consist in a "reading of the MT output to evaluate whether a reformulation is necessary" (Krings, 2001). These processes can be linked to the Post-Editor's ability to understand "what the machine is doing when it translates", evoking the need to develop instrumental competences, as described in the second section of our paper. In other words, training should not only help linguists become familiar with Post-Editing requirements but also stress how these requirements may differ depending on the type of engine that produced a specific output (statistical, rule-based etc.). In addition to identifying different types of issues, Post-Editors should be able to employ different techniques to tackle them. This error variation and the technical background knowledge required to identify and resolve them is one of the main challenges that any comprehensive training approach should focus on. To this we must add the temporal dimension of such issues, as training should also help linguists develop the ability to quickly identify and eliminate errors according to a pre-established quality level.

As for the technical and psychological effort commonly related to Post-Editing, that is, the effort needed to trigger all the technical skills and competencies to achieve maximum speed (technical) as well as the question of strategy and attitude (psychological), Post-Editing training should educate linguists on what Krings (2001) refers to as "Global task-related processes". This means, providing guidelines on how the Post-Editing procedure and specific subtasks should be organized and emphasizing the importance of providing feedback. Linguist feedback and comments on Post-Editing issues are of vital importance in optimizing both MT engines and the training framework itself (Krings, 2001) and should thus be considered as one of its constitutive elements.

The above analysis of training requirements is further supported by the survey results [Image 4], with language professionals reporting vague instructions regarding the management of MT output and scope of work (light compared to full Post-Editing), a lack of detailed guidelines and, consequently, exaggerated speed and productivity models as the most important challenges encountered in current training schemes.

To sum up, a comprehensive Post-Editing protocol should be able to address the linguists' currently most pressing questions; how should one go about Post-Editing, *how much* of the MT output should the Post-Editor use, which instances call for re-translation and which ones simply require correction, how may client specifications (client description, domain, text description, glossary, MT engine, specific MT used etc.) modify the Post-Editing process? More crucially, a successful Post-Editing protocol should establish the fact that the goal of Post-Editing is to finalize a translation, not to recreate it, and clearly define speed, productivity and compensation standards.

CONCLUSION

As empirical research demonstrates, despite the rising significance of Post-Editing as a legitimate and increasingly in-demand language service, the development of a specialized and widely available training protocol seems to lag behind. We argue that an efficient and balanced training approach entails an understanding of new roles and new tasks involved in the Post-Editing process on the part of all stakeholders, linguists, academics and researchers, as well as industry professionals alike. By presenting the different aspects of a Post-Editor's role, this paper attempts to shed light on the elements that any comprehensive training framework should cover, to help linguists develop the skills and competences required to perform Post-Editing tasks. A training protocol focused on developing linguistic skills, instrumental skills as well as core competences, combined with a thorough understanding of the type of effort that using these skills demands, will not only help to improve the linguistic quality of future Post-Editing projects in general, but also contribute to a change in attitudes among linguists themselves. To this end, it should also establish an unambiguous model for linguist compensation and clearly define speed and productivity standards.

References

Allen, J. (2003). Post-Editing. In H. Somers (Ed.), *Computers and Translation: a Translator's Guide* (pp. 297-317). Amsterdam: John Benjamins.

Aranberri, N. (2017). What Do Professional Translators Do When Post-Editing for the First Time? First Insight Into the Spanish-Basque Lanuage Pair. *HERMES - Journal of Language and Communication in Business*(56), 89-110.

Bentivogli, L., Bisazza, A., Cettolo, M., & Federico, M. (2016). Neural Versus Phrase-Based Machine Translation Quality: a Case Study. *Conference on Empirical Methods in Natural Language Processing*, (pp. 257-267). Austin, Texas.

Blanchon, H., & Besacier, L. (2017, September 21). *Comparing Statistical Machine Translation (SMT) and Neural Machine Translation (NMT) Performance.* Retrieved from http://lig-membres.imag.fr/blanchon/SitesEns/NLSP/resources/SMT-vs-NMT.pdf

Daems, J., Vandepitte, S., Hartsuiker, R. J., & Macken, L. (2017). Identifying the Machine Translation Error Types with the Greatest Impact on Post-Editing Effort. *Frontiers in Psychology, 8, 1282*. doi:https://doi.org/10.3389/fpsyg.2017.01282

de Almeida, G. (2013). Translating the Post-Editor: an investigation of Post-Editing changes and correlations with professional experience across two Romance languages. Dublin: Dublin City University.

Dino, G. (2017, December 18). *3 Reasons Why Neural Machine Translation is a Breakthrough.* Retrieved from Slator: Language Industry Intelligence: https://slator.com/technology/3-reasons-why-neural-machine-translation-is-a-breakthrough/

Gene, V. (2020, May 21). *The Management and Training Challenges of Post-Editing (Part 2).* Retrieved from GALA: Globalization and Localization Association: https://www.gala-global.org/ondemand/management-and-training-challenges-Post-Editing-part-2

Ginovart Cid, C., Colominas, C., & Oliver, A. (2020). Language Industry Views on the Profile of the Post-Editor. *Translation Spaces*. doi:https://doi.org/10.1075/ts.19010.cid

Guerberof Arenas, A., & Moorkens, J. (2019, January). Machine Translation and Post-Editing Training as Part of a Master's Programme. *The Journal of Specialised Translation*(31), 217-238.

Guerrero, L. (2018). From a Discreet Role to a Co-Star: The Post-Editor Profile Becomes Key to the Post-Editing Workflow. *Translating and the Computer - TC40.* London.

Hu, K., & Cadwell, P. (2016). A Comparative Study of Post-Editing Guidelines. *Baltic Journal of Modern Computing, 4*(2), 346-353.

Krings, H. P. (2001). *Repairing texts: Empirical investigations of machine translation Post-Editing processes.* Kent, Ohio: Kent State University Press.

Nunes Vieira, L., Alonso, E., & Bywood, L. (2019). Introduction: Post-Editing in practice – Process, product and networks. *The Journal of Specialised Translation*, 2-13.

O'Brien, S. (2002). Teaching Post-Editing: A Proposal for Course Content. *6th EAMT Workshop "Teaching Machine Translation"*, (pp. 99-106).

O'Brien, S., & Moorkens, J. (2014). Towards Intelligent Post-Editing Interfaces. In W. Baur, B. Eichner, S. Kalina, N. Kessler, F. Mayer, & J. Orsted (Ed.), *20th FIT World Congress*, (pp. 131-137). Berlin.

O'Brien, S., Roturier, J., & de Almeida, G. (2009). *Post-Editing MT Output; Views from the researcher, trainer, publisher and practitioner*. Retrieved from http://www.mt-archive.info/MTS-2009-OBrien-ppt.pdf

O'Brien, S., Winther Balling, L., Carl, M., Simard, M., & Specia, L. (Eds.). (2014). *Post-Editing of Machine Translation: Processes and Applications.* Newcastle: Cambridge Scholars Publishing.

Rico, C., & Torrejon, E. (2012). Skills and Profile of the New Role of the Translator as MT Post-Editor. *Tradumàtica: tecnologies de la traducció*, 166-178.

Saillard, J.-L. (2019, January 28). *Transitioning to a Post-Editing machine translation business model.* Retrieved from Smartcat: https://www.smartcat.com/blog/transitioning-to-a-Post-Editing-machine-translation-business-model/

TAUS. (2014, January 27). *Evaluating Post-Editor Performance Guidelines.* Retrieved from TAUS: the language data network: https://www.taus.net/academy/best-practices/postedit-best-practices/evaluating-Post-Editor-performance-guidelines

Tyler, S., Hertel, P., & McCallum, M. H. (1979). Cognitive Effort and Memory. *Journal of Experimental Psychology: Human Learning and Memory, 5*(6), 607-617.

Vashee, K., & Wiggins, D. (2014). *A 6 Step Plan to Create an Optimal MT Post-Editing Strategy.* Retrieved from https://omniscien.com/wp-content/uploads/2016/02-new/Asia-Online-Webinar-A-6-Step-Plan-to-Create-an-Optimal-MT-Post-Editing-Strategy2.pdf

Zainurrahman, S. (2010). *Five Translation Competencies.* Retrieved from Zainurrahman's Personal Journal: https://zainurrahmans.wordpress.com/2010/06/06/five-translation-competencies/

Zaretskaya, A. (2019). Optimising the Machine Translation Post-Editing Workflow. In I. Temnikova, C. Orasan, G. Corpas Pastor, & R. Mitkov (Ed.), *2nd Workshop on Human-Informed Translation and Interpreting Technology*, (pp. 136-139). Varna.

Zaretskaya, A., Vela, M., Corpas Pastor, G., & Seghiri, M. (2016). Measuring Post-Editing Time and Effort for Different Types of Machine Translation Errors. *New Voices in Translation Studies, 15*, 63-91.

Καλαντζή, Δ. (2016). Αυτόματη Μετάφραση και Post-Editing: Εισαγωγικά Στοιχεία και Σκέψεις. *5th Meeting of Greek Translatologists.*

Jogging, Swimming, Interpreting

At the same place, on the very same day. Impossible?

Patrick Lehner

The PL Project

patrick@theplproject.org

ABSTRACT

What is the relationship between these three activities you will say? Well, you can do all in a row if you are well organised. The first one to warm you, the second one to cool you and the last one to let you do your job relaxed and mentally well prepared.

But those activities have to be practiced at the same place, otherwise, one day will not be enough, especially if you have to take public transport or drive to the place where interpretation is requested.

Ever heard about the RSI – Remote Interpretation Services? No? But you know what the coronavirus is, surely? Well, there is a relationship between the two: in the past, hundreds of years ago, interpreters had to move to the place of interpretation (booth, exhibition, court, …) and would lose a lot of time and sometimes nerves, spend a little money for the night, food and show nice dress apparel on site. Then came a nasty virus which forced everyone to stay home for some weeks and a lot of events were cancelled. All of a sudden, event organisers, but also all companies in need of communication devices discovered that interpretation services can also be offered on-line with rather simple tools through the Internet.

Actually, RSI were created well before the COVID-19 crisis, but they were not very welcomed by parties on both sides: interpreters would refuse to use such a degrading device, arguing about the low quality of the equipment, the risk of connection interruption, the loss of quality of interpretation and criticise every colleague showing too much interest. No serious discussion could be started about RSI without hysteric recriminations by the “real professionals”.

And for the companies, though interested by the reduced cost of this service compared to on site interpretation, they doubted that the technique was reliable, that the interpreters were real professionals and that the audience would be satisfied and would appreciate such service.

Then, due to the fact that nearly anybody today in our Western societies has a smartphone that could support the download of a specific App and possibly use earbuds, RSI is today accessible nearly for free for the audience and at a reduced price for the customers.

How could such a change occur and what are the future prospects?

Digitalisation is applied in every domain, including in translation and interpretation businesses. Solutions have existed for a long time, but were not really considered in the past. The main opposition came from the professionals themselves, either in translation or in interpretation. A translation agency is considered a kind of devil stealing the heart and essence of the profession to resell garbage, useless documents, thus hiring the worst translators and taking the most profitable customers away from the market.

RSI underwent the same criticism in terms of recruitment: only incompetent interpreters would work for such horrible employers, equipment would be of bad quality, and unpredictable events could occur during transmission. So, why do broadcasting programs show World Cup finals or any event

if the quality of the devices were not good. Everyone should go to the stadium instead. Ah ah, but the match takes place on the other side of the world, so what is to be done to see the event?

What is the relationship between these three activities will you say? Well, you can do all in a row if you are well organised. The first one to warm you up, the second one to cool you down and the last one to let you do your job relaxed and mentally well prepared.

But those activities have to be practised at the same place, otherwise, one day will not be enough, especially if you have to take public transport or drive/fly to the place where interpretation is requested and sleep there overnight.

***

Have you ever heard about RSI (Remote Simultaneous Interpreting)? No? But you have heard about the COVID-19 and its economic impact, haven't you?

Well, there is today an even greater relationship between the two than ever: in the past, ("thousands of years ago"), interpreters had to move to the place of interpretation (booth, exhibition, court, …) – not to mention the ones working already for the Egyptian pharaohs or the Pope in Rome being on site or having to follow the armies to be wherever they were needed. If not killed afterwards because they brought or translated negative messages.

They would lose a lot of time and most often their nerves in railway stations, airports, buses, spend some money for the night, food and have to appear in nice dress apparel on site the next day. Costs, costs and costs. Could they invoice the transportation time? Probably not. Could they invoice the accommodation? Probably not either, it depends. So, how much remains in their pocket then at the end of the mandate?

Videoconferencing was fashionable back in the 90s when Internet became available and affordable. Companies would equip conference rooms for an expensive price, allowing their executives to be available 24/7 and willing to reduce their transportation time and cost. But the impact of remote interpretation remained as a last chance opportunity and only international organisations made their first tries with RSI.

Then came recently a nasty virus which forced everyone to stay home for some weeks/months/possibly years and a lot of events were cancelled. All of a sudden, event organisers, but also all companies in need of communication discovered that interpretation services could also be offered on-line with rather simple tools through Internet at a very affordable price. But by whom and at what quality level are you going to ask?

Actually, RSI has been created well before the COVID-19 crisis, but it was not very warmly welcomed on either side of the interpretation setup: interpreters would refuse to use such a degrading technique, arguing about the low quality of the transmission, the lack of equipment for interpreting locally (costs again), the risk of connection interruption, the loss of quality of interpretation and promising hell to every colleague showing too much interest in it. They would rather cast out their colleagues, and at the same time bury their head into the sand like the ostrich, hoping the danger (of communication development and technical improvements) would pass by without hurting them. It is a bit like Trump saying that the coronavirus would "vanish by itself".

As S. Braun[1] puts it: "the practice of remote interpreting was controversial among interpreters". And I would add, still is. No serious discussion can be started about RSI without hysteric recriminations on the side of the "real professionals". I can testify about those conversations, the last one taking place in November 2019 for the celebration of the ILO's 100th anniversary in Geneva. Colleagues went hysteric when we started a discussion with the CEO of Interprefy, Kim Ludwigsen, about the emergence of new tools and practices within the interpretation community.

Participating as a guinea pig in a study at the University of Geneva for an assistant's thesis in July 2020 about interpreters' reactions to visual stimuli when working, we had the same kind of discussion and the name "Interprefy" popped up (again) in the middle of the conversation. The devil came out of its box. We will come back to that.

On the companies' side, however, though interested by anything that can reduce the cost of this service compared to on-site interpretation, they doubted at first that the technique was reliable, that the interpreters were real professionals and that the audience could be attracted and like such a service. In the meantime, thanks to or because of the COVID-19 crisis, they realised that any interpretation service online would help them to overcome the meetings and transportation restrictions decided by the authorities. And some of them could, eventually, be convinced that RSI can allow them to access a multilingual information exchange market at a much lower cost than expected, creating a new demand for that kind of service, thereby creating niches in the interpretation market.

So, what is the situation today in the middle (where is it) of the covid-19 crisis?

***

Most interpreters are against it

The idea of remote interpreting is met with considerable resistance by professional conference interpreters, most visible in the discourse of the International Association of Conference Interpreters (AIIC). In its "Code for the use of new technologies in conference interpretation" published in 2000, the association warned that "*the temptation to divert certain technologies from their primary purpose e.g. by putting interpreters in front of monitors or screens to interpret at a distance a meeting attended by participants assembled in one place (i.e. tele-interpreting) is unacceptable*" (2000). The updated version of 2012 is unchanged in this respect[2].

But there is no further explanation about the reasons for this extreme position, which is just reinforcing the lack of motivation for accepting change from our colleagues and a mark of stubbornness. There is nothing mentioned, in particular, concerning health problems, technical drawbacks or financial disadvantages. It is just a sheer refusal. In principle.

As I participated in past meetings of the TRAFUT program chaired by Eulita, the European Association of Court Interpreters, numerous participants in the different meetings in Europe could testify that courts and police institutions, especially in France and the UK, but also in the Baltic republics, were investing massively in remote interpreting devices and facilities with a considerable cost advantage over transportation and security measures for bringing criminals from the jails they were sitting in to prosecutors' offices and courts with a minimum of quality loss in the work of interpreters or for no loss at all as demonstrated by recent research from the University of Essex3.

As Braun states[4], a "*survey among 200 legal interpreters conducted by Braun & Taylor (2012) shows a wide variety of attitudes towards videoconference-based interpreting. Although many*

1 Braun, S. (2015). Remote Interpreting. In H. Mikkelson & R. Jourdenais (Eds.) (2015). Routledge Handbook of Interpreting. London/New York: Routledge.
2 Braun, S. op. cit.
3 Erik Hertog, professor emeritus (Lessius Antwerp / K.U. Leuven) Trafut - Training for the Future – Final Report : DIRECTIVE 2010/64/EU OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 20 October 2010 on the right to interpretation and translation in criminal proceedings
4 Braun, S. op. cit.

interpreters perceive the introduction of video conferencing and interpreting as a cost-cutting exercise, some also have positive views, especially regarding the potential of videoconference technology to improve access to interpreting services and fairness of justice. Moreover, the survey reveals links between the interpreters' attitudes towards video conferencing and the situation in the country in which they work, in terms of the quality of the equipment and the general working conditions".

The main opposition comes from the professionals themselves, either in translation or in interpretation. A translation/interpretation agency is considered as a kind of devil which is stealing the heart and essence of the profession to resell garbage, useless documents, thus appealing to the worst translators and taking the most profitable customers away from the market. But it would be interesting to have recent figures about agencies' and platforms' market share compared to self-employed translators: I am sure the Pareto rule would be applicable again.

RSI went under the same criticism in terms of recruitment: only incompetent interpreters would work for such horrible slave-drivers and rely on the bad quality of the equipment, plus any possible hazardous event happening during the transmission. Is that the reality? After having practised for about 6 years now at a rate of nearly once per week on average, I doubt it. I have seen AIIC members coming down to the prosecutor's office to interpret at a much lower rate per hour than they normally request from conference organisers, so does the official AIIC's website position reflect reality? Its members have fewer problems with trying to find customers where they are. Besides, AIIC is the not the only representative of the interpretation business around the world.

Modern technology allows many accesses to online interpreting. Platforms such as Interactio, Interprefy, Zoom and many others allow RSI. And the post-COVID-19 economic situation is likely to reinforce the diversification of such tools and working methods. Not for today or tomorrow, but forever. Have you heard that Facebook proposed that its employees stay home and work from there up to 2021? And that Twitter requested its staff to stay and work from home **forever**?

So, while there are so many advantages to use it, why are professionals against it? Where is the trick?

***

Cost Cutting Advantage for Companies

There are very few studies about the real impact of online interpreting as a cost factor. Most of them deal with so-called health questions on the interpreters' side, but none is a true economic study. What a pity. We live in a world of digitalization, we have been using computers for a long time in the translation industry, what then are the obstacles to use them also with interpretation services?

Another major factor landmarking the technical landscape which is favourable to RSI is the fact that a lot of people today have a Smartphone that could support the download of a specific App and use earbuds to listen to remote interpreters. How do I know? Well, I am teaching it, 1/3 of the world's population has old mobile phone models with buttons and which cannot connect to Internet and another 1/3 has no mobile phone at all due to their lack of means. But in our Western societies, the equipment rate is close to 90%. And, for the listeners, RSI is reachable at no further cost within the targeted audience.

On top of it, there is no need to hire expensive intermediate audio and video companies and to set up an interpretation booth in an hotel early in the morning or the day before for as much as $5,000 per day. So that, at the end of the day, a newer market segment has appeared: SMEs can now afford the cost of RSI since a lot of the fixed costs are deleted. It's magical.

Interpreters would stay home and charge the normal interpretation amount per half-day or hour, without further costs. Companies would "rent" an IT platform to access the service for a short time and to get their audience connected. All in all, everyone would make some savings. And would also lose less time. You get the notion of the swimming pool now?

A lot of people speak foreign languages today, probably not enough to replace professional interpreters, but enough to have a "standard conversation", especially in the most commonly used languages. This effect is what we call, in business, the "Ikea Syndrome": people have the opportunity to solve basic problems themselves without requiring professional services. What is important is the cost. We have seen that in the translation industry where "automatic translation" is now common. There are some products also coming out on the market in the interpretation industry, mostly for Smartphone owners: people would talk together and try to get along together without any intermediary.

Even in the prosecutor's office or in front of the courts, most lawyers or judges would understand at least English or German, so why use a professional interpreter? But since RSI would reduce the cost for companies, a new middle-range-segment-market is now available for smaller companies which are not used to (and could not afford to) rely on professional interpreting services and that is a new chance for us.

The cost-cutting side of RSI is probably not the most important factor for interpreters since, in most cases, there is a possibility to negotiate a fee for transportation and accommodation. Food on site is generally available for the participants, so for the interpreters as well. The trickiest part is the time lost to go to the interpretation venue, which might be hours and miles away. Those practising translation and interpretation could do some translation during the trip, thus not completely wasting their time, but, as a matter of principle, "lost time" should also be compensated.

I have just made an offer for on-site interpretation at about 3 hours-drive from my office and the result is clear: the cost for the customer would double compare to online interpretation from my home.

In very few cases, transportation and accommodation are not provided and the interpreter has to pay for it. It happened to me last year for an interesting mandate in Copenhagen while the fee paid was low. But since I happened to be in Oslo the day before, I could change my flight for a reduced price and use my Expedia points to get a very nice hotel in Copenhagen for only $100. So, it was worth it.

Concerning interpreters' fees, there is also a trend to issue quotes at a reduced amount simply because the interpretation will be held "online". Someone answered me last week that my quote was "too expensive for online interpretation". I tried to convince the customer that, on-site or online, the job was the same. I wasted my time.

While, as mentioned, new customer segments might open themselves through the fixed-costs-reduction due to digitalization, there is no argument as to why interpreters should reduce their fees: the job is the same, just there is no physical presence. So, about the economic side of RSI: okay, here, something must be said about the tariffs: online interpreting, especially conference interpreting, must be paid the same price as on-site interpreting. That is a definitive general statement without which interpreters must refuse to accept working. Full stop.

As in the translation business, there will always be less-demanding people needing money so desperately that they will reduce their price; let it happen, there is nothing we can do against that.

***

Performance Issue?

Digitalisation is applied in every domain, also in translation and interpretation businesses. Solutions have existed for a long time[5], but were not really considered plainly as a major argument in

5 "The Australian immigration service is commonly credited with establishing the first service for telephone-based interpreting, the Telephone Interpreting Service (TIS), in 1973". Braun, S. op. cit.

the past. Exceptions apply to international organisations or courts. Early in the 2000s, London police authorities set up dozens of hubs in London to avoid too much time loss due to transportation, so that interpreters would go only to the nearest hub and work from there. When the criminal code was amended back in 2011 at the European level, courts everywhere in Europe started to equip themselves with digitalized distance devices so that prisoners would not have to be transferred from jail to court room, thus reducing transportation costs and increasing security measures.

Researchers have "demonstrated" a performance reduction of online interpreters as opposed to on-site interpreters[6]. Every case is different, for sure, but I wonder what they would have said when I met 11 police officers back in 2017 at 5.15 in the morning for a house search for a German-French hearing which lasted, with very short breaks, quasi non-stop up until 8.30 pm… The Zurich prosecutor wanted to make his journey profitable and to come back to the economic centre of Switzerland the same day with quite a large quantity of seized documents. So, online or on site, the tiredness of the work is the same.

Ko (2006)[7] and Lee (2007)[8] draw attention to the working conditions of telephone interpreters arguing that the generally high levels of dissatisfaction associated with telephone interpreting partly stem from the working conditions including low remuneration rather than from the use of the technology as such. The guidelines issued by the Wisconsin court authorities, for example, recommend a maximum of 15 minutes conversations[9]. Okay, to illustrate this point linked to my own experience, it must be said here that, without any visual cues, the required level of concentration is more intense, so that, yes, there can be more tiredness after a while. But, on the other side, a telephone conversation does not last 11 hours…

According to a performance analysis of the ITU/ETI study[10], one of the major differences between on-site and remote interpreting is that the interpreters' performance in remote interpreting declined faster than their on-site performance and this was explained by an earlier onset of fatigue in remote interpreting. With regard to stress, the interpreters participating in the study found remote interpreting more stressful, and their stress hormone values were higher in remote interpreting, although neither difference reached statistical significance.

Other studies provide similar information[11]: in an hospital-context-experiment, a comparison has been made between on-site, video conferencing and phone interviews where phone interviews were significantly shorter than in-person. Patients rated interpretation services highly no matter how they were provided, but experienced only the method employed at the time of the encounter. Providers and interpreters were exposed to all methods, but were more critical of remote methods, and preferred video conferencing to the telephone as a remote method. The significantly shorter phone interviews raise questions about the prospects of miscommunication in telephonic interpretation, given the absence of a visual channel, but other factors might have affected time results. Since the patient population studied was Hispanic and predominantly female, care must be taken in generalizing these results to other populations[12].

---

6 Moser-Mercer, B. (2003) Remote interpreting: assessment of human factors and performance parameters. Communicate! Summer 2003. https://aiic.net/page/1125/remote-interpreting-assessment-of-human-factors-and-performance-parameters/lang/1 (accessed 19/06/2020).

7 Ko, L. (2006) The need for long-term empirical studies in remote interpreting research: A case study of telephone interpreting. Linguistica Antverpiensia NS5, 325–338.

8 Lee, J. (2007) Telephone interpreting — seen from the interpreters' perspective. Interpreting, 2(2), pp.231–252.

9 (http://www.wicourts.gov/services/interpreter/docs/telephoneinterpet.pdf).

10 Moser-Mercer, B. (2003), op. cit.

11 Locatis, C., Williamson, D., Gould-Kabler, C., Zone-Smith, L., Detzler, I., Roberson, J., Maisiak, R. & Ackerman, M. (2010). Comparing In-Person, Video, and Telephonic Medical Interpretation. Journal of General Internal Medicine. 25. 345-50. 10.1007/s11606-009-1236-x.

12 Locatis, Craig et al, Op. cit.

Using trained interpreters to provide medical interpretation services is superior to services provided on an ad hoc basis, but little is known about the effectiveness of providing their services remotely, especially using video. There should be a comparison between remote medical interpretation services by trained interpreters via telephone and videoconference and those provided in-person. There was a quasi-randomized control study. Two hundred and forty-one Spanish speaking patient volunteers, twenty-four health providers, and seven interpreters. Patients, providers and interpreters each independently completed scales evaluating the quality of clinical encounters and, optionally, made free text comments. Interviews were conducted with 23 of the providers, the seven interpreters, and a subset of 30 patients. Time data were collected.

Encounters with in-person interpretation were rated significantly higher by providers and interpreters, while patients rated all methods the same. There were no significant differences in provider and interpreter ratings of remote methods. Provider and interpreter comments on scales and interview data support the higher in-person ratings, but they also showed a distinct preference for video over the phone.

This being said, in another study conducted by the European Parliament[13], the comparison of the interpreters' performance in the two conditions resulted in slightly lower rates for remote interpreting but the difference failed to reach statistical significance, and the authors believe that the differences "may be regarded as rather minor in practical terms"[14]. The interpreters participating in the study perceived remote interpreting as being significantly more stressful than on-site interpreting, but again no such differences were found in objective measures of stress in this study. In general, the most striking result of these studies seems to be the discrepancy between objective findings and subjective perception.

So, what conclusions can we draw about those studies? There is the psychological effect of being far away, not seeing all the physical cues and body language. But there is the positive effect of being in a well-known environment with familiar objects, a probably more comfortable chair, no-one alongside to pass you little notes about this and that which is perturbating, your familiar cup of coffee or tea, this gives reassuring messages when you are "in".

***

Technological Issue?

Causo[15] also highlights further conditions that may have an adverse effect on the sound quality and comprehension of videoconference-based interpreting, emphasising that "videoconferences are frequently linking standard offices unsuitable for this purpose, or have a poor set up, which means sound reverberation [...], simple omni directional microphones integrated in the table, etc.".

This might have been true years ago, but it is no more the case. 3G, 4G and in the future 5G connections are good enough to get a flow of audio and video streams that is as good as sitting 30 metres from the screen and the speakers. Wi-Fi is obviously the best possible connection basis today and most conference halls, stadiums and commercial offices are equipped, so that the connection is no longer a problem. In remote areas, it is possible to have a line drop for a few seconds, but that will not prevent the interpreter from understanding the meaning of the sentence since it is possible to restart the line within 1-3 seconds.

The interface of modern RSI tools is a perfect copy of the view you get from inside a booth without the heat and the noise of the fans, the cam(s) reflected in the window in front of you showing both the speaker and the presentation and there are enough in and out channels to support tens of different languages. There are two chat windows that allow the interpreter to exchange with his/her

13 Roziner, I. & Shlesinger, M. (2010) Much ado about something remote: Stress and performance in remote interpreting. Interpreting 12 (2), 214–247.
14 Braun, S. (2015), op. cit.
15 Causo, J.E. (2012) Conference interpreting with information and communication technologies. Experiences from the European Commission DG Interpretation. In S. Braun & J. Taylor (Eds), 227-232.

partner(s) and with the conference organiser when needed. The distribution of channels replicates the distribution of booths with only one pair of languages at a time plus a possible relay channel. Changing from one booth to another is done by clicking on the parameters: it happens quite often to me to jump from the French-English to the German-French booth to take over from someone when the budget was reduced or in case of absence or late arrival of a colleague. All that done from my office at home!

***

About the last issues for debate, there is the length of interpreter-mediated encounters that involve the use of communication technology. Given that research shows a faster onset of fatigue in remote interpreting (Braun 2013, Moser-Mercer 200[16]), an interpreter's working turn in remote and teleconference interpreting should be shorter than in traditional interpreting.

Another subject deals with preparation and briefing that an interpreter would normally receive; even with RSI, it should not be omitted because of the usage of communication technology. When deciding to go for RSI, institutions using teleconference or videoconference interpreting should develop procedures for deciding whether or not these methods of interpreting are suitable for a particular situation. Interpreters should be consulted where necessary. Furthermore, testing of the connections between the locations is crucial, especially when the equipment is used only occasionally.

As S. Braun mentions it nicely: "*To date there is no consensus about the quality of interpreting that can be achieved in remote and teleconference interpreting compared to the quality of traditional interpreting in comparable situations, and what exactly the relevant shaping forces are. The variation in settings, requirements for quality and research methods means that the findings from different fields of interpreting are difficult to compare.*"

And lastly, may the condition of remoteness or the lack of "presence" be a problem with remote interpreting? In my experience, it is not the case. Looking at a big screen in short pants or being on-site from inside a booth does not change anything.

You are "in" anyway!

Welcome to RSI!

16 Remote Interpreting: Assessment of human factors and performance parameters, Barbara Moser-Mercer, last seen on October 13, 2020 on https://aiic.net/page/1125/remote-interpreting-assessment-of-human-factors-and-performance-parameters/lang/1

On Air: How Can Terminology Extraction and Management Technology Help Language Professionals in Broadcast Media?

Daniya Khamidullina

University of Wolverhampton / University of Málaga

daniya.khamid@gmail.com

ABSTRACT

Modern broadcast media is characterised by a high degree of internationalisation. Most major media outlets are either multilingual by design (e.g. BBC, France 24, VOA, RT, etc.) or function in a plurilingual environment due to the globalised nature of the current news agenda. News organisations seek to ensure their viewers can access the latest developments with minimal latency, which is sometimes achieved through live on-air interpretation of speeches, announcements, press conferences and other events.

While broadcast media interpreting is intrinsically a technology-reliant process, technological solutions that specifically facilitate the work of media interpreters still appear to be relatively scarce. This paper proposes a prototype of a digital tool that could be used by interpreters working in multilingual broadcast media. The tool could serve as an aid during multiple stages of the interpreting process, including assignment preparation, in the booth and during the post-assignment de-briefing. Its key functions would include terminology extraction, terminology management and, optionally, speech recognition in the booth. The prototype combines several existing technological solutions and could interface with available tools offering these solutions.

Since the tool is designed to be used in multilingual media outlets, it considers the specific workflows in these contexts. In news media outlets, translation-related activities are often not considered as "essentially dissimilar from other tasks involved in the production of news" (Bielsa, 2007, p. 143). This highlights an interesting dichotomy: while interpreting as a process technically does not overlap with journalism per se, in terms of terminology, semantics and pragmatics, it is still interlaced with other activities that take place in a newsroom. Therefore, from the point of view of practicality and marketability, a technological tool designed for the media interpreter should ideally provide transferrable solutions that could be employed by both interpreters and other agents involved in plurilingual-context news production, such as translators, news writers or output editors. This paper therefore focuses on the applicability of such a tool in the area where the activities of interpreters and other language specialists could overlap the most: terminology extraction.

The prototype was tested on the Russian-English language pair using Vladimir Putin's annual, wide-ranging press conference, as complete official transcripts of all recent editions of this event are openly accessible online both in Russian and English, which facilitated the compilation of pilot corpora. A 268.000-word parallel corpus was created, which was then run through automatic terminology extraction tools in different configurations: as a bilingual corpus composed of complete unedited transcripts and as pre-processed subject-specific bilingual subcorpora, whereby information had to be categorised by topic manually. Tools provided by Sketch Engine and Terminotix were used in the tests. In our case, the solution that appeared to yield the most comprehensive results requiring the least amount of post-editing was automatic bilingual term extraction from subject-specific parallel subcorpora, carried out in Synchroterm by Terminotix. Indeed, a corpus comprising full unedited transcripts of previous press conferences might not be the best fit for automatic term extraction, as such texts contain both spontaneous speech and common language, which generate noise, as our tests have demonstrated. Pre-processed thematically arranged corpora appear to be a better option, since bilingual term lists automatically generated from this type of corpora require less post-editing.

As the pilot run has shown, although the proposed tool design could potentially facilitate various aspects of a broadcast media interpreter's work, the tool's other modules would need to be tailored carefully to the task at hand for best results. For terminology management, it should be easy to use the tool for a variety of types of search. In addition, results would ideally be shareable to allow for teamwork between different members of the news production team; a placeholder solution corresponding to these criteria was adopted for testing purposes. Although initial tests of Automatic Speech Recognition were conducted, further testing using professional hardware is necessary.
In conclusion, some components of the proposed tool could simplify the work of the interpreter and other language specialists in the media. Future work will test the actual applicability, practical benefits, interpreters' performance, and attitudes toward the prototype with media interpreters working in the Russian to English language pair.

* * *

Introduction

Modern broadcast media tends to be characterised by a high degree of internationalisation, which is particularly manifested in the context of global news production and distribution. Many major news media outlets are either multilingual by design (e.g. DW, France 24, VOA, etc.) or function in a plurilingual environment due to the globalised nature of the current news agenda. News organisations normally seek to ensure their viewers can access the latest developments with minimal latency, and sometimes this is achieved through live on-air interlingual interpretation of speeches, press conferences and other events which, in a slightly broader sense, is also known as *media interpreting* or *broadcast interpreting* (Pöchhacker, 2010, p. 224).

Live coverage of global events through interlingual interpreting is an integral part of modern journalism and mass media broadcasting (Castillo, 2015, p. 281). However, while both of the above are intrinsically technology-reliant processes, technological solutions that would specifically aid language specialists in mass media seem to be comparatively scarce. Despite the fact that there have been considerable advances in translation and interpreting technology in recent years, it appears that solutions designed specifically for interpreters and for meeting their needs are still lacking (Fantinuoli, 2017, p. 25; Corpas Pastor, Durán-Muñoz and Costa, 2018, p. 81; Goldsmith, 2020, p. 299) and the level of technology uptake within the professional community remains quite low (Corpas Pastor, Durán-Muñoz and Costa, 2018, p. 60). In the same vein, to the best of our knowledge, tools that could potentially be used in the relatively niche scenario of media interpreting do not seem to be task-specific, which is what prompted this exploratory study. In this paper, a prototype of a tool that could facilitate the work of news mass media language professionals is proposed.

The paper is structured as follows: firstly, a brief overview of the specifics of workflows in media outlets will be provided based on a review of existing publications on the subject. Then the design of the proposed tool prototype will be described. The key modules of the tool (terminology extraction, terminology management and automatic speech recognition (ASR)) will be discussed, with a special focus on the desired characteristics imposed by the media context. This will be followed by a section dedicated to presenting the results of pilot tests of the terminology extraction module in the Russian-English language pair; the methodology and

the description of preliminary findings will be outlined. Finally, conclusions will be presented, and possible future research avenues will be touched upon.

The specifics of interpreting in multilingual media outlets

Since the prototype proposed in this paper is tailored for use in multilingual media outlets, it is necessary to touch upon the characteristic features of workflows in such contexts. Existing publications on media translation and interpreting point out that in mass media, language-related tasks are not necessarily clearly delineated from other activities involved in the production of news, and the lines between the duties of different team members may be blurred (Bielsa, 2007, p. 143). While this is not always the case with interpreting, which is sometimes done by external specialists hired on an ad-hoc basis (e.g. as described by Andres and Fünfer (2011, p. 104)), overlapping duties and hybrid in-house interpreter roles are still commonplace in news production environments.

For instance, interlingual interpretation can be performed not only by freelance or in-house linguists, but also by other members of a team covering a given event such as reporters, expert commentators or desk editors (Castillo, 2015, p. 285–286). Furthermore, in certain setups, even the concept of *interediting* is applicable, as proposed by Arzık Erzurumlu (2019, p. 69); this term refers to the submodality of media interpreting carried out by in-house staff and accurately reflects the effect that the hybrid nature of the interpreter's occupation may have on the interpreting process. This convergence of roles, skill-sets and approaches in language-related work in the media has been discussed by scholars in the past (Kurz, 1997, p. 196; Mack, 2002, p. 208), and as Kurz put it (1990, p. 173), a media language specialist needs to be a 'hybrid' that possesses the skills of a translator, an interpreter and an editor at the same time.

All of the above highlight the extent to which interlingual interpreting tends to be interlaced with other newsroom activities. Therefore, from the point of view of practicality and marketability, a digital tool designed for the media interpreter should include transferrable solutions. Such a tool should ideally facilitate the work not only of interpreters, but also of other agents involved in news production in a multilingual setting: e.g. translators, news writers or desk editors. This factor was taken into account during the selection of the functional modules of the tool, which will be described in the next section.

The modules of the proposed tool

The proposed digital aid could help the interpreter at multiple stages of the process: during assignment preparation, in the booth and during post-assignment de-briefing. The tool would include three key modules performing the following functions: terminology extraction, terminology management and automatic speech recognition. In this section, the key desired features of each of the components will be outlined, and special focus will be placed on context-specific requirements.

The terminology extraction module would be used by the interpreter during assignment preparation. As this stage tends to be quite time-consuming (Fantinuoli, 2017, p. 25), the interpreter could benefit greatly from access to automated or semi-automated terminology extraction functionality that might speed up the process. This is highly relevant in the case of media interpreting, where it is quite common for assignments to be announced at very short notice: for example, when it comes to coverage of emergencies or press conferences (Kurz,

1997, p. 196; Pöchhacker, 2010, p. 225). In view of this, a speedy ad-hoc termbase compilation solution could be of great value. In addition, it could improve the interpreter's performance: as a recent study by Xu has shown (2018, p. 50), an assignment preparation procedure that involves automated term extraction can lead to higher terminological accuracy during interpretation as compared to traditional preparation options.

In terms of task-specific features, the terminology extraction tool for the media linguist should be able to support a wide variety of formats and possibly include a webpage import option such as, for instance, that provided by *Intragloss* (Goldsmith, 2017). This could prove to be extremely useful as termbases on news-related topics are likely to be built and updated using mostly online sources, including those generated by the media outlet in question. Therefore, the tool would ideally have a built-in text pulling function that would enable the interpreter to download text data from which terminology would then be extracted.

Terminology management tools, in turn, allow the interpreter to create, use and reuse termbases and glossaries, which might be done at multiple stages of the interpreting process: during preparation, in the booth and at the de-briefing stage. In addition, depending on the layout of the tool, terminology management software can help the user not only to memorise the terms or search for target-language term equivalents, but also to structure their subject knowledge in general. As Xu highlights, 'by learning the most relevant terms and the concepts behind them, an interpreter can acquire key features of the knowledge system' of a given field (2018, p. 30). As the news agenda is intrinsically highly dynamic, an efficient terminology management system could facilitate the media interpreter's acquisition of knowledge on new topics, the array of which in the news context is wide and ever-changing.

To be used successfully, the terminology management system might need to have such interpreting-specific functions as 'intuitive, incremental accent-insensitive search, integrated search of online sources, and workflow support during preparation' (Ruetten, 2017, p. 99). In addition, just like for the terminology extraction module, the online source access option would be highly beneficial in the media context. A log file feature, such as that implemented in *InterpretBank* as described by Ruetten (2017, p. 100), where changes and search queries made during an assignment can be tracked, could also be useful at the stage of de-briefing. Upon completion of an interpreting assignment, new entries could be added to the termbase or existing items could be updated; here, the results of terminological work of the wider team could be taken into account.

On a larger scale, collaborative terminology management systems can also foster terminological uniformity within the organisation in which they are employed. As Costa, Corpas Pastor and Durán-Muñoz observe, the possibility of sharing terminology is important as it 'allows users to improve terminology by enhancing term coverage and consistency within and throughout domains in a collaborative fashion' (2018, p. 80). In the media setting, the added benefit of this function would be the fact that it can help ensure terminological conformity across platforms (e.g. TV news reports, online articles, social media, etc.). A system where in-house terminology is stored could also prove useful during onboarding of newly hired interpreters or other language specialists, who could this way become acquainted with editorial conventions and stylistic guidelines in a rather efficient manner.

As for the automatic speech recognition module, there are several ways in which it could facilitate the work of the media interpreter. Firstly, it could be used in the booth: as pilot studies

have shown, automatic recognition of numbers in source audio and their display on a screen can improve the interpreter's performance (Desmet, Vandierendonck and Defrancq, 2018, p. 22). Another way that speech-to-text technology could be employed is as a documentation tool during assignment preparation (Gaber, Corpas Pastor and Ahmed, 2020): ASR can facilitate transcription of speeches available in audio or video format. This application appears to have particular relevance in the media context, as many news-related materials that interpreters could use for preparation exist in these formats but are not always available as text. In addition, the speech recognition functionality could be used by other members of the news team to transcribe interviews, vox pops and any other speech contained in raw footage. Bearing this in mind, it can be said that a speaker-independent variant of speech recognition technology would be the preferable option in the media due to the high variability in voices that might need to be processed (ibid.).

1. Terminology extraction module: pilot testing phase

To arrive at a better understanding of which type of solution architecture would be preferable in the given context, preliminary tests of the modules using placeholder components were carried out. In this paper, the results of the terminology extraction module test in the Russian-English language pair will be presented.

1.1. Dataset and methodology

The news-related assignment scenario adopted for the test was Russian-to-English simultaneous interpretation of a press conference. Publicly accessible bilingual transcripts of several editions of Vladimir Putin's annual wide-ranging press conference were used as the test dataset for terminology extraction. These materials were selected due to their public availability, relevance to the news agenda and the broad spectrum of topics covered, which made it possible to run tests in different configurations. Complete transcripts and their translations can be accessed via http://www.kremlin.ru and http://en.kremlin.ru, the official internet resources of the President of Russia.

In total, 10 transcripts of five press conferences (five original texts in Russian and their translations into English) held between 2015 and 2019 were downloaded and converted to plain text. At this stage, an ad-hoc solution that facilitated format conversion was used. A spreadsheet with links to source texts, webpage IDs and the names of corresponding plain text files was created to ensure traceability of sources. The files were then pre-processed manually: time and date information and tags were removed. This resulted in a preliminary set of files combining all 10 transcripts, which contained a total of 267,898 words.

To detect the most suitable terminology extraction modality in the given scenario, tests were run in two ways: extraction in bulk and extraction from thematic subcorpora. For the latter modality, complete conference transcripts were manually divided into thematic files to be used for subsequent domain-specific term extraction.

Since transcripts in Russian and English were originally accessed and saved as separate files, *YouAlign*, a free online document alignment solution, was used to create aligned bilingual transcripts for each of the editions of the conference. The aligned bitexts generated as .htm files were then cleaned up: automatically added tags were removed *([RUS], [ENG], [~]*).

The transcripts were then analysed manually: the topic of each question-answer unit (QAU) in the Russian transcript was identified, and then the QAU and its translation into English were copied into a separate thematic file. If a QAU covered more than one topic, such QAU was cross-added to all relevant thematic subcorpora. QAUs that did not contain specialised language were not added to the subcorpora.

In total, 17 key topics were identified within the dataset:

1. Agriculture and Aquaculture
2. Defence
3. Domestic Politics
4. Economy
5. Energy
6. Environmental Issues
7. Healthcare
8. Industry
9. International Relations – China
10. International Relations – Syria-Middle East
11. International Relations – Turkey
12. International Relations – Ukraine
13. International Relations – USA
14. International Relations – various
15. Social Affairs
16. Sports
17. Transport

For the tests, three of the resulting subcorpora were selected: International Relations – China (2,919 words), Economy (44,201 words) and Healthcare (5,838 words). These subsets were selected in accordance with two criteria: 1) length diversity; 2) recurrent nature of the topics.

Terminology extraction tests were then performed in the two aforementioned modes: extraction from complete transcripts vs. thematically arranged subcorpora. The process and the results will be described in detail in the next sections.

1.2. Results and discussion

1.2.1. Extraction in bulk: *OneClick Terms* by Sketch Engine

First, full transcripts of all press conferences were run through *OneClick Terms* as plain text in bulk. Not unexpectedly, due to the wide variety of subjects covered in the uploaded corpus, the resulting lists of terms were extremely thematically heterogeneous. Therefore, it would be quite difficult for an interpreter to create a glossary based on such output given the lack of semantic coherence within the generated termbase.

The list of suggested terms appears to contain numerous multi-word expressions (MWEs) that belong to common rather than specialised language. For instance, the top 10 entries on the English term list are as follows:

good afternoon, news conference, first point, second question, retirement age, defence industry, Russian economy, first question, news agency, tv channel

When it comes to Russian, many of the top suggested terms also belong to common language (literal translations are provided):

добрейший день (good afternoon), соединенный штат (a united state), средств массовой информации (mass media), пенсионный возраст (retirement age), следующим год (next year), лучший показатель (best result), экономический союз (economic union), уважаемый Владимир (dear Vladimir), центральный банк (central bank), Евразийский экономический союз (Eurasian Economic Union)

Another issue with term extraction in Russian is that some of the suggestions are non-lemmatised: e.g. '*средств массовой информации' ('of mass media')* in the genitive case form. Moreover, in certain cases, there are declension conflicts within the proposed entries: e.g. '*следующим год' ('next year',* where the adjective is in the instrumental case form, while the noun is in the nominative/accusative case form). Such suggestions would need to be corrected manually if the terms were to be added to a glossary, which would require extra effort from the interpreter.

All of the above makes extraction in bulk via *OneClick Terms* unlikely to be an interpreter's first choice as the output is too heterogeneous to serve as basis for a glossary and requires thorough manual clean-up, especially on the Russian side. In addition, as *OneClick Terms* does not offer bilingual term extraction, Russian and English term lists would need to be aligned manually, which is quite time-consuming. At any rate, due to the discrepancies between the lists in the two languages, most interpreters might prefer to just translate the terms from scratch in this case.

Nevertheless, the results of bulk extraction with *OneClick Terms* might have one particularly useful application in the given context – the term lists can help the interpreter determine the scope of the topics that tend to be covered during such press conferences. This knowledge can help them create a catalogue of subjects for subsequent compilation of ad-hoc domain-specific corpora.

1.2.2. **Extraction in bulk: *Synchroterm* by Terminotix**

The second tool tested was *Synchroterm* by Terminotix. Unlike *OneClick Terms*, this tool offers bilingual term extraction, which essentially allows the user to generate a bilingual termbase. *Synchroterm* aligns and processes input texts, and there are two termbase creation options: the user can either add entries to it manually or use the batch processing feature to produce a termbase automatically. In the present test, initially, the latter option was employed; in this setup, the degree of the user's intervention in the termbase generation process is similar to that observed in the case of *OneClick Terms*.

As was the case with *OneClick Terms*, the output of bulk processing via *Synchroterm* is extremely heterogeneous and unlikely to be useful as a glossary basis. Below is an excerpt of extraction output:

Source Entry life news
Target Entry life news

Source Entry вчера только
Target Entry just yesterday

Source Entry бюджетной сфере
Target Entry public sector

Source Entry газета «коммерсантъ
Target Entry kommersant newspaper

Source Entry ветеранов и инвалидов
Target Entry including disabled war veterans

Source Entry говорит москва
Target Entry govorit moskva

Source Entry внутренний спрос
Target Entry domestic demand

Source Entry гоменюк-кравцова
Target Entry maria gomenyuk-kravtsova

Source Entry военной разведки
Target Entry military intelligence

Source Entry долларов за баррель
Target Entry barrel

[Figure 1]

It can be observed that in some cases, alignment is slightly off (e.g. '*долларов за баррель' (lit. dollars per barrel)* – **'***barrel'*), and the suggested entries are often non-lemmatised ('*бюджетной сфере' (lit. public sector)* – prepositional case form); this means that batch extraction output would have to be edited carefully. However, it should be noted that the user of *Synchroterm* is warned about that explicitly in the 'batch processing' section of the programme manual (Terminotix inc., n.d., p. 25).

All in all, even putting the noise aside, it appears that bulk extraction output of both *OneClick Terms* and *Synchroterm* would not be exactly usable – partially due to the width of the range of topics covered in the source texts. In both instances, resulting term lists were semantically heterogeneous, which essentially defeats the purpose of creating a glossary in the first place.

In the next sections, the results of automatic term extraction from thematically arranged subcorpora will be described.

1.2.3. **Extraction from thematic subcorpora: *OneClick Terms* by Sketch Engine**

All in all, the results of domain-specific term extraction in *OneClick Terms* seem to be better suited for termbase creation that those described above; in this case, suggested term lists in both languages turned out to be much more homogeneous.

For instance, the top 10 entries of the generated term list on the topic of healthcare in English are as follows:

medical assistance, primary care, cancer treatment, own pharmaceutical industry, medical air service, system-wide solution, head doctor, medical air, air service, child mortality

Comparable levels of semantic coherence are observed for Russian:

тариф омс (Compulsory Health Insurance tariffs), первичное звено (primary care), уровень заработной платы (pay level), лекарственный препарат (medication), следующим год (next year), данные минфина (Finance Ministry data), строительство онкоцентров (construction of cancer treatment centres), рядовой врач (rank-and-file doctor), системное решение (system-wide solution), звено здравоохранения (medical care level)

Although these lists would still need to be cleaned up, the entries comprising them are mostly domain-relevant and therefore make a good term set to start from. Nevertheless, the entries would still have to be amended: just like in the cases mentioned in previous sections, terms and keywords in Russian are often non-lemmatised (e.g. '*ростом экономики' (lit. 'by economic growth'* – instrumental case form) or contain declension conflicts (e.g. *'многополярного* (genitive case form) *мир* (nominative/accusative case form)). In addition, there are numerous partial repetitions of MWEs (e.g. '*natural population growth'/'natural population'*; '*different healthcare management model'/'healthcare management model')*.

Finally, as *OneClick Terms* does not offer bilingual extraction, the interpreter would still have to find and add target-language matches manually. Nevertheless, on the whole, term extraction from thematic subcorpora with this tool is a viable solution provided that the interpreter has enough time to correct the term list and to search for translations.

1.2.4. **Extraction from thematic subcorpora: *Synchroterm* by Terminotix**

When terminology extraction from domain-specific subcorpora was performed in *Synchroterm*, initially, the batch processing feature was used so that the termbase would be generated automatically. However, the output for all three tested domains contained quite a lot of noise and misaligned entries. Below is an example of the 'Healthcare' subcorpus extraction:

Source Entry алмазовский центр
Target Entry addition to this hospital
Source Entry внутрь самой отрасли
Target Entry need to look
Source Entry вообще не останется
Target Entry change anything
Source Entry впервые включена
Target Entry put on that list
Source Entry врачей – совсем другая
Target Entry higher than ordinary doctors
Source Entry выездов к больным
Target Entry need to have a single
Source Entry границу наших лекарственных
Target Entry medicines abroad
Source Entry данным минфина
Target Entry finance ministry
Source Entry действующего тарифа омс
Target Entry buying medicine or equipment
Source Entry деле известны
Target Entry quite simple

[Figure 2]

Here, the resulting termbase includes many entries belonging to common language and is often misaligned (e.g. '*внутрь самой отрасли' (lit. inside the very sector) – 'need to look'*). Presumably, the former could have been caused by an insufficient volume of source material and/or the overlapping of several domains in the QAUs comprising the subcorpus. If this is the case, a possible solution could be to enrich the corpora with comparable domain-specific texts. This, however, would require further testing. As the output of batch extraction from the present dataset required extensive editing, *Synchroterm's* extraction modality involving manual term validation was then also tested so that results could be compared.

1.2.5. **Extraction from thematic subcorpora with manual term validation: *Synchroterm* by Terminotix**

In this mode, while terminology is extracted from bitexts automatically, the user has to approve the system's suggestions manually and, if necessary, make adjustments prior to adding entries to the termbase. The manual term validation feature allows the user to reject irrelevant suggestions, which means that the resulting termbase might not contain any noise at all. The final termbase may not need to be post-edited if this step is carried out carefully since here,

entry editing is performed in a semi-automated manner during validation. The procedure helps ensure that the result is a 'clean', ready-to-use bilingual term list:

1	ВВП	GDP
2	Евразийский экономический союз	Eurasian Economic Union
3	НДФЛ	personal income tax
4	Резервный фонд	reserve fund
5	Фонд национального благосостояния	National Welfare Fund
6	Центральный банк	Central Bank
7	высокотехнологичные сферы	high-tech industries
8	дефицит бюджета	budget deficit
9	диспропорции на рынке	market disproportions
10	доходы населения	income of the population

[Figure 3]

However, manual term validation is quite a time-consuming process as the user needs to filter out unwanted suggestions and correct the occasional lemmatisation inaccuracies. While performing term extraction in this modality might eliminate the need for extensive post-editing of the end product, some editing may still need to be done at the validation stage. Due to this, such a procedure might not be the preferred course of action in a time-pressured situation, which is often the case for media interpreters. Nevertheless, in less urgent circumstances, this setup allows the interpreter to compile a curated termbase in an efficient manner. In *Synchroterm*, this semi-automated approach also has a significant advantage over completely automated extraction: in a way, it facilitates assignment preparation on a semantic level (Ruetten, 2003), enabling the interpreter to view each of the suggested entries in context and thus potentially improve their grasp of the subject matter.

Overall, it appears that while this approach requires more time and effort on the part of the interpreter, it provides the most usable end result, which potentially makes semi-automated bilingual terminology extraction the preferred option out of all the tested modalities from the point of view of end output quality.

2. Conclusions

All in all, it can be said that while the proposed digital tool design could facilitate certain aspects of the media interpreter's work, its modules would ideally need to be tailored to the task in question. Although the key technological solutions for the suggested modules already exist, most of them might have to be equipped with an adjusted interface and task-specific features to become truly useful for simultaneous interpreters in the media. For best results, context-specific requirements of mass media workflows described above should be taken into account. In addition, given the high degree of hybridisation of interpreting-related roles in the media environment, it appears reasonable to have the tool designed in a way that could facilitate linguistic work beyond the scope of interpreting, if needed.

The pilot test of the terminology extraction module has indicated the following: it seems that in the given test scenario (preparation for Russian-English interpretation via extraction of terminology from a wide-ranging press conference dataset), semi-automated generation of bilingual term lists from thematically arranged subcorpora yields output that requires the least amount of post-editing of the end product.

Regarding the corpora used for terminology extraction, in our test scenario, complete unedited transcripts of press conferences do not appear to be the most suitable source material. In this case, terminology extraction output contains quite a lot of noise and is characterised by a high degree of heterogeneity, which might be caused by a large proportion of spontaneous speech and common language as well as high thematic diversity in raw transcripts. Extraction from thematically arranged subcorpora predictably yields more homogeneous and therefore more usable results. The analysis of the output of extraction in bulk vs. extraction from domain-specific text sets demonstrates that enriching transcript-based thematic subcorpora with additional comparable materials might improve the quality of terminology extraction output. This hypothesis, however, would require further testing.

3. Future work

In this section, additional future research avenues will be outlined. An adapted version of the four-stage system development approach proposed by Ruetten (2003) can be used in further research on this project, which could consist of the following four stages: conceptualisation, evaluation, prototype, testing.

At the conceptualisation stage, the proposed prototype could be reviewed and refined. Additional modules, such as a corpus management component, and functionalities, such as natural language processing (NLP) solutions (e.g. automatic topic analysis or named entity recognition applied to source corpora), could be added to the tool.

During the evaluation stage, the chosen solutions could be assessed to determine which types of architectures are the best fit for the task at hand. The groundwork has already been laid with the pilot test of the terminology extraction module, but more extensive tests are needed for conclusive remarks to be made.

The mock-up tool (prototype) combining the selected components could then be created based on the results of the first two steps. It would then be rather interesting to try implementing the procedure in practice and to assess interpreters' performance and attitude – possibly via remote interviews or think-aloud protocols. This could help shed light on what the practical benefits of automating different aspects of the interpreter's work are and acquire a better understanding of which solutions could become a welcome addition to the media language specialist's toolkit.

References

Andres, D. and Fünfer, S. (2011) TV interpreting in Germany: the television broadcasting company ARTE in comparison to public broadcasting companies, pp. 99–114.

Arzık Erzurumlu, Ö. (2019) Interediting: Revealing the Dual Role of Staff Interpreters on Turkish News Channels, *transLogos Translation Studies Journal*, **2**(1), pp. 68–84.

Bielsa, E. (2007) Translation in Global News Agencies, *Target*, **19**(1), pp. 135–155.

Castillo, P. (2015) Interpreting for the Mass Media, In *The Routledge Handbook of Interpreting*, *Routledge Handbooks in Applied Linguistics*, Mikkelson, H. and Jourdenais, R. (eds.), Routledge, pp. 280–301.

Corpas Pastor, G., Durán-Muñoz, I. and Costa, H. (2018) Assessing Terminology Management Systems for Interpreters, In *Trends in E-Tools and Resources for Translators and Interpreters*, Leiden, Brill, pp. 57–84, [online] Available from: https://brill.com/view/title/35221 (Accessed 8 November 2020).

Desmet, B., Vandierendonck, M. and Defrancq, B. (2018) Simultaneous Interpretation of numbers and the impact of technological support, In *Interpreting and technology*, Fantinuoli, C. (ed.), Berlin, Language Science Press, pp. 13–27.

Fantinuoli, C. (2017) Computer-assisted preparation in conference interpreting, *The International Journal of Translation and Interpreting Research*, **9**(2), [online] Available from: http://www.trans-int.org/index.php/transint/article/view/565 (Accessed 16 May 2020).

Gaber, M., Corpas Pastor, G. and Ahmed, O. (2020) Speech-to-Text technology as a documentation tool for interpreters: A new approach to compiling an ad hoc corpus and extracting terminology from video-recorded speeches, *Trans. Revista de Traductología.*

Goldsmith, J. (2020) Terminology extraction tools for interpreters, In *Interdependence and innovation in translation, interpreting and specialized communication*, Ahrens, B. (ed.), Frank & Timme: Verlag für wissenschaftliche Literatur, pp. 279–302.

Goldsmith, J. (2017) The Interpreter's Toolkit: Intragloss - a useful glossary-building tool for interpreters, [online] Available from: https://aiic.net/page/7886/the-interpreter-s- toolkit-intragloss-a-useful-glossary-tool/lang/1 (Accessed 2 April 2020).

Kurz, I. (1997) Getting the message across — Simultaneous interpreting for the media, In *Translation as Intercultural Communication*, Snell-Hornby, M., Jettmarová, Z., and Kaindl, K. (eds.), John Benjamins Publishing Company, pp. 195–205.

Kurz, I. (1990) Overcoming Language Barriers in European Television, In *Interpreting: Yesterday, Today, and Tomorrow*, Bowen, D. and Bowen, M. (eds.), Amsterdam, John Benjamins Publishing Company, pp. 168–175.

Mack, G. (2002) New Perspectives and Challenges for Interpretation. The Example of Television, In *Interpreting in the 21st Century: Challenges and opportunities*, Garzone, G. and Viezzi, M. (eds.), John Benjamins Publishing Company, pp. 203–214.

Pöchhacker, F. (2010) Media Interpreting, In *Handbook of Translation Studies*, Gambier, Y. and van Doorslaer, L. (eds.), John Benjamins Publishing Company, pp. 224–226.

Ruetten, A. (2017) Terminology Management Tools for Conference Interpreters – Current Tools and How They Address the Specific Needs of Interpreters, In *Proceedings of the 39th Conference Translating and the Computer*, London, pp. 98–103.

Ruetten, A. (2003) Computer-based information management for conference interpreters, or How will I make my computer act like an infallible information butler?, In London, Aslib.

Terminotix inc. (n.d.) Synchroterm 2019 Version User Guide, Terminotix.

Xu, R. (2018) Corpus-based terminological preparation for simultaneous interpreting, *Interpreting*, **20**(1), pp. 29–58.

Language Technology for Interpreters: The Vip Project

Gloria Corpas Pastor

University of Malaga (Spain)

University of Wolverhampton (United Kingdom)

ABSTRACT

Recent years have witnessed a tremendous interest in language technologies and digital resources for interpreters (Braun, 2019; Drechsel, 2019). Nowadays there is a pressing need to develop interpreting technologies, with practitioners increasingly calling for tools tailored to their needs and their new work environments. However, technology growth in the profession still appears rather limited and slow-paced, despite some evidence that the profession is heading towards a technological turn (Fantinuoli, 2018). While language technologies have already had a profound transformative effect in translation, they have not yet led to a paradigm shift to the interpreters' "digital workplace". Although interpreting has not yet benefited from technology as much as its sister field, translation, interest in developing tailor-made solutions for interpreters has risen sharply in recent years. With the advent of new technology, interpreters can work remotely, deliver interpreting in different modes (consecutive, simultaneous, liaison, etc.) and contexts (conferences, courts, hospitals, etc.), on many devices (phones, tablets, laptops, etc.), and even manage bookings and invoice clients with ease. But, unlike translation, interpreting as a human activity has resisted complete automation for various reasons, such as fear, unawareness, communication complexities, lack of tools tailored to interpreters' needs, etc. (Mellinger and Hanson, 2018).

Several attempts to meet interpreters' needs have been developed, mainly computer-assisted interpreting (CAI) tools and computer-assisted interpreting training (CAIT) tools but they are rather modest in terms of the support they provide (Wang and Wang, 2019). Nowadays, CAI tools basically encompass terminology management tools, corpora and note-taking applications (for an overview, see Corpas Pastor, 2018; Fantinuoli, 2017; Rütten, 2017; Xu, 2018; and Braun, 2019). There are almost no terminology tools to assist interpreters during interpretation or in the follow-up of interpreting assignments, nor can they be fully integrated in the interpreter's workflow. There is a severe lack of purpose-built tools that fulfill interpreters' needs and requirements. State-of-the-art tools suffer from further limitations in terms of platform-dependency, cross-platform access problems, integration and interoperability issues, low precision and recall, low degree of automation, lack of multiple format exchange, absence of robust cross-lingual NLP methodology and speech technology, among other problems.

This paper will present the results of a R&D project (VIP: Voice-text integrated system for interpreters) on language technologies applied to interpreting. Interpreters need to be equipped with tools which support new functionalities that can provide assistance during all phases of the interpretation process (both onsite and remote), including self-assessment and training. The VIP platform provides access to a wide range of tools and resources to assist interpreters in the preparation phase, during a given interpreting job and after the assignment (for training, life-long learning and follow-up purposes). VIP integrates terminology tools, corpora building and processing, automatic glossary building, automatic speech recognition and quality assessment applications, etc. VIP is freely accessible for researchers and practitioners.

The paper will be structured as follows. The first section will provide an overview of existing tools and resources for interpreters. The second section will describe the VIP tool, an environment designed to assist interpreters during the entire process (preparation phase, interpreting job and follow-up).

Introduction

Evidence of technological change, led by advances in digital technologies, is all around us and the field of interpreting is no exception. Recent years have witnessed a tremendous interest in language technologies and digital resources for interpreters (Braun, 2019; Drechsel, 2019). Nowadays there is a pressing need to develop interpreting technologies, with practitioners increasingly calling for tools tailored to their needs and their new work environments. However, technology growth in the profession still appears rather limited and slow-paced, despite some evidence that the profession is heading towards a technological turn (Fantinuoli, 2018). While language technologies have already had a profound transformative effect in translation, they have not yet led to a paradigm shift to the interpreters' "digital workplace".

Although interpreting has not yet benefited from technology as much as its sister field, translation, interest in developing tailor-made solutions for interpreters has risen sharply in recent years. With the advent of new technology, interpreters can work remotely, deliver interpreting in different modes (consecutive, simultaneous, liaison, etc.) and contexts (conferences, courts, hospitals, etc.), on many devices (phones, tablets, laptops, etc.), and even manage bookings and invoice clients with ease. But, unlike translation, interpreting as a human activity has resisted complete automation for various reasons, such as fear, unawareness, communication complexities, lack of tools tailored to interpreters' needs, etc. (Mellinger and Hanson, 2018).

Several attempts to meet interpreters' needs have been developed, mainly computer-assisted interpreting (CAI) tools and computer-assisted interpreting training (CAIT) tools but they are rather modest in terms of the support they provide (Wang and Wang, 2019). Nowadays, CAI tools basically encompass terminology management tools, corpora and note-taking applications (for an overview, see Corpas Pastor, 2018; Fantinuoli, 2017; Rütten, 2017; Xu, 2018; and Braun, 2019). There are almost no terminology tools to assist interpreters during interpretation or in the follow-up of interpreting assignments, nor can they be fully integrated in the interpreter's workflow. There is a severe lack of purpose-built tools that fulfill interpreters' needs and requirements. State-of-the-art tools suffer from further limitations in terms of platform-dependency, cross-platform access problems, integration and interoperability issues, low precision and recall, low degree of automation, lack of multiple format exchange, absence of robust cross-lingual NLP methodology and speech technology, among other problems.

This paper will present the results of VIP, a R&D project on language technologies applied to interpreting. Interpreters need to be equipped with tools which support new functionalities assisting during all phases of the interpretation process (both onsite and remote), including self-assessment and training. The VIP platform provides access to a wide range of tools and resources to aid interpreters in the preparation phase, during a given interpreting job and after the assignment (for training, life-long learning and follow-up purposes). VIP integrates terminology tools, corpora building and processing, automatic glossary building, automatic speech recognition and quality assessment applications, etc. VIP is freely accessible for researchers and practitioners.

The paper will be structured as follows. As a convenient background, the first section provides a general discussion on the rise of interpreting-related technologies in recent years. The second section is devoted to the VIP tool, an environment designed to assist interpreters during the entire process (preparation phase, interpreting job and follow-up).

Overview

Despite the actual interdependence between technology and the human skills in interpreting (Jekat, 2015), interpreting technologies are reportedly scarce (Costa, Corpas Pastor and Durán Muñoz, 2014) and have entered the profession only in recent years (Fantinuoli, 2018). Some studies suggest that interpreters are still largely unaware of them or even reluctant to use them (Corpas Pastor and Fern, 2016). Most current technological advances in interpreting differ so much from interpreters' work practice that they are perceived as irrelevant or useless to the interpretation task. Major concerns are the loss of quality and the dehumanisation of interpreting that allegedly tend to accompany technological developments (Jourdenais and Mikkelson, 2015). However, there is a growing interest for language technologies and digital resources in the field of interpreting. For instance, in the 8th AIIC Interpreters for Interpreters Workshop (Bonn, 15 September 2017) there were some papers on collaborative terminology management systems and new software for preparing for and follow-up of interpreting assignments[1]. Similarly, the 2017 edition of the well-established Conference Translating and the Computer (TC39), organised by AsLing in London, had a special session with a strong focus on technology tools for interpreters. In the Panel discussion on New Frontiers in Interpreting Technology[2], active interpreters of international renown emphasised the need to develop new and improved tools and resources for interpreters. Since then, interest in computer-assisted interpreting tools (or interpreting technologies) has been steadily increasing among practitioners and academia. For instance, the 2018 edition (TC40)[3] featured a Round Table on Language Technologies for Interpreters and a Workshop on Interpreters' Help; and the 2019 edition (TC41)[4] included a session on interpreterQ Media Player for interpreter training. Papers and workshops on interpreting technologies have also grown in number in the latest 2020 edition (TC42)[5], with three full papers on (i) computer-assisted language technologies and resources for interpreting, (ii) automatic speech recognition and machine translation of interpretated discourses, and (iii) remote interpreting, as well as one short paper on intermodal parallel corpora to investigate processes of interpreting and translation, and two workshops (practical introductions to VIP system for interpreters and to Televic hybrid translation and interpreting teaching platform) .

Language technologies have already changed the shape of interpreting services, interpreting training and interpreting research. This has a significant impact on the theoretical and empirical foundations of interpreting, the multifaceted cognitive processes underlying the various interpreting modes and tasks. Technology changes and developments have also paved the way for profound transformations in the discipline, and the academic debate is starting to address these changes, their implications and the challenges that lie ahead. Several research gaps have been already identified, such as insufficient empirical research on the impact of computer-assisted interpreting (CAI) tools and resources, limited range of CAI tools, lack of technology contents in interpreting curricula, lack of quality-driven and user-driven studies of interpreting technologies, etc. (Costa, Corpas Pastor and Duran Muñoz, 2018; Goldsmith, 2017), among others.

Interpreters' attitude to technology has also evolved in recent years. By contrast to the findings in the exploratory study conducted in 2014 by Corpas Pastor and Fern (2016), technology is no longer viewed with so much suspicion or reluctancy. New generations of

[1] https://aiic.de/event/8-dolmetscher-fuer-dolmetscher-workshop. (Last accessed: 05-01-2020).
[2] https://www.asling.org/tc39/?page_id=955 (Last accessed: 28-12-2019).
[3] https://www.asling.org/tc40/ (Last accessed: 20-11-2020).
[4] https://www.asling.org/tc41/?page_id=1094 (Last accessed: 19-11-2020).
[5] https://asling.org/tc42/?page_id=3326 (Last accessed: 10-11-2020).

interpreters (whether digital natives or digital immigrants) seem to be ready to embrace technology. For instance, in a later survey conducted in 2017 within the framework of the VIP project, interpreters showed a positive attitude towards technology in general (Corpas Pastor, 2018). For instance, to the question "To what extent do you think that technology tools would be of use to interpreters?" (rating 0 to 5 with 5 being the highest score)", 42.11% of interpreters rated 5, followed in increasing order by 4 (15.79%), 3 (13.16%) and 2 (5.26%). And to the question "Do you think that the use of technology tools can impact the quality of interpreting? respondents rated mostly 5 (31.58%), followed by 4 (21.05%) and 3 (13.16%).

Similar results were obtained from audience interaction[6] at my own TC42 presentations. 60% of all participants were either student interpreters (40%) or interpreters (20%). To the first question ("To what extent do you think that technology tools would be of use to interpreters?"), attendants rated 5 (44%), 4 (37%), 3 (11%) and 1 (7%). Most of them considered themselves "a tech-savvy individual ready to try new language technology" (70%), and to a lesser degree "somewhere in between both options" (26%); only 4% declared not to feel very confident when it comes to try new technology. When asked to list technologies that could be used for interpreting, the majority mentioned terminology-related technology and automatic speech recognition (ASR), followed by machine translation, summarisation, corpora, artificial intelligence (AI) applications and devices for note-taking, microphones, videoconference platforms. Figure 1 illustrates the results of the live wordcloud poll.

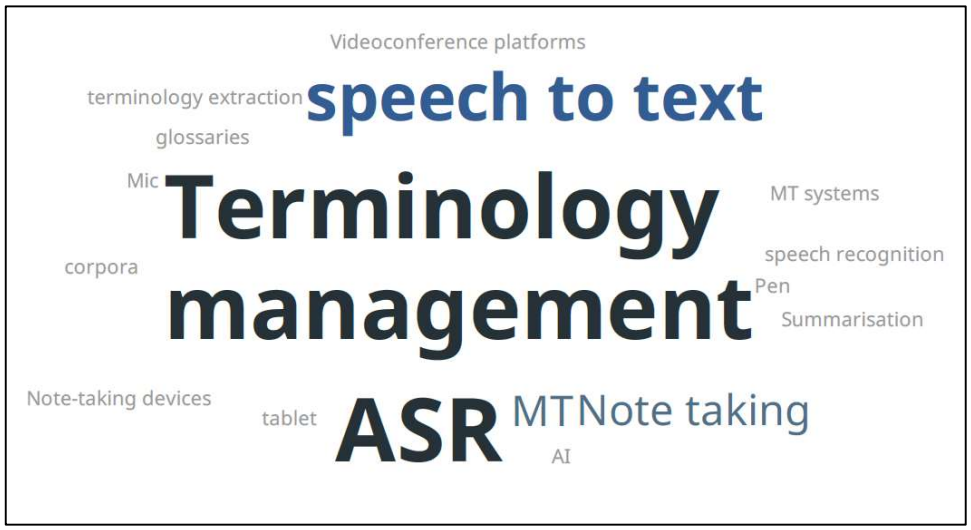


Fig. 1. *Wordcloud poll on interpreting-related technologies*

Finally, technology has already had a dramatic impact on the remote modalities of interpreting. In the words of Gaber and Corpas Pastor (2020: 58): "With the advent of new technology, interpreters can work remotely, deliver interpreting in different modes (consecutive, simultaneous, liaison, etc.) and contexts (conferences, courts, hospitals, etc.), on many devices (phones, tablets, laptops, etc.), and even manage bookings and invoice clients with ease." From a survey conducted in 2019-2020 on the use of remote technology in public services interpreting, the authors report interpreters' low degree of tech-savviness (cf. Kerremans et al., 2019), and yet public service interpreters seem to be fully aware of the

[6] For audience interaction, live polls were organised through Slido (https://www.sli.do/).

practical benefits of using remote technology (economic or business-like advantages) and point out the need to integrate more technology contents into interpreter training courses.

Lack of space leaves little room to provide a more detailed overview. Suffice to mention seminal contributions by Fantinuoli (2018a, 2018b), Mellinger and Hanson (2018), Braun (2019) or Drechsel (2019), and the papers in the volume edited by Rodríguez Melchor, Horváth and Ferguson (2020).

Voice-text integrated system for interpreters

The acronym VIP stands for "**V**oice-text **I**ntegrated system for inter**P**reters". The VIP[7] system encompasses a suite of tools to assist interpretation at all phases, plus an open catalogue of interpreting-related technologies (tools and resources). It is composed of two main components: a catalogue and a multifunctional platform. Figure 2 provides a diagram that illustrates the conceptual representation of the components and subcomponents of the VIP system.

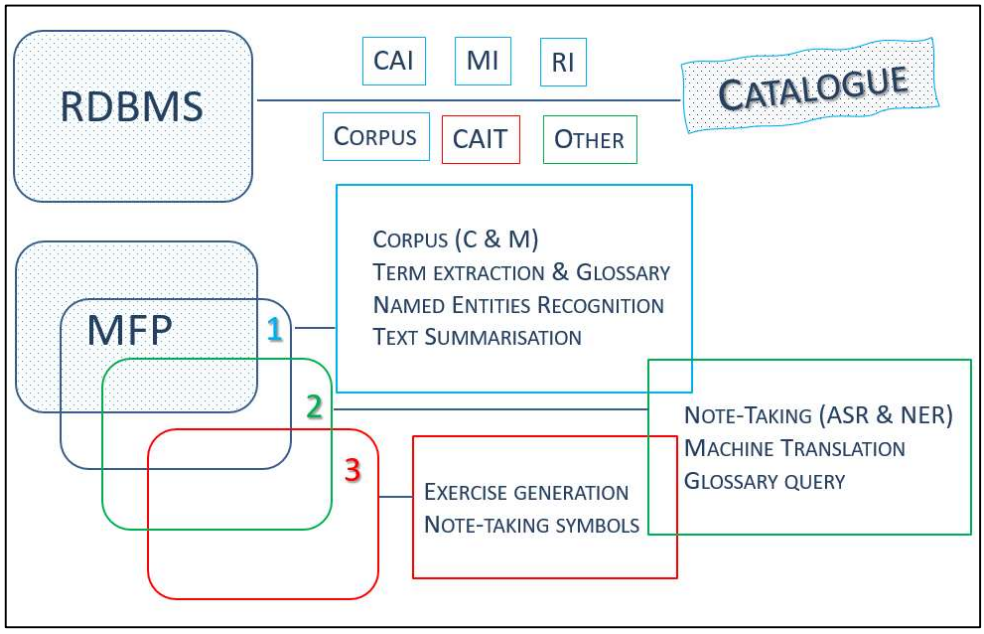


Fig. 2. *VIP system architecture*

Component A of the system is a relational database management system (RDBMS) that uses the SQL (Structured Query Language) to access the catalogue database. The VIP relational database contains an extensive catalogue of tools and resources aimed at helping interpreters and student interpreters to perform their tasks and to train their skills. The catalogue is based on the classification of technology tools and resources advanced by Costa, Corpas Pastor and Durán Muñoz (2014), and later refined by Corpas Pastor (2018) and Fantinuoli (2018a), among others. The VIP catalogue includes:

- computer-assisted (CAI) tools and resources, i.e., (language) technology intended to enhance interpreter's performance,
- remote interpreting (RI), i.e. technology aimed at delivering an interpretation remotely (mainly cloud-based),
- machine interpreting (MI), i.e., technology intended to replace human interpreters,

[7] http://www.lexytrad.es/VIP/index_en.php. (Last accessed: 05-01-2020).

- computer-assisted interpreting training (CAIT) tools, that encompass various oral or multimedia resources and e-training platforms intended for teaching or life-long learning, and
- other resources that can aid the interpreting process, with special focus on corpora.

This is an open and collaborative catalogue, as users can suggest more tools or resources and report any issues. The VIP database can be searched by individual categories and subcategories. For more refined searches, keywords can also be entered to locate tools and resources with specific features that appear in their description field. All tools and resources are first classified by their main category and subcategories. Then, they are further characterised through a general feature-based template: operating system and/or platform (Windows, Linux, MacOS, Android, iOS, Web, Java), languages (interface language, languages supported), and type of licence available (free, commercial, demo). Users can also click on a particular tool or resource to get more detailed information (short description, URL, screenshot, etc.).

CAI tools are divided into three categories (terminology management, note-taking and speech-to-text). The first category of CAI tools includes terminology management tools designed for and intended to be used by interpreters (Intragloss, InterpretBank, Interplex UE, Interpreters' Help, Flashterm, etc.). Besides the basic features mentioned above, specific information is also provided for each tool (documentation, available, export and import formats, author, further information). For instance, information provided for Interpreters' Help includes operating system/platform (Windows, MacOS, iOS, Web), interface language (English), language supported (multi-language), type of licence (free), documentation (Help & Documentation, Community), file export format (Word, Excel), URL (https://interpretershelp.com), authors (Yann Plancqueel and Benoît Werner), a screenshot and a short description: "*Interpreters' Help is a collaborative glossary management tool which supports several formats and is available for different platforms. This tool is designed to prepare glossaries before interpreting, allowing the access and collaboration between colleagues*."

Note-taking applications have not been specifically designed with interpreters in mind, but they are increasingly being used in digital consecutive interpreting and in hybrid interpreting modalities (SimConsec and SightConsec). They are further divided into standalone software and smart pens. Evernote is an example of note-taking multi-language software that runs in Windows, MacOS, iOS and Android. Authored by Evernote Corporation, exports PDF files and imports PDF, TXT and DOCX imports, and it is it is well-documented (Help & Learning, Support, Forum). Some examples of digital smartpens are Equil JOT, Livescribe3, Neo N2, etc.

Speech-to-text applications (also termed S2T and ASR, automatic speech recognition) are currently being used as a central component of CAI tools, either bundled or standalone (cf. Gaber, Corpas Pastor and Omer, 2020). Only S2T standalone applications are included in VIP (e.g., Voice Dictation, Dragon NaturallySpeaking, etc.). They are classified by means of the usual features and a short description is also provide. For instance, Voice Dictation (iOS), is a multi-language commercial application authored by Quanticapps.

Remote technology is represented in the catalogue by cloud-based interpreting systems (RI). Unlike telephone- and video-mediated interpreting, cloud-based RI usually involves two main components: (a) the Interpretation Management System, designed to schedule and manage interpreting assignments, and (b) the Interpretation Delivery Platform, designed to support the delivery of the interpretation. Some well-known examples are Headvox, Kudo, Linguali or Interprefy, among others.

The VIP catalogue also features CAIT (computer-assisted interpreting training) tools, that encompass various training materials (oral resources, digitised interpretations, videos, transcribed speeches, portals, research projects, institutional multimedia repositories, etc.), and virtual training platforms (e.g., IVY, Virtual Interpreting Environment or Melissi VS).

Machine interpreting cannot be considered interpreting-related technology proper. In fact, it involves several complex language technologies in a three-phase cascade: (a) speech-to-text conversion (automatic transcription), (b) machine translation (interlingual translation), and (c) text-to-speech synthesis (conversion of translated text into speech). Recently, artificial intelligence is beginning to explore the possibility to skip the second phase and deliver speech-to-speech translation. The VIP catalogue includes MI systems because they are currently used in real interactions between humans (this is the case of Ambassador or Skype Translator). Besides, MI is expected to have a significant impact on professional interpreters' work environment (akin to MT in translation).

A broad category of miscellaneous resources that can aid interpreters are also included. They encompass terminology management tools used by translators that could be useful for interpreters (e.g., Terminus, SDL Multiterm, TermSuite), units converters (e.g., ConvertUnits, Converto, Units), and other relevant speech technologies, like voice recording (e.g., Voice Pro, Audacity, iTalk Recorder). Corpus tools appear as a separate category in the catalogue due to their increasing importance for the preparation phase of an interpreting assignment and most specifically for glossary building. They encompass tools for building, tagging, parsing and managing comparable corpora (Bootcat, Corpógrafo, SketchEngine, FreeLing, AntConc, etc.) and multi-lingual parallel corpora (Bitextor, Strand, ParaConc, ParaVoz, AntPConc, etc.).

The VIP database allows for advanced searches. General advanced searches query the database through the categories and subcategories and the basic feature template (platform, languages, license available) and keywords. For instance, a general search [comparable corpus management + Windows + Spanish, English, German + Free] retrieves 16 results (MonoConc Easy, UAM Corpus Tool, TextStat, Corpus Presenter, MonoConc, etc.); whereas a general search [Speech-to-text + Windows + English, Spanish + Text (*punctuation*)] retrieves only one (Voice Dictation).

Specific advanced searchers can be performed for Terminology management systems and Cloud-based interpreting systems. In the first case, search criteria are further expanded to encompass relevant specific features, namely, multiple glossary management, document handling, Unicode compatibility, embedded online search for translation candidates, remote glossary exchange and documentation. Specific tools can be retrieved and compared according to the features selected. Thus, the search [manages multiple glossaries, handles documents, embedded online search for translation candidates, remote glossary exchange, Windows + Spanish, English, Italian] retrieves five results, which can be then compared as regards the whole set of specific extended features (see Figure 3).

	InterpretBank	SDL Multiterm Desktop	AnyLexic	Lingo	SDL Multiterm Extract
Manages multiple glossaries	✔	✔	✔	✔	✔
Handles documents	✖	✖	✖	✖	✖
Unicode compatibility	✔	✔	✔	✔	✔
Embedded online search for translation candidates	✔	✖	✖	✖	✖
Remote glossary exchange	✖	✔	✔	✖	✔
Well documented	✔	✔	✔	✔	✔

Fig. 3. *Advanced searches (Terminology management systems)*

In the second case (cloud-based RI systems), features have been added according to users: conference organisers (suitable for silent meetings, wifi supplied by VIT company), individual clients (recordings, speech recognition), and language service providers or LSPs (schedule interpreter in advance, schedule interpreter for onsite, interpreters on demand, interpreter availability (live), automated scheduling). Other features relate to branding (incorporate brand logo, incorporate brand guidelines) and to the interpretation management system component (location, invoice generation, payments, speed test, built-in CRM to manage customers, interpreter employment management subject matter, pre-call data). For illustrative purposes, a search with [schedule interpreter in advance + automated scheduling + multi-language] retrieves five results (Headvox, Interprefy, Kudo, Linguali and Boostlingo), and the search [recordings + speech recognition + multi-language] retrieve only four systems (Headvox, Kudo, Linguali and Boostlingo). Both searches allow comparison of those RI systems according to all general and specific features, similarly to what is illustrated in Fig. 2.

Component B is a multifunctional platform that integrates various functionalities and technologies in three different modules (see Fig. 2. VIP system architecture). Module I has been designed to be used in the preparation for an interpreting assignment. It comprises four main functionalities: (i) corpus management, (ii) glossary management, (iii) named entity recognition (NER), and (iv) automatic text summarisation. Figure 4 illustrate automatic corpus compilation about Covid-19 in English with the following seed words: *coronavirus*, *Covid-19*, *vaccine*. The web-crawled corpus contains 100 documents and 255,073 tokens.

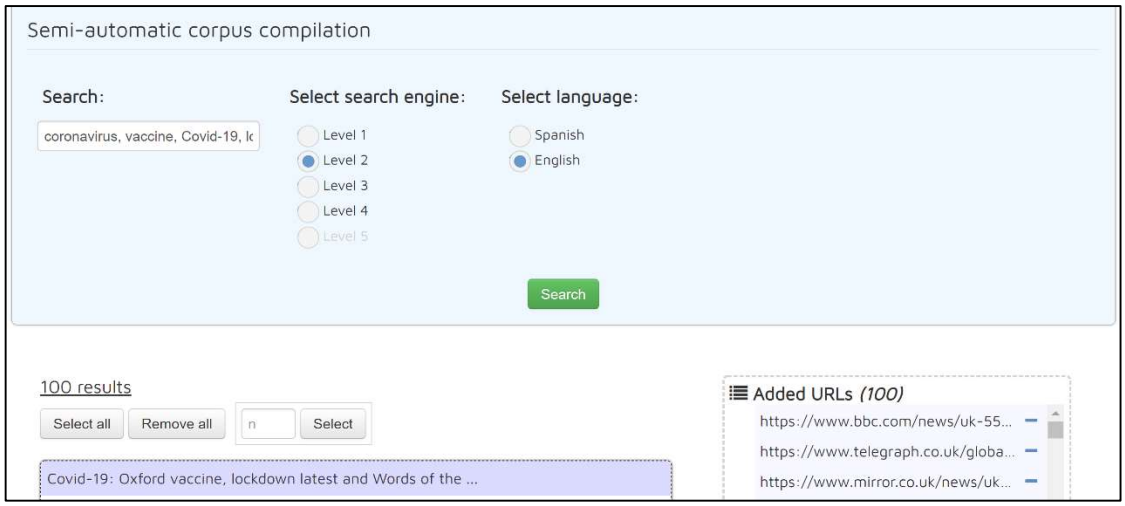


Fig. 4. *Automatic corpus compilation*

Corpus management offers different functionalities related to corpora: automatic and user-assisted corpus compilation (webcrawling), uploading of corpora, and corpus query (concordances, right/left sorted KWIC, n-grams, patterns, candidate terms). Glossaries can be created from corpora or manually compiled. Dictionary and glossary management allows users to create, upload and delete glossaries, perform external searches to locate translation equivalents or, else, translate terms automatically by using machine translation and post-editing. Figure 5 illustrates automatic bilingual glossary creation of multiword terms and postediting through external searches.

Fig. 5. *Automatic bilingual glossary creation*

NER extracts named entities (NEs) by pasting a text or uploading a corpus. NEs are then retrieved according to a set of predefined categories: location (LOC), person (PER), organisation (ORG), etc., and highlighted within the text or, else, as tables. They can be also added automatically to a given glossary.

Text summarisation allows users to produce a domain survey on any topic automatically, either by uploading texts or by selecting several documents. The domain survey can be also downloaded as a corpus and managed as such. This option is particularly relevant to extract key terms for a particular topic. Key terms can be then used as seed terms for automatic corpus compilation or added automatically to glossaries, etc.

Module II is intended to be used when delivering an interpreting job. It includes: (i) automatic note-taking, (ii) machine translation and (iii) glossary query. Automatic note taking incorporates speech recognition and automatic transcription. The system detects NEs and numbers, including physical magnitudes (e.g., 25 tons) and common nouns (e.g. 2 rockets). Glossaries created in Module I can be searched with Glossary query. This functionality provides instant access to terms, NEs and multiword expressions (MWEs), and to their translation equivalents (either by typing the first three letters or orally through the microphone). Machine translation is provided in case an instant draft equivalent is needed on the spot. VIP includes Translate Shell and VIP translator, a neural experimental system. While automatic note-taking could be easily used in sight translation and consecutive interpretation, in simultaneous modalities it would be more convenient for the interpreter's booth mate.

Module III has been primarily designed for training of student interpreters or for life-long learning purposes. It includes (i) a training module with exercises that are automatically generated, and (b) symbols for practising note-taking (experimental). Anticipation exercises enable users to practice terminology and phraseology from a selected corpus. They are basically fill-in-the gap exercises where the user can select the number of exercises, the corpus and even the type of gap: just anything, only terms from the glossaries selected by the user, or even patterns with the blank space in the position indicated by the user (this is particularly useful for collocations and MWEs). Exercises with numbers are also customisable as regards range (e.g., from 1 to 1000), decimals, and language (Spanish, American English or British English). Based on ASR, this type of exercises enables users to practice listening/reading random numbers and then typing or saying the answers. The system indicates mistakes and provides the correct answer. Sight-translation exercises also make use of ASR techniques. See screenshot in Fig. 6.

Fig. 6. *Exercises for sight translation*

Users can select the number of exercises, the language configuration (including Spanish and English) and the specific corpus they want to use (EuroParl, GlobalVoices, EU Bookshop, Multi United Nations, TildeMODEL and Wikipedia). The system combines parallel corpora

management and ASR. The user is presented with a fragment of the subcorpus 1 (in the source language selected, for instance, Spanish), and produces the sight translation orally in the target language variety (for instance, British English). Then, the system recognises the speech automatically (British variety), transcribes the user's spoken utterances automatically and checks the accuracy of his/her output against the aligned subcorpus 2 (in the target language, British English in this case). The accuracy rate is approximate as it compares user's output with the actual bitext in English (synonyms, term variations, syntactic transformations or paraphrases are considered errors by the system, similarly to standard translation memory systems). In addition, some errors could be due to the ASR system in place.

The fourth type of exercises in this category are intended to practice terms, multiword terms, and multiword expressions. Glossary exercises allow selection of the glossaries, language configuration and number of exercises desired, and enable users practice with their bilingual glossaries. For instance, users are presented with a series of terms or MWEs in the source language (written or spoken) and they provide the corresponding equivalents, either orally or by typing. The system checks and provides correct answers.

Finally, note-taking training exercises combine speech technologies and artificial intelligence for image recognition. The tool displays terms/concepts randomly (spoken, but also written if theoption is selected) for users to draw the corresponding symbol. The image is then processed automatically and a checking bar indicates the percentage of accuracy of the symbol with regard to the displayed concept. If it is correct, the process is restarted by clearing the drawing area and displaying a new term. If incorrect, the tool displays a reference image to guide the user. See a screenshot in Figure 7.

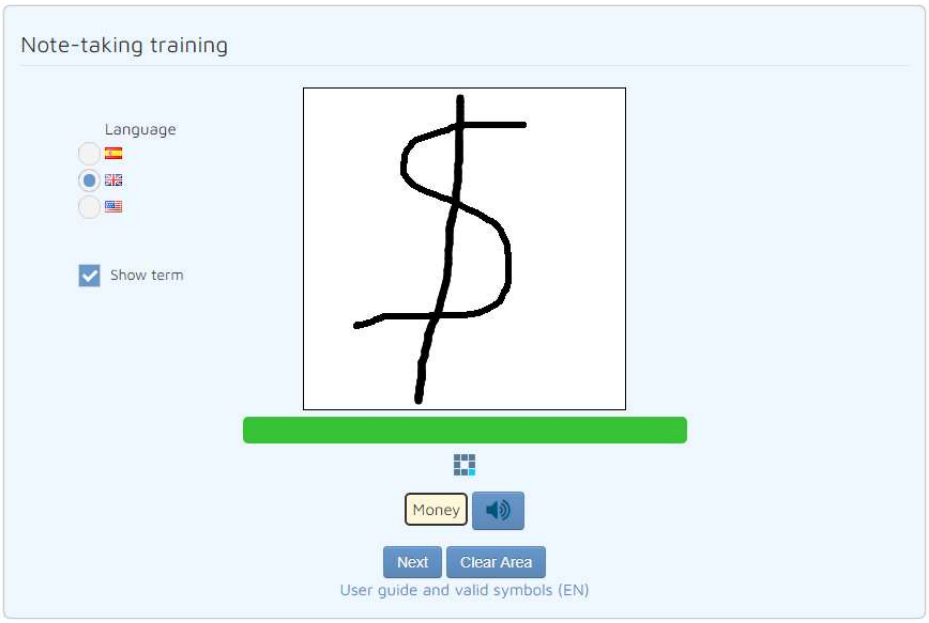


Fig. 7. *Exercises for note-taking (symbols)*

Assessment and conclusion

Complementary to the VIP presentation, a Workshop was organised as part of TC42. In this hands-on session participants could use the tool at ease to prepare for a blind interpretation and then provide feedback. In general, the VIP systems was rated either useful or very useful as a tool to prepare for an interpretation in a 0-5 Lickert scale (3=33%; 4=33%; and 5=22%) and quite intuitive and user-friendly (3=22%; 4=56%, 5=22%). And to the question, "which

exercises do you like most?", participants selected Glossaries (42%), followed by Numbers (33%) and Anticipation exercises (25%).

Then, individual modules and functionalities were also evaluated as regards their usefulness in the preparation phase. The average rating was 4 out of 5. The results obtained are provided below:

- Corpus compilation: 3=20%; 4=40%; 5=40%.
- Corpus management: 4=55%; 5=45%.
- Text summarisation: 3=25%; 4=75%.
- Glossary creation: 3=8%; 4=58%;5= 33%.

Finally, participants were asked to provide feedback in order to improve VIP. The most repeated suggestions mentioned were adding more languages (the present version of VIP only supports English and Spanish), quick glossary search, inclusion of abbreviations, plus some technical issues, like browser compatibility, increased website capacity and future site maintenance.

A preliminary conclusion of this study is that interpreters already see technology as a key asset, that their attitude towards interpreter-related technology has undergone a positive development and that they are willing to use technology in their daily practice. The question now is whether academia and developers are prepared to tap into interpreters' needs and provide them with the appropriate tools and resources.

Funding: This work was supported by the Spanish Ministry of Research (Ref. no. FFI2016-75831-P, 2017-2020; EUIN2017-87746) and the European Regional Development Fund (ERDF) (Ref. no. UMA18-FEDERJA-067).

References

Braun, S. 2019. Technology and interpreting. In O'Hagan, M. (Ed.). *The Routledge Handbook of Translation and Technology*. Routledge.

Costa, H., G. Corpas Pastor and Durán Muñoz, I. 2014. Technology-assisted Interpreting. *MultiLingual Computing* 143, 25(3): 27-32.

Costa, H.; Corpas Pastor, G. and Durán Muñoz, I. 2018. Assessing Terminology Management Systems for Interpreters. In G. Corpas Pastor and I. Durán Muñoz (eds.) *Trends in e-tools and resources for translators and interpreters*. Leiden: Brill. 57-84.

Corpas Pastor, G. 2018. Tools for Interpreters: the Challenges that Lie Ahead. *Current Trends in Translation Teaching and Learning* E, 5. 157-182

Corpas Pastor, G and Fern, L. 2016. *A survey of interpreters' needs and practices related to language technology*. Technical report [FFI2012-38881-MINECO/TI-DT-2016-1]. University of Málaga.

Corpas Pastor, G. and Gaber, M. 2020. Remote interpreting in public service settings: technology, perceptions and practice. *Skase Journal of Translation and Interpretation*. 13 (2). 58-78.

Drechsel, A. 2019. Technology literacy for the interpreter. In Sawyer, D. B., Austermühl, F., and Enríquez Raído*, V.* (Eds.). *The Evolving Curriculum in Interpreter and Translator Education: Stakeholder perspectives and voices*. Amsterdam: John Benjamins. 259-268.

Gaber, M., Corpas Pastor, G. and Omer, A. 2020. Speech-to-Text technology as a documentation tool for interpreters: A new approach to compiling an ad hoc corpus and extracting terminology from video-recorded speeches. *TRANS: Revista de Traductología*, 24. 1-18.

Goldsmith, J. A. 2017. Comparative User Evaluation of Tablets and Tools for Consecutive Interpreters. In *Proceedings of the 39th Conference Translating and the Computer, London, UK, November 16-17, 2017*. 40–50.

Fantinuoli, C. 2018a. Computer-assisted Interpreting: Challenges and Future Perspectives. In G. Corpas Pastor and I. Durán Muñoz (eds.) *Trends in e-tools and resources for translators and interpreters*. Leiden: Brill. 153-174.

Fantinuoli, C. 2018b. Interpreting and technology: the upcoming technological turn. In Fantinuoli, C. (Ed.). *Interpreting and Technology*. Berlin: Language Science Press, 1–12.

Jekat, S. 2015. Machine Interpreting. In F. Pöchhacker (ed.) *Routledge Encyclopedia of Interpreting Studies*. London and New York: Routledge, 239-242.

Jourdenais, R. and Mikkelson, H. 2015. Conclusion. In Mikkelson, H. and Jourdenais, R. (Eds.). *The Routledge Handbook of Interpreting*. London/New York: Routledge. 447-450.

Kerremans, K., Lázaro Gutiérrez, R., Stengers, H., Cox, A. & Rillof, P. 2019. Technology Use by Public Service Interpreters and Translators: The Link Between Frequency of Use and Forms of Prior Training. *FITISPos International Journal* 6(1). 107-122

Mellinger, C. D. and Hanson, T.A. 2018. Interpreter traits and the relationship with technology and visibility. *Translation and Interpreting Studies* 13 (3). 366-392.

Rodríguez Melchor, M. D., Horváth, I. and Ferguson, K. (Eds.). 2020. *The Role of Technology in Conference Interpreter Training*. (New Trends in Translation Studies 31). Berlin, Bern, Brussells, New York, Oxford, Warszawa, Wien: Peter Lang.

Rütten, A. 2017. Terminology Management Tools for Conference Interpreters – Current Tools and How They Address the Specific Needs of Interpreters. *Proceedings of the 39th Conference Translating and the Computer*. 98–103. London, UK, November 16-17, 2017.

Wang, X. and Wang, C. 2019. Can Computer-assisted Interpreting Tools Assist Interpreting? *Transletters: International Journal of Translation and Interpreting*. 2 (1). 109-139.

Xu, R. 2018. Corpus-based terminological preparation for simultaneous interpreting. *Interpreting Research*. 20 (1). 29-58.

Assessing Cross-lingual Word Similarities Using Neural Networks

Rafał Jaworski

XTM International

rjaworski@xtm.cloud

ABSTRACT

One of the classic problems in natural language processing is word-level alignment. An algorithm is given a bilingual corpus – a set of sentence pairs in which the source sentences are paired with their corresponding translations – target sentences. For each of these pairs the algorithm has to decide which words from the source and target sentences are each other's counterparts.

Basic applications

The problem of automatic word-level alignment was tackled primarily by statistical algorithms (see IBM models). This approach to word alignment was well suited for its main purpose This approach to word alignment was well suited for its main purpose (1) – statistical machine translation (SMT). The standard workflow for SMT involved a training phase during which, among others, the word alignments were computed. By assumption, the training phase was time consuming and so were the statistical word alignment algorithms. However, the information about word alignments between a sentence and its translation can be used in several other applications. In our scenario it is CAT (Computer-Assisted Translation) tools which can use this information in order to perform tasks of automatic correction of text.

Main challenges

In order to enable automatic correction tasks, it is necessary to predict word alignments for a given sentence pair on the fly, i.e. without the lengthy training phase. All automatic corrections in a CAT tool must not take more than a few seconds (ideally – no more than one second), otherwise such mechanisms would not be able to speed up the translation process. The standard statistical word alignment computation must be run on the whole corpus to ensure the quality of results. This process, however, can take hours.

The standard statistical word alignment computation must be run on the whole corpus to ensure the quality of results. This process, however, can take hours.

Apart from speed, the quality of alignments is another challenge in our CAT scenario. The IBM statistical models, even when trained on very large corpora, exhibit observable shortcomings. These shortcomings were not disruptive in the machine translation training but are not acceptable in automatic text correction where the highest quality of word alignments is desired. The third challenge is multilingualism – we require the word alignment mechanism to operate on over a hundred languages and all possible language pairs within these languages. Since the statistical approach requires separate training for all translation directions the support for 100 languages would involve 100 times 99 = 9900 separate training operations. Not to mention the necessity of acquiring bilingual corpora for each of these operations.

Our approach

As a step towards automatic prediction of word alignments we introduced a mechanism for predicting the similarity between words in different languages. The input for this mechanism is a pair of words – one in the source language, the other in the target language. The output is a real number between 0 and 1 indicating the probability of whether the target word is the translation of the source word.

The technique of computing interlingual similarity scores between words relies on data provided by Facebook (2) and BabelNet (3). The data is processed using the following procedure.

Firstly, the Facebook vectors are downloaded as text files. These files are dictionaries (one language per file), containing the top 2 000 000 words with their vector representations. Word representations (also called word embeddings) were obtained by Facebook using an auto-encoder neural network on

the CommonCrawl corpus. It was observed that by examining mathematical relations between these vectors, it is possible to find words which are semantically similar (4). This, however, only applies to word similarities within one language. In order to examine similarities between words across languages, an additional operation was needed – alignment of vectors (5) between languages by computing a transformation matrix. The matrix is trained by a bilingual dictionary for which we used licensed resources provided by BabelNet. In order to handle multiple languages we assumed the following procedure: for English, we use the vectors provided by Facebook directly. For all other languages, we first compute the transformation matrix from this language into English using data from BabelNet dictionaries. Thus we are able to support every language direction, for which Facebook provides source and target language vectors. Sample results of similarity calculation between Italian and English:

English word	Italian word	Similarity
cat	gatto	0.696
cat	fatta (female cat)	0.552
cat	giorno (day)	0.164 (as expected, low similarity)
day	giorno	0.692
day	fuoco (fire)	0.193 (as expected, low similarity)
fire	fuoco	0.590

* * *

Introduction

One of the classic problems in natural language processing is word-level alignment. An algorithm is fed a bilingual corpus: a set of sentence pairs in which the source sentences are paired with their corresponding translations or target sentences. For each of these pairs the algorithm has to decide which words in the source and target sentences are each other's counterparts.

The problem of automatic word-level alignment was tackled primarily by statistical algorithms (see IBM models[1]). This approach to word alignment was well suited for its main purpose, statistical machine translation (SMT). The standard workflow for SMT involved a training phase during which the word alignments were computed. Because of the design of the workflow, the training phase was time-consuming and so were the statistical word alignment algorithms.

However, the information about word alignments between a sentence and its translation can be used in several other applications. In our scenario it is CAT (Computer-Assisted Translation) tools which can use this information to perform automatic text correction tasks, among others.

Main challenges

To enable automatic correction tasks, it is necessary to predict word alignments for a given sentence pair on the fly, i.e. without a lengthy training phase. All automatic corrections in a CAT tool must not take more than a few seconds (ideally - no more than one second), otherwise such mechanisms would not speed up the translation process. The standard statistical word alignment computation must be run on the whole corpus to ensure the quality of results. This process can take hours, however.

Apart from speed, the quality of alignments is another challenge in our CAT scenario. The IBM statistical models, even when trained on very large corpora, exhibit shortcomings. These shortcomings were not disruptive in machine translation training but are not acceptable in automatic text correction where the highest quality of word alignments is desired.

The third challenge is multilingualism - we require the word alignment mechanism to operate on over a hundred languages and all possible language pairs within these languages. Since the statistical approach requires separate training for all translation directions the support for 100 languages would involve 100 times 99 = 9900 separate training operations, not to mention the necessity of acquiring bilingual corpora for each of these operations.

Our approach

As a step towards automatic prediction of word alignments we introduced a mechanism for predicting the similarity between words in different languages. It is called the Inter-Language Vector Space. The input for this mechanism is a pair of words: one in the source language, the other in the target language. The output is a real number between 0 and 1 indicating the probability of whether the target word is the translation of the source word.

Inter-Language Vector Space (ILVS) is a technology based on deep learning, neural networks and algebraic algorithms for supervised learning of vector transformations. The construction of the ILVS consists of three phases. The first phase uses large-scale monolingual text corpora for different languages fed to a neural network with a task of predicting the context of a given word. Internally, the neural network computes 300-dimensional vector representations (word embeddings) for all the words in the corpus using its hidden layer of 300 neurons. The second phase consists in training transformation matrices to convert these vectors between languages. The last phase involves building a disk-based index to store the converted word vectors in multiple languages. ILVS makes it possible to compute the similarity of a pair of words between any two of the languages by retrieving vector representations of the words and applying algebraic functions to those vectors.

Similar solutions

Solutions to the problem of assessing cross-lingual word similarities are also found in other projects, each of them assuming a slightly different approach. The “Similarity” open source library[2] developed by the Systran company is based on direct neural network training to compute word similarities across languages. This requires pre-training on word-aligned corpora (which are not available for many languages). Additionally, this approach can leverage part-of-speech taggers (provided that they are available for a given language pair). In the proposed solution, the calculation of similarity requires a considerable amount of data to be loaded into RAM for each language pair.

On the other hand, the company Babylon Health introduced the idea of normalizing vector spaces in their fastText_multilingual[3] project. Here again, the calculation of similarity requires a considerable amount of data to be loaded into RAM. The authors suggest that the vector space normalization step can benefit from bilingual dictionary data, although no data of this kind is used in their experiment.

There are also algorithms for computing similarities between whole text spans. In this case the similarity is calculated not at the word level but between phrases. A statistical approach to this problem is presented in a US patent[4].

Creation of ILVS

Training of word embeddings

The process of training word embeddings is aimed at converting words found in text into their vector representations. Each word is represented by a single 300-dimensional vector of real numbers from the range [-1, 1]. To obtain these representations a shallow 2-layer feed forward neural network is used. Its training objective is the skip-gram model - the network is given a task of predicting words surrounding the input word. To do this the network is presented with a large text corpus containing an abundance of contexts for each distinct word. In our implementation the corpus was extracted from a crawl of the entire explorable Internet, separately for each language.

When the training of the neural network is finished, the network can predict the context of a given input word according to the skip-gram model. In its hidden layer it uses a representation of the input word on multiple neurons. This representation is used to construct the vector representation of the word within the language. *[Figure 1]*

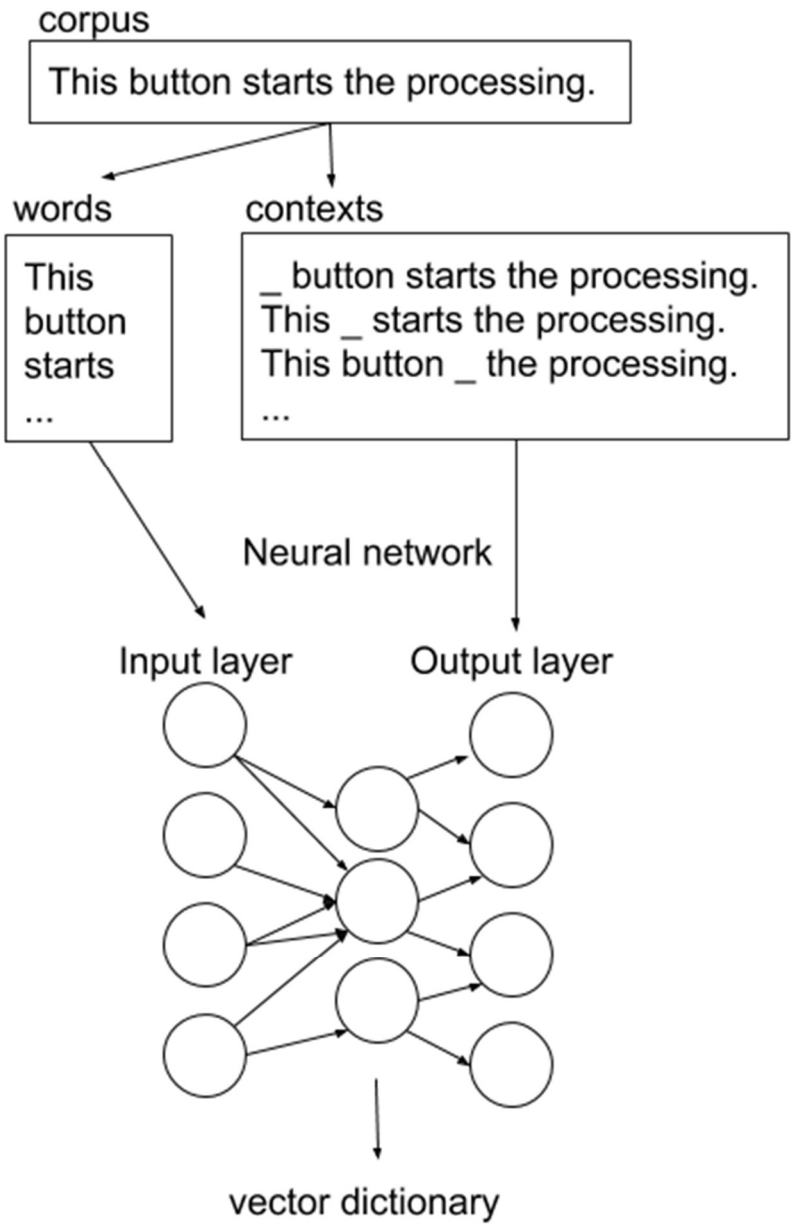


Figure 1. Constructing vector representations of words.

This vector representation makes it possible to assess similarity of words within one language. The calculated cosine similarity between vectors for semantically similar words (such as “street” and “road”) yields high similarity scores. On the other hand, cosine similarity calculated on vectors for distant words (such as “table” and “sun”) yields low scores. This technique therefore models the concept of contextual similarity of words using their vector representation and mathematical vector similarity scores.

The results of this stage are dictionaries mapping distinct words into their 300-dimensional vector representations, one dictionary per language.

Alignment of vector spaces

The vector representation of words obtained in the previous step is used to calculate similarity in only one language. Our solution however makes it possible to assess word similarities across languages. To make this possible, the vector spaces for individual languages are aligned. Without this step, the vector for the word "cat" in the English vector space does not exhibit any similarity to the vector for the word "chat" in the French vector space. This is due to the fact that English and French vector spaces were created in separate training processes using different texts from which no information about correspondence between the English word "cat" to French "chat" could be inferred.

The aim of the vector space alignment operation is to enable the comparison of similar words across languages. After this operation the vector for the English word "cat" is similar (on the basis of cosine similarity) to the French vector for "chat", Spanish vector for "gato", Russian vector for "кошка", etc.

To align the vector space for the language *A* to the vector space for the language *B*, the following procedure is executed. First, a transformation matrix is calculated according to the Singular Value Decomposition (SVD) algorithm. An input for this method is a list of pairs (*a*,*b*), where $a \in A$ and $b \in B$. Points *a* and *b* effectively represent words in language *A* and *B* respectively. It is important that the pairs *(a,b)* represent words that correspond to each other, i.e. their expected inter-language similarity is high. To identify such pairs, a list of homographs found in languages *A* and *B* is used. These homographs typically include proper names, such as "London" or "Paris". Before alignment of the vector spaces, the cosine similarity between the vectors for the word "London" in vector spaces *A* and *B* is typically low. It is expected to increase after the vector space alignment is finished. The SVD learns to convert the vector for the word "London" (and all other examples) so they resemble their respective equivalents in the target vector space. To enhance the quality of this training, the list of homographs is enriched with a list of word equivalents from a digital bilingual dictionary for the language direction *A*-*B*. Note that this step is crucial in ensuring high performance of the Inter-Language Vector Space.

The SVD method is used to compute a 300x300 transformation matrix with the following inputs:

- list of pairs *(a,b)*
- vectors in the vector space *A*
- vectors in the vector space *B*

The matrix is used to align vector spaces *A* and *B* by providing a transformation to be applied to vectors in *A* to make them comparable with vectors in *B*. Although the transformation is trained on the homographs and words appearing in the bilingual dictionary, it can be applied to any word in the source language to make it comparable with target vectors. For this reason, the Inter-Language Vector Space is capable of assessing the probability of a translation of two words even when neither of the words was found in the dictionary, i.e., for words that have never been assigned as translations by human translators. *[Figure 2.]*

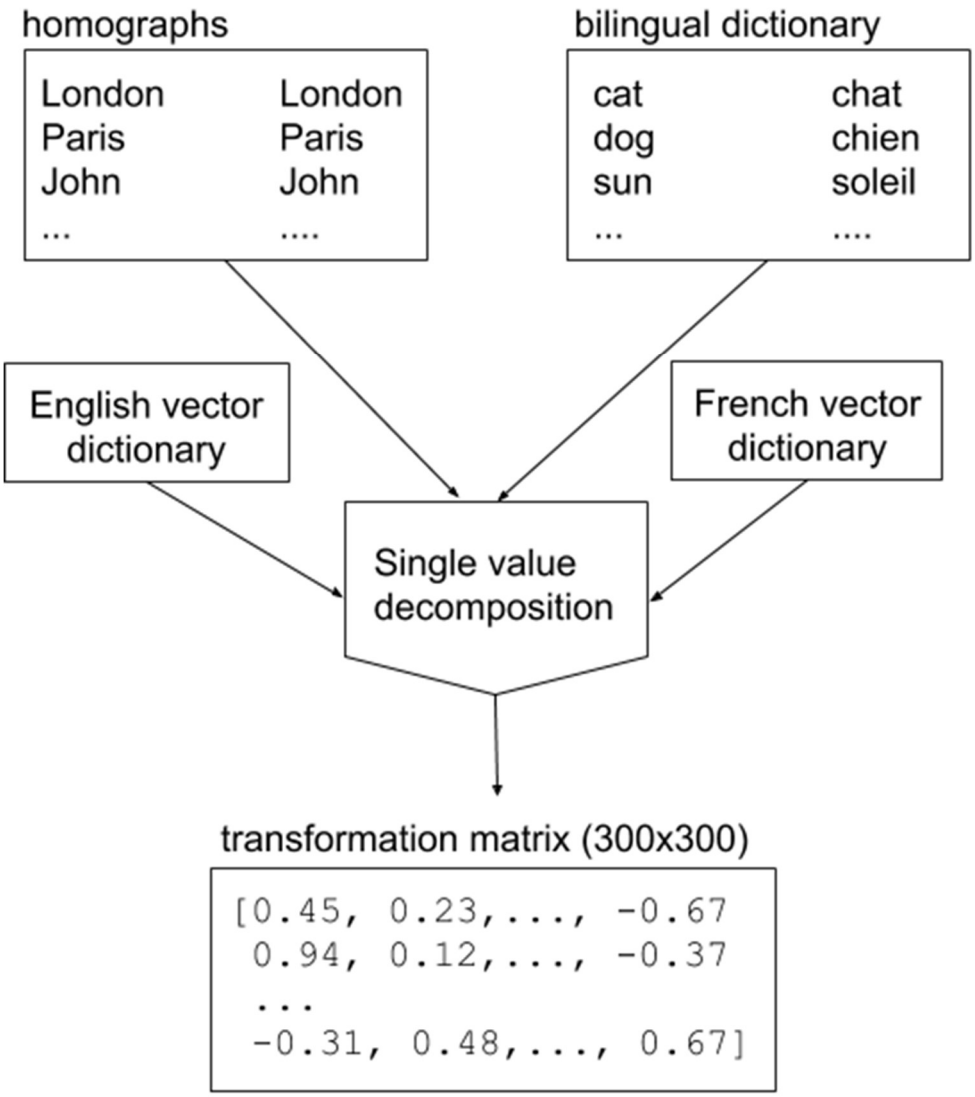


Figure 2. Training of the transformation matrix used to align vector spaces.

The Inter-Language Vector Space currently supports 250 languages. This is to say that word vectors in any two of the supported languages can be compared to each other. Theoretically, this would require performing vector space alignment for all 31125 language pairs which can be composed using the set of 250 languages. Because the step of aligning the vector spaces using SVD takes about 10-15 minutes on a high-performance server machine, alignment in all these directions would be impractical. Hence, the following approach was assumed: the vector space for English became the central, reference vector space. Vector spaces for all other languages are aligned to English. Therefore, adding support for a single new language to the Inter-Language Vector Space makes it possible to compare vectors in this language to any of the languages previously supported. All vectors are compared in the English vector space.

Experiments in several language pairs (e.g. French and Turkish) have demonstrated that the approach of aligning separately French to English and Turkish to English performs even better than aligning French to Turkish directly. This is because the English vector space was trained on the largest amount of resources and therefore is the richest linguistically.

Transformation and indexing of the vector dictionaries

The last step in the Inter-Language Vector Space creation process is the transformation of the vectors using transformation matrices obtained in the previous step. Since the vectors in all vector spaces are 300-dimensional, they can be viewed as matrices of the size [1 x 300]. These are then multiplied by an appropriate transformation matrix, e.g. French vectors are multiplied by the [300 x 300] French to English transformation matrix. As a result, a new [300 x 1] matrix is computed which is viewed as the transformed 300-dimensional vector.

The vector transformation operation is performed on every word in the vector dictionary in a single vector space. This creates a new vector dictionary aligned to the English vector space. This dictionary is then indexed using a disk-based index software.

When calculating the similarity between two words in any of the supported languages, the disk-based vector dictionaries are first queried to obtain the vectors for these languages. Then, as the vector spaces for these vectors were previously aligned, the computed cosine similarity of the vectors obtained reflects the inter-language similarity of the words. *[Figure 3]*

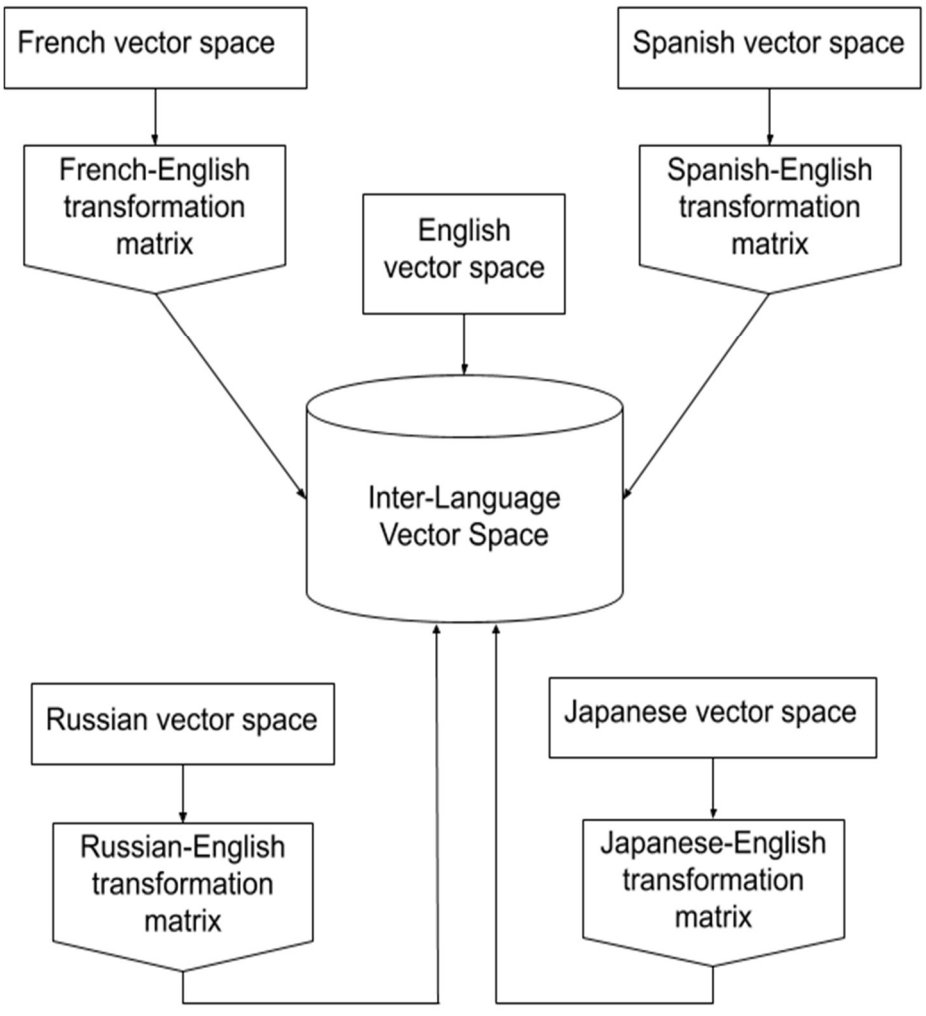


Figure 3. The Inter-Language Vector Space.

Sample results of similarity calculation between Italian and English:

English word	Italian word	Similarity
cat	gatto	0.696
cat	gatta (female cat)	0.552
cat	giorno (day)	0.164 (as expected, low similarity)
day	giorno	0.692
day	fuoco (fire)	0.193 (as expected, low similarity)
fire	fuoco	0.590

Applications

Inter-Language Vector Space can be used to assist in the translation of a source text into a target language. It can reduce the time and effort required in the translation process itself as well as helping to improve the quality and consistency of the translation. It can also assist, in the following translation-related functions:

1. Bilingual corpus alignment

An input to this process is a text in the source language and its translation in the target language. The output is a structured document where source and target sentences (referred to as segments) are aligned to each other. Such a list of segments is then ready to use as a translation memory, or bilingual corpus. The key operation in identifying correspondence between segments in two languages is the assessment of similarity between words in these segments. This can be done with the use of Inter-Language Vector Space.

2. Bilingual terminology extraction

Bilingual terminology extraction automatically creates a list of domain-specific terms with their translation from a translated text. The input for this process is a translation memory or a bilingual corpus. After the terminology is extracted from the source side, the Inter-Language Vector Space helps to identify the translation of each extracted source term on the target side.

3. Automatic placement of non-textual inline element placeholders in a target segment

The automatic inline transfer is a mechanism that handles the copying of elements identified as inline elements (e.g. HTML tags) from a source sentence to a translated target sentence.

The goal of the inline transfer mechanism is to fully automatically copy the Inline elements from the source sentence to the translated target sentence in the correct positions. Thus, the translator need only concentrate on translating the meaningful content of the text. The inline element copying relieves the translator of the technical and time-consuming process of copying non-translatable elements from the source sentence to the translation. To make this possible, the Inter-Language Vector Space is used to identify the corresponding words in the source and target sentences.

4. Automatic assessment of machine translation quality

By providing similarity scores for words in different languages, the Inter-Language Vector Space can be used to develop a score for inter-language sentence similarity. This in turn can be used to compare the output of machine translation to the original source. A low similarity score would indicate unsatisfactory performance of the machine translation, whereas high similarity would indicate high quality of the translation.

5. Highlighting potential translation errors

The Inter-Language Vector Space can also be used to highlight words in the source sentence which do not have similar counterparts in the translation. This situation reveals a potential translation error. The technique of highlighting such errors facilitates the review of the translation.

6. Syntactic analysis of source and target translation

By providing similarity scores for every word pair from the source and target segment, the Inter-Language Vector Space can be used to identify syntactic chunks and their counterparts in the translation. This is based on the fact that contextual similarity captured by the Inter-Language Vector Space is known to model syntactic relationships between words.

7. Semantic analysis of source and target translations

Since contextual similarity captured by the Inter-Language Vector Space also models semantic relationships between words, it is also possible to perform semantic analysis of the source and target translation.

8. Identify synonyms across languages

The Inter-Language Vector Space captures the similarity between words and their synonyms (as being similar to the original word). It is therefore possible to identify synsets in one language and their corresponding synsets in another language.

9. Automatic language detection

Automatic language detection is a technique for predicting the language of a longer text (at least one paragraph) using automatic methods. By providing individual vector spaces for each language, the Inter-Language Vector Space makes it possible to compare words from the input text to all supported vector spaces. The language of the vector space containing the largest number of similar words to those from the input text would be predicted as the language of the input text.

Conclusions and future work

The results obtained in the experiments involving Inter-Language Vector Space have demonstrated its robustness, scalability and accuracy in assessing cross-lingual similarities of

words. Future work plans involve a further increase in the languages covered and experiments with different functionalities derived directly from the ILVS.

We also plan to use similar techniques for phrase similarity assessment. With the help of vector representations it is possible to compute similarity at the phrase level.

References

Peter F. Brown and Stephen A. Della-Pietra and Vincent J. Della-Pietra and Robert L. Mercer (1993): The Mathematics of Statistical Machine Translation, Computational Linguistics

MinhQuang Pham, Josep Crego, Jean Senellart and François Yvon. Fixing Translation Divergences in Parallel Corpora for Neural MT. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018. October 31–November 4. Brussels, Belgium.

Samuel L. Smith, David H. P. Turban, Steven Hamblin and Nils Y. Hammerla. Offline bilingual word vectors, orthogonal transformations and the inverted softmax ICLR 2017

US Patent 8,077,984: "Method for computing similarity between text spans using factored word sequence kernels"

Terminology: Towards a systematic integration of semantics and metadata (2)

Denis Dechandon

Publications Office of the European Union

Maria Recort Ruiz

International Labour Office

ABSTRACT

Aware of the challenges faced when producing, sharing and reusing information, and of the potential of the Internet of Things, the Publications Office of the European Union (PO) and the International Labour Office (ILO) presented at the 41st Translating and the Computer Conference the results of the first phase of a collaborative project designed to support a systematic and harmonised use of metadata.

The initial purpose was to build on various ISO and W3C standards to improve the quality of terminology and semantic assets and to move towards linked data. The ultimate ones were to make huge amounts of discoverable and reusable data available to authors, linguists, colleagues and citizens, and to promote the systematic integration of metadata throughout the entire authoring-translation-publishing (ATP) chain.

As the results achieved were further investigated and went deeper, the project proved to be fourfold. As well as an overall asset enhancement, an easier access to information and a systematic use of metadata, a further impact appeared related to the modernisation of public services through the contribution to the definition and implementation of interoperability frameworks and reference architectures.

This presentation focuses on the results of phase 2 of the project and the plan for phases 3 and 4.

1. Introduction

In a time when technological advancements foster the modernisation of public services and support rapidly growing information exchanges between public administrations, across borders and sectors, the need for interoperability is bigger than ever and can be subdivided into four layers[1] (legal, organisational, semantic and technical). To attain interoperability and avoid the risk of creating new digital barriers for administrations, businesses and citizens, it is necessary for public services to adopt a common terminology, define stable interfaces between them, and be aware of existing solution building blocks that have been developed by others.

The tip of the iceberg for citizens, and the majority of officials and public service staff members, in the form of bits and pieces of information disseminated across the internet and intranets, is linked to their basic expectations, i.e. availability, findability, retrievability and

[1] See New European Interoperability Framework, EIF - European Interoperability Framework | Joinup, https://youtu.be/g-CzHHJ0ZTM

reusability. In turn, this is basically linked to the creation of information, its understandability, portability, management and sharing.

Behind the scenes, the metadata used in the systems built to store, share and disseminate information are described in various ways. They are grouped in lists of keywords, taxonomies, thesauri and other kinds of controlled vocabularies or knowledge organisation systems (KOS), defined to describe a particular subject content and used for subject indexing. The idea of harmonising the metadata used in all domains on a global scale is a pleasant dream, as is that of having a world population with internet access, able to discover information on the web using harmonised lexicalisation of concepts grouped in KOS as mentioned above.

Where such harmonisation does not exist, enormous efforts are made to introduce semantic relations between ("map") concepts in different controlled vocabularies across geographic and linguistic areas. Mapping concepts in major KOS greatly facilitates access to information in these domains.

Linguists, terminologists and professionals in the field of knowledge organisation share similar resources and objectives: to make the information available for users in documents and on the web. The use of vocabularies is paramount. The objectives of this project, that started in 2019 between the PO and the ILO, are to improve the quality of terminology and semantic assets, as well as move towards the systematic use of semantics in the field of translation and build on linked data.

In this paper, we explain the different phases of the project, from the first steps involving the alignment of our vocabularies and the creation of mappings, and their validation, to an analysis of the improvements observed in our respective collections, followed by the new phase where we will start working on the structuration of unstructured data. The results of the first phase and the objectives of the second were presented at the 41st edition of the Translating and the Computer Conference.

2. The project - phase 1 - automatic (lexical) alignment

a. Background

While the use of metadata and reference data is one of the pillars of the semantic web and the definition and use of controlled vocabularies is a well-known practice in the librarian and information management professional communities, metadata is also an element available in most computer-assisted translation and terminology tools used by linguists. Considering the linguistic assets made available to translators and linguists in general, we observed that monolingual, multilingual and aligned corpora, just like terminology collections and translation units are kept in different tools based on diverse technologies, formats and, to some extent, the use of attributes or metadata.

In addition, metadata tends to play an increasing role in tools used by authors, such as XML-editors and in particular in view of the growing implementation of Akoma Ntoso[2]:

Akoma Ntoso ("linked hearts" in the Akan language of West Africa) defines a set of simple technology-neutral electronic representations in XML format of parliamentary, legislative and judiciary documents.

[2] See Akoma Ntoso | Akoma Ntoso Site

The XML schemas of Akoma Ntoso make explicit the structure and semantic components of the digital documents so as to support the creation of high value information services that deliver the power of Information and Communication Technologies (ICTs) and increase efficiency and accountability in parliamentary, legislative and judiciary contexts.

Against similar backgrounds, characterised by impaired access to information due to coexisting technology silos, it is however possible to observe a horizontal component, i.e. metadata. This component makes data more discoverable, reusable and reproducible. Metadata is indeed commonly defined as "data about data". On closer scrutiny, its main characteristic appears to be the ability to predicate data from numerous angles. The type of metadata we consider here, describes concepts with specific meanings and hence links with semantics.

These characteristics lie at the intersection between authors, linguists, publishers and information management specialists, who are all stakeholders in the authoring-translation-publishing chain but not always aware of the need for a unified workflow with transversal features, like standard formats and metadata.

b. When information management specialists meet terminologists

Following the JIAMCATT[3] annual meeting in 2019, the PO and the ILO Information management specialists, alongside the ILO Document Services Coordinator and Terminology Manager, embarked on a controlled vocabulary alignment project with several milestones and objectives. The first phase included the following objectives:

- automatically align (lexically) all ILO vocabularies with EuroVoc,

- align ILO vocabularies with each other, including those differing in scope, language and structure,

- define direct mappings between concepts, terms and category labels within each vocabulary and with those in each other vocabulary.

The aim being to ensure consistency and understanding within and between the organizations, enrich the content of each vocabulary, enable the insertion of mappings and structured terminology assets and implement these enriched vocabularies and mappings to improve website search features.

c. Alignment of datasets and mapping concepts between KOS - preparatory work and analysis

For this test, the EuroVoc[4] and the ILO thesaurus[5], taxonomy[6] and glossary on the gig economy were selected under this framework:

- EuroVoc is a multilingual, multidisciplinary thesaurus covering the activities of the EU. It contains terms in 23 EU languages (Bulgarian, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hungarian, Italian, Latvian, Lithuanian, Maltese, Polish, Portuguese, Romanian, Slovak, Slovenian, Spanish and Swedish), plus in three languages of countries which are candidate for EU accession: Republic of North Macedonia,

[3] See JIAMCATT
[4] See eurovoc - EU Vocabularies - Publications Office of the EU
[5] See ILO Thesaurus
[6] See ILO Taxonomy

Albania and Serbia. EuroVoc is managed by the PO, which moved forward to ontology-based thesaurus management and semantic web technologies conformant to W3C recommendations as well as the latest trends in thesaurus standards. EuroVoc users include the European Union Institutions, the PO, national and regional parliaments in Europe, plus national governments, and private users around the world.

- Developing and maintaining the thesaurus is ensured by the *Groupe Interinstitutionnel Lex - Groupe EuroVoc*, consisting of one or more staff members of any EU institution or body, all of them being librarians, terminologists, or semantic and IT experts. Its content and structure are maintained using VocBench[7] that builds on semantic web standards[8].

- The ILO thesaurus ensures comprehensive coverage of the ILO library resources. It contains over 4000 terms grouped in 19 main subject categories or facets relating to the world of work in English, French and Spanish.

- The ILO taxonomy is used to search the ILO website[9]. Reflecting the ILO work programme, it comprises 400 labour-related terms arranged into 25 subject groups and is well suited for use as a subject classification system to arrange collections in small libraries or labour information centres. Both thesaurus and taxonomy are managed and maintained by the ILO Library team with PoolParty[10].

The glossary on the gig economy used in this project was prepared in the framework of research on non-standard forms of employment. As this is a field in constant evolution, it requires regular updating. The glossary is used by internal ILO staff and most of its terms can be consulted in ILOTERM[11], used as a basis to update the taxonomy.

While EuroVoc was already available in RDF/XML, all ILO's vocabularies were in Excel format and were transformed into RDF/XML using RDF/SKOS as the presentation format.

Once available, the KOS were analysed as RDF files. In short:

- All 3 ILO files contain labels in English, French and Spanish,

- Except for the comments mentioned above, no other issues were identified in terms of content,

- Each KOS contains at least one scheme (3 for the thesaurus) covering the complete list of concepts,

- All KOS were validated in both Prestige and VocBench and no major flaws impeding the alignment process were identified,

- An inconsistent use of alternative labels (skos:altLabels[12]) in all ILO vocabularies led to the conclusion that they were not suitable for alignment purposes,

[7] See VocBench3 | Joinup
[8] See Semantic Web Standards published by the W3C
[9] See International Labour Organization (ILO)
[10] See PoolParty
[11] ILOTERM is the multilingual terminology database of the ILO. It contains more than 24,000 entries and 100,000 terms in seven official and working languages of the Office. Most of the terms are related to labour issues, human rights and other UN-related fields.
[12] See https://www.w3.org/2009/08/skos-reference/skos.html#altLabel

- The use of skos:scopeNote[13] in the ILO thesaurus is widespread but a preliminary evaluation indicated its use was not particularly relevant in the foreseen alignment workflow.

The gig economy glossary, still under construction at the time of the transformation, included some imprecise entries that impacted on the final SKOS version[14].

Results of the analysis[15]:

	EuroVoc	ILO Glossary	ILO Taxonomy	ILO Thesaurus
Axioms / Number of Triples:	3 904 783	716	3 787	6 137
Logical axiom count	1 309 331	111	474	4 833
Class count	33	2	2	2
Individual count	446 788	111	474	4 833
Annotation Property count	34	7	8	17
ClassAssertion	446 143	111	474	4833
AnnotationAssertion	2 595 328	605	3 313	56 534

d. Aligning datasets and mapping concepts between KOS - using a lexical aligner

At this stage, clarifications were needed:

- an automatic lexical aligner would be used to map concepts lexicalised in English,

- whatever the results of the validation work, it would not be possible to infer that existing mappings between concepts lexicalised in English are applicable as such to the same concepts lexicalised in other languages,

- instead of "translation" we speak of "provision of equivalents in other languages", as a given concept in a language-speaking community does not necessarily have an exact equivalent in that of another community,

- as several controlled vocabulary management applications and aligners were available, hence different alignment algorithms, this exercise should be run with each of them, comparing the various alignment results and expanding the set of validated mappings,

- mappings are established bilaterally: they are created between concepts from vocabulary A found in vocabulary B as well as from vocabulary B found in vocabulary A, thus these bilateral links are not necessarily symmetrical,

[13] See https://www.w3.org/TR/skos-reference/#notes
[14] See Translating and the Computer 41
[15] See OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax (Second Edition)

- only 1-to-1 mappings were considered.

Silk[16] was used to automatically align all four KOS used. Through this process, we observed the following:

- Differences in the use of lower/upper case characters in the EuroVoc and the ILO thesaurus could be overcome using the Silk alignment methodology.

- Some consistency issues observed in the ILO glossary are detailed below:

a. prefLabels are often but not always present in all 3 languages,

b. differences in the level of details in each language versions can be considerable,

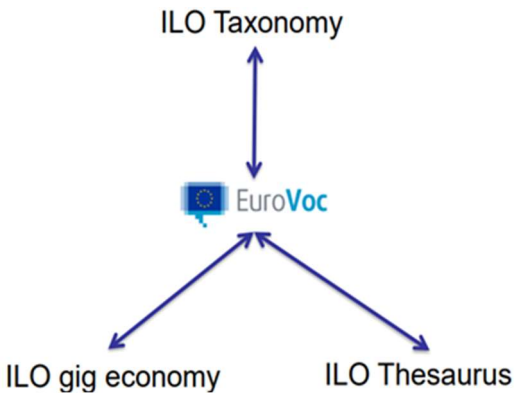
c. several skos:prefLabels contain notes, but the use of notes is not common (as in the ILO thesaurus).

- Some labels may be too detailed for an actual concept.

Since those inconsistencies could affect the alignment results, it was decided to align all 3 languages in parallel to counteract such effects.

The "ILO KOS to EuroVoc" alignment was performed using Silk with the following configuration:

1. Paths to source and target set to: skos:prefLabel
2. Transformation: Lower case
3. Comparator: Equality
4. Link type to skos:exactMatch

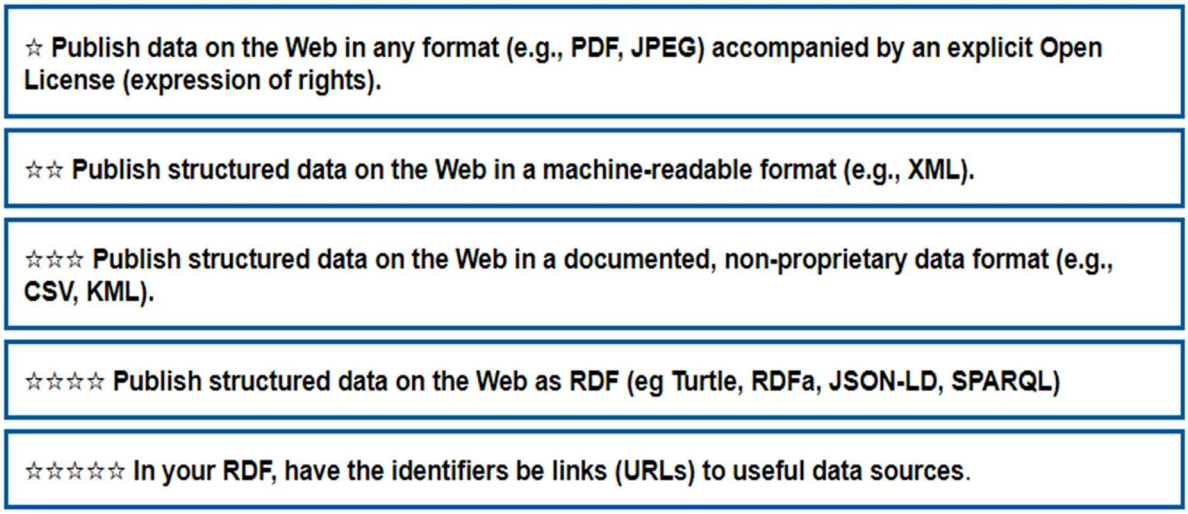


3. The project - phase 2 - automatic (semantic) alignment (same vocabularies)

Despite a well-defined work plan, our joint activities were turned upside down by this year's vast turmoil, delaying the 2nd phase of the project, i.e. the alignment of the same KOS with automated semantic aligners. The objectives which remain to be met in 2021: application of automated semantic aligners (integrated in PoolParty and in the PMKI platform[17], whose development is led by the PO) and assessment of their reliability, advantages and drawbacks.

[16] See Silk - The Linked Data Integration Framework
[17] Public multilingual knowledge management infrastructure, see Overcoming language barriers | ISA²

Nevertheless, we reassessed our process, deepened our analysis of the first sets of automatically suggested mappings (phase 1) and embarked on a major activity, relying on the principles of linked open data[18] (LOD):



If large amounts of the translations are made accessible online worldwide, they are mostly unstructured data. In light of this, and in view of another PO project, i.e. the definition and implementation of an XML-based content generation and maintenance platform (in particular for the *Interinstitutional Style Guide*[19]), we dedicated efforts to the definition of steps required to give structure to this huge bulk of unstructured data (see below under "4. **Impact of the use of metadata and reference data on knowledge management and the Authoring-Translation-Publishing chain"**).

a. Review of the process

Starting from a SKOS schema overview, the links between concepts (aka mappings) are of different types: skos:closeMatch, skos:exactMatch, skos:broadMatch, skos:narrowMatch and skos:relatedMatch. These SKOS properties are used to state mapping (alignment) links between SKOS concepts (lexicalised in assets) in different concept schemes ("an aggregation of one or more SKOS concepts"[20]), where the links are inherent in the meaning of the linked concepts[21]:

- Equivalence links -

The property skos:closeMatch is used to link two concepts that are sufficiently similar that they can be used interchangeably in some information retrieval applications. In order to avoid possibilities of "compound errors" when combining mappings across more than two concept schemes, **skos:closeMatch is not declared to be a transitive property.**

[18] See http://linkeddata.org/ and https://dvcs.w3.org/hg/gld/raw-file/default/glossary/index.html#linked-open-data
[19] See Interinstitutional style guide
[20] See https://www.w3.org/TR/skos-reference/#schemes
[21] See Methodology and tools for Metadata Governance - Joinup

The property skos:exactMatch is used to link two concepts, indicating a high degree of confidence that the concepts can be used interchangeably across a wide range of information retrieval applications. **skos:exactMatch is a transitive property**, and is a sub-property of skos:closeMatch.

- Hierarchical links -

The properties skos:broadMatch and skos:narrowMatch are used to state a hierarchical mapping link between two concepts

- Associative links -

The property skos:relatedMatch is used to state an **associative mapping link between two concepts.**

	ILO	EuroVoc	Description
1	child worker	young worker	Related match
2	social policy, social protection and social security		No match
3	housing	community development	broader match
4	economic development	economic and social development	broader match
5	economy	economy	Exact match

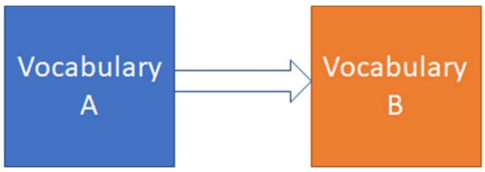
b. Mapping validation

Validating mappings is mainly a human-based activity and requires precision and attention as a concept may be lexicalised with different terms, while a lexicalisation may be associated to different concepts. Beyond the languages used to create controlled vocabularies to be aligned, the focus is also on all texts surrounding lexicalisations, i.e. on definitions (skos:definition), scope notes (skos:scopeNote), history notes (skos:historyNote), editorial notes (skos:editorialNote), hierarchical structures and synonyms (skos:altLabel[22])[23].

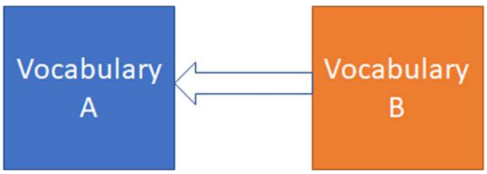
Following the automatic generation of mappings between the concepts in all four KOS in our project, a first human analysis and validation of the results was performed by a semantic web specialist at the PO. Mappings were suggested between English lexicalisations of concepts, and only those between EuroVoc concepts and ILO concepts were validated by the PO. Thus, the central question when assessing the suggested mappings was "which concept(s) in ILO

[22] See https://www.w3.org/TR/skos-reference/#labels
[23] See https://www.nomos-elibrary.de/10.5771/0943-7444-2014-1-66.pdf

assets is (are) an exact or close match(es) of EuroVoc concepts as lexicalised in English". To this end, all the notes and internal relationships of a concept must be taken into account.



Once generated and checked by the team maintaining EuroVoc, alignment files and suggested mappings were made available to ILO colleagues, a terminologist and an information management specialist, for a first analysis. All stakeholders were given access to the controlled vocabulary management application, VocBench, used by the PO. Following this analysis, the alignment files were imported into another controlled vocabulary management application, PoolParty, newly implemented at the ILO. As both applications build on semantic technologies, no data transformation was necessary and ILO colleagues could start the mapping validation from their perspective. This meant that the central question was "which concept(s) in EuroVoc is (are) an exact or close match(es) of concepts in ILO assets, as lexicalised in English".



Beyond this, a basic strategy was agreed upon. As such, the alignments run by the PO only produced close and exact matches. Following human validation, exact and close matches and rejected mappings were submitted to the assessment of ILO colleagues. Basically, it could not be assumed that the same kinds of mappings between the same concepts would be valid from the ILO perspective. This meant that an exact EuroVoc-ILO match might remain the same as an ILO-EuroVoc match, but it might also be a close match or a match of another kind, or even be rejected as a match in the case of erroneous definitions or lexicalisations.

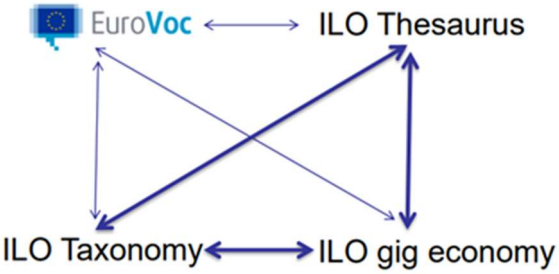
Equality matching gave the following statistics in the 3 ILO KOS:

	Eurovoc	ILO Taxonomy	ILO Thesaurus
ILO Glossary - Equality	X	X	X
ILO Taxonomy - Equality	X	-	X
ILO Thesaurus - Equality	X	-	-
ILO Glossary - Jaccard	-	X	-

	ILO glossary	ILO taxonomy	ILO thesaurus
Number of source entities:	7243	7243	7243
Number of target entities:	110	473	4830
Number of links:	7	202	2067
Valid links:	7*	201*	2030*

**All links were manually validated and were correct (except for inconsistencies in the concept label definition either in the ILO KOS or EuroVoc).*

As for the alignments between the ILO KOS, the exercise was completed to establish whether these KOS shared concepts between them and, if so, to what extent. The alignments were performed using the same Silk configuration.



Equality matching gives the following statistics:

Source	Target	Number of source entities:	Number of target entities:	Number of links	Valid links
ILO thesaurus	ILO glossary	4830	110	9	9
ILO thesaurus	ILO taxonomy	4830	473	523	523
ILO taxonomy	ILO glossary	473	110	5	5

To ensure full coverage of the alignment, the links targeting both broad and narrow concepts with the same labels were preserved as valid.

As a result, we could conclude the following:

- The Taxonomy is well integrated in the Thesaurus,

- The Glossary has a much lower level of direct integration.

To identify the reason behind the distance between the Glossary and the Thesaurus an alternative exercise was performed using a token-based approach and the Jaccard algorithm[24]:

[24] See Jaccard index

Source	Target	Number of source entities:	Number of target entities:	Number of links	Valid links
ILO thesaurus	**ILO glossary**	4830	110	204	16

c. **Asset quality improvement and LOD**[25]

Building on the sets of automatically generated mappings suggested and on the basic rules mentioned above, both organisations validated two different sets of mappings, made a cross-validation and identified some quality issues in all the vocabularies in terms of lexicalisations and definitions.

Based on the issues found, some corrections were considered, notably for the glossary. Furthermore, it was suggested that adding more glossary terms to the thesaurus could offer advantages but the need for this depends on the actual use of each KOS by the ILO.

One of the main challenges was the highly technical level of some ILO concepts compared to Eurovoc ones.

Example - concepts: Labour legislation and Labour law

While in EuroVoc labour law and labour legislation are synonyms, or at least refer to the same concept, in the ILO Thesaurus each refer to a different concept.

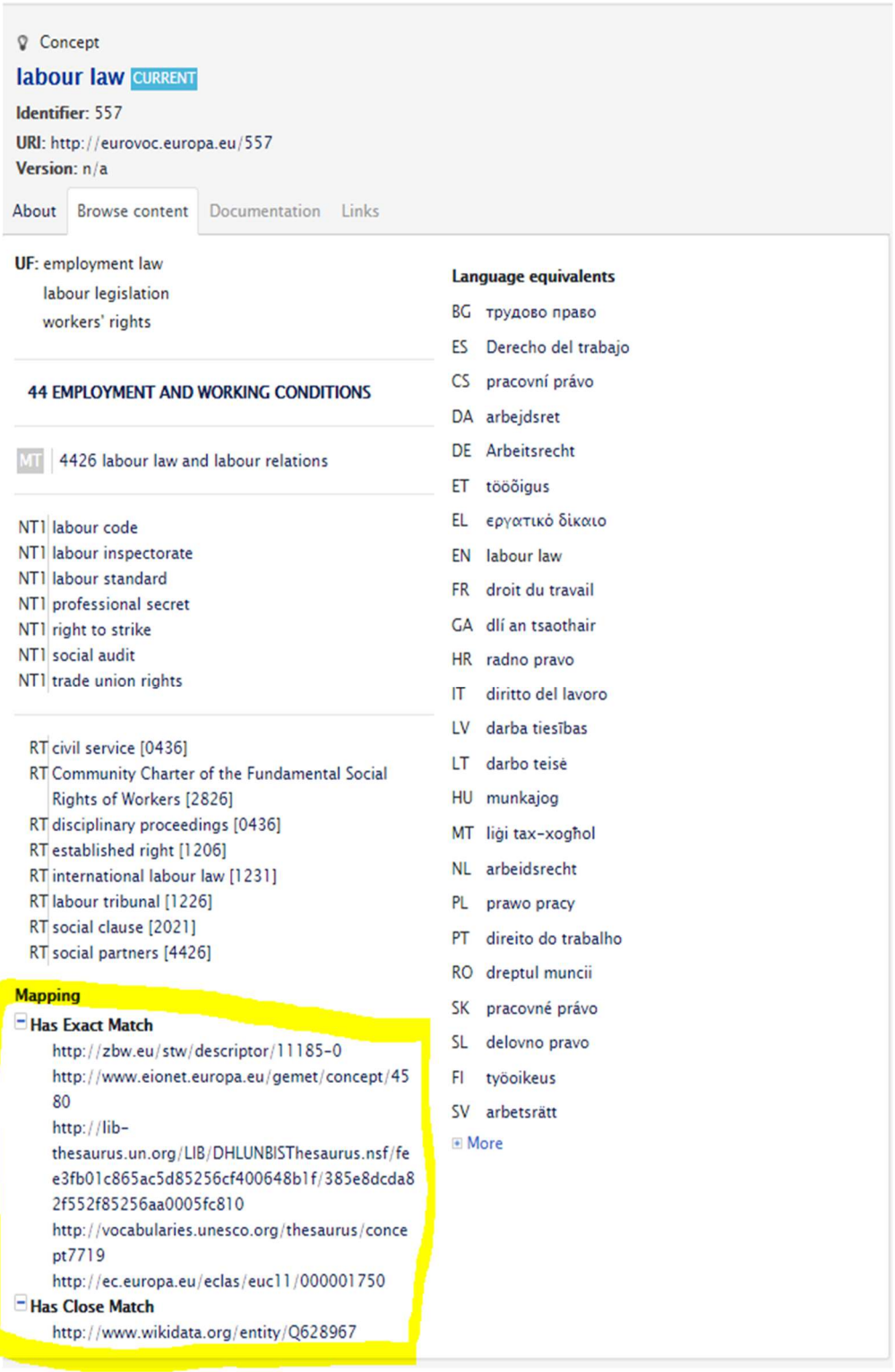
EuroVoc	ILO Thesaurus
Labour law is in the facet 44 - Employment and working conditions > 4426 Labour law and labour relations	**Labour law** is under 04. Law, human rights, government and politics > 04.01 Law
Labour legislation is a non-preferred label for Labour law, i.e. they refer to the same concept	**Labour legislation** is a preferred label, with a distinct scope note to distinguish it from labour law

On the other hand, the manual validation of the mappings between EuroVoc concepts and ILO thesaurus concepts by an information management specialist at the PO gave the opportunity to assess EuroVoc in a different context and some issues were observed.

If the EuroVoc thesaurus can indeed be considered as generic, all ILO KOS are very specialised. It appeared that several EuroVoc labels (skos:prefLabels and skos:altLabels) needed to be modified, while a few definitions were considered for improvement. Issues, mainly related to the equivalent labels in some languages and to possible misinterpretation of concepts during the "translation" process, were reported to the team maintaining EuroVoc in VocBench.

As such content is not machine-readable, mappings are done at the identifier level (Uniform Resource Identifier[26] (URI)) by human beings.

[25] Linked Open Data
[26] See Uniform Resource Identifier



Following the aforementioned validations, both organisations improved the consistency and quality of their respective vocabularies.

This done, both organisations planned a further step, i.e. sharing alignment files to allow external users the benefit of validated mappings. So, the EuroVoc -> ILO KOS mappings should be published with the 4.13 release of EuroVoc and disseminated through the EU

Vocabularies[27] and EU Open Data Portal[28] websites, as well as on the PMKI platform once implemented.

4. Increased discoverability of information, enhanced interoperability and the potential impact on the Authoring-Translation-Publishing chain

The setting up and running of interoperable systems is intended to ensure effective communication between digital components, e.g. devices, networks and data repositories. Such systems provide more efficient connections across borders between entities[29].

a. Discoverability, retrieval and reuse of information

The sharing of controlled vocabularies and of alignment files served to create mappings[30] between concepts included in different vocabularies and ultimately to improve distributed search and enrich vocabularies. Thus, mappings facilitate/enhance interoperability among systems and organisations.

They are also helpful when there are significant differences between vocabularies, as they allow for an integrated information search across different platforms. As such, mappings improve and expand access to vocabularies, in specific domains and between domains. This means that with a single query in a particular field, users can retrieve information from different databases. For each concept used in a search, all internal relationships are taken into account.

As stated above, mappings are done at the URI level. As the files made available by the ILO did not contain any URI, the information management specialists at the PO added fake URIs to align all datasets. The ILO moved their controlled vocabularies to PoolParty in 2019 and fake URIs were replaced during the validation process so that mapping could be implemented.

Nevertheless, for mappings to be effective and consistent over time, they need to be reviewed regularly and updated every time changes are inserted into a KOS selected for alignment. This is also a way to reassess the relevance of concepts in specific vocabularies and maintain or improve the discoverability of information. In this way, information is retrievable and can be reused.

b. Enhancing interoperability

In companies and large organisations, metadata tends to be defined according to requirements and in operational silos. As well as technical interoperability, the focus needs to be on semantic interoperability and, ideally, on the definition of rules and roles for the creation and maintenance of data, metadata and reference data.

Thus, it is necessary to break down silos and harmonise the sets of metadata used in the various IT applications and information dissemination platforms involved in operational entities as well as between entities collaborating closely. When you rely on knowledge

[27] See Alignments - EU Vocabularies - Publications Office of the EU
[28] See https://data.europa.eu/euodp/en/data/dataset?q=alignments&ext_boolean=all
[29] See New European Interoperability Framework
[30] "[alignment] links between SKOS concepts in different concept schemes, where the links are inherent in the meaning of the linked concepts", see https://www.w3.org/TR/skos-reference/#mapping

management, it is advisable to identify reference data, which are fundamental building blocks of most information systems and support interoperability.

This is applicable to the various tools used by translators, since they face similar challenges and questions:

- Which metadata are used in which tools?
- Who decides upon their creation and modification?
- Are metadata created based on a data model?
- Can duplicates and overlaps be observed?
- To what extent are these metadata sets identical?
- How to reuse reference data and maintain mappings over time?

All these questions are meant to tackle one of the biggest challenges faced by companies and large organisations wanting to transform an inconsistent landscape where IT specialists may not always understand how the management of metadata is organised, who is responsible for their maintenance, how decisions on changes are made, and how and when new, modified and deprecated metadata are managed in the various sectors of the organisation employing them. Another related aspect is the definition of a governance for the creation and maintenance of data, metadata and reference data.

This becomes obvious when it comes to the modernisation of public services and engaging in a digital transformation. As for large organisations, for the creation and sharing of information and knowledge by means of data exchange between the ICT systems, the definition and implementation of interoperability frameworks are extremely useful, even if challenging, and should be linked to reference architectures. An example of this is the *European Interoperability Framework*[31] and the *European Interoperability Reference Architecture*[32].

c. Potential impact on the authoring-translation-publishing chain?

The rise of the semantic web shed light on the crucial need for metadata, controlled vocabularies and ontologies, the latter being "primarily designed for machine-readability of knowledge"[33].

Ontologies have proven useful for many purposes. Writing consistent concept definitions within a specific domain is at least an informal ontology definition. In other words, building an ontology is exactly that, when the proper methodology is at the core of the task. It is therefore a basic premise of this paper that the principles of rigorous terminology work are useful for building consistent ontologies.

Developers of IT systems can encounter severe problems when they neglect the need to develop a proper ontology (concept model) before they develop a conceptual data model as a basis for a given IT system. As a result, ontology construction should ensure interoperability between systems and facilitate reuse of the components. This is one of the reasons we argue

[31] See New European Interoperability Framework
[32] See About European Interoperability Reference Architecture (EIRA) | Joinup
[33] See (PDF) Terminology meets the multilingual Semantic Web: A semiotic comparison of ontologies and terminologies

that an initial data modelling stage should always include the development of an *ad hoc* ontology.

As many similarities can be observed between concept modelling by terminologists and the construction of formal ontologies, terminologists are increasingly involved in IT development activities. Ontology construction has been shown to benefit from sound methodology and, in our opinion, this means that terminologists must be involved.

To better explain the premises of this paper we include some examples of mapping between ontologies and conceptual data models. We also believe that future research will reveal to what extent it will be possible to set up rules for automatic mapping of ontology concepts into classes and attributes of a conceptual data model.

Considering the authoring-translation-publishing chain as a whole, what would be the benefits of building an ontology to describe it, i.e. from the creation of a document to its publication, passing through the translation phase? In terms of metadata use, it would have an impact on authoring tools, computer-assisted translation and terminology tools as well as knowledge management systems, as all the metadata they include would be integrated in this ontology, while being listed and maintained as KOS. As such a switch cannot happen overnight, an intermediary step would consist in:

- harmonising or aligning the various metadata sets used in these tools and systems,

- implementing mappings to ease automatic retrieval of all linguistic assets relevant to authors and linguists,

- defining a metadata creation and maintenance governance to ensure consistency and effectiveness and, as a consequence, to enhance the efficiency of authors and linguists.

Adopting this approach, streamlining efforts in the ATP chain would take place following the definition of the defined ontology and the systematic implementation of metadata and semantics.

In addition, data structuration would be necessary to make it machine-readable. Building on standardised terminology, phraseology, document templates and advanced XML[34]-based text editors, documents would benefit from an architectural structure, i.e. an XML-schema structure and semantics, and thus be standardised.

But what about existing corpora of documents? They can be considered as mainly unstructured data or, in other words, data that is neither organised according to a predefined data model (i.e. ontology) nor structured (XML-based, as described above). As such, unstructured data is difficult to analyse, which implies that making sense of it merely implies examining individual documents to discern potential features.

Such features will most likely be present in other pieces of data ("documents") within the same corpora. If each document can be manually XMLised, this delicate operation proves to be efficient and beneficial to repetitive and partially standardised documents (in terms of type, format and layout, as well as terminology, standardised wordings and sentence patterns). In this case again, standardised documents (see above) help pave the way towards structuring data available in corpora and using artificial intelligence (AI).

[34] See https://www.w3.org/standards/xml/core

To define the purposes of using AI, structuring documents makes them machine-readable and more likely help identify unsuspected patterns that further enhance data structuration.

In the framework of the ATP chain, multiple benefits can be foreseen, like the streamlined authoring of new documents, the simplified and accelerated pre-processing of documents to be translated, a harmonised use of metadata throughout the whole process as well as other uses.

d. Structuring unstructured data

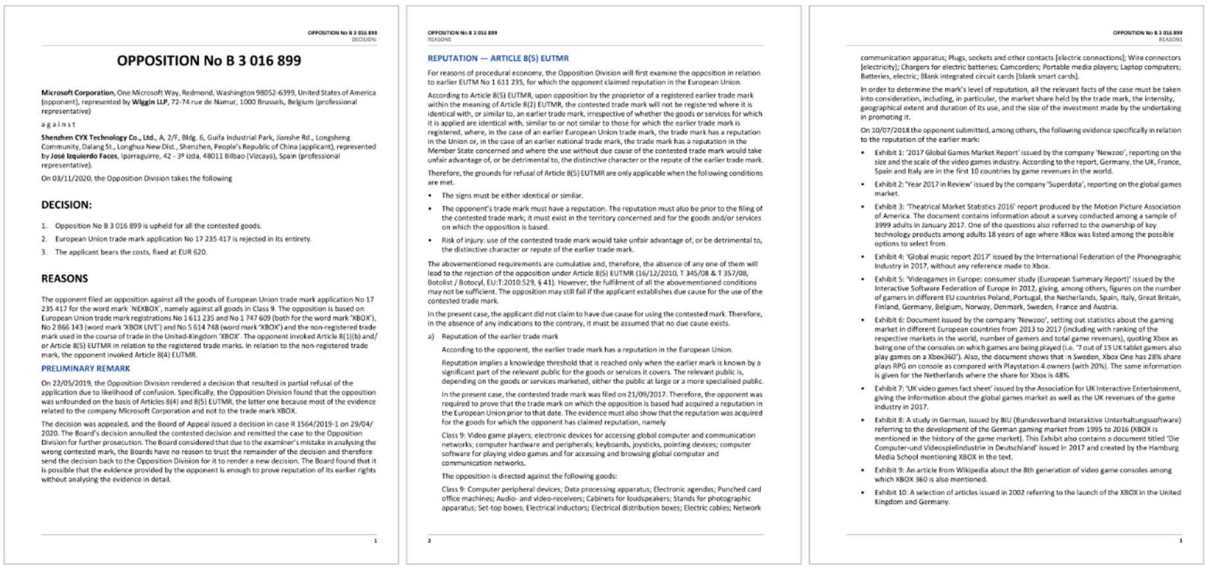
Before jumping into examples related to structured documents, it is important to structure our understanding of what comes into play with the following three issues:

- First of all, how can the result be exploited once our documents are structured?

- The methodology to be chosen: richness of the semantics to be embedded.

- Finally, how will the structuring model fair over time and what is the technological evolution going to favour?

To start discussing all three issues we need to see a document that is initially only designed for readability.



OPPOSITION No B 3 016 899

DECISION

OPPOSITION No B 3 016 899

Microsoft Corporation, One Microsoft Way, Redmond, Washington 98052-6399, United States of America (opponent), represented by **Wiggin LLP**, 72-74 rue de Namur, 1000 Brussels, Belgium (professional representative)

against

Shenzhen CYK Technology Co., Ltd., A, 2/F, Bldg. 6, Guifa Industrial Park, Jianshe Rd., Longsheng Community, Dalang St., Longhua New Dist., Shenzhen, People's Republic of China (applicant), represented by **José Izquierdo Faces**, Iparraguirre, 42 - 3ª izda, 48011 Bilbao (Vizcaya), Spain (professional representative).

On 03/11/2020, the Opposition Division takes the following

DECISION:

1. Opposition No B 3 016 899 is upheld for all the contested goods.
2. European Union trade mark application No 17 235 417 is rejected in its entirety.
3. The applicant bears the costs, fixed at EUR 620.

REASONS

The opponent filed an opposition against all the goods of European Union trade mark application No 17 235 417 for the word mark 'NEXBOX', namely against all goods in Class 9. The opposition is based on European Union trade mark registrations No 1 611 235 and No 1 747 609 (both for the word mark 'XBOX'), No 2 866 143 (word mark 'XBOX LIVE') and No 5 614 748 (word mark 'XBOX') and the non-registered trade mark used in the course of trade in the United-Kingdom 'XBOX'. The opponent invoked Article 8(1)(b) and/ or Article 8(5) EUTMR in relation to the registered trade marks. In relation to the non-registered trade mark, the opponent invoked Article 8(4) EUTMR.

PRELIMINARY REMARK

On 22/05/2019, the Opposition Division rendered a decision that resulted in partial refusal of the application due to likelihood of confusion. Specifically, the Opposition Division found that the opposition was unfounded on the basis of Articles 8(4) and 8(5) EUTMR, the latter one because most of the evidence related to the company Microsoft Corporation and not to the trade mark XBOX.

The decision was appealed, and the Board of Appeal issued a decision in case R 1564/2019-1 on 29/04/ 2020. The Board's decision annulled the contested decision and remitted the case to the Opposition Division for further prosecution. The Board considered that due to the examiner's mistake in analysing the wrong contested mark, the Boards have no reason to trust the remainder of the decision and therefore send the decision back to the Opposition Division for it to render a new decision. The Board found that it is possible that the evidence provided by the opponent is enough to prove reputation of its earlier rights without analysing the evidence in detail.

1

OPPOSITION No B 3 016 899

REASONS

REPUTATION — ARTICLE 8(5) EUTMR

For reasons of procedural economy, the Opposition Division will first examine the opposition in relation to earlier EUTM No 1 611 235, for which the opponent claimed reputation in the European Union.

According to Article 8(5) EUTMR, upon opposition by the proprietor of a registered earlier trade mark within the meaning of Article 8(2) EUTMR, the contested trade mark will not be registered where it is identical with, or similar to, an earlier trade mark, irrespective of whether the goods or services for which it is applied are identical with, similar to or not similar to those for which the earlier trade mark is registered, where, in the case of an earlier European Union trade mark, the trade mark has a reputation in the Union or, in the case of an earlier national trade mark, the trade mark has a reputation in the Member State concerned and where the use without due cause of the contested trade mark would take unfair advantage of, or be detrimental to, the distinctive character or the repute of the earlier trade mark.

Therefore, the grounds for refusal of Article 8(5) EUTMR are only applicable when the following conditions are met.

- The signs must be either identical or similar.
- The opponent's trade mark must have a reputation. The reputation must also be prior to the filing of the contested trade mark; it must exist in the territory concerned and for the goods and/or services on which the opposition is based.
- Risk of injury: use of the contested trade mark would take unfair advantage of, or be detrimental to, the distinctive character or repute of the earlier trade mark.

The abovementioned requirements are cumulative and, therefore, the absence of any one of them will lead to the rejection of the opposition under Article 8(5) EUTMR (16/12/2010, T 345/08 & T 357/08, Botolist / Botocyl, EU:T:2010:529, § 41). However, the fulfilment of all the abovementioned conditions may not be sufficient. The opposition may still fail if the applicant establishes due cause for the use of the contested trade mark.

In the present case, the applicant did not claim to have due cause for using the contested mark. Therefore, in the absence of any indications to the contrary, it must be assumed that no due cause exists.

a) Reputation of the earlier trade mark

According to the opponent, the earlier trade mark has a reputation in the European Union.

Reputation implies a knowledge threshold that is reached only when the earlier mark is known by a significant part of the relevant public for the goods or services it covers. The relevant public is, depending on the goods or services marketed, either the public at large or a more specialised public.

In the present case, the contested trade mark was filed on 21/09/2017. Therefore, the opponent was required to prove that the trade mark on which the opposition is based had acquired a reputation in the European Union prior to that date. The evidence must also show that the reputation was acquired for the goods for which the opponent has claimed reputation, namely

Class 9: Video game players; electronic devices for accessing global computer and communication networks; computer hardware and peripherals; keyboards, joysticks, pointing devices; computer software for playing video games and for accessing and browsing global computer and communication networks.

The opposition is directed against the following goods:

Class 9: Computer peripheral devices; Data processing apparatus; Electronic agendas; Punched card office machines; Audio- and video-receivers; Cabinets for loudspeakers; Stands for photographic apparatus; Set-top boxes; Electrical inductors; Electrical distribution boxes; Electric cables; Network

2

OPPOSITION No B 3 016 899

REASONS

communication apparatus; Plugs, sockets and other contacts [electric connections]; Wire connectors [electricity]; Chargers for electric batteries; Camcorders; Portable media players; Laptop computers; Batteries, electric; Blank integrated circuit cards [blank smart cards].

In order to determine the mark's level of reputation, all the relevant facts of the case must be taken into consideration, including, in particular, the market share held by the trade mark, the intensity, geographical extent and duration of its use, and the size of the investment made by the undertaking in promoting it.

On 10/07/2018 the opponent submitted, among others, the following evidence specifically in relation to the reputation of the earlier mark:

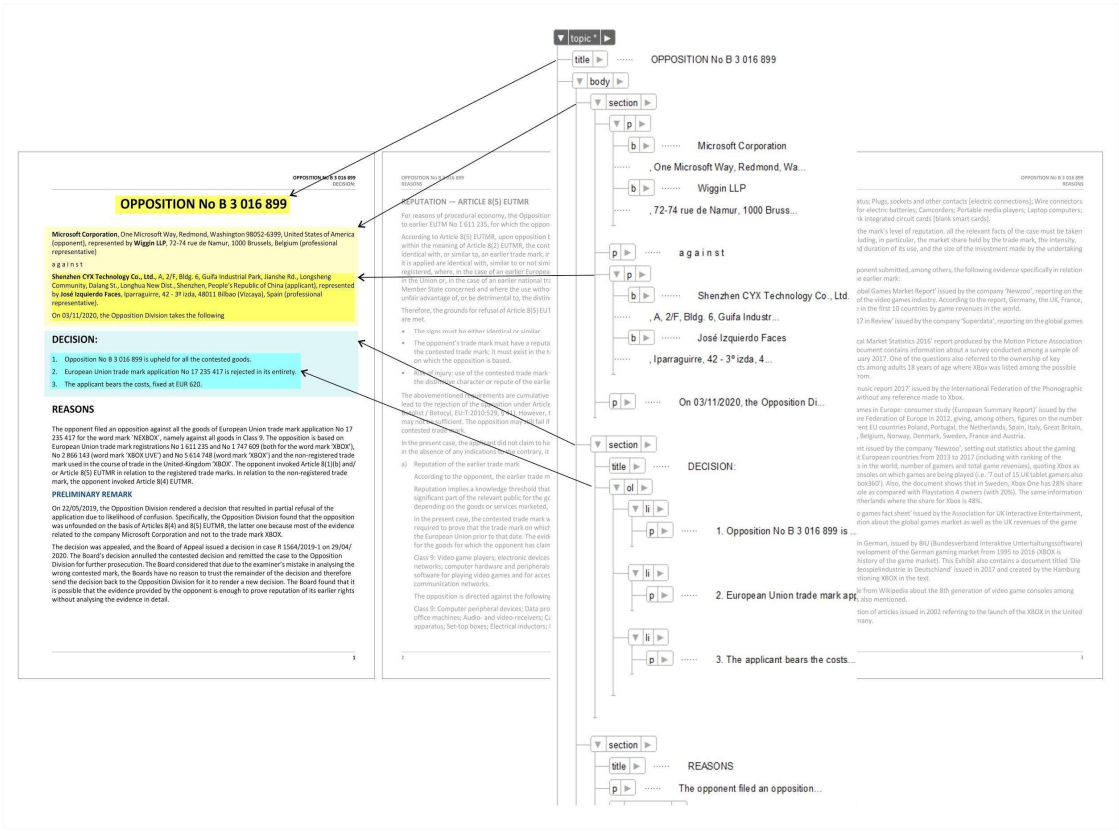
- Exhibit 1: '2017 Global Games Market Report' issued by the company 'Newzoo', reporting on the size and the scale of the video games industry. According to the report, Germany, the UK, France, Spain and Italy are in the first 10 countries by game revenues in the world.
- Exhibit 2: 'Year 2017 in Review' issued by the company 'Superdata', reporting on the global games market.
- Exhibit 3: 'Theatrical Market Statistics 2016' report produced by the Motion Picture Association of America. The document contains information about a survey conducted among a sample of 3999 adults in January 2017. One of the questions also referred to the ownership of key technology products among adults 18 years of age where XBox was listed among the possible options to select from.
- Exhibit 4: 'Global music report 2017' issued by the International Federation of the Phonographic Industry in 2017, without any reference made to Xbox.
- Exhibit 5: 'Videogames in Europe: consumer study (European Summary Report)' issued by the Interactive Software Federation of Europe in 2012, giving, among others, figures on the number of gamers in different EU countries Poland, Portugal, the Netherlands, Spain, Italy, Great Britain, Finland, Germany, Belgium, Norway, Denmark, Sweden, France and Austria.
- Exhibit 6: Document issued by the company 'Newzoo', setting out statistics about the gaming market in different European countries from 2013 to 2017 (including with ranking of the respective markets in the world, number of gamers and total game revenues), quoting Xbox as being one of the consoles on which games are being played (i.e. "7 out of 15 UK tablet gamers also play games on a Xbox360"). Also, the document shows that in Sweden, Xbox One has 28% share plays RPG on console as compared with Playstation 4 owners (with 20%). The same information is given for the Netherlands where the share for Xbox is 48%.
- Exhibit 7: 'UK video games fact sheet' issued by the Association for UK Interactive Entertainment, giving the information about the global games market as well as the UK revenues of the game industry in 2017.
- Exhibit 8: A study in German, issued by BIU (Bundesverband Interaktive Unterhaltungssoftware) referring to the development of the German gaming market from 1995 to 2016 (XBOX is mentioned in the history of the game market). This Exhibit also contains a document titled 'Die Computer-und Videospielindustrie in Deutschland' issued in 2017 and created by the Hamburg Media School mentioning XBOX in the text.
- Exhibit 9: An article from Wikipedia about the 8th generation of video game consoles among which XBOX 360 is also mentioned.
- Exhibit 10: A selection of articles issued in 2002 referring to the launch of the XBOX in the United Kingdom and Germany.

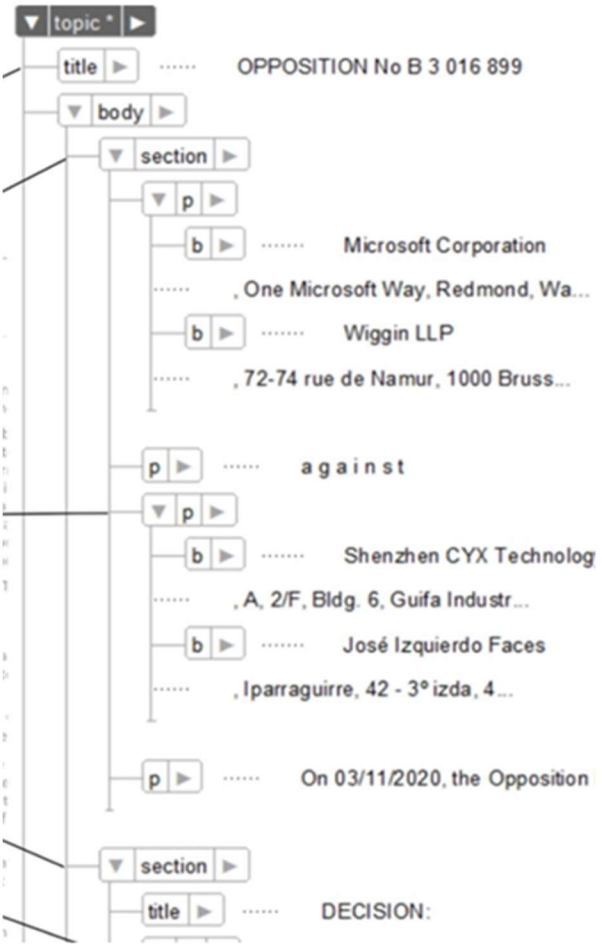
3

If we say that this document is already structured but only for the human eyes and understanding, such assertion would be correct no matter what is happening behind the scenes to format the results we see. The structure is visually coded in a way that readers can understand it at a glance. We see titles, paragraphs, lists and hierarchy. Unfortunately, none of this will be sufficient if we care about the three issues previously listed.

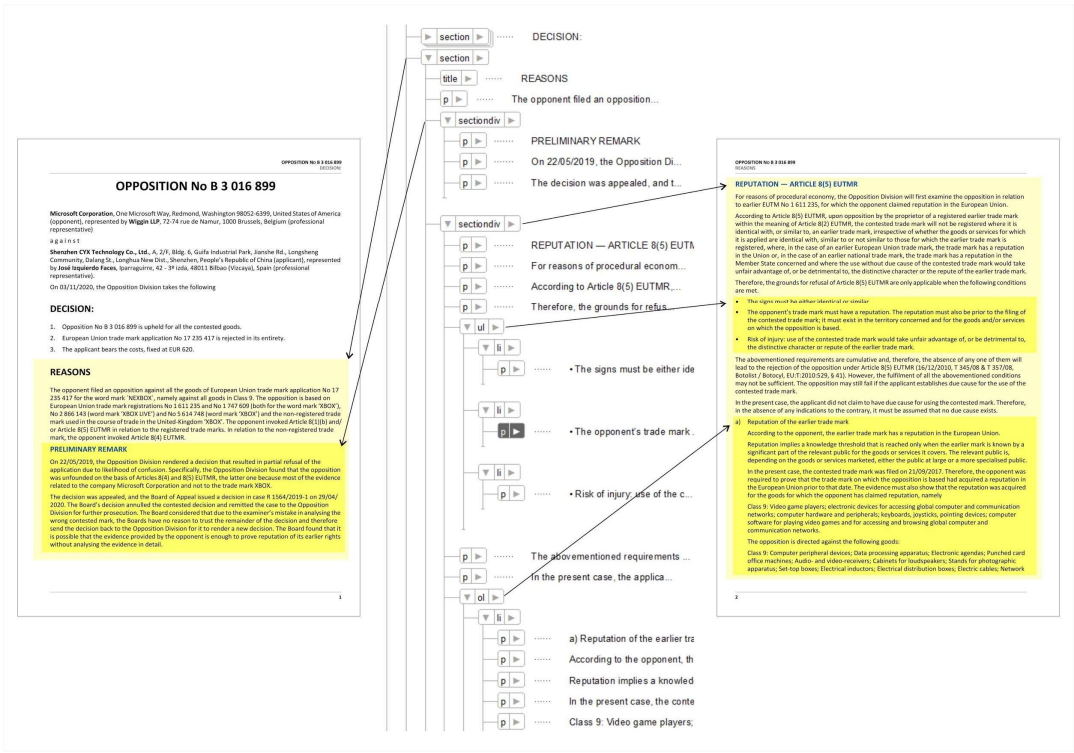
First, we need to assume that the digital file is better than the hardcopy, and then we can add metadata to the document as a whole. This will help us to reference and retrieve the full document and then readability will be the only factor to be considered.

However, are there parts of the document that are important on their own, so could it be beneficial to retrieve them independently or for separate purposes other than that of the whole monolithic document? An affirmative answer resolves the first issue listed above: there is now a purpose and there are attainable benefits.


OPPOSITION No B 3 016 899
DECISION:
1. Opposition No B 3 016 899 is upheld for all the contested goods.
2. European Union trade mark application No 17 235 417 is rejected in its entirety.
3. The applicant bears the costs, fixed at EUR 620.
REASONS
PRELIMINARY REMARK
topic *
title
OPPOSITION No B 3 016 899
body
section
Microsoft Corporation
, One Microsoft Way, Redmond, Wa...
Wiggin LLP
, 72-74 rue de Namur, 1000 Bruss...
against
Shenzhen CYX Technology Co., Ltd.
, A, 2/F, Bldg. 6, Guifa Industr...
José Izquierdo Faces
, Iparraguirre, 42 - 3º izda, 4...
On 03/11/2020, the Opposition Di...
DECISION:
1. Opposition No B 3 016 899 is ...
2. European Union trade mark ap...
3. The applicant bears the costs...
REASONS
The opponent filed an opposition...

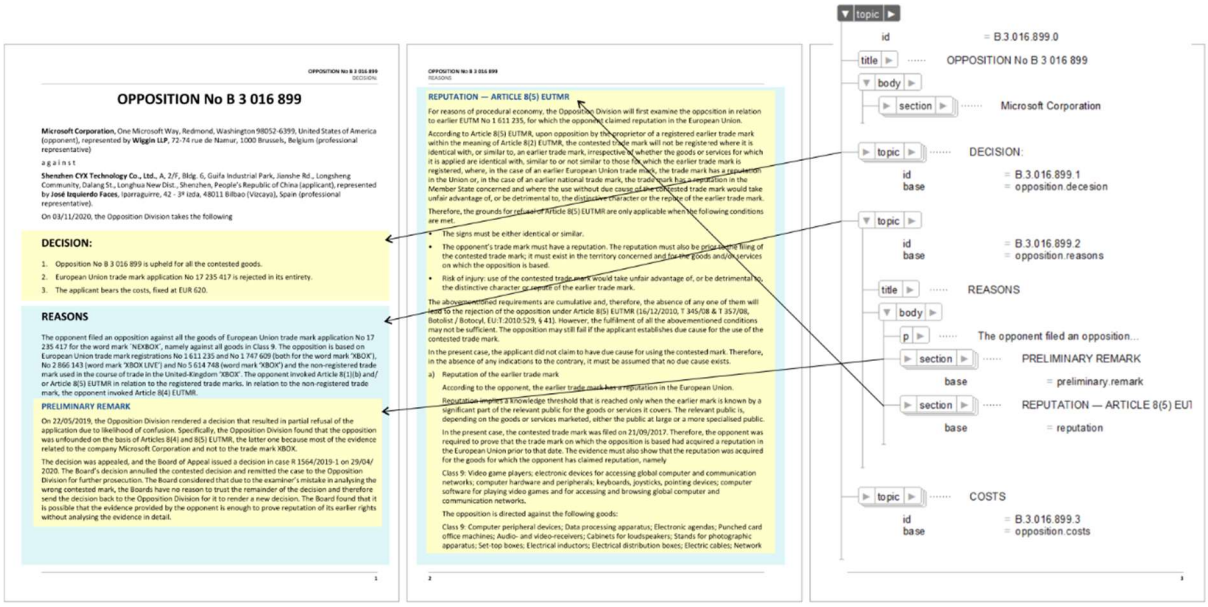

topic *
title
OPPOSITION No B 3 016 899
body
section
Microsoft Corporation
, One Microsoft Way, Redmond, Wa...
Wiggin LLP
, 72-74 rue de Namur, 1000 Bruss...
against
Shenzhen CYX Technolog
, A, 2/F, Bldg. 6, Guifa Industr...
José Izquierdo Faces
, Iparraguirre, 42 - 3º izda, 4...
On 03/11/2020, the Opposition
DECISION:

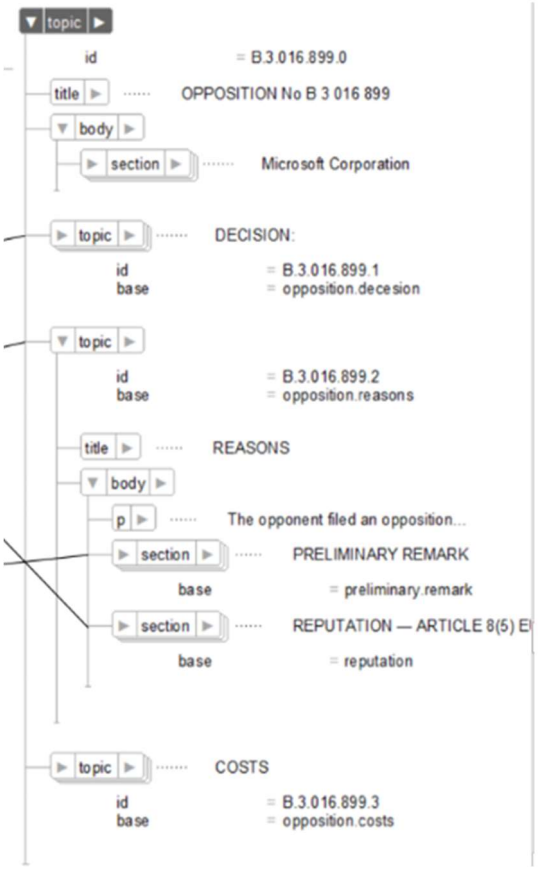
In this new view of our document there is embedded structure but, above all, there is a basic level of semantics within the obvious hierarchy of the proposed structure. The titles, paragraphs, lists and grouping of content are now explicit and duly embedded for the “eyes” of information technology to “see”. This structured model is only based on “elements” and how they relate to one another.



However, the embedded semantic information is not extraordinarily rich. For instance, if we need to retrieve a specific “section” it would be necessary to dig into the “title” element contained, identifying which one we want. This is a big improvement but probably not sufficient for certain purposes.

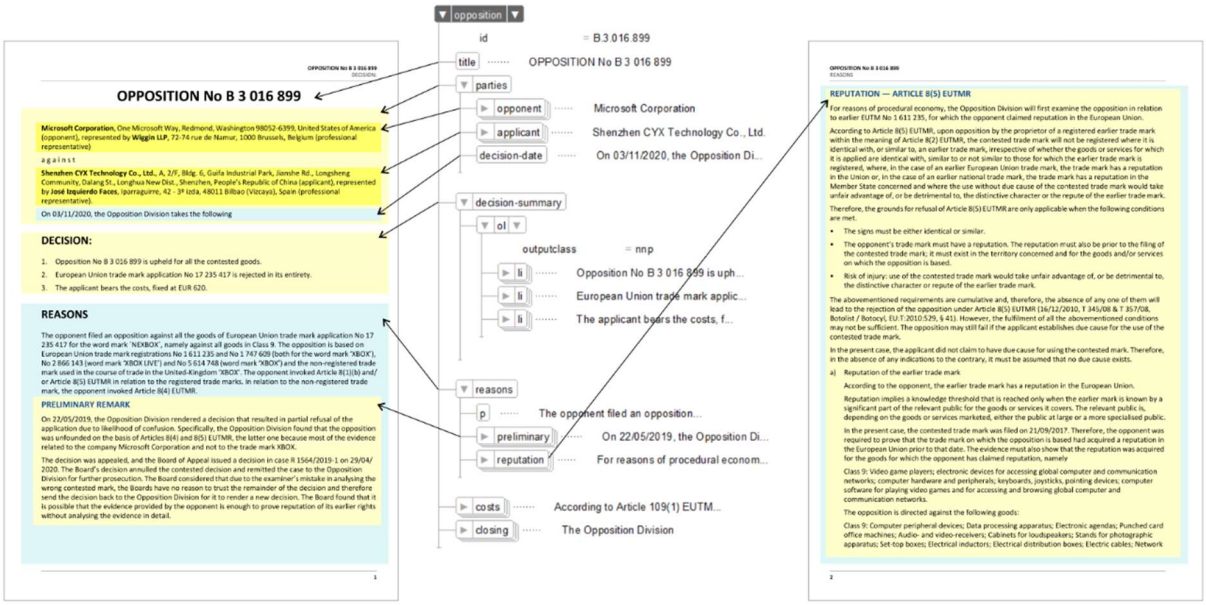
We have talked about adding metadata to the document to enable retrieval of the whole. Now we should also think about adding metadata to the structural components too. This will have a favourable impact on the embedded semantics.

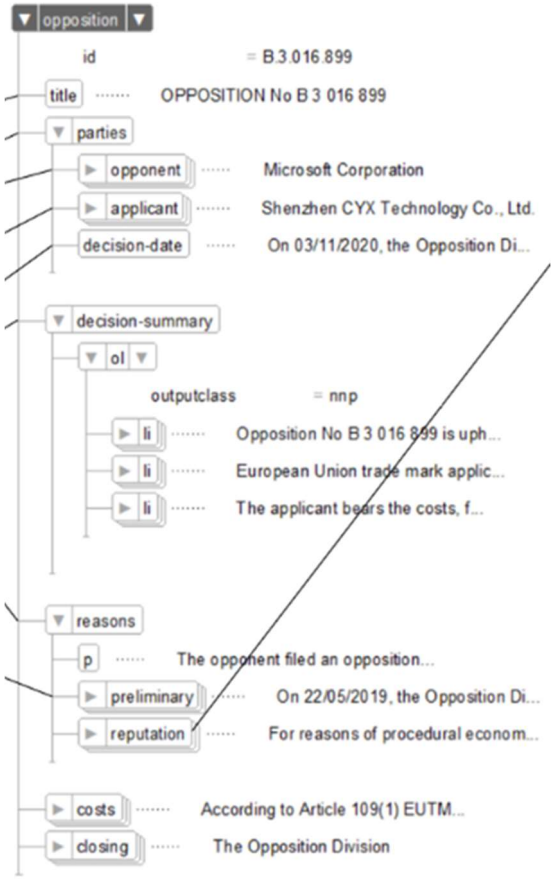




We have now added metadata to “elements” and the components that carry the metadata are called “attributes”. The value(s) inside an attribute creates a new combined level of semantic information. By enriching the semantics, we have responded to the second issue in our initial list.

Nevertheless, we have not defined a very detailed model of semantics and hierarchy. It can run much deeper and be much more specific. For that, we need to create a model that is less generic.





In this new version of our document, we have gone beyond a generic model. In fact, the new model can only be applied to a specific type of document.

The understanding of document structuring is based on forty years of science that has dealt with precisely this issue more than the other two. Being very specific in the definition of a document type is always a possibility and it has pros and cons. Technological evolution has led to tools and environments that favour generic over specific. This is a lesson learnt the hard way with decisions that have not aged well.

On the bright side, these lessons have brought about proposals for widening specific models[35]. Also, there are general-purpose document definition models that can and have already evolved in parallel with technology[36].

e. And the future?

We have already addressed the definition of an ontology for the ATP chain, using semantic web technologies and XMLising documents with a view to deep structuration of unstructured data. Based on the example provided above, this structuring might be applicable if several criteria are met. The willingness to turn data into information and information into knowledge can be facilitated by the existence of standardised documents, terminology and phraseology, layout templates, and the willingness and readiness to invest efforts in defining relevant and meaningful metadata, as well as purposes.

If the harmonisation, maintenance, streamlining and systematic reuse of metadata/reference data now seem to be understandable and attainable objectives, the use of XML-based formats and content generation and maintenance platforms could be linked to further harmonisation.

Here we come back to one PO project, mentioned above (under 3. **The project - phase 2 - automatic (semantic) alignment (same vocabularies)**), and a current side-activity (see **d. Structuring unstructured data** above). The XML-based content generation and maintenance platform used *inter alia* for the streamlined maintenance and generation of the *Interinstitutional Style Guide* in various formats for different dissemination platforms enables the structuration of text (unstructured data in mainstream content editors).

The use of XML tags could then be linked to the various classes, individuals and relations of the ontology, as shown above, while the content of style guides, like the *Interinstitutional Style Guide*, could be transcribed into restrictions, rules, axioms and events to be inserted, as rules at the document level, into the same ontology[37], which would lead to further automation (e.g. intrinsic validation).

Such rules can be classified into various categories[38]: integrity (or constraints), derivation (conditions resulting in conclusions), reaction (event, condition, action, alternative action, post-condition), production n (condition, action) and transformation (change of state).

[35] See Akoma Ntoso | Akoma Ntoso Site
[36] See OASIS Darwin Information Typing Architecture (DITA) TC | OASIS
[37] See SWRL: A Semantic Web Rule Language Combining OWL and RuleML
[38] See A Ontology-based Approach for Business Process Compliance Checking

5. Conclusion

As stated during the Translating and the Computer conference in 2019, the "main purpose of semantics is to empower knowledge extraction from enormous sets of raw data in various formats, as it offers faster and more cost-effective access to meaningful and accurate data, for analysis and transformation into knowledge. [...]

Reinforcing bridges between terminology assets and KOS, between KOS and corpora[39], and using standard formats, we can foresee in the short-term new tools assembling assets for linguists and authors, using semantic web technologies and reusing online contents with a click. In short, the prerequisites would be further development of both KOS and terminologies, and interlinking KOS (also with terminology assets) ."[40]

The first two phases of our project laid the foundations for creating synergies and collaboration between two communities of professionals, enhancing assets and tools and broadening horizons. An improved terminological and semantic interoperability further improved semantic and terminology assets and steps towards linked (open) data came to light.

Benefiting from various external impulses and a growing interest in metadata, semantic technologies and integrated semantics, the subsequent phases of the project were re-examined and led to ontology building and XMLisation experiments that will be further investigated in the next phase of our project, as preliminary promising results have already been produced.

Acknowledgements

We would like to thank Ms. Christie Damnet (proofreading) for her patience, Ms. Hala El-Gohary and Ms. Verónica Parsiale (examples), Mr. Juan Souza (XMLisation and structuring of unstructured data) and Mr. Mihai Paunescu (KOS analysis and alignments) for all the knowledge passed on.

[39] See Browse by EuroVoc
[40] See Translating and the Computer 41

References

- **Compliance**

Gailly, F., Geerts, G., Ontology-Driven Business Rule Specification. *Journal of Information Systems* (2013) 27 (1): 79–104. https://doi.org/10.2308/isys-50428.

Pham, T., Le Thanh, N., A Ontology-based Approach for Business Process Compliance Checking. IMCOM '16 - the 10th International Conference on Ubiquitous Information Management and Communication, Sungkyunkwan University, Korea and Universiti Kuala Lumpur, Malaysia, Jan 2016, Da Nang, Vietnam. pp.1 - 6, 10.1145/2857546.2857603. hal-01401769

Rubiera, E., de Sainte Marie, C., ONTORULE: from business knowledge to ontology- and rules-based applications. CASCON '11: Proceedings of the 2011 Conference of the Center for Advanced Studies on Collaborative Research, November 2011, pp. 327–329

Corea, C., Delfmann, P., Detecting Compliance with Business Rules in Ontology-Based Process Modeling. Wirtschaftsinformatik und Angewandte Informatik, 2017.

Berstel-Da Silva, B., Chniti, A., Detection of Inconsistencies in Rules due to Changes in Ontologies: Let's Get Formal, International Conference on Web Reasoning and Rule Systems, 2013, pp. 198-203.

- **Interoperability frameworks**

EIF - European Interoperability Framework | Joinup

https://youtu.be/g-CzHHJ0ZTM

New European Interoperability Framework

- **Metadata**

Apostolou, C., The role of metadata in managing knowledge | Request PDF, 2009

Sicilia M.-Á., Metadata, semantics, and ontology: providing meaning to information resources Miguel-Ángel Sicilia, International Journal of Metadata, Semantics and Ontologies, 2006.

Horning, C., Handbook of Metadata, Semantics and Ontologies, Cataloging & Classification Quarterly 54, 2016, pp. 504 - 505, https://doi.org/10.1142/7077.

Kempf, A. O. et al., New Ways of Mapping Knowledge Organization Systems: Using a Semi-Automatic Matching Procedure for Building up Vocabulary Crosswalks, *Knowledge Organization* 41, 2014, pp. 66-75.

- **Ontologies**

Rubiera, E., Marie, C., ONTORULE: from business knowledge to ontology- and rules-based applications, CASCON, 2011.

Noy, N., Ontology Development 101: A Guide to Creating Your First Ontology, 2001

- **Semantic tools**

Fiorelli, M. et al., VocBench 3: A collaborative Semantic Web editor for ontologies, thesauri and lexicons, Semantic Web 11, 2020, pp. 855-881.

Guidelines for mapping into SKOS, dealing with translations, 2010.

- **Standards**

W3C. Semantic Web, https://www.w3.org/2001/sw/wiki/Main_Page

ISO 23081-1:2017 - Information and documentation -- Records management processes -- Metadata for records -- Part 1: Principles, 2006. https://www.iso.org/standard/73172.html

ISO/DIS 23081-2 - Information and documentation — Managing metadata for records — Part 2: Conceptual and implementation issues. https://www.iso.org/standard/81600.html

ISO/TR 23081-3:2011 - Information and documentation — Managing metadata for records — Part 3: Self-assessment method. https://www.iso.org/standard/57121.html

ISO 25964-1:2011, Information and documentation — Thesauri and interoperability with other vocabularies — Part 1: Thesauri for information retrieval. https://www.iso.org/standard/53657.html

ISO 25964-2:2013, Information and documentation — Thesauri and interoperability with other vocabularies — Part 2: Interoperability with other vocabularies. https://www.iso.org/standard/53658.html

Terminology ISO standards, https://termcoord.eu/terminology-iso-standards/

Polleres, A., Steyskal, S., Semantic Web Standards for Publishing and Integrating Open Data, 2015.

Hitzler, P. et al., OWL 2 Web Ontology Language Primer (Second Edition), 2012.

- **Structuring**

Day, D. et al. "Introduction to the Darwin Information Typing Architecture Toward portable technical information, 2017.

Vitali, F. and Palmirani, M., Akoma Ntoso: Flexibility and Customization to Meet Different Legal Traditions, 2019.

Sikos, Leslie F., Mastering Structured Data on the Semantic Web: From HTML5 Microdata to Linked Open Data, 2015.

- **Terminology**

Terminological concept modelling and conceptual ... - Semantic Scholar

(PDF) Terminology meets the multilingual Semantic Web: A semiotic comparison of ontologies and terminologies

- **Miscellaneous**

Conclusions and recommendations of JIAMCATT 2017

2018 UN GENEVA ANNUAL REPORT

Dechandon, D., Gerencsér, A., Recort Ruiz, M., Terminology: Towards a Systematic Integration of Semantics and Metadata, Translating and the Computer 41 Conference, London, 21-22 November 2019, pp. 10-26.

Paschen, U. et al., Artificial intelligence: Building blocks and an innovation typology, Business Horizons 63, 2020, pp. 147-155.

APE-QUEST, or how to be picky about Machine Translation?

Tom Vanallemeersch
Crosslang

Sara Szoc
Crosslang

Abstract

APE-QUEST (Automated Post-Editing and Quality Estimation, 2018-2020) proposes a Quality Gate aiming at automatically estimating the quality of domain-specific sentence translations produced by a generic machine translation (MT) system and to channel sentences with a low quality estimation (QE) score to a human post-editor or to an automatic post-editing (APE) system. The Quality Gate aims to achieve the desired translation quality (assimilation or dissemination use case) in a time- and cost-efficient way using MT and human or automatic correction. The consortium has trained QE and APE systems for three language pairs (English into French, Dutch and Portuguese) on domain-specific post-edited data (texts relating to the legal domain, procurement, and online dispute resolution). The data (approximately 10,000 sentences per language pair) are made publicly available through the ELRC-SHARE repository. During the final stage of the project, human evaluations were performed to assess the impact of the Quality Gate on the translation workflow. The results evidence the potential of the QE component in the Quality Gate for reducing cost and time while preserving high translation quality. However, tests using APE have shown the difficulty of producing automatic corrections that are both precise and exhaustive.

1. Introduction

While machine translation (MT) quality has dramatically progressed in recent years, especially because of the shift to deep learning, the data behind the MT system remains a crucial factor. Therefore, the risk of errors (e.g. terminological errors) is especially high when a generic MT engine is applied to domain-specific text, as there is a disparity between the data on which the system is trained and the data to which it is applied. An example of such a mismatch is shown in Table 1 (the problematic parts are marked in bold). In addition, sentence translations produced by MT systems tend to have a variable quality, regardless of the domain in question. This variability relates to factors such as the length of a sentence, the distance between source and target language, the presence of ambiguous words with multiple possible translations, etc.

	In-domain (privacy statement)	**Out-of-domain (consumer complaint)**
Source sentence	This privacy statement explains the reason for the processing of your personal data, the way we collect, handle and ensure protection of all personal data provided …	XXXX XXXX -- -- I am not aware **of this debt I would like this debt** validated or removed from my credit report.
Translation	La présente déclaration de confidentialité explique les raisons du traitement de vos données à caractère personnel, la façon dont toutes les données à caractère personnel fournies sont recueillies, traitées et protégées …	XXXX XXXX — -- Je n'ai pas connaissance **de cette dette que je souhaiterais que cette dette** soit validée ou retirée de mon rapport de crédit.

Table 1: Domain mismatch in MT

To cope with domain specificity, various strategies can be applied, in isolation or in combination. One such strategy is the customisation of the MT system built from generic data (consisting of "general-language" sentence pairs and pairs from domains other than the one under consideration). Customisation involves adding in-domain sentence pairs to the generic data and retraining the MT system (from scratch or, in case of a neural MT system, through continued training of the existing neural network). Another possible strategy consists of the automatic quality estimation (QE) of a translation produced by an MT system, i.e. the automatic prediction of a quality score for such a translation. Based on this score, a professional translator or a tool can select which MT translations may be kept as such, should be post-edited, or should be translated from scratch. Yet another strategy consists of the automatic post-editing (APE) of sentences as a complement to human post-editing. For instance, APE may be applied to sentences selected based on their QE score.

In this article, we report on the Quality Gate built within the APE-QUEST project, a proof-of-concept environment which combines MT with both QE and APE. The Quality Gate addresses not only the issue of domain specificity, as explained above, but also that of the variable quality of MT output, as QE scores are also helpful for pinpointing sentences that are erroneous based on factors other than the domain.

This article is structured as follows. In Section 2, we explain the concept of the Quality Gate. We elaborate on its components in Section 3 and provide details on the evaluation of its results in Section 4. Finally, conclusions are presented in Section 5.

2. APE-QUEST Quality Gate

The APE-QUEST project (Automated Post-Editing and Quality Estimation, duration 2018-2020) is a collaboration between CrossLang (coordinator), Unbabel, and the University of Sheffield, funded by the Connecting Europe Facility programme of the European Commission (EC). It investigates the combination of an MT system with a QE system to channel a selection of sentence translations to a human post-editor. In addition, the project investigates the integration of an APE system, as a complementary means of correcting MT translations. Channelling takes place through a Quality Gate, as illustrated in Figure 1. This is a workflow that applies automatic QE to the translation of a sentence produced by MT and decides, based on QE thresholds, whether the translation should be kept (high QE score), post-edited automatically (moderate score), or reviewed by a human post-editor (low score).

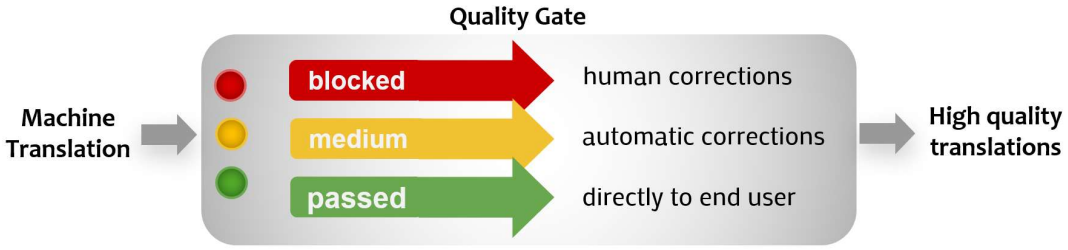


Figure 1: The APE-QUEST Quality Gate

Figures 2 and 3 provide an example of how QE and APE components are applied to the MT output for the out-of-domain sentence in Table 1. As shown in Figure 2, the predicted QE score is 0.75 (the maximum score being 1); more details on how QE scores are calculated are given in Section 3.

Figure 2: Application of QE to MT output

If the threshold of the Quality Gate considers that a score is high enough for the sentence to be post-edited automatically rather than manually (e.g. if the threshold is 0.7), the MT output is submitted to the APE component, as shown in Figure 3. To set an appropriate threshold value for a specific translation environment, tests with human raters should be conducted, as explained in Section 4.

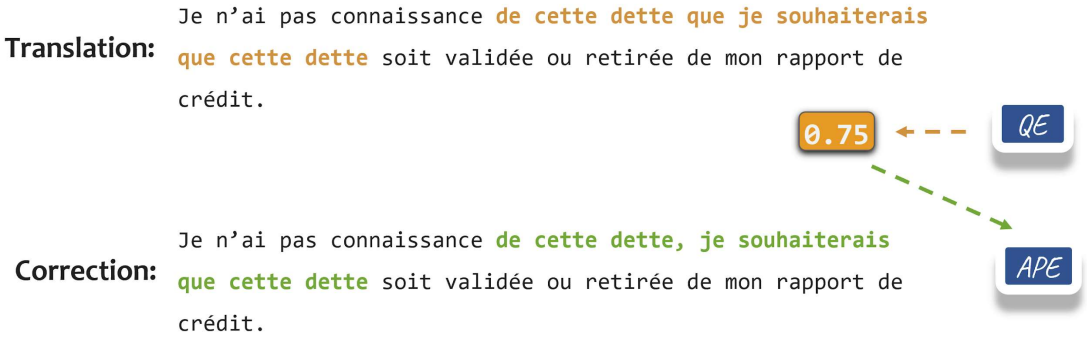


Figure 3: Application of APE to MT output

3. Component training and testing

State-of-the-art QE systems are trained on human post-editing through deep learning (previous systems made use of "classical" machine learning methods).

We selected a dataset of about 10,000 English domain-specific sentences for each of the language pairs in the project, i.e. English-French, English-Dutch and English-Portuguese. The sentences were extracted from texts relating to the legal domain, procurement, and online dispute resolution. Each of them was translated into the target language in question by a generic MT system[1] based on deep learning (neural MT). The MT output was manually post-edited by professional translators. The resulting datasets with post-editings have been made publicly available by the consortium through the ELRC-SHARE repository.[2]

Based on sentence triples consisting of the source sentence, the MT output, and the post-edited MT output, training examples were created for the QE system: they link the triples to the HTER score (Snover et al. 2006). This score expresses the post-editing effort, i.e. the Human Translation Edit Rate (0 means that no post-editing was necessary, and scores are higher if it

1 See Ive et al. (2020).
2 https://elrc-share.eu, name of the resource: "Post-editing corpus English to Dutch/French/Portuguese, legal domain"

was). A training example is shown in Table 2 (the difference between the MT output and its post-edited version is marked in bold).

Two state-of-the-art QE frameworks were applied to create models from the training data: OpenKiwi[3] and TransQuest.[4] When applied to unseen MT output, the models predict a HTER score, which is then converted into a QE score (as opposed to HTER scores, the QE scores get better as they get higher, making them more intuitive for the end-user).

Source sentence	I am not aware of this debt I would like this debt …
MT output	Je n' ai pas connaissance de cette dette **que** je souhaiterais que cette dette …
Human post-edition	Je n' ai pas connaissance de cette dette **,** je souhaiterais que cette dette …

Table 2: QE training example

State-of-the-art APE systems equally involve deep learning. We used the post-edited sentences as training data for the University of Sheffield's systems. Tests showed that it is very difficult to automatically correct MT output both in a precise and exhaustive way: the trained models were either too aggressive or too conservative. In fact, many sentence translations produced by state-of-the-art MT have such a high quality that APE tends to overcorrect them. APE can be set up to avoid such overcorrection, as in the CopyCat system[5] (Ive et al., 2019), but then its corrections may become too limited and insufficient. Therefore, we decided not to include the APE component in the evaluations described in Section 4.[6]

The post-edited data can also be used as MT training data, if an in-house MT system is used in the Quality Gate instead of an external (online) system like eTranslation, which is the MT system developed by the EC. This can further improve the MT system and make human post-editing more efficient.

Besides an MT, QE and APE system, the Quality Gate also has a user interface component. It allows the user to request the MT output for a sentence and shows the QE score and (if applicable) the APE output or existing human translation for the sentence produced by the MT system. This component is an adaptation of the open-source tool MateCat (Federico et al., 2014) by CrossLang.[7]

[3] https://github.com/Unbabel/OpenKiwi
[4] https://github.com/TharinduDR/TransQuest
[5] https://github.com/ImperialNLP/CopyCat
[6] However, there is still potential in APE, as shown by the recent WMT shared task (Chatterjee et al., 2020). APE systems give a clear improvement in case the domains of the training and test data are very close and pre-trained representations are used, i.e. models that are built on a large amount of data in order to learn the behaviour of words in context.
[7] https://github.com/CrossLangNV/MateCat

4. Evaluation

We examined the applicability and benefits of introducing a Quality Gate into the MT workflow in scenarios where MT is used to reduce the translation effort while maintaining an acceptable overall translation quality. Although the translation expert has to decide what *acceptable* means precisely, depending on the context, two use cases are considered conventionally: the first one is *assimilation*, which involves gisting (i.e. getting the *gist* of the translated document) or creating in-house translations for internal communication purposes. In this context, translation errors may occur as long as they do not prevent the end user from getting an approximate idea of the meaning of the text. The second one is *dissemination*, which involves the publication of translated documents for them to be used externally. Here, MT is used as an instrument to achieve high-quality publication-ready translations, meaning that any potential errors produced by the MT system must be corrected by a human professional. In the assimilation case, it is likely that there are more unedited MT translations which satisfy user needs than in the dissemination case.

To evaluate the Quality Gate in a realistic scenario, we consulted with an organisation working on online dispute resolution to set up experiments for both the dissemination and assimilation use cases. We were provided with two English datasets, merely for testing purposes, one in the privacy legislation domain (dissemination), the other one in the more informal domain of consumer complaints (assimilation). Together, the datasets accounted for approximately 1000 sentences.[8] Two specific questions were addressed: (1) When compared to a traditional translation workflow, where all the machine-generated translations are manually checked and edited as needed, does the Quality Gate help improve speed (time to obtain final translation) and reduce cost (number of translations to be manually post-edited), while maintaining an acceptable translation quality? (2) When compared to a purely automatic MT-only workflow, implying no human intervention at all, does the Quality Gate help improve translation quality?

To answer the above questions, a two-step evaluation was conducted for two language pairs (English into Dutch and English into French), based on the availability of human raters (two raters per language pair). During the first phase of the evaluations, the entire datasets were machine-translated using eTranslation and human post-edits (HPEs) were collected for all the translations. Secondly, all MT outputs and their corresponding HPEs were annotated as *Acceptable/Unacceptable* (for assimilation) or *Publishable/Unpublishable* (i.e. acceptable for dissemination) without the human rater knowing the origin (MT or human post-editing) of the translations. This evaluation procedure allowed us to collect figures for measurable criteria such as translation quality, cost and time for different QE thresholds.

Table 3 compares the evaluation criteria of *cost* (i.e. the percentage of sentences that require human post-editing), *time* (i.e. the percentage of time needed to produce the final translation compared to the traditional workflow) and *quality* (i.e. the percentage of sentences considered as acceptable for the given use case) for different scenarios: (i) the traditional translation workflow involving MT where all sentences would go to human post-editing; (ii) the MT-only workflow where no human post-editing is involved; (iii) the Quality Gate

[8] The consumer complaints used for our tests are not publicly available information. Therefore, the example in Table 1 originates from another source, https://catalog.data.gov/dataset/consumer-complaint-database, a dataset from the U.S. General Services Administration, Technology Transformation Service.

workflow where a selection of sentences go to human post-editing based on the QE score assigned to each translation (in Table 3, only the QE thresholds of 0.90 and 0.80 are considered, i.e. each translation falling below this threshold goes to human post-editing). Note that Table 3 relates to the language pair English into Dutch; the results for English into French are similar.

Threshold	Assimilation			Dissemination		
	Cost (%)	Time (%)	Quality (%)	Cost (%)	Time (%)	Quality (%)
Traditional	100.00	100.00	97.67	100.00	100.00	97.89
QE < 0.90	55.52	64.18	83.87	24.21	33.02	94.74
QE < 0.80	9.48	8.80	59.16	3.16	0.75	90.53
MT-Only	0.0	0.0	54.07	0.0	0.0	90.53

Table 3: Cost, Time and Quality measures for traditional, MT-only and Quality Gate workflows for the assimilation and dissemination use cases (English into Dutch)

We observe an important difference between the two use cases. The quality gains compared to an MT-only workflow are much lower for the dissemination use case, because the initial quality of the MT outputs is much higher than for the assimilation dataset. The lower MT quality of the assimilation data can be explained by a domain mismatch between the more informal style of the translated dataset and the MT system that was trained to produce translations for more formal content. Overall, the results of the evaluations show that using the Quality Gate can result in important cost and time savings without strongly compromising the quality of the translation, for both target languages and both use cases. While we used the predicted QE scores to evaluate the current state of the Quality Gate, experiments using oracle QE scores computed on the human post-editing show that there is still room for improving the QE models, meaning there are higher potential gains for both use cases.

5. Conclusions

We described the Quality Gate built within the APE-QUEST project to address the variability of MT output and the domain mismatch between the text to translate and the MT training data. The Quality Gate is a proof-of-concept implementation involving four components (MT, QE, APE, and user interface). It aims to achieve the desired translation quality in a time- and cost-efficient way by using MT and human or automatic corrections. It routes translations to different workflows based on the estimated translation quality: if the MT output has a low QE score (according to a threshold), it is channelled to a human post-editor, if it has a moderate score, it is submitted to an APE system, and in case of a high score, the translation is considered acceptable in the given context.

We created domain-specific post-edited datasets for three language pairs (English into French, Dutch and Portuguese) and made them publicly available through ELRC-SHARE. We trained QE and APE systems using this data. The APE component did not appear to produce

satisfactory results, as neural MT generally produces high-quality output, and it is difficult to produce automatic corrections that are both precise and exhaustive. Therefore, when performing tests on two use cases, we only used the MT system (eTranslation) and the QE system in the Quality Gate. In the first use case, assimilation, which involved consumer complaints, a translation is considered acceptable if it is understandable. In the second case, dissemination, involving text from web pages related to online dispute resolution, a translation is only acceptable if it is publishable. The evaluation shows a clear difference in the results for dissemination and assimilation, caused by the domain mismatch between the MT training data and the consumer complaints. The evaluation evidences the potential of the Quality Gate for reducing cost and time without compromising translation quality.

6. Acknowledgements

APE-QUEST is funded by the EC's CEF (Connecting Europe Facility) Telecom programme (project 2017-EU-IA-0151).

7. References

Chatterjee, R., Freitag, M., Negri, M., & Turchi, M. (2020). "Findings of the WMT 2020 Shared Task on Automatic Post-Editing". In: *Proceedings of the 5th Conference on Machine Translation (WMT)*, pp. 644–657.

Federico, M., Bertoldi, N., Cettolo, M., Negri, M., Turchi, M., Trombetti, M., Cattelan, A., Farina, A., Lupinetti, D., Martines, A., Massidda, A., Schwenk, H., Barrault, L., Blain, F., Koehn, P., Buck, C., & Germann, U. (2014). "The MateCat Tool". In: *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: System Demonstrations*, pp. 129–132.

Ive, J., Specia, L., Szoc, S., Vanallemeersch, T., Van den Bogaert, J., Farah, E., Maroti, C., Ventura, A., & Khalilov, M. (2020) "A Post-Editing Dataset in the Legal Domain: Do we Underestimate Neural Machine Translation Quality?" In: *Proceedings of LREC 2020*, pp. 3692–3697.

Ive, J., Madhyastha, P., and Specia, L. (2019). "Deep copycat networks for text-to-text generation". In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3218–3227, Hong Kong, China, November. Association for Computational Linguistics.

Snover, M., Dorr, B., Schwartz, R., Micciulla, L., & Makhoul, J. (2006). "A Study of Translation Edit Rate with Targeted Human Annotation". In: *Proceedings of Association for Machine Translation in the Americas*, pp. 223–231.

Szoc, S., & Depraetere, H. (2020). "Quality Estimation", In: Jörg Porsiel (Hg., 2020): *Maschinelle Übersetzung für Übersetzungsprofis*. Berlin: BDÜ Fachverlag. Sammelband mit Beiträgen in deutscher und englischer Sprache. 384 Seiten, ISBN: 9783946702092, pp. 198–208.

Van den Bogaert, J., Depraetere, H., Szoc, S., Vanallemeersch, T., Van Winckel, K., Everaert, F., Specia, L., Ive, J., Khalilov, M., Maroti, C., Farah, E., & Ventura A. (2019). "APE-QUEST: an MT Quality Gate". In: *Proceedings of Machine Translation Summit XVII Volume 2: Translator, Project and User Tracks*, pp. 110–111.

Van den Bogaert, J., Depraetere, H., Szoc, S., Vanallemeersch, T., Specia, L., Ive, J., & Khalilov, M. (2020). "APE-QUEST: a Quality Gate for Routing MT". In: *Proceedings of EAMT 2020*, pp. 473–474.

Author Index