Bloomsbury Advances in Translation

Translation, Interpreting and Technological Change

Innovations in Research, Practice and Training

Edited by
Marion Winters,
Sharon Deane-Cox and Ursula Böser

BLOOMSBURY

Translation, Interpreting and Technological Change

BLOOMSBURY ADVANCES IN TRANSLATION SERIES

Series Editor:
Jeremy Munday, Centre for Translation Studies, University of Leeds, UK

Bloomsbury Advances in Translation publishes cutting-edge research in the fields of translation studies. This field has grown in importance in the modern, globalized world, with international translation between languages a daily occurrence. Research into the practices, processes and theory of translation is essential and this series aims to showcase the best in international academic and professional output.

A full list of titles in the series can be found at:
www.bloomsbury.com/series/bloomsbury-advances-in-translation

Titles in the series include:

Celebrity Translation in British Theatre, Robert Stock
Collaborative Translation, edited by Anthony Cordingley and Céline Frigau Manning
Community Translation, Mustapha Taibi and Uldis Ozolins
Corpus-Based Translation Studies, edited by Alet Kruger, Kim Wallmach and Jeremy Munday
Extending the Scope of Corpus-Based Translation Studies, edited by Sylviane Granger and Marie-Aude Lefer
Genetic Translation Studies, edited by Ariadne Nunes, Joana Moura and Marta Pacheco Pinto
Global Trends in Translator and Interpreter Training, edited by Séverine Hubscher-Davidson & Michał Borodo
Institutional Translation for International Governance, Fernando Prieto Ramos
Intercultural Crisis Communication, edited by Federico M. Federici and Christophe Declercq
Islamic State in Translation, Balsam Mustafa
Music, Text and Translation, edited by Helen Julia Minors
The Pragmatic Translator, Massimiliano Morini
Quality in Professional Translation, Joanna Drugan
Retranslation, Sharon Deane-Cox
Sociologies of Poetry Translation, Jacob Blakesley
Systemic Functional Linguistics and Translation Studies, edited by Mira Kim, Jeremy Munday and Zhenhua Wang

Theatre Translation, Massimiliano Morini
Translating For Singing, Ronnie Apter and Mark Herman
Translating Holocaust Lives, edited by Jean Boase-Beier, Peter Davies, Andrea Hammel and Marion Winters
Translating the Poetry of the Holocaust, Jean Boase-Beier
Translating in Town, edited by Lieven D'hulst and Kaisa Koskinen
Translation Solutions for Many Languages, Anthony Pym
Telling the Story of Translation, Judith Woodsworth
Vladimir Nabokov as an Author-Translator, Julie Loison-Charles
What Is Cultural Translation? Sarah Maitland

Translation, Interpreting and Technological Change

Innovations in Research, Practice and Training

Edited by
Marion Winters, Sharon Deane-Cox and
Ursula Böser

BLOOMSBURY ACADEMIC
LONDON • NEW YORK • OXFORD • NEW DELHI • SYDNEY

BLOOMSBURY ACADEMIC
Bloomsbury Publishing Plc
50 Bedford Square, London, WC1B 3DP, UK
1385 Broadway, New York, NY 10018, USA
29 Earlsfort Terrace, Dublin 2, Ireland

First published in Great Britain 2024

Cover design: Elena Durey

A catalogue record for this book is available from the British Library.

A catalog record for this book is available from the Library of Congress.

ISBN: HB: 978-1-3502-1294-7
ePDF: 978-1-3502-1295-4
eBook: 978-1-3502-1296-1

Series: Bloomsbury Advances in Translation

Typeset by Deanta Global Publishing Services, Chennai, India

To find out more about our authors and books visit www.bloomsbury.com and sign up for our newsletters.

Contents

Figures

Tables

Contributors

Khetam Al Sharou is a researcher in machine translation and multilingual crisis communication at the University of Antwerp, Belgium, and an honorary research associate at Imperial College London, England. Her research lies at the intersection of translation studies, computer science and natural language processing, producing work with academic and industrial impact in tool development and user-experience. Her research interests include human-machine interaction and its wider applications, translation pedagogy, didactics in HE and intercultural training in healthcare. Her contributions include several cross-disciplinary projects, working alongside developers and users. Her recent publications focused on machine translation for user-generated content and language technology use in medical settings.

Ursula Böser is Emeritus Professor of Languages and Intercultural Studies of Heriot-Watt University, Edinburgh, Scotland. Her most recent research is in public service interpreting. In particular, she has published on interpreter-mediated interactions with adults and minors in a policing context. She has been a member of EU-funded international and interdisciplinary research projects on police interpreting (ImPLI – improving police and legal interpreting) and CO-Minor-IN/QUEST (cooperation in interpreter-mediated questioning of minors).

Lynne Bowker is Full Professor at the University of Ottawa, Canada, where she holds a cross-appointment between the school of translation and interpretation and the school of information studies. Her teaching and research interests focus on translation technologies, corpus linguistics, terminology and lexicography and specialized translation. She has published on all of these subjects, and her most recent books include *Machine Translation and Global Research: Towards Improved Machine Translation Literacy in the Scholarly Community* (2019) and the open-access book *De-mystifying Translation: Introducing Translation to Non-translators* (2023).

Sharon Deane-Cox is Senior Lecturer in Translation and Interpreting at the University of Strathclyde, Scotland. Her current interdisciplinary research focus

is on how translation mediates memory and trauma on textual, interpersonal, intersemiotic and ethical levels. She has published on retranslation, Holocaust testimony translation, memorial museum audio-guide translation and heritage translation. She recently co-edited the *Routledge Handbook of Translation and Memory* (2022) and is associate editor of the journal *Translation Studies*.

Bart Defrancq is Associate Professor at Ghent University, Belgium, and course leader of the interpreting programmes. His research focuses on simultaneous conference interpreting and police interpreting. He advocates corpus-based research methods and technological tools to support interpreters.

Claudio Fantinuoli is Senior Lecturer at Johannes Gutenberg University, Germany, in the Interpreting Department. His research focuses on augmented human interpretation and machine interpretation. He has published on the design principles of augmented tools for professional interpreters and on the evaluation of speech translation systems.

Damien Hansen is a PhD candidate in translation studies at the University of Liège (CIRTI) Belgium and in computer science at the Grenoble Alpes University (LIG/GETALP) France. The subject of his thesis mainly focuses on literary machine translation, but other and past research interests include computer-assisted literary translation, the evolution and reception of translation technologies, as well as machine translation for game localization and the semiotics of video games.

Dorothy Kenny is Full Professor of Translation Studies at Dublin City University, Ireland. Her current research interests include corpus-based analyses of translation and translator style, literary applications of machine translation and approaches to the teaching of translation technology. Her most recent book is the edited volume *Machine Translation for Everyone: Empowering Users in the Age of Artificial Intelligence* (2022). She is co-editor of the journal *Translation Spaces*.

Jeffrey Killman is Associate Professor of Spanish at the University of North Carolina at Charlotte, USA. His research focuses primarily on the usability of translation technologies and on the challenges faced by translators and interpreters providing their services in specialized fields such as the law. He has published a series of articles and book chapters in these areas and recently co-edited a special issue on legal translation and technology (*Revista de Llengua*

i Dret, *Journal of Language and Law*). He is also the current vice president of the American Translation and Interpreting Studies Association (ATISA).

Maarit Koponen is Professor of Translation Studies at the University of Eastern Finland. Her research interests focus on the use of machine translation and other translation technologies both in professional translation workflows and in non-professional settings. She has published on topics such as machine translation post-editing and the use of machine translation for information purposes. She is one of the co-editors of the volume *Translation Revision and Post-editing: Industry Practices and Cognitive Processes* (2021).

Mary Nurminen is University Instructor at Tampere University and a post-doctoral researcher at the University of Eastern Finland. Her research focuses on the different ways non-translators use raw machine translation and her 2021 thesis was titled 'Investigating the Influence of Context in the Use and Reception of Raw Machine Translation'. Currently Mary is working in a project focused on investigating how linguistic minorities in Finland access, use and interact with knowledge in different languages.

Hanna Risku is Professor for Translation Studies and Head of the Research Group Socio-Cognitive Translation Studies (socotrans) at the University of Vienna, Austria. Her research areas include translation and cognition, translation workplace and network research, and translation expertise. She has published on sociological and cognitive approaches in translation studies and the ethnographic study of translation processes. She has worked at different universities in Finland, Sweden, Denmark and Austria. In 2023, she was CETRA chair professor at KU Leuven and DOTTSS guest professor at Tampere University. She is a member of the Finnish Academy of Science and Letters.

Regina Rogl is Post-doctoral Researcher at the Centre for Translation Studies, University of Vienna, Austria, where she is a member of the Research Group Socio-Cognitive Translation Studies (socotrans). Her research lies at the intersection of translation studies, science and technology studies, and media studies. Her main research interests include translation and digital media, non-professional translation/interpreting, and workplace research. She has published on field research in translation and interpreting, volunteer translation/interpreting, and online translation communities.

Helena Snoeck is Doctoral Researcher at Ghent University, Belgium. She previously worked on the EABM project in the area of technology for simultaneous interpreting and is currently preparing a PhD on police interpreting.

Marion Winters is Senior Lecturer in Translation Studies at Heriot-Watt University, Edinburgh, Scotland. She is founding editor of *New Voices in Translation Studies* and a professional member of the German and Irish translators' associations (BDÜ, ITIA). Her main research interests lie in the areas of corpus-based approaches to translator style, translation and autobiography, and more recently how the use of machine translation might affect literary translators.

Introduction

Technological Innovations and Their Implications in Translation and Interpreting Research, Practice and Training

Marion Winters, Sharon Deane-Cox and Ursula Böser

The observation, made over thirty years ago, that 'our existence is technologically textured' (Ihde 1990: 2) is perhaps all the more salient today; technology is imbricated in every aspect of our private, social, cultural and professional lives, shaping how we act in and interact with the various spaces we inhabit. The seemingly inexorable reach of technology, especially with regard to language and communication, is then reason enough for us to pause and take stock of how recent technological changes have been impacting the discipline of translation and interpreting studies across research, practice and training. Indeed, things are moving apace: the 'technological turn', posited by Michael Cronin in 2010, has since been fully realized, to the extent that 'technology is now an integral part of the discipline of TS . . . and will continue to be so for years to come' (Jiménez-Crespo 2020: 332). Accordingly, the disciplinary significance and embeddedness of technology sets the backdrop for this book and its exploration of how and to what effect the processes, the products, as well as our conceptual and applied understanding of translation and interpreting, are all being shaped by innovations in software, artificial intelligence and the increasing distribution and portability of digital services.

The initial momentum for this book stemmed from the very successful CIUTI (Conférence Internationale Permanente d'Instituts Universitaires de Traducteurs et Interprètes) conference on 'Translation and interpreting in an era of demographic and technological change', hosted by the Centre for Translation and Interpreting Studies in Scotland (CTISS) at Heriot-Watt University, Edinburgh, in May 2018. In turn, the richness and diversity of perspectives that characterized the event now inform the present chapters, with international contributions from across the world that bring to light insights from researchers,

trainers and practitioners who span different career stages and language pairings, and who hold diverse stakeholder positions with regard to technology. Together, these contributions reflect prominent trends in the field today, as approached from the standpoints of research, practice and training, and with a focus on shifting human-machine interactions (Part I), shifting methods and models (Part II) and shifting pedagogies (Part III). At the centre of each chapter is the drive to explore the implications that technological developments are having and will continue to have on translation and interpreting activities.

Contributions to this volume were written before the start of the pandemic and it has not been possible to cover related developments here. Undoubtedly, the sudden adaptations and changes which came about in the wake of an epidemic which enforced remoteness on all aspects of social life will provide fertile ground for future research. Indeed, the volume endeavours to play its part in advancing the growing body of research on translation technologies and its attendant insights into areas as diverse as literary translation, risk assessment, the developments from statistical machine translation (SMT) to neural machine translation (NMT) and training.

The content of the volume is weighted towards translation technology. This is a reflection of current patterns of usage and the research ecology of the discipline. Yet the changing technological basis of social practices might be well illustrated by the fact that technology was instrumental in the inception of the simultaneous mode (SI) of interpreting for a wider audience in the 1920s. A more recent 'technological turn' has been unfolding with advances in digital Information and Communication Technology (ICT), and the increasing mobility of its usage due to distributed devices.

In the field of translation technologies, the research foci today remain on computer-assisted translation (CAT), notably on the main CAT tool of translation memory systems, and machine translation (MT), in particular due to the recent significant advance to neural technology, while a growing body of research emerges on human-machine interaction.

In interpreting technologies, notwithstanding a degree of flux regarding terminology in an evolving field, a broad distinction can be made between machine interpreting (MI), technology-mediated interpreting and technology-assisted interpreting (Braun 2019). It is the latter on which contributions in this book focus. This reflects the relative significance of technology-supported interpreting in the wider context of the impact of technology on aspects of interpreting. Computer-assisted interpreting (CAI) is defined as 'a form of oral translation in which a human interpreter makes use of computer software

designed to support and facilitate some aspects of the interpreting task with the goal to increase quality and – to a minor extent – productivity' (Fantinuoli 2018: 3). It includes digital tools which assist with preparation, pre-interpreting information extraction and management. To varying degrees, the use of functionalities afforded by search engines as well as text mining or glossary building tools is now an established part of the interpreter's work. By contrast, computer-based on-site real-time support for the interpreter is still in its early development stages.

The recent 'technological turn' has raised a set of concerns and research questions regarding its various manifestations in an interpreting context. An issue that has particular bearing on CAI is the integration of real-time support and its impact on cognitive load and the overall performance (Frittella 2022; Fantinuoli 2018; Fantinuolo 2017). Some recurring issues in other fields of application are the physiological, psychological and cognitive impact of technologically mediated remoteness on the interpreter and the quality of their performance (Roziner and Shlesinger 2010; Mouzourakis 2006; Moser-Mercer 2003). Most fundamentally, the quality of sound and visuals, and their perception by interpreters, are of importance in this context (Braun 2020). The alteration of interactional dynamics features particularly in settings for remote consecutive or dialogue interpreting (Licoppe,Verdier and Veyrier 2018; Napier, Skinner and Braun 2018; Braun 2017). The 'technological turn' will also affect users of interpreting services and the options that will be available in different settings (Price et al. 2012). On the part of interpreters, futureproofing will undoubtedly include preparing for a work context that is increasingly shaped by technology and that will require the development of new skill sets (Davitti and Braun 2020; Alley 2012).

Technology-assisted translation and interpreting as well as machine translation and interpreting have the potential to improve both translation and interpreting performance and efficiency. However, such opportunities come with risks. The challenge to the translation and interpreting community is one of constructive engagement with the conceptualization and realization of a social practice that may become increasingly technology-mediated or assisted, but must not be driven by either technology or a primacy of cost-saving concerns.

There is always the danger that discussions of translation technologies, their workings, applications and economic benefits, might supersede considerations of the users themselves, and so in this volume emphasis has been placed where possible on the human side of the technological encounter; and by rendering the user more visible, it is hoped that they might become further empowered.

The very need for empowerment stems partly from the well-reported negative consequences of technological advances, such as reduced pay rates (see, for example, Vieira 2020), the requirement to use specific translation memory software, their increasing complexity (see Schneider, Zampieri and van Genabith 2019) and unsatisfactory product design that ignores translators' needs (Olohan 2020). Technology now also appears to be encroaching on areas that were previously considered unconquerable: MT is becoming an increasingly viable tool in literary translation (see Toral and Way 2014) and in other creative texts, alongside CAT tools (see, for example, Hadley et al. 2022).

While there has been resistance among the translator community to MT, there are also signs of increasing engagement now, not least in literary contexts. This shift was captured in *Counterpoint*, the magazine of the Conseil Européen des Associations de Traducteurs Littéraires (CEATL), which published a special feature on machine translation and literature and included reports from literary translators on their use of MT (Oeser 2020) and translation memories (Zakrajšek 2020). Although still in the minority, members of the literary translation community have clearly started to explore translation technologies, including MT, and a UK survey of literary translators further suggests that something of a sea change is taking place, perhaps along generational lines: 'Attitudes towards technology are extremely complex, however, having undertaken technology training resulted in more positive attitudes and higher levels of confidence with technology. The youngest generation of respondents was also the most positive and confident with technology' (Ruffo 2020). It may well be the case that the extent of such explorations is greater than we might think. At a recent seminar on the use of MT in literary translation held in the German Interpreters' and Translators' Association's (BDÜ) seminar series *Die Zukunft ist jetzt . . .* (Winters and Kenny 3 March 2022), the audience consisted in large part of literary translators at different points in their careers. While there seemed to be little appetite among the audience for using MT to actually translate literature, over 80 per cent of the 125 participants who answered the poll question said they used MT to assist them in their literary translation. Even 30 per cent responded that this was often the case. Of course, this provides just one snapshot, but it nevertheless serves to nuance the idea that resistance to MT is the default position.

In turn, the shifting position of professional translators towards engagement with translation technologies, including MT, adds weight and urgency to the value of foregrounding the role of the translator in our research, while simultaneously offering a counterpoint to the productivity-oriented agendas of

publishers and agencies. This view is shared within the translation profession itself:

> I for my part, shall continue to avail of electronic tools but, being conscious of the dangers arising to my artistic autonomy, only to spot-check and not over a wide area of text. As an organised community, however, we must strive to resist and reject any attempt by publishers (some of whom are already rumoured to contemplate steps in that direction) to transform, as part of a cost-minimising exercise, from machine-aided human translation to human-aided machine translation that which is rightly our work. We will have to be the Luddites of the humanities! (Oeser 2020)

Shining a more nuanced light on the capabilities and challenges of MT from a user-driven perspective has the potential to lead to a number of beneficial outcomes, not least a better understanding of the varied and complex labours of the translator and other stakeholders. As Landes (2020) argues, there is scope for the public perception of MT to come closer to reality in order to offset any further devaluation of human translation.

A significant number of chapters in this volume align themselves with this turn towards the human agent in various research, practice and pedagogical contexts. The volume opens in Part I with an emphasis on the shifting boundaries of human and technology interaction. Rogl and Risku explore in Chapter 1 how our understanding of the role of tools in the under-researched area of translation project management can be enhanced by approaching processes from a socio-cognitive perspective. The findings of a longitudinal study provide the basis for a new conceptualization of those tools as cognitive artefacts and as socio-cognitive boundary objects, thereby opening up valuable new lines of enquiry that emphasize, among other dynamics, how technology use is contingent on the evolving context of its users and how it can facilitate collaboration among actors.

The cognitive load of the interpreter becomes the focal point in Chapter 2 by Defrancq, Snoeck and Fantinuoli. The authors draw on a study on interpretation quality assessment to investigate how performance, with specific reference to number interpretation, is impacted by the support of a CAI tool. In parallel, the chapter also sheds new methodological light on the viability of using the fundamental frequency of the participants' voices to measure cognitive load.

Kenny and Winters sustain focus on the experience and actions of the technology user in Chapter 3 but move us into the innovative space where MT and literary translation intersect. Starting from a critical review of how customization, personalization and style have hitherto been framed in reference

to literary MT, the authors make the compelling argument that approaches to style in this context should be informed by literary studies. They then go on to characterize literary post-editing as a kind of downstream translator-specific personalization, exemplifying the approach with reference to one literary translator's use of DeepL.

Computer-assisted literary translation (CALT) is also the focus in Chapter 4 by Hansen. He draws parallels between resistance of the literary translators against technological developments in the early days when computers became a widely accessible commodity and more recent resistance against CAT tools and MT. Hansen takes a broader look at CAT tools, beyond their translation memory function, and argues for their usefulness in literary translation as tools for text analysis and corpus exploration.

Methods and models frame Part II of the volume, where Chapter 5 presents a nuanced consideration of the potential risks and mitigating factors associated with the end-user consumption of information delivered via raw MT in organizational settings. In concrete terms, Koponen and Nurminen offer a framework for assessing and managing the risks around raw MT use and reception; the specific domain of patent machine translation is used as a testbed, but the best practice recommendations are generalizable across organizations with interlingual communication needs, and they also serve to advance limited scholarly discussion in this area.

Killman's comparative study of statistical and neural machine translation output in Chapter 6 then turns our attention towards MT quality assessment in a Spanish-English legal context. Human evaluation of terminology points towards a small but observable advantage in terms of accuracy on the part of SMT, thereby shining new and perhaps unanticipated light on the capabilities of NMT, especially in relation to context-bound terms and in contrast to its predecessor. The chapter also suggests the need for more and varied assessments of NMT performance in legal settings in order to better establish where and how the technology might play a more integral role in legal translation workflows.

Questions around the precedence of NMT carry over into the pedagogically oriented reflections of Part III. As Al Sharou illustrates in Chapter 7, it is open-source SMT technology that lends itself most productively to student training in an Arabic-speaking higher education setting. The author reports on a curriculum-building endeavour designed to explore if and how translation students might acquire competence in building customized MT engines, alongside developing their confidence as autonomous users of the technology. In highlighting the

range of challenges and opportunities encountered, the chapter also serves as a call to empower translators through access to supported training.

The volume is then rounded off by Bowker in Chapter 8 who further foregrounds the advantages of teaching machine translation literacy, in this instance to non-translation students at a Canadian university. The author charts the fast-paced development of machine translation and notes that, in view of the increasingly prevalent use of MT tools, students should be provided with the pedagogical scaffolding which enables them to become critical, informed and responsible users of that technology. The detailed depiction of the module, including its content, pedagogical underpinnings and student feedback, attests to its potential for replicability and to its educational and social relevance beyond its original starting point.

It is safe to assume that translation and interpreting technologies will play an increasingly central, salient and embedded role in language service provision and training, as well as in allied- and non-professional settings. The seemingly uninterrupted forward momentum of those technologies will undoubtedly yield a wealth of new research directions, along with the attendant (re)conceptualization and recommendations they bring to the discipline, to industry, and to education more broadly. The angles and developments captured in this volume will hopefully go a long way to driving and informing future work in the area, not least in ways that serve a more pro-human and pro-social understanding of the relationship between user and technology.

References

Alley, E. (2012), 'Exploring Remote Interpreting', *International Journal of Interpreter Education*, 4 (1): 111–19.

Braun, S. (2017), 'What a Micro-analytical Investigation of Additions and Expansions in Remote Interpreting Can Tell Us about Interpreter's Participation in a Shared Virtual Space', *Journal of Pragmatics*, 107: 165–77. https://doi.org/10.1016/j.pragma.2016.09.011

Braun, S. (2019), 'Technology and Interpreting', in M. O'Hagan (ed.), *The Routledge Handbook of Translation and Technology*, 271–88, London: Routledge.

Cronin, M. (2010), 'The Translation Crowd', *Tradumática*, 8. https://doi.org/10.5565/rev/tradumatica.100

Davitti, E. and Braun, S. (2020), 'Analysing Interactive Phenomena in Video Remote Interpreting in Collaborative Settings:Implications for Interpreter Education',

The Interpreter and Translator Trainer, 14: 279–302. https://doi.org/10.1080/1750399X.2020.1800364

Fantinuoli, C. (2017), 'Computer-Assisted Interpreting: Challenges and Future Perspectives', in G. C. Pastor and I. Duran-Munoz (eds), *Trends in E-tools and Resources for Translators and Interpreters*, 153–74, Leiden: Brill.

Fantinuoli, C. (ed.) (2018), *Interpreting and Technology: The Upcoming Technological Turn*, Berlin: Language Science Press.

Frittela, F. M. (2022), 'CAI Tool-Supported SI of Numbers: A Theoretical and Methodological Contribution', *International Journal of Interpreter Education*, 14 (1): 32–56. https://doi.org/10.34068/ijie.14.01.05

Hadley, J. L., K. Taivalkoski-Shilov, C. S. C. Teixeira, and A. Toral (eds) (2022), *Using Technologies for Creative-Text Translation*, London: Routledge.

Ihde, D. (1990), *Technology and the Lifeworld: From Garden to Earth*, Bloomington: Indiana University Press.

Jiménez-Crespo, M. A. (2020), 'The "Technological Turn" in Translation Studies: Are We There Yet? A Transversal Cross-Disciplinary Approach', *Translation Spaces* 9 (2): 314–41. https://doi.org/10.1075/ts.19012.jim

Landes, D. (2020), 'Maschinelle Übersetzung und die Lage der Übersetzungsbranche', *DVÜD*, 18 October. Available online: https://dvud.de/2020/10/maschinelle-uebersetzung-und-die-lage-der-uebersetzungsbranche-teil-1/

Licoppe, C., M. Verdier,and C. A. Veyrier (2018), 'Voice,Power and Turn-taking in Multilingual, Consecutively Interpreted Courtroom Proceedings with Video Links',in J. Napier, R. Skinner,and S. Braun (eds), *Here or There: Research on Interpreting via Video Link*, 299–322. Washington, DC: Gallaudet University Press.

Moser-Mercer, B. (2003), 'Remote Interpreting: Assessment of Human Factors and Performance Parameters,' Communicate! Summer 2003. http://aiic.net/ViewPage.cfm?page_id=1125

Mouzourakis, P. (2006), 'Remote Interpreting: A Technical Perspective on Recent Experiments', *Interpreting*, 8 (1): 45–66. https://doi.org/10.1075/intp.8.1.04mou

Napier, J., R. Skinner,and S. Braun (2018), 'Interpreting via Video Link: Mapping of the Field',in J. Napier, R. Skinner,and S. Braun (eds), *Here or There: Research on Interpreting via Video Link*, 11–35. Washington, DC: Gallaudet University Press.

Oeser, H.-C. (2020) 'Duel with DeepL. Literary Translator Hans-Christian Oeser on Machine Translation and the Translator's Voice', *Counterpoint*, 4: 19–123. Available online: https://www.ceatl.eu/wp-content/uploads/2020/12/Counterpoint_2020_04_article_05.pdf

Olohan, M. (2020) 'Technology, Translation', in M. Baker and G. Saldanha (eds), *Routledge Encyclopedia of Translation Studies*, 574–9, London: Routledge.

Price, E., E. Pérez-Stable, D. Nickleach, M. López,and L. Karliner (2012), 'Interpreter Perspectives of In-person, Telephonic, and Videoconferencing Medical Interpretation in Clinical Encounters', *Patient Education and Counseling*, 87 (2): 226–32.

Roziner, I. and M. Shlesinger (2010), 'Much Ado about Something Remote: Stress and Performance in Remote Interpreting', *Interpreting*, 12: 214–47. https://doi.org/10.1075/intp.12.2.05roz

Ruffo, P. (2020), *In-Between Role and Technology: Literary Translators on Navigating the New Socio-Technological Paradigm*, PhD Thesis, Heriot-Watt University, Edinburgh.

Schneider, D., M. Zampieri, and J. van Genabith (2019) 'Translation Memories and the Translator', *Babel*, 64 (5/6): 734–62.

Toral, A. and A. Way. (2014), 'Is Machine Translation Ready for Literature?', *Proceedings of Translating and the Computer 36. Translating and the Computer*, 36: 174–6, London. Available online: https://aclanthology.org/2014.tc-1.23.pdf

Vieira, L. N. (2020), 'Automation Anxiety and Translators', *Translation Studies*, 13 (1): 1–21. https://doi.org/10.1080/14781700.2018.1543613

Winters, M. and D. Kenny (2022), 'DeepL – Feind oder Freund der Literaturübersetzung', Paper presented at the BDÜ seminar series *Die Zukunft* ist *jetzt* …, 3 March.

Zakrajšek, K. (2020), 'CAT Tools – The Literary Translator's New Assistant?', *Counterpoint*, 4: 24–5. Available online: https://www.ceatl.eu/wp-content/uploads/2020/12/Counterpoint_2020_04_article_06.pdf

Part I

Shifting Boundaries of Human and Technology Interaction

1

Cognitive Artefacts and Boundary Objects

On the Changing Role of Tools in Translation Project Management

Regina Rogl and Hanna Risku

Introduction

With translation services increasingly being delivered in global production networks (Abdallah 2012) using industrial-like business models, translation processes and the demands on the participants and their work environments have changed significantly. Mass production, disproportionately larger order volumes and the growing standardization of work processes are all symptomatic of this industrialization of translation and are accompanied by outsourcing and the subcontracting of orders in increasingly complex production chains (Rodríguez-Castro 2013: 38). Translation production processes have likewise become more technology-intensive, thus making some phases more automated and facilitating or encouraging work in decentralized virtual teams (Stoeller 2011). Moorkens (2021) views this scenario as typical of current neoliberal production policies that have also led to an adoption of the model of digital taylorism (Moorkens 2020) in the translation industry.

As Rodríguez-Castro (2013: 39–40) notes, project managers (PMs) play a key role in this complex and seldom conflict-free configuration:

> Global virtual teamwork has resulted in new team dynamics and a work environment characterized by a lack of interpersonal relationships, a lack of face-to-face communication, a lack of social events to build trust, and a lack of close supervision, among other factors. The role of the PM has become increasingly important as the 'glue' that holds the team together.

Although not hitherto a major topic in translation research, scholars are now increasingly exploring the crucial role of PMs in steering and framing translation processes (Alonso 2016; Dunne and Dunne 2011; Olohan and Davitti 2017; Plaza-Lara 2020; Rodríguez-Castro 2013; Sakamoto 2018, 2019, 2021; Sakamoto and Foedisch 2017). This role generally requires them to perform a 'balancing act' (Olohan and Davitti 2017), serving both as the interface between clients, translators, revisers and other participants in the network and handling diverse tasks like administration, project, quality and account management, or sales (Rodríguez-Castro 2013: 39).[1] In other words, they have to mediate between different expectations, information requirements and loyalties (see Risku 2016: 219–41 for a discussion of PMs' affiliations to different communities).

Prior research on translation project management (TPM) has therefore often focused on topics like communication and cooperation in heterogeneous, usually virtual, production networks (Rodríguez-Castro 2013; Sakamoto and Foedisch 2017; Stoeller 2011; Tsvetkov and Tsvetkov 2011), the role of interpreting agencies as 'third clients' of interpreters (alongside the parties involved in an interpreting situation [Ozolins 2007]), or (mis)trust between the different participants (Alonso 2016; Dong and Turner 2016; Olohan and Davitti 2017; Risku, Milošević and Rogl 2021).

With interaction and work processes in TPM now increasingly technology-based or even automated, researchers are placing greater emphasis on the social and structuring role of technology (e.g. Kenny 2017; O'Brien 2012; Olohan 2020). So far, however, they have mostly approached this topic from the perspective of translators, who seem to experience uncertainties above all in areas where technology is replacing the mediating role of PMs, such as order processing via online marketplaces or crowdsourcing systems (Sakamoto 2018). A focus group study by Alonso (2016) also shows that communicating solely via technology in translation production networks creates risks of asymmetric information flows, non-communication, mutual lack of understanding or unfulfilled expectations. This is something to keep in mind as automated project management approaches that strive to minimize or avoid any human involvement – such as low/no touch and lights-out project management (Esselink 2020) – start to gain a foothold in the language industry.

Past findings seem to indicate that most translators and PMs view the use of technology in the translation process as positive (Koskinen and Ruokonen 2017: 13; LeBlanc 2017: 47; Marshman 2014: 389; Sakamoto 2019: 62). Marshman (2014) also delivers relevant insights into its coordinative and cooperative role in translation management by investigating the extent to which translators feel

that the use of language technologies gives them greater or less control over their work. The translators who participated in her study indicated that the use of technology gave them greater control in many areas – such as work volumes, quality, order planning or work methods – and only reported a reduction in control when it came to earnings. However, their assessments were least clear for areas that extend beyond their own work, including cooperation with customers and clients, that is also PMs (Marshman 2014: 396–7). They felt that, while language tools helped them to adapt to client expectations and standardize their work processes (thus saving time but also limiting their flexibility), they also prevented them from being able to offer certain services if they did not have the corresponding tools. They also lamented having to purchase large numbers of different tools such as specific translation memory or localization software in order to be able to build up a broad customer base. When such tools were supplied by clients, they were frequently found to be lacking in quality and forced translators to adhere to externally prescribed procedures. Like Alonso (2016), Marshman's study (2014: 397) also pointed to the insufficient exchange of information with clients. The participating translators were likewise frustrated by the different understandings of the purpose of tools held by the various actors in the translation process. While they themselves see language technologies as a way of ensuring quality, clients see them primarily as a means of improving cost-efficiency (Marshman 2014: 399). Marshman (2014: 391) attributes the discrepancies in translators' views to their diverse work realities, which make it more difficult to generalize.

This further confirms the importance of workplace studies (e.g. Risku, Rogl and Milošević 2020) for investigating the role of technology in actual practice. Little is still yet known about the coordinative and cooperative role of technology in heterogeneous translation networks, especially from the perspective of PMs, that is, those whose actual task is to mediate between the different expectations and demands of the various actors. Furthermore, recent studies have tended to exclude analogue artefacts (such as printed checklists, handwritten notes or physical filing systems) and focus on the role of digital technologies. Yet cognition-based translation research shows that such artefacts have not been totally replaced by digital alternatives and can still play an instrumental role (Risku 2016). Last but not least, the rapid developments in the language industry – both in technology and in work and management practices – underscore the need for longitudinal studies, which are still rare in our field.

We sought to combine all these issues in a longitudinal research project that studied long-term developments in TPM from 2002 to 2014 (see also Risku

2016). In this chapter, we present our corresponding findings on the developing role of artefacts in translation project management as perceived by PMs. We concentrate on two specific research questions:

1. What role do artefacts play in the PMs' daily work practices in the translation agency under study and how did this role evolve from 2002 to 2014?
2. How can the role of these artefacts and the shifts in their use be conceptualized?

The overarching theoretical framework for our study was based on situated approaches to translation and cognition. In what follows, we therefore start by explaining how such conceptualizations can enhance our understanding of the role of tools in TPM. In the course of our data analysis, our attention was repeatedly drawn to one particular aspect that we suspected might yield particularly rich insights, namely the social, coordinative and arguably conflictual role of tools shared by heterogeneous groups. To do justice to this aspect, we decided to complement our situated understanding of artefacts with a concept that is still relatively new in translation studies, namely that of *boundary objects* (Bowker and Star 1999; Star 2010; Star and Griesemer 1989; Star and Ruhleder 1996). By discussing possible ways of applying this concept to the tools encountered and analysed in our study, we seek to provide a more detailed understanding of the socio-cognitive functions of artefacts and their coordinative and cooperative role in translation production networks. In doing so, we also strive to explain the reasons for the changes in the use of tools based on the socio-technical and organizational developments in the translation agency under study.

Theoretical framework: Situated cognition and cognitive artefacts

The special role attributed to tools in the situated cognition approach is tied to developments in cognitive theories. From the 1980s onwards, cognitive science and related fields grew increasingly dissatisfied with the notion of human thought as computation. In the decades that followed, alternative approaches were developed, including the situated views on cognition (for an overview from a translation studies perspective, see Risku and Rogl 2020). These maintain that the environment in which human cognition takes place and interaction with it have to be considered in order to understand intelligent behaviour because

cognition is not an exclusively mental process (i.e. something that occurs solely 'in the head') but is instead produced through the interaction of brain, body and world (see Clark 1998).

Situated views of cognition emphasize that intelligent behaviour is embodied (enacted in bodily movements and preconditions), embedded (interactively coupled with the environment), enacted (constituted in and by action) and extended (partly located in the environment, especially as external tools) (Clark and Chalmers 1998). To understand cognition, we have to look beyond what goes on in individual brains or individuals and think of it as an 'interaction effect' (Robbins and Aydede 2009: 6). Cognitive processes span the boundaries that separate individuals from their social and material environment. We offload cognitive work onto the environment and thus keep internal processing to a minimum (Clark 1998). Instead of trying to remember, retrieve, represent and process information, we store and manipulate parts of the cognitive process physically in the environment or our bodily activities. This saves us from having to remember all the details or complex situations and allows us to rely on the environment 'as its own model' (Brooks 1991: 140) and 'as an outside memory' (O'Regan 1992: 461).

The use of external tools is an important factor in situated views of cognition. Artefacts are used to reconfigure the cognitive space (Suchman 2007) and serve as 'amplifiers of intelligence and repositories of achieved knowledge and wisdom' (Clark 1998: 244). Suchman's (2007: 13) descriptions of plans also apply to artefacts: they are 'cultural resources produced and used within the course of certain forms of human activity'. Artefacts also have a strong social dimension: they are formed and embedded in complex systems with multiple actors and objects.

Artefacts that are used as environmental offloads are often referred to as 'scaffolds' due to their ability to support cognitive activities (see Clark 1998). Scaffolds give the volatile, dynamic, non-linear and ephemeral mental processes an external structure to lean on, which not only facilitates complex problem solving but also helps to coordinate social processes and explain one's thinking both to oneself and to others. Hutchins' (2005) concept of objects as 'material anchors' that stabilize mental structures also points to this phenomenon.

Some artefacts enable activities that would otherwise not be manageable; others can simply be ad hoc reminders like writing a note or deliberately misplacing an object so as not to forget something (e.g. placing the garbage bag at the door). They are also often needed as scaffolds when learning new abilities. These can range from rhymes to checklists and dictionaries that are consulted

extensively in the learning phase but then slowly become internalized, thus removing the need for the external prompt (the so-called 'ladder' function of artefacts; see also Wittgenstein 1990).

We continue to include specific artefacts in our cognitive processes because of their ability to extend the scope of our activities (Clark 1998: 193). Cognitive supports like calculators, databases, browsers, translation memories and TPM systems enable processes that exceed the capacities of the human brain, such as remembering large amounts of detail, retrieving information exactly as it was first learned and stored or keeping an overview of a complex system with many interdependencies or dynamic subsystems.

Clark (1998: 200) maintains that this trading of external representation against labour- and time-intensive internal computation is not a side-effect of occasional tool use but one of the very foundations of human intelligence. We are born into a social world in which we manipulate, categorize and label the rich environment around us and benefit cognitively from interacting with it. The role given to artefacts in the cognitive process thus also emphasizes the cultural and social dimension of situated cognition: situatedness refers not only to the here and now of cognition but also to artefacts as 'cultural resources' (Suchman 2007: 13) that incorporate cultural and social knowledge.

Case study and method

To investigate the role of artefacts in TPM, we conducted an ethnographic longitudinal case study in an Austrian translation agency. Its typical customers are small and medium-sized companies with vast translation needs, mainly for technical translation and into multiple languages. The agency employed four PMs in 2002, nine in 2007 and thirteen in 2014. In 2002, all employees still both translated in-house and managed outsourced translation projects, whereas by 2007, all translation services were outsourced, and the in-house employees concentrated on TPM.

We collected the data for our study (see Table 1.1) during extended observation periods in the aforementioned years. During our on-site visits, we shadowed the PMs at their workplaces, observed and documented their actions (physical and computer-mediated actions, bodily movements), wrote down their explanations of what they were doing (and how or why), took notes on how they interacted with their colleagues, clients and vendors (both in the physical world and via digital communication) and made an inventory of the physical and virtual tools

Table 1.1 Overview of data gathered for the case study

Year	Duration of field stay (in weeks)	Documented observation hours	Interviews
2002	4	47.5	during observation hours
2007	1	17.5	during observation hours
2014	4	170.75	10 additional interviews (26.4 hours)

they used. We transcribed our notes after each session and supplemented them with example screenshots of the various software applications.[2]

These workplace observations were augmented by qualitative semi-structured interviews with the PMs and CEOs. This allowed us to collect a combination of elicited and (mostly) non-elicited data that we hoped would not only provide us with the PMs' (potentially more abstract) rationalizations of their (inter)actions but also grant us insights into their actual actions and reactions in concrete work situations. While it is often difficult for participants to explicate something they perceive as mundane, everyday or self-evident in interviews, the observation setting allowed us to pinpoint (typical and atypical) situation-specific issues as they arose.

We analysed the interview transcripts and observation protocols using the qualitative content analysis method proposed by Gläser and Laudel (2009), beginning with a set of deductive categories derived from Risku, Windhager and Apfelthaler (2013) and adding further inductive categories as we went along. To retrace the roles and meanings that artefacts acquired in the different situations and activities observed, we conducted an artefact analysis (Lueger and Froschauer 2018). In this analysis, which is of particular relevance here, we focused above all on documenting the tools used by PMs (individually or collaboratively; in typical or atypical work processes) and how they talked about them. This provided us with valuable insights into the variety of meanings and contexts of use that the different artefacts could acquire in a translation production network.

Findings: Macro trends and technological change

We do not seek in this chapter to identify new technological trends in the language industry. Instead, we use our comparison of the data from the different observation periods to explore the social context of technological change in

greater depth. We thus focus not only on how the use of tools evolved over time but also on the organizational developments or policies that caused, encouraged or hindered technological change, and how users adapted to such changes.

From 2002 to 2007: Digitalization and standardization

The developments between 2002 and 2007 reflect a growing digitalization and standardization of the agency's work processes. These went hand in hand with major organizational changes: the agency expanded, hired more staff, was restructured, moved to larger premises and increasingly outsourced its translation and proofreading tasks. This was accompanied by a clear digital transformation, in particular in the (amount of) digital artefacts used in TPM. As the agency grew, processes that had previously been handled by a single PM were now shared (e.g. to make it easier for PMs to cover for colleagues who were on holiday or sick leave), thus requiring more coordination among staff. Tasks that staff had previously handled in their own particular ways and on paper were (also) digitalized in the new standardized work processes. The cardboard folders that had hitherto held all documentation on a project now only contained printouts of the digital workflow (emails, purchase orders, etc.).

The rapid developments in software were evident in all areas of the agency's work. The search for new translators shifted to their web presence and to online translator networks. Contacts to translators were standardized and recorded in an in-house database. Order volumes for technical translations rose sharply. Source texts became more multimedia and were delivered in a variety of file formats, increasing the relevance of desktop publishing and conversion tools. To meet these demands, the agency intensified its cooperation with an affiliated technology company which supplied customized tools for the PMs.

A customized project management tool was introduced and used to process and manage all orders. At this stage, the project management system was only used internally and not (yet) shared with clients, translators and reviewers. Its introduction led (as intended) to a standardization of the agency's work processes. Our workplace observations nonetheless revealed variations in the way it was used. While some PMs used it to oversee all the steps in the lifecycle of a translation project, others still preferred to make additional written checklists. The technological developments thus did not supplant analogue artefacts: Post-it® notes, personal checklists and overviews remained important planning tools and reminders.

From 2007 to 2014: Customization and sharing

Another shift in the role of technology for the PMs was evident between 2007 and 2014. As the company had grown, some PMs sought to specialize in particular areas (e.g. quality management and software) in addition to their regular PM work, while also concentrating on specific clients. The corresponding shift towards more flexible work processes that focused more strongly on the needs of certain clients was likewise accompanied by changes in the use of artefacts. For one thing, the PMs now relied on an increasing number of specific tools for each task (Figure 1.1), many of which themselves necessitated further tools. As one project manager recounts:

> You find yourself thinking: Oh man, I need one tool just to enter something into another. I use one tool to analyse something, another to process the analysis, which I then save in another tool so that I can produce a purchase order [laughs]. (PM2)[3]

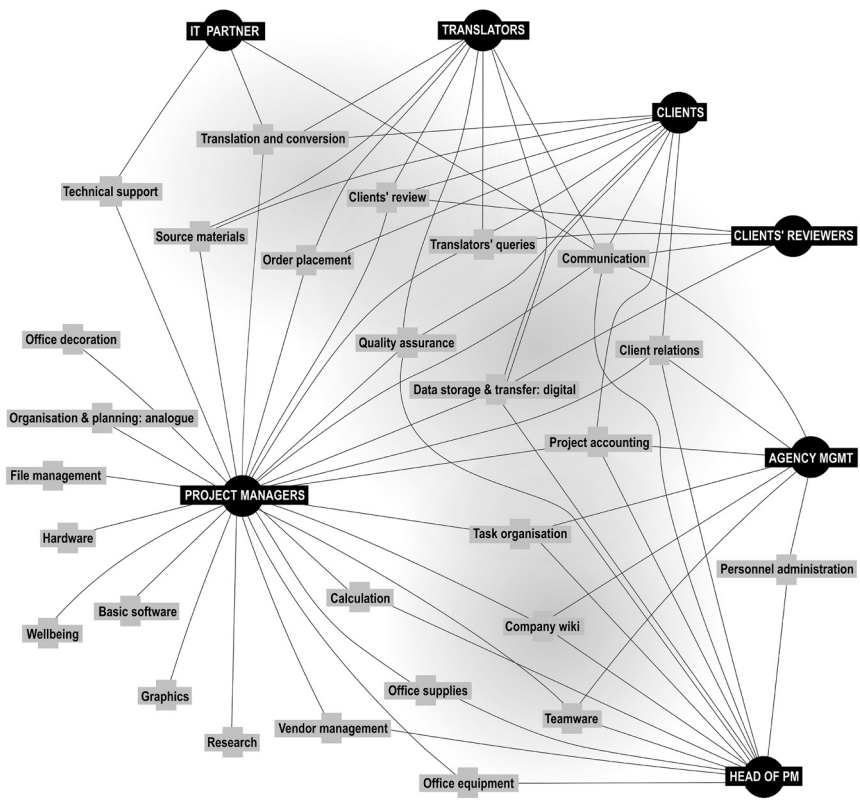

Figure 1.1 Artefacts and their primary users from the PMs' perspective as of 2014.

In contrast to the standardization process observed in the previous period, this development could even be described in some instances as a fragmentation process in terms of the technology used. Indeed, our interviews revealed some ambiguity towards this development:

> It's always a bit of a halfway house. Most tools are practical and help you with your work. But you sometimes have to work round them or can't use all the functions, the whole idea, because it simply doesn't fit your process. You always have to adapt. (PM2)

The changes in this particular period also extended to the functions of the artefacts. Software tools previously only used by one user group (usually PMs) were increasingly expanded and opened to multiple groups involved in the translation process (e.g. translators, clients or their designated validators). These shared tools included the project management system (where customers could now create projects and monitor status), a platform for questions from translators (that was managed by the PMs and could be accessed by clients), a review platform for client feedback to PMs/translators, and shared wikis.

Figure 1.1 shows the primary actors in this translation network along with the artefacts mentioned by the PMs in our interviews or used during observation sessions. For purposes of clarity, the artefacts have been grouped by function. The network diagram shows which groups of artefacts are used primarily by individual PMs and those which had been opened for shared use by 2014. Figure 1.2 shows which specific tools these groups include.

Interestingly, a detailed analysis revealed that the shift towards a more inclusive work process prioritized client information needs over those of translators. While the artefacts stipulated for use in each work process are standardized, clearly defined and rarely omitted in work with translators, the PMs are more flexible in their dealings with clients. The same tendency can be observed for 'redundancies' in the use of tools by PMs, most of which are at the behest of clients. The agency has designated software tools for most steps in its work processes. However, the PMs often agree to use alternative tools to accommodate the preferences of clients who refuse to use a collaborative platform, for instance, to manage translator queries. In other cases, clients' unwillingness to use the agency's standard tools results in double or triple communication (system-generated email, personal email, and phone call) to inform a client about new developments or ask them to do something:

Task organisation PM software, paper calendar, task management software, email flags, email templates for different tasks, task organization and visualization functions in email program, shared email inbox, digital team calendar, colleague's list of ongoing projects and relevant information, individual to do lists	**Order placement** translation commission template for clients, job email template for translators, purchase order, quality assurance list to be filled in by translators, TM configurations, SDL Multiterm databases, reference materials by client, reference documents for DTP projects	**Review by client** fact sheet for reviewers, comments function, track-changes function, email template to inform about translation changes, email template to ask clients for review, collaborative platform for client-sided review, WebYep, automatic emails generated by project management software	**Client relations** newsletter, monthly activity report to clients, non-disclosure-agreement, purchase order, checklist and template for translation commission, business cards, company website, information brochure on services and prices, out-of-office messages, chocolate with Christmas wishes
Teamware MS Excel template for PMs' feedback on their prospective work capacity, team calendar, online chat function (test run), internal wiki (messaging, calendar, client information, meeting minutes, to do lists…), shared email inbox, uniform folder structure, Virtual Machine	**Data storage & transfer: digital** USB, email attachments, FTP server, client wiki, internal company wiki, SDL WorldServer, Microsoft SharePoint Server, shared or individual lists (e.g. for passwords, email header conventions), project folder (digital), Virtual Machine	**Translation and conversion tools** SDL Trados Studio & Passolo, tools for testing compatibility with SDL Studio, specialized conversion tools (e.g. Excel-SDLxliff-conversion, Trados log files,…), collaborative terminology management system, analysis tools for preparing SDL Studio projects and pricing calculation	**Basic software** operating system, MS Office, Internet access (also for cloud-based and remote work), different Internet browsers with automatic configuration of tabs, screen capture tools, Google Translate
Calculation MS Excel, cost calculation function in project management software, specialized analysis tools for preparing Studio projects and pricing calculation	**Organization & planning: analogue** handwritten notes, lists and drawings, notebook, index cards, phone call memos, printout documents for archiving, project folder (paper)	**Communication** phone (memos), email, online chat (test run), automatic forwarding of inquiries by using email rules function, web meetings, social media (LinkedIn, Xing), messaging function in company wiki	**Office equipment** desk, ergonomic chair, desk lamp, phone with speaker, calculator, print dictionaries (used to support the computer screen), magazine holder, projector and screen, kitchen
Quality assurance quality assurance list (to be filled in by translators), Verifika QA check, DTP checklist, spell checking tool, SDL Studio term check, commenting and track-changes function, process documentation folder	**Project accounting** price list, purchase order, specialized analysis tools for preparing Studio projects and pricing, project billing and closure documents, online banking, accounting folder (e.g. for bank statements)	**Translators' queries** queries on ongoing translation projects by email or phone, online form for queries (for specific clients), collaborative online platform for translators' queries, commenting and track-changes function	**Office supplies** pin board with company phone list, pens, highlighters, post-its (e.g. blank, with keyboard short keys, phone numbers, abbreviations for language combinations…), paper tray, packages of blank paper
File management uniform folder structure, bookmarked folders, Total Commander, specific folder structure in email program	**Technical support** email form for IT support	**Source materials** texts or tables in different data formats, image films, technical documentation	**Research** web search, online dictionaries, online route planner for meetings with clients
Graphics graphics and DTP software, FinePrint, screen capture tools	**Vendor management** application form, overview of currently absent translators	**Hardware** two screens, keyboard, shared printer, mouse, headset, laptop	**Personnel administration** software for the electronic recording of working hours, staff-related files
	Wellbeing stress ball, hand cream, ergonomic wrist pad, sitting ball instead of chair, automated reminder ("Your eyes deserve a rest"), drinks, good lighting, kitchen	**Office decoration** wall art, motivational quotes, figurines, plants, children's drawings, crystal stones, personal photos, decorative desk calendar	

Figure 1.2 Artefacts used or mentioned by PMs in 2014 (grouped by function).

> All translators have to use the [review platform] but not all clients use it. So one of my clients uses it – I can assign him questions in the platform – and for another I have to export the questions – there's an option for that in the tool – and send them to him in Excel. (PM9)

Our observations thus pointed towards an increasing use of technology to digitalize and standardize work processes between the first and second data collection periods. In contrast, a shift towards a more collaborative and customized use of technology was evident between the second and the third. In some instances, this development led to a certain fragmentation in terms of the technology used.

Reconceptualization: From socio-cognitive artefacts to boundary objects

As shown earlier, the role of some of the artefacts used in the agency has changed fundamentally over time. From a situated cognition perspective, their role in our first two observation periods is as tools that facilitate complex individual problem solving. Examples include artefacts used to process orders (e.g. text processing software, conversion programmes, research tools, dictionaries) or to manage workflow (e.g. email flagging and calendaring functions). Many can also be regarded as external cognitive supports for the human memory. The PMs use them not only to augment their own cognitive abilities (Hutchins 1995: 153–5) but also to 'transform the task the person has to do by representing it in a domain where the answer or the path to the solution is apparent' (Hutchins 1995: 155), (e.g. calculation, spellchecking or CAT tools).

Since situated views on cognition also look beyond the individual, they allow us to take a more precise look at the social structuring role of technology. In our case study, they help to analyse how artefacts scaffold and modify on-site work processes – as evidenced by the profound changes induced by the new project management system. They also help to explain the broader coordinative function of artefacts, especially for distributing expertise among team members (e.g. the shared email inbox for PMs, project folder, team checklists or early versions of the project management software).

As our study progressed, we became increasingly intrigued by the connective role of technology as a coordination artefact between different groups in distinct professional realms, as a common denominator among contrasting interests

and expectations, or even as a facilitator of collaboration without the need for consensus. We thus decided to experiment with a new way of conceptualizing this connective role in this chapter and complement our situated cognition approach to artefacts with a theoretical approach that focuses on precisely these aspects. We will therefore now discuss our findings against the backdrop of *boundary objects*, a concept introduced by sociology of scientific knowledge scholars Susan Leigh Star, James Griesemer and Geoffrey Bowker (see Star and Griesemer 1989; Bowker and Star 1999), in which boundary objects were originally defined as

> those objects that both inhabit several communities of practice and satisfy the informational requirements of each of them. Boundary objects are thus both plastic enough to adapt to local needs and constraints of the several parties employing them, yet robust enough to maintain a common identity across sites. They are weakly structured in common use and become strongly structured in individual-site use. (Bowker and Star 1999: 297)

The concept has since been widely applied in science and technology studies and extended into other fields such as organizational studies, computer-supported cooperative work and information science (for an overview, see Lee 2007: 308). While it is often applied to non-physical (perceived or immaterial) objects (e.g. cancer as a conceptual artefact [Fujimura 1992]; classification systems [Bowker and Star 1999]), it has also proved useful for objects in the material world (as in our study). As Bowker and Star (1999: 286) explain, many tools can also be seen as having 'both material and symbolic' dimensions.

The boundary objects concept is still relatively new in translation studies, although notions of boundary crossing, liminal spaces, interstitial spaces or third spaces are at the core of cultural theory based debates on translation (e.g. Batchelor 2008; Pym 2001; Wolf 2000). There are similarities between the boundary objects and boundary work concepts, whereby the latter stresses demarcation processes instead of the bridging of boundaries. Parallels can be found when it comes to standardization and categorization – both also key elements in the boundary object approach (Star and Griesemer 1989). The boundary work concept has been used to describe the boundaries between different groups, for example, professional and non-professional translators (En and En 2019; Grbić 2010), or in the development of translation studies (Grbić and Kujamäki 2019; Koskinen and Dam 2016).

The following phenomena from our case study demonstrate some typical characteristics of boundary objects.

Overlapping communities of practice

Boundary objects are elements that build and sustain a bridge between different, overlapping communities of practice (CoP) (see Wenger 1998). Examples of such communities in our study include the PMs, the staff at a particular client or the individual translators in their occupational group.[4] While people can be members of multiple CoPs at the same time, we follow Lee (2007: 316) and focus primarily on CoPs in the 'functional unit' sense.

The connective role of boundary objects

Boundary objects are those that become a common point of reference for the different CoPs and enable them to work together (in our case on translation projects) despite their diversity. In our study, the tools that gradually integrate more user groups into the translation process can be viewed as boundary objects. This connective potential was also the reason why the agency's management expedited the purchase and development of such software. A variety of tools were developed and introduced over time, covering virtually all steps in the work process. The boundary objects discussed here thus constitute a comprehensive, connected software infrastructure (for more on boundary infrastructures, see Star and Ruhleder 1996), which one PM described as a 'veritable software maze' (PM5). The head PM referred to the connective role of the shared review platform, describing it as 'the tool that connects the reviewers and the translators – the tool through which they communicate' (HPM).

Boundary objects and learning

The link between the objects in a CoP and how its members learn to use them is a crucial element in the boundary object concept: 'Strangers and outsiders encounter infrastructure as a target object to be learned about. New participants acquire a naturalized familiarity with its objects as they become members' (Star and Ruhleder 1996: 113). The need for such a learning process was frequently mentioned during our interviews and observation sessions – both for new PMs and new clients:

> I would say that we are well-equipped when it comes to technology, which means that new PMs have a relatively long familiarization period. (PM13)

> A new client will just send 100 files and say: 'Please start!' So you have to find out for yourself where any problems might lie. After all, the client can't know where our technology will work and where we reach our limits. (PM5)

Acquiring knowledge about the use of such boundary objects is therefore essential for becoming a member of a community of practice.

The mutability of boundary objects

This does not mean that shared boundary objects are also always of equal relevance for all user groups. They are, after all, used to different extents by different groups. Although they have the same overarching function for all actors, each user adapts them to their own specific needs.

Boundary objects are designed to be able to assume a variety of meanings depending on the situation, practices and users. While the project management software is the primary planning tool for the PMs, it provides clients primarily with greater transparency and a checking option. The differences in meaning should naturally not be too great or the objects will lose their cohesive power and thus their applicability for a given group (Star and Griesemer 1989: 393).

Star (2010: 602) explains that this 'interpretive flexibility' applies in principle to every artefact. Of particular interest from a boundary objects perspective is how the different meanings of an artefact are handled in concrete 'work arrangements' (2010: 602) and how these can change over time. Boundary objects constantly develop new meaning, they never really 'close' because practices, too, are in constant transition. Such changes in meaning were also evident in our data. They show that the tools and work processes were constantly replaced or adapted – from the development and introduction of tools specific to the agency's work to the involvement of clients via a software tool (e.g. the shared platforms for questions or aspects of the project management software).

Coordinating colliding viewpoints

Special emphasis is placed in the boundary objects concept on the role of objects in coordinating colliding viewpoints in cooperative processes. As our data confirms, the members of the different groups attribute meaning to a boundary object based on their own particular areas of praxis (Bowker and Star 1999: 293). This can also extend to how strongly the different groups feel obliged to use a boundary object – as exemplified here with regard to the level of responsibility clients should – or can be expected to – assume in the translation process:

> We send new [client] reviewers an info sheet [on how to use the review software]. They usually have no idea what to do with a translation, and then an external agency sends them texts to review. This is not a planned part of their work, so it can lead to conflicts. (PM2)

As this example suggests, if cooperating CoPs have different viewpoints and levels of knowledge, boundary objects can serve as a means of coordinating and aligning their specific information needs. The aforementioned review platform serves, for instance, to align the agency's quality assurance *process* with the client's *product* quality requirements. It establishes a shared quality assurance procedure that allows the agency to obtain a review from the client prior to delivery of the translation and thus take better account of any specific product or terminology requirements by involving the client's own experts in this review. Star and Griesemer (1989: 389) themselves also actually use the term 'translation' in this context to refer to the 'reconciliation' or 'translation of the concerns' of one group into those of another via boundary objects. Boundary objects thus become mediating agents that enable different user groups to collaborate without always having to re-negotiate the rules for the process.

Cooperation without consensus

This is also linked to the core idea in the boundary objects concept, namely that consensus is not necessarily essential either to facilitate cooperation or to guarantee the success of shared work (Star and Griesemer 1989: 388).

In our translation network, cooperation was achieved through shared artefacts and was mostly a result of habitualized and standardized working procedures. Different user groups participated in different phases of the translation process: they focused on different aspects thereof and needed different levels of insight into it. By digitalizing some of their tasks, the PMs' traditional mediating role shifted in some areas more towards the boundary objects. These were implemented to tap into the clients' subject matter knowledge, provide the various actors with information on progress, and obtain client feedback both during and after a translation process. Our data, however, point less to a lack of consensus between the various CoPs and more to a lack of insight into their different needs and working methods. The latter is circumnavigated through the use of boundary objects but not necessarily nullified. Interestingly, the PMs nonetheless feel that the cooperation works.

Thomas, Hardy and Sargent (2007: 8) also point to the power dimension that is inevitably encountered in such shared processes. Actors rarely have equal access to resources and equal authority in a concrete situation. There is

often no attempt – or perhaps even possibility – to balance these asymmetries, which Thomas, Hardy and Sargent (2007: 8) also refer to as 'political differences'. Star and Griesemer (1989: 414) do not, however, see this as an 'imposition of one world's vision on the rest', since this would essentially nullify all exchange. They instead treat such asymmetries as primarily negotiation problems that are played out via boundary objects as 'anchors or bridges, however temporary' (1989: 414). The PMs frequently mentioned such negotiation processes in our interviews, especially between the agency – with its corresponding expertise and as supplier of numerous software-based communication tools – and its clients, who ultimately pay for the service and are thus often seen as the more powerful party yet can also demonstrate deep-rooted resistance by rejecting some of the boundary objects proposed by the agency.

Plasticity vs. robustness

Our data reveal two processes that our intuition would suggest must be mutually exclusive. A standardization process was particularly evident from the second observation phase (2007) onwards. This was clearly expedited by agency management and set clear tool and process guidelines that were applicable both internally and for work with freelance translators. Yet personalization processes could also be seen in the way these tools were used and the work processes were arranged. These deviations from the standard occurred when the PMs adapted their workflow and tools to client demands, for example, swapped certain tools for easier-to-use software for reasons of practicability. Aside from these recurring cases of customization, we also noted that the PMs often worked around some tools if they did not quite correspond to their own way of working or personalized them to best suit their individual tasks:

> You can personalize the PM software so that you see all current orders on the start page. All ongoing projects, all orders that should already have been delivered, and all that have been delivered – they turn a nice shade of red. (PM13)

This balancing act between standardization and personalization could be explained by the notion of the *plasticity* or *robustness* of boundary objects. To fit the specific needs of individual user groups, boundary objects must have some 'plasticity' (Star and Griesemer 1989: 393). In other words, they must be flexible enough to be able to connect multiple groups, who can then develop their own ways of working with them or to allow new things to be discovered in them (e.g. new ways of using physical objects) (Roßler 2016: 33).

But boundary objects also need to have a stable meaning. This is closely linked to the standardization concept, that is, situations where their robustness is particularly important (Roßler 2016: 33). In the translation agency we studied, this robustness leads to the boundary objects exerting influence on how work processes are handled, the order of the steps taken, and which variables are prescribed or controlled by people or technology, respectively. The following examples illustrate how an interplay between standardization and personalization can emerge over time, for instance when over-standardization causes a familiarization effect:

> We use job emails, and they look very similar. And that naturally has an effect. (PM9)

> I think it's good that we've standardized them [job emails to translators]. But I also think there is a risk that the translators then become blasé and think: 'OK, that's just the same old stuff.' Because we often write the same in them. [. . .] I've switched to writing the important stuff right at the top, on the first page, because I think the translators might still read that. (PM13)

Because they are constantly subject to negotiations of meaning and attempts at control on the part of the various user groups, such balancing acts between standardization and personalization might be at the core of any boundary object.

Conclusion

Our longitudinal study revealed a series of major developments over the observation period with regard to both the technology and the work processes employed by the PMs. Whereas a change towards the increasing use of technology to digitalize and standardize work processes was observed between the first and second data collection periods, a shift towards a more personalized, customized and collaborative role of technology was evident between the second and the third. On a more theoretical note, we complemented our description of the role of technology as a cognitive artefact that served to scaffold and harmonize on-site work processes with the concept of a socio-cognitive boundary object that allowed translation PMs to work with other, mostly external, actors in the process.

The results of our analysis mirror a range of important insights in technology research. Technological change is contingent: the developments in our first observation periods neither continued linearly in the later phases nor could

they have ultimately been predicted. Our findings also show that technology in organizations 'is both engine and barrier for change; both customizable and rigid; both inside and outside organizational practices. It is product and process' (Star and Ruhleder 1996: 111) and that 'the discontinuities are not between system and person, or technology and organization, but rather between contexts' (Star and Ruhleder 1996: 118).

The changes observed are clearly related to the ways work environments have changed in an increasingly globalized network economy. But they are also linked to specific company policies and the way translation is viewed at this particular agency, that is, as a highly collaborative process that requires high levels of expertise and can only be successful when work processes and tools closely involve other relevant actors. Gaining insights into both individual work practices and organizational perspectives can thus be a valuable approach in the quest to better understand the role of technology in translation networks.

Our analysis also shows that some aspects of the classic mediating role of PMs have increasingly been shifted to software. Accordingly, a closer investigation of the social role of technology and non-human agency (see Olohan 2020; Mihalache 2021) could prove beneficial in future research.

Our study has one important limitation: a description of collaboration processes in CoPs should ideally draw on data from various groups. Our study focuses on PMs and thus only allows us to give limited consideration to aspects like conflicting views on tools. Since only a few translation studies scholars have as yet sought to combine data material from multiple actors (e.g. translators, PMs, clients, or others) equally in one study (see e.g. Risku, Milošević and Rogl 2021; Risku, Pein-Weber and Milošević 2016), this could also be an exciting avenue for future research.

The concept of boundary objects proved particularly useful in this research as it drew our attention to the use, further development and adaption of artefacts in a specific production network. This addition to our initial theoretical framework seems to complement the technology concepts used in the cognitive sciences. Indeed, there are many points of common ground and parallels between the two like the social meaning that artefacts can adopt, their embeddedness in larger systems and conventions, and how they can reconfigure cognitive space (Suchman 2007). Boundary objects, however, also allow us to focus more on the in-between, the perceived boundaries and the negotiation spaces – all aspects that could prove particularly interesting to translation scholars.

Notes

1 PMs' typical tasks may include planning and coordinating multiple language projects, negotiating rates/prices, preparing quotes, assigning and managing teams of translators, negotiating deadlines, tracking and documenting the progress of a translation project, communicating project changes and updates, ensuring document conversion and file compatibility, advising the client on possible technological solutions, coordinating the use of shared technology in the workflow, building and maintaining client-specific terminology and knowledge databases, TMs and so on, managing and forwarding translator queries, coordinating the proofreading process, managing the project's budget and invoicing, and managing customer and vendor relationships.

2 Prior to our field visit, we agreed with the agency that everything we learned in the course of the interviews and during the scheduled observation appointments (i.e. observations at the PMs' work stations or conversations with/between them) could be included in our data. In some cases, we were also invited to observe meetings. However, our non-disclosure agreement naturally stipulates that, in addition to anonymizing personal data, any company and customer information shall remain protected and confidential (especially project volumes and totals or any information on rates or costs).

3 All quotes from the data were translated into English by the authors.

4 Communities of practice can, of course, also be heterogeneous in nature and bring together people with different functions and backgrounds. In this sense, we could also describe this whole production network as one single community of practice (Wenger, McDermott and Snyder 2002: 25f.). In this chapter, the boundary objects concept allows us to stress issues such as colliding viewpoints, mutual non-transparencies and work in environments where consensus is not always achieved or even intended. Like in any social group, this may equally occur in a single CoP. However, in the case of our study, these asymmetries largely arose because each of the groups in question (PMs, translators, clients, etc.) is governed by their own business and operational logic. They do not necessarily work towards the same goal and often have little insight into the work of the other groups. This is why we prefer to view this production network as a set of overlapping CoPs. It could, however, prove valuable to explore how boundary objects may initiate a process of convergence across previously separate CoPs, bringing them together into a single CoP over time.

References

Abdallah, K. (2012), *Translators in Production Networks: Reflections on Agency, Quality and Ethics*, Joensuu: University of Eastern Finland.

Alonso, E. (2016), 'Conflict, Opacity and Mistrust in the Digital Management of Professional Translation Projects', *International Journal for Translation and Interpreting Research*, 8 (1): 19–29.

Batchelor, K. (2008), 'Third Spaces, Mimicry and Attention to Ambivalence', *The Translator*, 14 (1): 51–70.

Bowker, G. C. and S. L. Star (1999), *Sorting Things Out: Classification and its Consequences*, Cambridge: MIT Press.

Brooks, R. A. (1991), 'Intelligence Without Representation', *Artificial Intelligence*, 47 (1–3): 139–59.

Clark, A. (1998), *Being There: Putting Brain, Body, and World Together Again*, Cambridge: MIT Press.

Clark, A. and D. J. Chalmers (1998), 'The Extended Mind', *Analysis*, 58 (1): 7–19.

Dong, J. and G. H. Turner (2016), 'The Ergonomic Impact of Agencies in the Dynamic System of Interpreting Provision', *Translation Spaces*, 5 (1): 97–123.

Dunne, K. J. and E. S. Dunne (eds) (2011), *Translation and Localization Project Management: The Art of the Possible*, Amsterdam: John Benjamins.

En, M. and B. En (2019), '"Coming out" … as a Translator? Expertise, Identities and Knowledge Practices in an LGBTIQ* Migrant Community Translation Project', *Translation Studies*, 12 (2): 213–30.

Esselink, B. (2020), 'Multinational Language Service Provider as User', in M. O'Hagan (ed.), *The Routledge Handbook of Translation and Technology*, 109–26, London and New York: Routledge.

Fujimura, J. (1992), 'Crafting Science: Standardized Packages, Boundary Objects, and "Translation"', in A. Pickering (ed.), *Science as Practice and Culture*, 168–211, Chicago: University of Chicago Press.

Gläser, J. and G. Laudel (2009), *Experteninterviews und Qualitative Inhaltsanalyse als Instrumente rekonstruierender Untersuchungen*, 3rd edn, Wiesbaden: VS Verlag.

Grbić, N. (2010), '"Boundary Work" as a Concept for Studying Professionalization Processes in the Interpreting Field', *Translation and Interpreting Studies*, 5 (1): 109–23.

Grbić, N. and P. Kujamäki (2019), 'Professional vs. Non-Professional? How Boundary Work Shapes Research Agendas in Translation and Interpreting Studies', in H. V. Dam, M. N. Brøgger, and K. K. Zethsen (eds), *Moving Boundaries in Translation Studies*, 113–31, London: Routledge.

Hutchins, E. (1995), *Cognition in the Wild*, Cambridge: MIT Press.

Hutchins, E. (2005), 'Material Anchors for Conceptual Blends', *Journal of Pragmatics*, 37: 1555–77.

Kenny, D. (ed.) (2017), *Human Issues in Translation Technology*, London: Routledge.

Koskinen, K. and H. V. Dam (2016), 'Academic Boundary Work and the Translation Profession: Insiders, Outsiders and (Assumed) Boundaries', *Journal of Specialised Translation*, 25: 254–67.

Koskinen, K. and M. Ruokonen (2017), 'Love Letters or Hate Mail? Translators' Affective Responses to Technology', in D. Kenny (ed.), *Human Issues in Translation Technology*, 8–24, London: Routledge.

LeBlanc, M. (2017), '"I Can't Get no Satisfaction!": Should We Blame Translation Technologies or Shifting Business Practices?', in D. Kenny (ed.), *Human Issues in Translation Technology*, 45–62, London: Routledge.

Lee, C. P. (2007), 'Boundary Negotiating Artifacts: Unbinding the Routine of Boundary Objects and Embracing Chaos in Collaborative Work', *Computer Supported Cooperative Work*, 16 (3): 307–39.

Lueger, M. and U. Froschauer (2018), *Artefaktanalyse: Grundlagen und Verfahren*, Wiesbaden: VS Verlag.

Marshman, E. (2014), 'Taking Control: Language Professionals and Their Perception of Control When Using Language Technologies', *Meta*, 59 (2): 380–405.

Mihalache, I. (2021), 'Human and Non-Human Crossover: Translators Partnering with Digital Tools', in R. Desjardins, C. Larsonneur, and P. Lacour (eds), *When Translation Goes Digital: Case Studies and Critical Reflections*, 20–43, Cham: Palgrave Macmillan.

Moorkens, J. (2020), 'A Tiny Cog in a Large Machine: Digital Taylorism in the Translation Industry', *Translation Spaces*, 9 (1): 12–34.

Moorkens, J. (2021), 'Translation in the Neoliberal Era', in E. Bielsa and D. Kapsaskis (eds), *The Routledge Handbook of Translation and Globalization*, 323–36, London and New York: Routledge.

O'Brien, S. (2012), 'Translation as Human–Computer Interaction', *Translation Spaces*, 1: 101–22.

Olohan, M. (2020), 'Sociological Approaches to Translation Technology', in M. O'Hagan, Minako (ed.), *The Routledge Handbook of Translation and Technology*, 384–97, London and New York: Routledge.

Olohan, M. and E. Davitti (2017), 'Dynamics of Trusting in Translation Project Management: Leaps of Faith and Balancing Acts', *Journal of Contemporary Ethnography*, 46 (4): 391–416.

O'Regan, J. K. (1992), 'Solving the "Real" Mysteries of Visual Perception: The World as an Outside Memory', *Canadian Journal of Psychology*, 46 (3): 461–88.

Ozolins, U. (2007), 'The Interpreter's "Third Client": Interpreters, Professionalism and Interpreting Agencies', in C. Wadensjö, B. Englund Dimitrova, and A.-L. Nilsson (eds), *The Critical Link 4: Professionalisation of Interpreting in the Community*, 121–31, Amsterdam: John Benjamins.

Plaza-Lara, C. (2020), 'How Does Machine Translation and Post-Editing Affect Project Management? An Interdisciplinary Approach', *Hikama*, 19 (2): 163–82.

Pym, A. (2001), 'Alternatives to Borders in Translation Theory', in S. Petrilli (ed.), *Athanor: Semiotica, Filosofia, Arte, Letteratura, Anno XII., nuova serie, n. 4,* 172–82, Rome: Meltemi.

Risku, H. (2016), *Translationsmanagement: Interkulturelle Fachkommunikation im Informationszeitalter*, 3rd edn, Tübingen: Narr Francke Attempto.

Risku, H., Milošević, J. and Rogl, R. (2021), 'Responsibility, Powerlessness and Conflict: An Ethnographic Case Study of Boundary Management in Translation', in E. Monzó Nebot and O. Carbonell Cortés (eds), *Translating Asymmetry – Rewriting Power*, 145–68, Amsterdam: John Benjamins.

Risku, H., C. Pein-Weber, and J. Milošević (2016), '"The Task of the Translator": Comparing the Views of the Client and the Translator', *International Journal of Communication*, 10: 989–1008.

Risku, H. and R. Rogl (2020), 'Translation and Situated, Embodied, Distributed, Embedded and Extended Cognition', in F. Alves and A. L. Jakobsen (eds), *The Routledge Handbook of Translation and Cognition*, 478–99, London: Routledge.

Risku, H., R. Rogl, and J. Milošević (2020), 'Researching Workplaces', in E. Angelone, M. Ehrensberger-Dow, and G. Massey (eds), *The Bloomsbury Companion to Language Industry Studies*, 37–62, London: Bloomsbury.

Risku, H., F. Windhager, and M. Apfelthaler (2013), 'A Dynamic Network Model of Translatorial Cognition and Action', *Translation Spaces*, 2: 151–82.

Robbins, P. and M. Aydede (2009), 'A Short Primer on Situated Cognition', in P. Robbins and M. Aydede (eds), *The Cambridge Handbook of Situated Cognition*, 3–10, Cambridge: Cambridge University Press.

Rodríguez-Castro, M. (2013), 'The Project Manager and Virtual Translation Teams', *Translation Spaces*, 2: 37–62.

Roßler, G. (2016), *Der Anteil der Dinge an der Gesellschaft*, Bielefeld: Transcript.

Sakamoto, A. (2018), 'Disruption in Translator-Client Matching: Paid Crowdsourcing Platforms vs. Human Project Managers', *Tradumàtica: Tecnologies de la traducció*, 16: 85–94.

Sakamoto, A. (2019), 'Unintended Consequences of Translation Technologies: From Project Managers' Perspectives', *Perspectives*, 27 (1): 58–73.

Sakamoto, A. (2021), 'The Value of Translation in the Era of Automation: An Examination of Threats', in R. Desjardins, C. Larsonneur, and P. Lacour (eds), *When Translation Goes Digital: Case Studies and Critical Reflections*, 231–55, Cham: Palgrave Macmillan.

Sakamoto, A. and M. Foedisch (2017), '"No News Is Good News?": The Role of Feedback in the Virtual-Team-Style Translation Production Network', *Translation Spaces*, 6 (2): 333–52.

Star, S. L. (2010), 'This Is Not a Boundary Object: Reflections on the Origin of a Concept', *Science, Technology & Human Values*, 35 (5): 601–17.

Star, S. L. and J. R. Griesemer (1989), 'Institutional Ecology, "Translations" and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology', *Social Studies of Science*, 19 (3): 387–420.

Star, S. L. and K. Ruhleder (1996), 'Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces', *Information Systems Research*, 7 (1): 111–34.

Stoeller, W. (2011), 'Global Virtual Teams', in K. J. Dunne and E. S. Dunne (eds), *Translation and Localization Project Management: The Art of the Possible*, 289–318, Amsterdam: John Benjamins.

Suchman, L. (2007), *Human-Machine Reconfigurations*, Cambridge: Cambridge University Press.

Thomas, R., C. Hardy, and L. Sargent (2007), 'Artifacts in Interaction: The Production and Politics of Boundary Objects', *AIM Research Working Paper Series*. Available online: http://www.aimresearch.org/uploads/pdf/working_papers/052rtpaper.pdf (accessed 19 April 2020).

Tsvetkov, N. and V. Tsvetkov (2011), 'Effective Communication in Translation and Localization Project Management', in K. J. Dunne and E. S. Dunne (eds), *Translation and Localization Project Management: The Art of the Possible*, 189–210, Amsterdam: John Benjamins.

Wenger, É. (1998), *Communities of Practice: Learning, Meaning and Identity*, Cambridge: Cambridge University Press.

Wenger, É., R. McDermott, and W. M. Snyder (2002), *Cultivating Communities of Practice: A Guide to Managing Knowledge*, Boston: Harvard Business School Press.

Wittgenstein, L. (1990), *Tractatus Logico-Philosophicus*, Leipzig: Reclam.

Wolf, M. (2000), 'The Third Space in Postcolonial Representation', in S. Simon and P. St-Pierre (eds), *Changing the Terms: Translating in the Postcolonial Era*, 127–45, Ottawa: University of Ottawa Press.

2

Interpreters' Performances and Cognitive Load in the Context of a CAI Tool[1]

Bart Defrancq, Helena Snoeck and Claudio Fantinuoli

Introduction

The aim of this chapter is to investigate the effect of a particular in-booth tool on both the cognitive load of students of simultaneous interpreting and their performances. Computer-assisted interpreting (CAI) tools for use during interpretation are a recent development and their effects on interpreters are still unknown. From the previous literature we know that CAI tools based on automatic speech recognition (ASR) technology have the ability to improve interpreter accuracy on specific target items, such as numbers and terminology (Defrancq and Fantinuoli 2021; Pisani and Fantinuoli 2021), and appear to offer interpreters some sort of psychological support. However, the potential downsides of such tools are largely unknown, mainly because the research has so far only focused on tasks that the CAI tool is meant to facilitate: interpreting numbers and/or terms. Increased accuracy in the rendition of those items is obviously an improvement, but not if it overburdens the interpreter or has a negative effect on overall interpreting quality.

The data reported here are drawn from the experiment described by Defrancq and Fantinuoli (2021), where the authors reported significant accuracy gains in students' renditions of numbers when they were offered an ASR-based support tool in the booth. This chapter complements that work by focusing on cognitive load and overall interpreting performance. It also introduces a novel method to assess cognitive load in interpreters, based on the fundamental frequency (*F0*) of their speech, that is, the lowest frequency of vibration of the vocal folds. Existing methods for measuring cognitive load are often non-specific or too invasive to be ecologically valid in the context of simultaneous interpreting, making alternative, ecologically more valid methods worthwhile exploring.

In the following sections of the chapter, we will focus on CAI tools and their potential effects on interpreters. Particular attention will be paid to effects on the cognitive load of the interpreter and on their overall interpreting performance. The research questions that aim to quantify those effects will be answered by a re-analysis of an existing experimental dataset that the authors created. Performance will be measured on two dimensions: accuracy and acceptability, while cognitive load will be determined using a new method based on *F0*.

Computer-assisted interpreting

Computer-assisted interpreting covers the technological tools developed to support interpreting. In a broad sense CAI significantly overlaps with computer-assisted translation (CAT); tools developed for the latter – terminology management systems in particular – are of course also suitable for interpreters. If we narrow down our perspective to tools that are specifically geared towards interpreting, two trends become apparent. The first is the long-established trend of physically separating interpreters from the speakers they interpret, while making their spoken output available for their audience. The advent of simultaneous interpreting marked an initial and important stage in this process, when interpreters no longer shared the stage with the speaker but were relegated to interpreting booths. This trend has now developed to include online meeting platforms equipped with interpreting facilities. The second, very recent trend, is the development of tools that help interpreters record or deliver the message: digital pens for consecutive interpreting (Orlando 2010), interpreter-specific terminology management systems (Fantinuoli 2018), and in-booth CAI tools delivering information to simultaneous interpreters. It is the last category that is of interest to us in this chapter. It covers technologies capable of automatically retrieving information from the spoken source text, while displaying it on a screen in the booth.

When developing software-based support systems, a delicate balance must be struck between potential advantages and downsides, all of which depend on how humans interact with the support system to form what is sometimes called the joint cognitive system (Hollnagel and Woods 2005). Cognitive ergonomics is the field that studies how that interaction takes place and how it can be optimized. From a cognitive ergonomics point of view the question whether software-based support systems reduce or increase cognitive load is an important one. This is especially relevant to interpreters. There is general consensus among

interpreting scholars that simultaneous interpreting is an extremely demanding task in terms of cognitive processing. It is commonplace in this context to refer to Gile's (1999) tightrope argument, which runs along the following lines: simultaneous interpreting requires careful balancing between multiple cognitive demands, the sum of which is close to a given interpreter's point of saturation. This makes the activity cognitively challenging and prone to failure. Therefore, errors are purported to be inherent to the interpretation process and to occur even in the absence of identifiable load triggers. Even though interpreting and translation tasks share many similarities, the addition of existing CAT tools into the simultaneous interpreting process then becomes untenable: the nature of the information provided and the way it is displayed make the resource virtually inaccessible to interpreters: segments retrieved from a translation memory are too long to be read and inserted into the interpretation; and term bank suggestions are usually displayed in small boxes on the periphery of the screen. Research even suggests that CAT tools are ergonomically unsuited to their target community, that is, translators (Ehrensberger-Dow and O'Brien 2015; Lavault-Olléon 2016; O'Brien et al. 2017).

Therefore, it is paramount that developers of CAI tools focus on particular items that are known to trigger cognitive load or to reduce the quality of performance: numbers, terminology and named entities. Experimental prototypes, such as InterpretBank (Fantinuoli 2017) and SmarTerp (Rodriguez et al. 2021), are currently being developed. Student performance in the rendition of the targeted items has been found to benefit from CAI support (Desmet, Vandierendonck and Defrancq 2018; Defrancq and Fantinuoli 2021; Van Cauwenberghe 2020; Pisani and Fantinuoli 2021). Questions of how interpreters interact with the tools and how those tools affect cognitive load and overall performance remain under-researched. Our overview of the literature on cognitive load and performance, therefore, draws not only on the scant body of work on CAI tools but also on the wider literature related to visual support in simultaneous interpreting, that is, simultaneous interpreting with text and multimodality in interpreting.

Visual support and interpreter performance

There is evidence to suggest that visual support boosts interpreter performance. In one of the earliest empirical studies on simultaneous interpreting with text, Lamberger-Felber (2001) assesses interpreting quality across three conditions:

interpretations without text, interpretations with text but without preparation time and interpretations with both text and preparation time. The quality assessment focuses on the accuracy with which targeted items (numbers and proper names) are rendered, as well as on the overall accuracy of the interpretations. The study shows that interpreters make fewer errors on targeted items when they are given source texts and even fewer when they have time to prepare. Their general performance levels show the same tendency. However, Lamberger-Felber (2001) also underlines high inter-subject performance variability. It should also be noted that no method is presented to determine whether the interpreters involved in the experiment actually consulted the written materials.

A number of studies have analysed interpreter performance while interpreting with visual support for specific items. Stachowiak-Szymczak and Korpal (2019) and Korpal and Stachowiak-Szymczak (2018, 2020) adopt a method whereby interpreters are shown slides with the highlights of each speech segment, including all target numbers. Desmet, Vandierendonck and Defrancq (2018), Van Cauwenberghe (2020), and Pisani and Fantinuoli (2021) all use automatic speech recognition (or a simulated version thereof) that displays only target items such as numbers or terms. Defrancq and Fantinuoli (2021) let an automatic speech recognition software display a running transcript of the source speech in which numbers were highlighted. In all cases where interpreters have access to written material, the target items are conveyed significantly more accurately when compared against cases where such support is absent.

However, in none of these studies is the rendition type mapped on to the gazing patterns. It is therefore impossible to determine whether participants actually looked at the screen when they rendered items correctly or looked away when they made mistakes. The analysis by Stachowiak-Szymczak and Korpal (2019) includes eye-tracking data, but all data, on both rendition type and fixations, are aggregated at the group level (trainees vs. professionals). So, even though the accuracy data shows significant improvement when written information is available, a causal relation between the two cannot be determined. Defrancq and Fantinuoli (2021) describe how they used booth cam recordings to analyse participant gaze, distinguishing three conditions: (i) simultaneous interpreting without visual input, (ii) simultaneous interpreting with visual input ignored by the participants and (iii) simultaneous interpreting with visual input gazed at by the participants. On the group level, the third condition yields more accurate renditions than the other conditions,

even though between-subject variability is considerable and the intake of visual information is only presumed. Surprisingly, even the condition where interpreters ignore available input turns out to yield more accurate renditions than the condition without input. The authors ascribe this to a placebo-like effect: participants seem to be reassured by the availability of a back-up system, which improves their performances, even when the back-up system is not consulted.

In a study on the potential adverse effects of visual input, in particular written input that is incongruent with the auditory input, Chmiel, Janikowki and Cieślewicz (2020) record longer fixation times on items that are accurately displayed than on items that are not. This seems to indirectly confirm the general positive effects of visual input. The main conclusion of the study, however, is that incongruent visual input most often leads to incorrect renditions, confirming visual dominance in multimodal processing.

A lack of general assessment of interpreting quality is characteristic of almost all these studies. Stachowiak-Szymczak and Korpal (2019) do include an accuracy analysis of those items in the immediate vicinity of the target items, concluding that there is a high correlation between the accuracy rates of both types of items. However, Yang's (2019) thesis on the overall quality of simultaneous interpreting with text at different delivery rates tends to show that more intensive reliance on the written input, for example, longer and more numerous fixations and saccades[2], correlates with poorer performance.

A comprehensive performance analysis is therefore needed. Experimental conditions may very well push interpreters to focus specifically on the rendition of targeted items. For instance, in Desmet, Vandierendonck and Defrancq (2018), Van Cauwenberghe (2020), and Defrancq and Fantinuoli (2021), student interpreters were informed beforehand that they would be given visual support for the numbers occurring in the source speech. This could have motivated them to focus on the rendition of numbers to the detriment of overall performance. Likewise, in Korpal and Stachowiak-Szymczak (2018, 2020) and Stachowiak-Szymczak and Korpal (2019), presentation slides contained some limited information in addition to the targeted items; this could also have boosted interpreters' performance across all those items, while turning their attention away from the remaining bulk of the speech. Theoretically, the failure to include a comprehensive analysis of the performance could yield misleading conclusions: if participants get all the target items right, but none of the remainder of the running text, their performance would be rated perfect.

Cognitive load in simultaneous interpreting (with text)

The issue of cognitive load is a crucial one in the development of CAI tools. Simultaneous interpreting is reputedly a cognitively challenging task; adding a further task, such as the consultation of information provided by a CAI tool, is likely to increase cognitive load in interpreters (Prandi 2018). The concept of cognitive load originally stems from cognitive psychology, where it has been used to account for experimental data that provide evidence of delayed and altered behavioural performances when different tasks are combined. It has since been enthusiastically embraced in interpreting studies, where various attempts have been made at modelling the processes that contribute to it (Gile 1995; Seeber 2011). However, cognitive load induced by interpreting-related processes is still an elusive concept in interpreting research. There is no established methodology or research protocol to measure it, with the result that a wide range of methods have been employed, from the most intrusive ones such as eye-tracking and pupillometry (Seeber and Kerzel 2012; Korpal and Stachowiak-Szymczak 2018; Yang 2019) to purely observational ones such as the analysis of various features of performance (Gile 1999; Shlesinger 2003; Plevoets and Defrancq 2016, 2018; Lv and Liang 2019). To this day, no method is entirely satisfactory: observational methods suffer from data scarcity and inherent limitations in explanatory power, while intrusive methods require high levels of experimental control, thereby reducing the ecological validity of the research design.

In this chapter we will put forward an alternative method, based on observational data drawn from a previous experiment. The fundamental frequencies of the subjects' voices will be analysed at particular segments of the interpretation. The method was previously used by Korpal (2017) to determine stress in interpreters but can be used as a proxy of cognitive load, according to the psychological literature (see section on 'Pitch as proxy for cognitive load'). At the time of writing there is no reported research into cognitive load related to the use of a CAI tool. As mentioned earlier, the closest proxy that has sparked some interest from the research community is simultaneous interpreting with text. Based on a theoretical estimate of cognitive demands, Seeber (2015) concludes that total cognitive load in simultaneous interpreting with text is 'considerably above the one predicted for simultaneous interpreting without text' (2015: 473), while further calling for empirical research to verify his claims. There is evidence that the Ear-Voice-Span in simultaneous interpreting with text is longer than in interpreting without text (Lamberger-Felber 2001;

Baxter 2016), suggesting higher cognitive load. However, it is also known that when acoustic and visual signals convey redundant information, processing is facilitated (Kinchla 1974; Miller 1982), even when one stimulus is delayed relative to the other. CAI tools presenting interpreters with written information that is redundant with the speaker's input and delivered with a reasonable latency could therefore balance the load created by the processing of written information. Previous research has shown that simultaneous interpreters do consult written information strategically. In a recent eye-tracking study, Seeber, Keller and Hervais-Adelman (2020) conclude that visual processing lags behind auditory processing, as simultaneous interpreters are found to focus on the printed version of a target item after hearing it. They suggest that interpreters use visual input from a text mainly to support the production of the target text, rather than to support comprehension of the source text. In an older study by Seeber (2012), subjects are found to consult written numerical information predominantly when they are interpreting long, rather than short, numbers. Yang (2019), Korpal (2017) and Stachowiak-Szymczak and Korpal (2020) show that interpreters consult visual support significantly more when the source text is delivered at a fast pace, suggesting that the additional load involved with consulting written input is felt to be offset by cognitive gains in particularly challenging conditions.

The picture that emerges is complex: on the one hand, consulting visual input can only increase cognitive load in interpreters; on the other hand, appropriate visual input can facilitate the process of interpreting and, in so doing, reduce cognitive load. Interpreters seem to develop strategies to optimize the cognitive investment to be made, based on the expectation that consulting visual input is most efficient when cognitive resources are under pressure.

Assuming simultaneous interpreting supported by a CAI tool is similar to simultaneous interpreting with text in terms of the cognitive load it imposes, we are most likely to find little evidence of an increase. The visual input offered by the CAI tool is expected to both increase and reduce cognitive load. However, the nature of the visual input could be more challenging in the case of the CAI tool: the likelihood of errors is higher than in the case of texts used in simultaneous with text, which could result in higher load. Given Seeber's (2012) findings, the area in which an effect of visual input is most likely to occur is the area corresponding to the rendition of the number: interpreters tend to use visual input to support production, rather than comprehension.

Research questions and hypotheses

Considering the research gaps identified earlier, we put forward the following theoretical and methodological questions:

(1) While ASR seems to improve performance in interpreting numbers, does ASR-based support improve overall performance?
(2) As numbers are assumed to trigger increased cognitive load in interpreters, does F0 data reveal increased cognitive load when the interpreter renders a number?
(3) The consultation of visual input during simultaneous interpreting both increases and reduces cognitive load. Does F0 data reflect these mutually cancelling trends?

In light of existing literature, the answer to the first research question is likely to be affirmative in so far as accuracy is concerned. Lamberger-Felber (2001) shows that the curves for target item accuracy and overall accuracy run parallel: the availability of written support boosts both. However, in Lamberger-Felber's experiment, participants possessed an accurate transcript of the entire speech before the start of the experiment, while an ASR-based support systems delivers a running transcript with errors.

Again, theoretical literature suggests that the answers to the second and third research questions are also likely to be affirmative. Processing numbers and consulting visual input mobilize significant cognitive resources. However, visual input also reduces load if visual and auditory information are congruent.

Data and methods

Data

The data come from an experiment organized in 2019 and described in Defrancq and Fantinuoli (2021). For technical details on the experiment, we refer to that publication. The experiment involved an ASR-based central support system that produced running transcripts of four English speeches given by a near-native speaker. While the running text of the transcript was displayed in black, numbers were highlighted in red and displayed in a bigger font. The transcript

was made visible in three out of six booths via a video link that connected the screen of the table PC running the software to the booth screens. In the other booths, the screen displayed the camera image of the speaker. The experiment participants were six students, all female, on a postgraduate programme in conference interpreting; each had approximately sixty hours of experience in simultaneous interpreting.

The speeches were relatively short, ranging between 6 minutes and 6.5 minutes. ASR support was only given for the first 5.5 minutes. The speeches contained between twenty-one and forty-five numbers and a warm-up period of at least half a minute without numbers. Speeches had similar readability scores according to the Flesch Reading Ease and Gunning Fox indexes; however, speech 2 turned out to score slightly higher on readability than the other speeches. For each speech, three participants interpreted with ASR support and three without. They switched booths after every speech, so that the first group (Students 1.1, 1.2 and 1.3) interpreted speeches 2 and 4 with ASR support and speeches 1 and 3 without, while the second group (Students 2.1, 2.2 and 2.3) interpreted speeches 1 and 3 with ASR support and speeches 2 and 4 without. All source speeches and interpretations were recorded. A video recording was made in each booth to study the interactions between student and tool. All oral data were transcribed using EXMARaLDA and aligned. All numbers were given a time stamp at the onset and end of the corresponding acoustic signal.

The data was not collected with a view to studying cognitive load and so the experimental conditions were not specifically tailored to that purpose. This study should therefore be considered exploratory, rather than explanatory. Students were allowed to write down numbers in both conditions, and some can be seen to do so. This means that there is no strict control condition; the notes are obviously also a visual support in the booth that can create additional load independently from the availability of an ASR-based CAI tool.

Performance assessment

General performance assessment is a real conundrum in interpreting studies (Gile 1999; Koby and Lacruz 2017). No widely agreed protocols exist and the choice of a particular methodology correlates with the object of study (Pöchhacker 2001). In the area of experimental studies, two approaches can be distinguished (Han 2017): a so-called atomistic approach based on error detection and a more holistic approach based on analytic rating scales. After submitting a transcript of

a single interpretation to eighty-one assessors, forty-eight of whom also listened to a recording, Gile (1999) concludes that assessments using both methods do not correlate and that ratings show less variation than error-based assessments.

Different analytical frameworks and scoring systems have been used across studies to determine the effects of visual input on overall interpreting quality. Lamberger-Felber (2001) categorizes renditions as correct or incorrect, with number approximations being considered as incorrect. For general quality assessment, renditions of content words are checked and categorized as correct or as semantic deviations. Omissions of entire text segments are analysed as semantic deviations but counted twice. Yang (2019), in contrast, instructs two raters to use a rating system that consists of eleven categories[3], with ten-point scales for each of the categories.

We have opted for a close error analysis based on ten categories, including not only accuracy variables, but also acceptability variables. A close error analysis seemed to be the best option in our case, because a holistic rating would probably have been influenced too much by issues with the rendition of the numbers. Interpretations were clearly more hesitant than normal due to the presence of an artificially high number of numerals. This is also the reason why disfluencies other than self-repairs and filled pauses were not included in the error category list in Table 2.1.

A few comments are in order with regard to specific categories. First of all, errors in the rendition of numbers have not been included in the count, as they were the subject of a separate study (Defrancq and Fantinuoli 2021). Secondly, collocation errors and lexical discrepancies were analysed in the same way, even though collocation errors are more likely to affect the acceptability of the target

Table 2.1 Error categories used in the close error analysis

Accuracy		Acceptability	
Tag	**Meaning**	**Tag**	**Meaning**
XX	a complete sentence is missing	RR	self-repair
YY	a phrase or clause is missing	HH	unclear reference
SS	semantic discrepancy between source and target extending over a phrase or clause	GG	word order error or syntactic structure error
LL	lexical discrepancy[4] between source and target or unsuitable lexical item in collocation	CC	agreement error (subject-verb; adjective-noun; etc.)
AA	addition	FF	filled pause

text. However, there were few cases of collocation error and some of them also reduced the accuracy of the rendition. Finally, filled pauses' effects are dual: they harm the interpretation's acceptability because they reduce fluency. However, they are also considered in a number of studies (Plevoets and Defrancq 2016) to be indicators of cognitive load. As such, filled pauses might have been used here as a way of triangulating the F0 data discussed later. We chose not to do so because their observed frequencies are too low and too concentrated in the output of two students (*c.* 60 per cent of the data) to be meaningfully applied as an indicator of load.

The close error analysis was first carried out independently by the first two authors of this chapter. After the first round a second analysis was conducted to reach consensus on the cases that had been overlooked or analysed differently during the first round. Chi-square testing was then used to establish significance.

Fundamental frequency as a proxy for cognitive load

The response variable for cognitive load used in this chapter is fundamental frequency (F0). Fundamental frequency, vocal intensity, and speech rates have been used as indicators of stress induced by cognitive processes or emotions in cognitive science (Johnstone and Scherer 2000; Rothkrantz et al. 2004). Rothkrantz et al. (2004) let subjects perform a Stroop test while measuring F0 and other voice-related parameters. Every minute, the pace at which words appeared on the computer screen increased by 20 per cent (relative to the first minute). Significant F0 increases were found after the first and the third minute. Scherer et al. (2002) found a highly significant increase in speech rate and a marginally significant increase in F0 and intensity in a condition where stress was induced solely by cognitive load (to the exclusion of emotional stress). However, not all subjects presented the same pattern of increased F0. A study carried out by Huttunen et al. (2011) reports on an experiment involving thirteen military pilots taking a challenging test flight in a simulator, finding a mean F0 increase of 7 Hz and a mean intensity increase of 1 dB in cognitively challenging flight phases. Based on a broad overview of existing studies, Van Puyvelde et al. (2018) conclude that increases in the mean F0 and smaller variation in F0 values tend to correlate with increased cognitive load.

F0 is not unknown in interpreting research. Most research focuses on deviant prosodic patterns in interpreted speech and their effect on the recipients of the interpretation. Shlesinger (1994) and Ahrens (2004) point to what is generally assumed to be an interpreter-specific F0 pattern, that is

rise-level intonation at the end of informational units. Ahrens (2004) relates it to interpreter uncertainty as to whether information units are complete or not while they are being produced, while Shlesinger (1994) concludes that it has detrimental effects on listener comprehension. F0 variability has also been investigated in several papers on comprehensibility and perceived quality of interpretation. Collados Aís (2001) points to lower comprehension scores in the case of monotonous interpretation, although results do not reach statistical significance. Zwischenberger (2011) shows that interpreters rate the quality of an interpretation with lively intonation higher than that of monotonous interpretation. Recent research by Lenglet and Michaux (2020) finds no significant effect of interpretation prosody on listener comprehension. F0 variability due to cognitive load or stress in interpreters has not been the object of investigation so far. A study by Daró (1990) showed that one interpreter's F0 was lower after reading a text in her L1 (Italian) than after reading texts in other languages of which she had mastery. Daró (1990) relates this to the higher levels of emotional tension that come with speaking a language other than L1. Finally, F0 was used in Korpal (2017) to study stress in simultaneous interpreters and relate it to performance. A high delivery rate of the source speech is put forward as a stressor, which Korpal (2017) finds to correlate positively with mean F0 and negatively with performance. It could be argued that high delivery rate is not only a stressor, but also, and perhaps most of all, a source of higher cognitive load in simultaneous interpreters. In accordance with the psychological literature reviewed in Van Puyvelde et al. (2018), we will hold mean F0 to be indicative of cognitive load.

In this study, F0 data were collected from the recordings of the twenty-four interpretations using Praat, a free speech analysis software programme, and its standard default pitch settings (75–500 Hz). The sampling rate was set at 0.01 seconds. The following samples were selected for analysis:

(1) All samples from the warm-up segment: devoid of numbers, the warm-up segment was taken as the baseline condition.
(2) Samples corresponding to the rendition of the number by the student.

To reduce the number of confounding variables, the following cases were excluded:

(1) Overlapping source and target data: a student starts rendering the number before the speaker finishes it. Cognitive load is likely to be higher in these cases due to the student's interpreting strategy.

(2) Overlapping data for several numbers: when numbers follow each other in rapid succession, the rendition of number 1 is sometimes postponed until after number 2 is uttered by the speaker. The overlap is likely to cause additional cognitive load as two numbers need to be kept in short-term memory.
(3) Data corresponding to the phase after ASR was no longer available; as ASR was only available for 5.5 minutes, there was a lack of support for some numbers. Including these cases would present a distorted picture of cognitive load in the ASR condition. They cannot be easily aggregated with the non-support condition for these students as the source texts are different.

Overall, roughly two-thirds of the F0 data were excluded based on one or a combination of these criteria. The F0 results shown in the following section are drawn from 216 instances of rendered numbers. Only nineteen of these concern highly complex numbers (> 4 numerical units). It was therefore impossible to analyse the effect of number complexity on F0. Considering that samples are different from one student to another, only a within-subject analysis will be carried out. F0 data come in massive amounts (several tens of thousands of samples for each interpretation). Therefore, statistical testing based on p-values is likely to lead to Type I errors, whereby even very small and trivial differences erroneously appear to be statistically significant. Effect size analyses and visual inspection of the data are preferable. As the samples are unequal in size, Hedges' *g* effect size will be computed. Effect sizes will be interpreted according to Cohen (1977): small effect size: 0.2; medium effect size: 0.5; large effect size: 0.8.

Results

Performance

The results of the performance analysis are summarized in Figure 2.1.

The total number of error tags across the conditions is very similar: 338 error tags occur in interpretations supported by ASR, while 348 error tags occur in the unsupported conditions. ASR seems to have had no effect on the overall quality of the students' performances. There is a tendency towards higher accuracy with ASR support and higher acceptability without. This is not really surprising: watching a screen to consult the transcript is logically more likely to help render the source text more accurately and to interrupt the flow of speech. However, the tendencies are not significant (df = 1, chi-squared = 1.168; p = 0.28). It should

Figure 2.1 Total number of acceptability and accuracy error tags for both conditions.

Table 2.2 Number of error tags per error category in both conditions

	ASR	Non-ASR
Accuracy		
XX	21	20
YY	73	80
LL	26	38
AA	7	4
SS	30	34
Total Accuracy	157	176
Acceptability		
RR	106	109
HH	9	14
GG	16	11
CC	10	9
FF	40	29
Total Acceptability	181	172
Total error tags	338	348

be recalled that ASR support had a significant positive effect on the rendition of numbers (Defrancq and Fantinuoli 2021). On balance, ASR support seems to positively affect student performance.

Table 2.2 shows the frequencies of individual error types. None of them significantly depart from the general trend.

It is worthwhile pointing out in Table 2.2 that the main contribution to lower acceptability in the ASR condition comes from filled pauses (FF), which are more frequent in the ASR condition (N = 40), than in the unsupported condition

(N = 29). As mentioned before, the data should be handled with caution in this case, as almost 60 per cent of the filled pauses were produced by only two students. Grammatical errors also seem to occur slightly more frequently in the ASR condition. As for the accuracy scores, it seems that small omissions and lexical errors occur more frequently in the unsupported condition.

When we look at the individual students (Table 2.3), ASR support is associated with fewer error tags for both accuracy and acceptability in two students and with more error tags for both parameters in two other students (**p < 0.01; *p < 0.05, chi-squared with general trend).

Strikingly, all students who received ASR support during the first and third speeches (ID tags starting with '2') saw their accuracy scores decrease. This may be related to the order in which the texts were administered. In the first text to be interpreted, students 2.1, 2.2 and 2.3 had to deal with both the numbers and the visual input of the CAI tool. The combination of those tasks also seems to have influenced the results reported in the next section.

Fundamental frequency

The F0 data are summarized in Figures 2.2 and 2.3. ASR-supported data are not limited to instances where students watched the screen, but encompass all instances where ASR was available. Conversely, some students took notes while

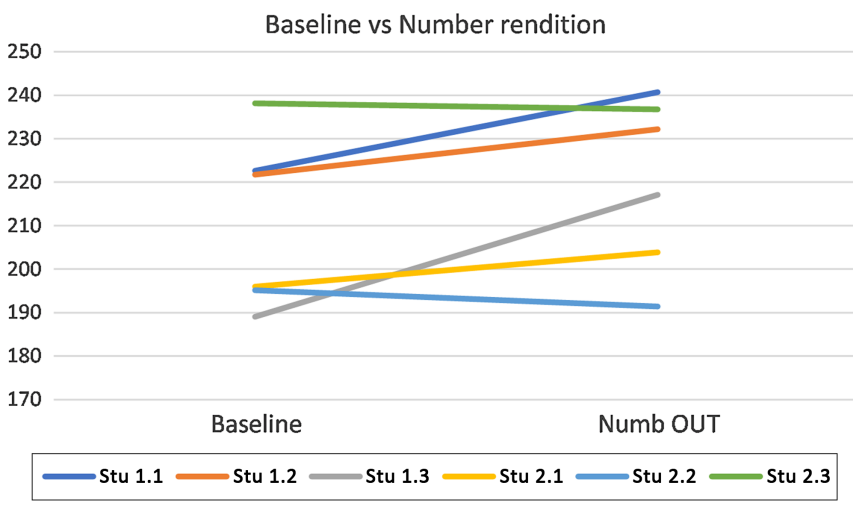

Figure 2.2 Mean of F0 in baseline interpreting and unsupported interpreting of numbers.

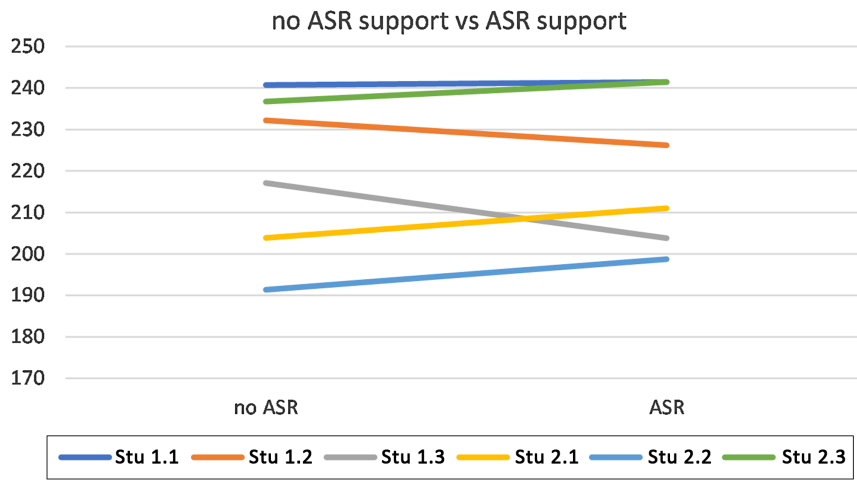

Figure 2.3 Mean of F0 in ASR-supported and unsupported conditions.

interpreting. Thus, in the unsupported condition, there may have been self-generated visual support.

Figure 2.2 compares the unsupported baseline condition (no numbers) and the unsupported interpretation of numbers. In four out of six interpreters the mean F0 value is higher when they interpret numbers than in the baseline condition. As Table 2.4a shows, F0 rises by 8 up to 28 Hz, effect sizes range between 0.18 (Student 2.1) and 0.70 (Student 1.3). Two students (2.2 and 2.3) show a small drop in mean F0, but effect sizes are minute in these cases. It thus seems that a rise of mean F0 is at least marginally associated with the cognitively more challenging task.

Figure 2.3 compares ASR-supported interpretation of numbers with unsupported interpretation of numbers. The overall picture is complex: three students show higher mean F0 in the ASR-supported condition; two present a lower mean F0 in that condition; and in one student there is no difference.

As Table 2.4b shows, all effect sizes are small. The expected additional load associated with the availability of visual output generated by an ASR-based support system can therefore not be substantiated.

It is important to point out that students in the same group show the same tendencies. Students 2.1, 2.2 and 2.3 were offered ASR-based support for the first and the third speech and had to interpret the second and the fourth speech without support. Considering the results shown in Tables 2.3 and 2.4a, F0 did not rise, or rose only marginally in this group for the task of interpreting numbers, while the other group consistently shows larger increases in F0. In contrast, students 2.1, 2.2 and 2.3 are the only ones to show a small increase in F0 in the ASR-supported condition, as compared to the unsupported condition.

Table 2.3 Accuracy and acceptability patterns per student

Student	Accuracy trend with ASR	Acceptability trend with ASR
1.1	Higher	Lower
1.2	Higher	Higher
1.3	Higher**	Higher
2.1	Lower**	Lower**
2.2	Lower*	Lower*
2.3	Lower	Higher

Table 2.4a Mean F0 differences and effect sizes in baseline interpreting (BL) and unsupported interpreting of numbers

	ΔF0 (µF0 Numb – µF0 BL)	Hedges' *g*
Student 1.1	18 Hz	0.45
Student 1.2	11 Hz	0.28
Student 1.3	28 Hz	0.70
Student 2.1	8 Hz	0.18
Student 2.2	−4 Hz	0.11
Student 2.3	−2 Hz	0.03

Table 2.4b Mean F0 differences and effect sizes in ASR-supported and unsupported number rendition

	ΔF0 (µF0 ASR – µF0 no ASR)	Hedges' *g*
Student 1.1	1 Hz	0.02
Student 1.2	−6 Hz	0.17
Student 1.3	−13 Hz	0.15
Student 2.1	8 Hz	0.19
Student 2.2	7 Hz	0.24
Student 2.3	5 Hz	0.13

The results may therefore be associated again with the order in which ASR-based support was provided. As students were unfamiliar with ASR-based support, those who received it for the very first speech seem to have experienced more cognitive load than when no support was provided. This effect may have been mitigated in the other group because they had already interpreted a text before they were offered a text with support.

Conclusion

The aim of this chapter was twofold. It first sought to assess the quality of interpretations delivered in two conditions: with support of a CAI tool based

on automatic speech recognition and without such support. It then aimed to determine whether F0 data yield reliable results as a proxy for cognitive load. With regard to the first aim, it seems to matter little for overall quality of performance whether students are offered a CAI tool or not: the number of error tags in both conditions is very similar. Accuracy seemed to fare slightly better with the CAI tool, but acceptability was slightly lower. This seems to concur with Lamberger-Felber's (2001) findings on the effects of visual input on interpreter performance and with the intuition that consulting a source of visual information may affect interpreter fluency. The results should be seen against the backdrop of significantly increased accuracy in the rendition of the target items of the research, namely numbers, as reported in Defrancq and Fantinuoli (2021). On balance, offering a CAI tool in the booth does have a positive effect on student performance.

As for the F0 data, there seems to be limited evidence of an association between increased F0 and the cognitively challenging task of interpreting numbers, and there is no evidence of an association between increased F0 and the availability of ASR-based support in the booth while interpreting numbers. The latter affords an affirmative answer to our third research question: the cognitive downsides and upsides of ASR-based support seem to cancel each other out. However, the potential of F0 to offer an alternative measure of cognitive load is not fully evidenced by the data collected for the second research question: the cognitive effects triggered by the interpretation of numbers can only be substantiated for two out of six participants on the basis of mean F0. For two other participants, the effect is noticeable, but the effect sizes are small. In the two remaining participants, the effect is absent. The results seem to confirm that F0 data are promising, but that further research is needed, not only in more controlled conditions, but also on aspects of F0 other than the mean F0: standard deviation, peaks and ranges.

This study relied on a re-analysis of existing data collected in conditions that were not specifically designed for the analysis of the parameters of interest. It is therefore paramount to repeat the data collection under more tightly controlled conditions, controlling for number complexity and type, pace of number delivery, lexical variation in the context of the number and so on. Professional interpreters' performances should also be explored to check whether expert participants show the same levels of increased accuracy and increased cognitive load. Being more used to performing simultaneous interpreting with text, it could very well be that professional interpreters are less likely to manifest the fluency issues observed in this study. Finally, careful cross-validation of

methods is required as a next step; this can be realized by building a dataset that combines eye-tracking and pupillometry data with recordings of participant interpretations to check whether pupil dilation and fixation patterns correspond to F0 patterns.

Notes

1 This research was carried out in the framework of the FWO Hercules Grant I003618N. None of the research was sponsored by the InterpretBank company, except for the offer to use a prototype version of the software for the purpose of the experiment. The authors also wish to thank An Baeyens wholeheartedly for preparing and delivering the speeches.

2 Fixations are time spans during which the subject's eyes fix particular segments of the visual input; saccades are rapid eye movements between fixations (Cassin et al. 1990).

3 Sense consistency with original, logical cohesion, completeness, correct terminology, correct grammar, appropriate style, fluency of delivery, lively intonation, pleasant voice, synchronicity and native accent.

4 Lexical discrepancies are items used in the interpretation that are semantically suited for the particular context, but not accurate in the sense that they are not situated on the same level of the lexical structure. Examples are hyponyms or hypernyms of the source item, *pars pro toto* and downtoning.

References

Ahrens, B. (2004), *Prosodie beim Simultandolmetschen*, Frankfurt am Main: Peter Lang.

Baxter, R. (2016), 'Exploring the Possible Effects of Visual Presentations on Synchronicity and Lag in Simultaneous Interpreting', *Sendebar: Revista de la Facultad de Traducción e Interpretación*, 27: 9–23.

Cassin, B., S. Solomon, and M. L. Rubin (1990), *Dictionary of Eye Terminology*, Gainesville: Triad.

Chmiel, A., P. Janikowki, and A. Cieślewicz (2020), 'The Eye or the Ear? Source Language Interference in Sight Translation and Simultaneous Interpreting', *Interpreting*, 22 (2): 187–210.

Cohen, J. (1977), *Statistical Power Analysis for the Behavioral Sciences*, New York: Academic Press.

Collados Aís, Á. (2001), 'Efectos de la entonación monótona sobre la recuperación de la información en receptores de interpretación simultánea', *Trans*, 5: 103–10.

Darò, V. (1990), 'Voice Frequency in Languages and Simultaneous Interpretation', *The Interpreters' Newsletter*, 3: 88–92.

Defrancq, B. and C. Fantinuoli (2021), 'Automatic Speech Recognition in the Booth. Assessment of System Performance, Interpreters' Performances and Interactions in The Context of Numbers', *Target*, 33 (1): 73–102.

Desmet, B., M. Vandierendonck, and B. Defrancq (2018), 'Simultaneous Interpreting of Numbers and the Impact of Technological Support', in C. Fantinuoli (ed.), *Interpreting and Technology*, 11–24, Berlin: Language Science Press.

Ehrensberger-Dow, M. and S. O'Brien (2015), 'Ergonomics of the Translation Workplace: Potential for Cognitive Friction', *Translation Studies*, 4 (1): 98–118.

Fantinuoli, C. (2017), 'Speech Recognition in the Interpreter Workstation', in J. Esteves-Ferreira, J. Macan, R. Mitkov, and O.-M. Stefanov (eds), *Proceedings of Translation and the Computer 39*, 25–34, Geneva: Tradulex.

Fantinuoli, C. (2018), 'Interpreting and Technology: The Upcoming Technological Turn', in C. Fantinuoli (ed.), *Interpreting and Technology*, 1–10, Berlin: Language Science Press.

Gile, D. (1995), *Regards sur la recherche en interprétation de conférence*, Lille: PUL.

Gile, D. (1999), 'Variability in the Perception of Fidelity in Simultaneous Interpretation', *Hermes*, 22: 51–79.

Gile, D. (1999), 'Testing the Effort Model's Tightrope Hypothesis in Simultaneous Interpreting – A Contribution', *Hermes*, 23: 153–72.

Han, C. (2017), 'Using Analytic Rating Scales to Assess English–Chinese Bi-directional Interpreting: A Longitudinal Rasch Analysis of Scale Utility and Rater Behaviour', *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 16: 196–215. doi:10.52034/lanstts.v16i0.429.

Hollnagel, E. and D. Woods (2005), *Joint Cognitive Systems. Foundations of Cognitive Systems Engineering*, Boca Raton: CRC Press.

Huttunen, K., H. Keränen, E. Väyrynen, R. Pääkkönen, and T. Leino (2011), 'Effect of Cognitive Load on Speech Prosody in Aviation: Evidence from Military Simulator Flights', *Applied Ergonomics*, 42: 348–57.

Johnstone, T. and K. Scherer (2000), 'Vocal Communication of Emotion', in M. Lewis and J. Haviland (eds), *The Handbook of Emotions*, 220–35, New York: Guilford.

Kinchla, R. (1974), 'Detecting Target Items in Multielement Arrays. A Confusability Model', *Perception and Psychophysics*, 15: 149–58.

Koby, G. and I. Lacruz (2017), 'The Thorny Problem of Translation and Interpreting Quality', *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 16: 1–12. DOI:10.52034/lanstts.v16i0.487.

Korpal, P. (2017), *Linguistic and Psychological Indicators of Stress in Simultaneous Interpreting*. Poznań: Wydawnictwo Naukowe UAM.

Korpal, P. and K. Stachowiak-Szymczak (2018), 'The Whole Picture. Processing of Numbers and their Context in Simultaneous Interpreting', *Poznan Studies in Contemporary Linguistics*, 54 (3): 335–54.

Korpal, P. and K. Stachowiak-Szymczak (2020), 'Combined Problem Triggers in Simultaneous Interpreting: Exploring the Effect of Delivery Rate on Processing and Rendering Numbers', *Perspectives*, 28 (1): 126–43.

Lamberger-Felber, H. (2001), 'Text-oriented Research into Interpreting: Examples from a Case-study', *Hermes*, 26: 39–63.

Lavault-Olléon, E. (2016), 'Traducteurs à l'œuvre: Une Perspective ergonomique en traductologie appliquée', *Revue de l'Institut des langues et cultures d'Europe, Amérique, Afrique, Asie et Australie*, 27: 1–9.

Lenglet, C. and C. Michaux (2020), 'The Impact of Simultaneous-interpreting Prosody on Comprehension', *Interpreting*, 22 (1): 1–34. doi:10.1075/intp.00032.len.

Lv, Q. and J. Liang (2019), 'Is Consecutive Interpreting Easier than Simultaneous Interpreting? – A Corpus-based Study of Lexical Simplification in Interpretation', *Perspectives*, 27 (1): 91–106.

Miller, J. (1982), 'Divided Attention: Evidence for Coactivation with Redundant Signals', *Cognitive Psychology*, 14: 247–79.

O'Brien, S. M., M. H. Ehrensberger-Dow, and M. Connolly (2017), 'Irritating CAT Tool Features that Matter to Translators', *Hermes Journal of Language and Communication in Business*, 56: 145–62. doi:10.21256/zhaw-3622.

Orlando, M. (2010), 'Digital Pen Technology and Consecutive Interpreting: Another Dimension in Note-Taking Training and Assessment', *The Interpreters' Newsletter*, 15: 71–86.

Pisani, E. and C. Fantinuoli (2021), 'Measuring the Impact of Automatic Speech Recognition on Number Rendition in Simultaneous Interpreting', in C. Wang and B. Zheng (eds), *Empirical Studies of Translation and Interpreting. A Post-structuralist Approach*, 181–97, London: Routledge.

Plevoets, K. and B. Defrancq (2016), 'The Effect of Informational Load on Disfluencies in Interpreting: A Corpus-based Regression Analysis', *Translation and Interpreting Studies*, 11 (2): 202–24.

Plevoets, K. and B. Defrancq (2018), 'The Cognitive Load of Interpreters in the European Parliament. A Corpus-based Study of Predictors for the Disfluency uh*(m)*', *Interpreting*, 20 (1): 1–29.

Pöchhacker, F. (2001), 'Quality Assessment in Conference and Community Interpreting', *Meta*, 46 (2): 410–25. doi:10.7202/003847ar

Prandi, B. (2018), 'An Exploratory Study on CAI Tools in Simultaneous Interpreting: Theoretical Framework and Stimulus Validation',in C. Fantinuoli (ed.), *Interpreting and Technology*, 29–59. Berlin: Language Science Press.

Rodríguez, S., R. Gretter, M. Matassoni, D. Falavigna, A. Alonso, O. Corcho, and M. Rico. (2021), 'SmarTerp: A CAI System to Support Simultaneous Interpreters in Real-time', in R. Mitkov, V. Sosoni, J. C. Giguère, E. Murgolo, and E. Deysel (eds), *Proceedings of the Translation and Interpreting Technology Online Conference*, 102–9, Shoumen: INCOMA Ltd.

Rothkrantz, L., P. Wiggers, J. W. van Wees, and R. van Vark (2004), 'Voice Stress Analysis', in P. Sojka, I. Kopeček, and K. Pala (eds), *Proceedings of Text, Speech and Dialogue*, 7, 449–56, Berlin: Springer.

Scherer, K., D. Grandjean, T. Johnstone, G. Klasmeyer, and T. Bänzige (2002), 'Acoustic Correlates of Task Load and Stress', in Z. Tan and P. Dalsgaard (eds), *International Conference on Spoken Language Processing 7 – Interspeech 2002, 2017–2020*, Denver: ISCA.

Shlesinger, M. (1994), 'Intonation in the Production and Perception of Simultaneous Interpretation', in S. Lambert and B. Moser-Mercer (eds), *Bridging the gap: Empirical Research in Simultaneous Interpretation*, 225–36, Amsterdam: John Benjamins. doi:10.1075/btl.3.17shl

Shlesinger, M. (2003), 'Effects of Presentation Rate on Working Memory in Simultaneous Interpreting', *The Interpreter's Newsletter*, 12: 37–49.

Seeber, K. (2011), 'Cognitive Load in Simultaneous Interpreting: Existing Theories – New Models', *Interpreting*, 13 (2): 176–204.

Seeber, K. (2012), 'Multimodal Input in Simultaneous Interpreting: An Eye-tracking Experiment', in L. Zybatov, A. Petrova, and M. Ustaszewski (eds), *Proceedings of the 1st International Conference TRANSLATA, Translation and Interpreting Research: Yesterday – Today – Tomorrow*, 341–47, Frankfurt: Peter Lang.

Seeber, K. (2015), 'Multimodal Processing in Simultaneous Interpreting', in J. Schwieter and A. Ferreira (eds), *The Handbook of Translation and Cognition*, 461–75, Hoboken: Wiley.

Seeber, K. and D. Kerzel (2012), 'Cognitive Load in Simultaneous Interpreting: Model Meets Data', *International Journal of Bilingualism*, 16 (2): 228–42.

Seeber, K., L. Keller and A. Hervais-Adelman (2020), 'When the Ear Leads the Eye – The Use of Text during Simultaneous Interpretation', *Language, Cognition and Neuroscience*, 35 (10): 1480–94. doi:10.1080/23273798.2020.1799045

Stachowiak-Szymczak, K. and P. Korpal (2019), 'Interpreting Accuracy and Visual Processing of Numbers in Professional and Student Interpreters: An Eye-tracking Study', *Across Languages and Cultures*, 20 (2): 235–51. doi:10.1556/084.2019.20.2.5

Van Cauwenberghe, G. (2020), *La reconnaissance automatique de la parole en interprétation simultanée. Étude expérimentale de l'impact d'un soutien visuel automatisé sur la restitution de terminologie spécialisée*, MA dissertation, Ghent University, Gent.

Van Puyvelde, M., X. Neyt, F. McGlone, and N. Pattyn (2018), 'Voice Stress Analysis: A New Framework for Voice and Effort in Human Performance', *Frontiers in Psychology*, 9, A01994. doi:10.3389/fpsyg.2018.01994

Yang, S. (2019), *Investigating the Effect of Speech Rate on the Cognitive Load in Simultaneous Interpreting with Text*, PhD dissertation, University of Macau, Macau.

Zwischenberger, C. (2011), *Qualität und Rollenbilder beim simultanen Konferenzdolmetschen*, PhD dissertation, University Vienna, Vienna.

3

Customization, Personalization and Style in Literary Machine Translation

Dorothy Kenny and Marion Winters

Introduction

Literary machine translation has come in from the cold. Once the preserve of the avant-garde (see Lennon 2014; Kenny 2020), it is now a focus of attention in mainstream machine translation (MT) and translation studies circles. As interest in the field has grown, so too have the variety of questions addressed and the range of methodologies adopted. Some studies follow well-trodden paths in MT research,[1] focusing on the automatic and human evaluation of MT outputs (e.g. Toral and Way 2018). Others also draw on familiar traditions in translation studies, homing in, for example, on putative universal features of post-edited machine-translated texts (Toral 2019; Castilho and Resende 2022). But only a small body of research has so far integrated the figure of the literary translator, to consider how they might experience literary MT, how the use of MT might affect their style, and what a shift to MT might mean for their oeuvre. In a previous study (Kenny and Winters 2020), we began this line of enquiry. In this chapter we continue in this vein by laying the groundwork for further investigations of translator/post-editor style in scenarios where MT is used. In our ongoing research (Kenny and Winters 2020; Winters and Kenny 2024), we are particularly interested in how the style of one prominent literary translator, Hans-Christian Oeser, is affected when he works in post-editing mode. Our broad investigation obviously relies on empirical observation of texts post-edited by Oeser, but it also requires us to have a clear understanding of what we mean by 'style' to begin with. At the same time, we are keenly aware that since we began our studies of translator style (see, e.g. Kenny 2001; Winters 2007), the world has changed: MT has improved to the extent that its use in literary translation

is no longer principally associated with the production of disfluencies and comic effects (see Kenny 2020); MT researchers have become more and more interested in the translation of literature (e.g. Voigt and Jurafsky 2012; Toral and Way 2018); and there is increasing focus in general on the customization of MT engines to make them more suitable for specific tasks, such as the translation of specific literary genres (Hansen 2021) or even the translation of works by a specific author, based on previous translations by a specific translator (Hansen et al. 2022). As the horizons of MT researchers expand, so too does their interest in notions like style, and their understanding of this concept begins to feed into an already crowded space. Meanwhile, long-established understandings of style in the field that we call home, namely corpus-based translation studies, also need to be revisited given the more expansive treatment the concept has recently received in digitally informed literary studies (Herrmann, van Dalen-Oskam and Schöch 2015). In short, before we proceed to further empirical investigation of translator style in post-edited texts, we need to take stock.

We thus begin our exploration in this chapter by surveying recent work in literary MT, drawing a tentative map of the various approaches taken in the field. We home in on studies that address the related concepts of customization, personalization and style in the contexts of machine and human translation, before making the case for the adoption of a literary studies approach to style in MT research and characterizing literary post-editing as a kind of downstream translator-specific personalization. We conclude, by way of illustration, by giving brief details of one study which has already taken this approach.

Literary machine translation

We understand 'literary machine translation' as an emerging interdisciplinary field that embraces a range of phenomena related to the application of MT to the translation of literary texts. It touches upon or intersects with, among other areas: speculative and empirical enquiry into the features that make literary texts particularly amenable to or (more often) difficult for MT (e.g. Voigt and Jurafsky 2012; Taivalkoski-Shilov 2019a); computational techniques that might meet the attendant challenges (e.g. Van de Cruys 2020); the application of generic MT to literary texts (e.g. Fonteyne, Tezcan and Macken 2020); the customization of MT systems for use with literary texts (e.g. Toral and Way 2018; Kuzman, Vintar and Arčan 2019; Hansen 2021; Hansen et al. 2022); the systematic evaluation of literary texts translated by MT (e.g. Toral and Way 2018; Moorkens et al. 2018;

Matusov 2019); literary translators' interactions with MT (Kenny and Winters 2020; Oeser 2020); reaction to the use of MT in the literary translation profession (Oeser 2020); readers' reception of machine-translated literature (Guerberof-Arenas and Toral 2020, 2022); style in human, raw and post-edited literary MT (Kenny and Winters 2020; Lee 2021); and studies of so-called 'universals' in post-edited literary texts (e.g. Castilho and Resende 2022). It would clearly be impossible to do justice here to all these interweaving strands. In what follows we focus on studies that engage in particular with questions of customization, personalization and style in literary MT.

Customization, personalization and style in literary MT

The current dynamism of the field shows that literary translation is now seen as a valid object of study in MT circles, having previously been considered beyond the reach of the technology. The increased attention given to literary MT coincided with the step change in quality that accompanied the transition first from rule-based to statistical MT and then, more rapidly, from statistical to neural MT in the mid-2000s and mid-2010s, respectively. A detailed discussion of how and when these transitions came about is beyond the scope of this chapter. The reader is referred instead to treatments in Bentivogli et al. (2016), Kenny (2018), Hansen (2021) and Hansen et al. (2022), among others. Suffice it to say here that statistical MT (SMT) and neural MT (NMT) not only offered superior quality to their respective predecessors (based on both human and automatic evaluations of their outputs), they also allowed customization. MT engines could now be trained specifically on literary texts from which they would presumably learn mathematical models of literary translation. Some early research in the area (e.g. Toral and Way 2015a, 2015b) focused on adapting generic SMT engines by adding relatively modest quantities of literary text to their training data, yielding 'literary-adapted MT' (Toral and Way 2015b). As efforts scaled up, and deep learning became embedded in MT, larger corpora of literary texts — usually novels — and their translations were used to train NMT engines that could now claim to be doing fully fledged 'literary MT', even without the support of 'out-of-domain' (i.e. non-literary) data.[2] At the time of writing, custom NMT engines designed for use with literary texts may be trained on both generic and literary text or on literary text alone. The NMT engines reported on in Guerberof and Toral (2022) serve as a case in point: the English-to-Catalan engine is trained on both out-of-domain and in-domain

(literary) text, while the English-to-Dutch engine is trained exclusively on literary text.[3]

Customization can take even more specific forms: texts written by particular authors and translations by particular translators can be integrated into the training data for a given engine, steering literary MT towards increasing personalization. What we call 'translator-specific personalization' is mooted in Oeser (2020: 23), while Kuzman, Vintar and Arčan (2019) experiment with both author- and translator-specific training data in literary MT, thus yielding both author- and translator-specific personalization, albeit on a very small scale using a single translated text as parallel data, and with limited success. Likewise, Hansen et al. (2022) report on an MT engine trained on six novels produced by one particular author and translated by one particular translator working in the heroic fantasy genre. The genre appears to be especially problematic for MT, whether generic or custom (2022), but Hansen et al. still conclude that early experiments point to the potential of translator-specific personalization in literary MT. Such personalization based on training data appears to us to be a promising way to give 'textual voice' (Alvstad et al. 2017) to human translators in the context of literary MT, following an ethical imperative that we have explored in previous work (Kenny and Winters 2020).

Generic systems have, meanwhile, proved capable of competing with and sometimes outperforming customized literary systems – even (modestly) personalized ones – in research settings (Kuzman, Vintar and Arčan 2019; Matusov 2019), no doubt because of the sheer size of the training data they learn from. More tellingly perhaps, the improved performance that accompanied the rise of NMT has meant that even generic MT has reached a level of quality at which some literary translators can begin to countenance experimenting with it (see Kenny and Winters 2020: 133; Oeser 2020: 20). This opens the door to an alternative approach to translator-specific personalization in our view: rather than changing the training data in an upstream effort to produce translations that reflect the preferences or style of a particular translator (although we do not deny the merit of this approach), it is also feasible to personalize downstream. In this case a literary translator would post-edit the output of a generic MT engine, so that it conforms to their expectations of a what a good translation of the source should look like. In so doing, we hypothesize, they will put their own 'stamp' on the final translation. What we are suggesting is that a literary translation created by a machine and post-edited by an experienced human literary translator will contain elements that reflect the style of that translator. If we have independent evidence as to what constitutes the translator's style, then we should be able to

show empirically whether their style as a post-editor coincides with their style as a translator. We characterize our approach as involving 'downstream translator-specific personalization'.

Before describing precisely how we proceed, we revisit some key concepts in literary MT, addressing the motivations behind 'literary-adapted' MT, and especially translator-specific personalization based on training data, problematizing in particular the understanding of style in MT research.

The interaction between corpus, domain and style in machine translation research

The basic idea behind any kind of customization or 'domain adaptation' in MT is that the best translation engine will be the one that has been trained (or fine-tuned) on 'relevant' data (Koehn 2020: 239).[4] Relevance here suggests that the texts in the corpus used to train the system share characteristics with the texts the system will subsequently be used to translate.[5] The term 'domain' is used very broadly in MT research, however: texts from the same domain might be expected to share 'similar topic[s], style, level of formality, etc.' according to Koehn (2020: 239), but in practice it is typically the provenance of a corpus that marks it as belonging to a given domain, or in Koehn's (2020: 239) words, 'domain . . . typically means a corpus that comes from a specific source'. The assumption in much of the work in MT seems to be that similar 'topic, style, level of formality, etc.' (2020: 239) can be inferred from provenance.

In this light, the treatment of 'literary texts' as a domain in literary MT seems problematic: there is no single source (e.g., author, publisher, or even country) from which the 133 English-language novels in the parallel corpus used in Toral and Way (2018) might be said to hail, for example. The same can be said of the 1,000 books in the Catalan and the 1,600 books in the English monolingual corpora used in their study. It also seems unlikely that the texts in the corpora in question can be united by topic or level of formality, as per Koehn's description. We are left with the assumption that the texts in each corpus perhaps share a certain literary style, although this is not explicitly pursued in Toral and Way (2018) or related studies.[6]

The picture gets fuzzier still when one considers how the term 'style' itself is understood in MT research. It seems that the general vagueness around the term 'domain' is mirrored in the treatment of 'style'. Korotkova, Del and Fishel (2018: 2), for example, treat the term 'text style' 'loosely as covering concepts

like text domain, genre, formality and other text characteristics', making domain an element of 'style'. The three 'styles' included in their study are OpenSubtitles, Europarl and JRC-Acquis, where 'style' effectively becomes coterminous with 'corpus', in the same way as 'domain' became coterminous with 'corpus' in Koehn (2020).[7] Korotkova, Del and Fishel (2018) provide examples of how sentences in French and German are automatically translated into English in the three different 'styles' under consideration. The French string '*on se lance?*', for example, is translated as 'let's go' in the OpenSubtitles 'style', 'are we getting started?' in the Europarl 'style', and 'are we going?' in the JRC-Acquis 'style'. Conceptual vagueness, it turns out, does not prevent the machine learning techniques used in NMT from learning (at least some of) those features of each corpus that mark it as stylistically different from the other two.[8]

Michel and Neubig (2018) also attempt to create NMT outputs in varying styles, but this time the focus is on mimicking the style of individual source-text speakers in a corpus of translated TED talks. Wang, Hoang and Federico (2021) pursue the related aim of mimicking the styles of individual translators in a similar corpus of translated TED talks. In both cases, the authors treat each individual (speaker or translator) as a 'domain' and add discrete speaker or translator tokens to the training data to create personalized NMT engines.[9] Interestingly, Michel and Neubig (2018) do not refer to 'style' as such. Rather they use the phrases 'speaker traits' and 'speaker-related variations' to capture what becomes labelled as 'speaker and author style' in Wang, Hoang and Federico (2021). Wang, Hoang and Federico (2021) commit to the term 'style', however, and claim that their 'style-augmented translation models are able to capture the style variations of translators and to generate translations with different styles on new data' (2021: 1). They also claim that 'translator information has more impact on NMT than the speaker information', although one possible explanation for this is that there are fewer translators than speakers in their training data, as some translators translate speeches by several different speakers. Hence there is more training data per individual translator than per individual speaker, making the translator signal stronger (2021: 5). They note that style-augmented translation outperforms the baseline in both human and automatic evaluations, and that 'human evaluation confirms that observed differences are all about style and not translation quality' (2021: 1–2). Wang, Hoang and Federico (2021), like some of the above-mentioned researchers, thus suggest that there is merit in translator-specific personalization in MT research, and that this merit resides in the ability of personalized MT output to capture translators' styles. They do not define 'style' formally, however, choosing instead to give examples of what has

been considered to constitute 'style' in the MT literature, mentioning 'verbosity', 'formality', 'politeness', 'demography' and 'personal traits' (2021: 2). In their own analysis they exemplify 'style' using three criteria: verbosity (how many words a translator uses), and lexical and grammatical choices. For a study that purports to be interested in translator style, however, references to the translation studies literature on the subject are conspicuous by their absence.

Overall, the MT literature is often vague or non-committal when it comes to defining style. This vagueness means that it is difficult to distinguish between style and related concepts in MT research, and that it would be difficult to draw on definitions from the area in any attempt to operationalize style in investigations such as ours. Rather than rely on the under-conceptualized notion of style encountered in much MT research, we thus turn for guidance to a field that has been concerned with literary style for centuries (Herrmann, van Dalen-Oskam and Schöch 2015: 25) and that has influenced corpus-based translation studies, namely literary studies.

Style in literary studies

There is a long history of scholarship on style in literary studies. Herrmann, van Dalen-Oskam and Schöch (2015) provide a comprehensive overview of contributions made since 1945, taking in hermeneutic, structuralist, pragmatic and digital humanities approaches, among others. The definitions of style encountered in these traditions, according to Herrmann, van Dalen-Oskam and Schöch (2015: 26) cover:

> style as revealing a higher-order aesthetic value, as the holistic 'gestalt' of single texts, as an expression of the individuality of an author, as an artefact presupposing choice among alternatives, as a deviation from a norm or reference, or as any formal property of a text.

In the interests of establishing 'a common ground for literary scholars and stylisticians when talking about style' (2015: 28), Herrmann, van Dalen-Oskam and Schöch go on to propose 'an operational definition of style that incorporates a minimal common ground for interdisciplinary empirical research and the application of new, digital methods' (2015: 28). Their subsequent definition of style is one that resonates with us, given its commitment to empiricism and the ease with which it can accommodate digital methods. It is also broad and non-normative, in the sense that it does not promote coherent style as a sign

of aesthetic quality. Nor does it suggest that style is unique to literary texts. It presents style as an empirical textual category, that is, as something that can be observed in texts, where observation can be primarily based on quantitative or qualitative methods, or a mixture of both. Herrmann, van Dalen-Oskam and Schöch's definition of style is as follows:

> *Style is a property of texts constituted by an ensemble of formal features which can be observed quantitatively or qualitatively*. (2015: 44, original emphasis)[10]

This definition is largely consistent with that adopted in digital corpus stylistics, one of the areas on which Herrmann, van Dalen-Oskam and Schöch (2015) draw, and which has been influential in corpus-based research into style in translation studies (e.g. Baker 2000; Saldanha 2011). However, the definition does not require style to be identified contrastively, an approach often taken in corpus-based translation studies, but neither does it exclude such an approach. In quantitative stylistic analyses, in particular, inter-corpus investigations are common. In such cases:

> Frequencies of selected elements can be quantitatively related *across* a collection of individual texts, for example by using the collection itself as the relative norm, or using a large(r) reference corpus as a backdrop. (Herrmann, van Dalen-Oskam and Schöch: 47, original emphasis)

The fact that style does not have to be identified contrastively *a priori* is another factor that draws us to Herrmann, van Dalen-Oskam and Schöch's definition. In the past we have used Saldanha's (2011) influential definition of 'translator style' (Kenny and Winters 2020), which integrates contrast in the form of 'deviance from a norm' as a necessary condition for the identification of an individual translator's style. More specifically, a translator's style is identified partly on the basis of how their work differs from that of other translators. Such contrast is not a sufficient condition for the recognition of style in Saldanha's definition, as translator style also has to meet other criteria. Namely, it is

> A 'way of translating' which
>
> - is felt to be recognizable across a range of translations by the same translator,
> - distinguishes the translator's work from that of others,
> - constitutes a coherent pattern of choice,
> - is 'motivated', in the sense that it has a discernible function or functions, and
> - cannot be explained purely with reference to the author or source-text style, or as the result of linguistic constraints. (Saldanha 2011: 31)

The 'contrast with other translators' condition seems to us now to be somewhat problematic in that it suggests that two translators cannot have the same style, or share elements of the same style, which in our view is an unnecessarily restrictive position to take. Investigations of style in literary translation do not have the same priorities as investigations of style in forensic linguistics, for example, where aims like author (or translator) attribution mean that techniques that distinguish individuals from each other are paramount.[11]

Herrmann, van Dalen-Oskam and Schöch's (2015) definition also differs from Saldanha's (2011) in dropping the requirement for 'coherence', which they associate with 'more normative views of style which see coherent style as a sign of aesthetic quality' (2011: 44). They fall back on the notion of an 'ensemble' of formal features instead:

> By 'ensemble' we mean that style is constituted by the combination of many possible features and should be seen as a complex system, with features situated at different linguistic levels. Such an ensemble does not necessarily exhibit a coherent unity; rather, it can have various degrees of unity or harmony, or, on the contrary, contrasts or incoherence. (Herrmann, van Dalen-Oskam and Schöch 2015: 44)

This approach seems to us to not only allow for very broad descriptive explorations of style that are not unnecessarily restricted from the outset, but that also allow for the fact that an individual translator's style might vary over a range of texts or over time, even if some features of their style remain constant. That said, there is some evidence that individual author styles can persist over decades and do so despite changing editorial environments. This is the case, for example, with the judges of the US Supreme Court investigated in Jockers, Nascimento and Taylor (2020). Similarly, this persistence of individual style over time can also apply to translators, as we have demonstrated for Oeser (Kenny and Winters 2020).

Causality and style in literary machine translation

Herrmann, van Dalen-Oskam and Schöch's general approach also has the merit of allowing for *causality* to be integrated into discussions of style in many ways:

> Even in the absence of conscious intentions, causal relationships may be hypothesized: genre can cause style (e.g. by means of conventions: form and themes), authors can cause style (e.g. by means of idiosyncrasies), theme and topic can cause style. The interpretability of style relative to categories such as

> authorship, literary genre, or literary period, is hence paramount. This means that any stylistic phenomenon can ultimately be considered the trace of or the index towards such categories (or, in other terms, may be 'characteristic' of them). (2015: 46)

To Herrmann, van Dalen-Oskam and Schöch's list, and drawing on the literature in corpus-based translation studies and literary MT, we would add the following hypotheses, some of which have been tested in the field.

Translation can cause style

The very fact of translation may influence the formal features of a target text, regardless of which translator performed the translation, what the source language was and so on. This is the basic idea behind research into so-called 'universals' of translation (see Baker 1993).

Translators can cause style

The formal features of a target text may be seen as a trace of a particular translator (Baker 2000; Kenny 2001; Saldanha 2005, 2011; Winters 2009, 2013).

Machine translation can cause style

Even if we allow for the fact that the particular texts on which an engine is trained, the genres to which they belong, and the language pair involved can be reasonable causes of style, we still have to admit that the algorithms used in contemporary machine translation themselves can cause style, most notably by amplifying characteristic features of texts in the domain in question, leading to an overuse of those features – for example, contractions like 'I'll' in subtitle translation (Korotkova, Del and Fishel 2018) – or to gender and other biases in their outputs (Hovy, Bianchi and Fornaciari 2020; Vanmassenhove, Shterionov and Gwilliam 2021).

Post-editing can cause style

Previous research has shown that post-editing – or more accurately, and for obvious reasons, the combination of machine translation and post-editing, also known as 'PEMT' – can lead to target texts that differ in systematic ways from human translations. Toral (2019), for example, uses a number of different corpora

involving several different language pairs to show that there are statistically significant differences between post-edited texts and human-translated texts on certain easily computed scores (e.g. type-token ratio and lexical density). In general, however, it is difficult to factor out post-editing from machine translation itself as a cause in this kind of study. Čulo and Nitzke (2016), who focus on terminological variation, thus find evidence of machine translation 'shining through' in PEMT texts, while Farrell (2018) similarly concludes that the PEMT texts in his study contain 'machine translation markers'. The latter two studies serve as examples of work that combines both quantitative and qualitative perspectives, identifying as they do the particular terms or expressions that prove to be characteristic of PEMT texts.

Post-editors can cause style

Although 'post-editing' is considered the condition of interest in many hypothesis-testing studies, some studies depart from what we might call a 'post-editing as monolith' position by pointing to differences between post-editors, among whom individual variation has been described as 'inevitable' (O'Brien and Simard 2014: 160). This is despite the fact that, for years, post-editors have been instructed to change as little of the machine translation output as possible (O'Brien 2022), so one would expect their room for manoeuvre to be limited. To date, however, there have been few, if any, studies that set out to investigate how stylistic phenomena can be considered 'the trace of or the index towards' (Herrmann, van Dalen-Oskam and Schöch 2015: 46) a particular post-editor. The work of Kenny and Winters (2020) is one exception and is described briefly in the next section.

The literary translator as post-editor: A case of downstream translator-specific personalization

As already indicated, we are interested in the potential of post-editing as a type of downstream translator-specific personalization in literary MT. Such personalization could be said to occur in cases where, by editing a machine-translated text in a particular way, a literary translator is shown to assert their own style in the post-edited text. In Kenny and Winters (2020), we attempted to capture data on such personalization in an experiment in which Hans-Christian Oeser post-edited a short text translated by DeepL. The text in

question was a chapter from F. Scott Fitzgerald's *The Beautiful and Damned* (1922), which Oeser had himself translated (from scratch, i.e. without the use of MT) some twenty years earlier (*Die Schönen und Verdammten*, 1998). His post-edited version was compared with his from-scratch translation in order to ascertain whether, as a post-editor, Oeser put his own stylistic stamp on the MT output. By tracking individual edits made on the short text, we found that Oeser's style, as familiar from previous corpus-driven studies, was somewhat attenuated in his post-editing. The study had some shortcomings, however. The translation commission was fictitious, and the use of MT was not the only variable that might have affected the outcome. Oeser himself commented on the passage of time as a factor affecting his performance as a translator/post-editor. In Winters and Kenny (2024) we address both of these issues. In this case the commission is a real one and we make comparisons with a corpus of translations published by Oeser around the same time as he completed the post-editing task in question, as well as with a reference corpus of texts originally written in German at roughly the same time. The passage of time is thus not a confounding variable.[12] Space constraints prevent us from giving fuller details of this study. In what follows we confine ourselves to its general design and some principal findings.

In Winters and Kenny (2024) we analyse a version of Christopher Isherwood's (1954) novel *The World in the Evening* that was post-edited by Hans-Christian Oeser on the basis of a machine translation into German produced by DeepL (free version used in 2019). Oeser's post-edited version was subsequently published as *Die Welt am Abend* (Isherwood 2019). As elaborated upon in Winters and Kenny (2024), Oeser's work on the DeepL translation does not conform to the image of 'post-editing' promulgated in many language industry sources. In fact, Oeser himself does not describe the process as post-editing but rather as one of 'painstaking retranslation', adding that 'there was hardly a sentence that did not have to be thoroughly revised and rebuilt' (Oeser 2020: 20). We received from Oeser the full raw machine-translated version of the novel and his own post-edited version, both in digital form, and can attest to the extent of the changes made to the MT output, which in turn suggested that there would be ample data to study both the style of the MT and the style of Oeser's subsequent post-edited version. What interests us most in our study is the extent to which, in editing the MT output, Oeser not only fixes errors, which is the main purpose in conventional post-editing (O'Brien 2022), but also makes changes that have the effect, whether intended or not, of imposing his style on the final version.

From previous quantitative and qualitative empirical studies (see especially Winters 2007, 2009) we have a comprehensive picture of Oeser's style as it then stood (see also Winters 2015). We have also created a reference corpus, *Oeser 12*, containing twelve of Oeser's recent translations (novels and novellas translated as sole translator between 2016 and 2021[13]) and a reference corpus of fifty-seven works originally written in German over a similar period. Our data sets allow us to study the style of the machine-translated text and the post-edited text in their own rights, by observing 'an ensemble of formal features' (Herrmann, van Dalen-Oskam and Schöch 2015: 44) in each. Knowing how both texts came into being, we can then hypothesize MT and post-editing as 'causes' of the various features in the latter text. This part of the analysis will certainly draw on contrastive analysis between the machine-translated and the post-edited texts, but it can also allow us to see similarities between these two texts. Meanwhile, by comparing the style of his post-edited text with what we already know about Oeser's style as a translator and with what we can glean about this style from the contemporaneous reference corpus of his translations, we can work out how consistent his 'style when post-editing' is with his 'style when translating'. If there is a high level of consistency, then the argument that literary MT post-editing can be seen as a kind of downstream translator-specific personalization becomes more compelling. Finally, comparisons with the reference corpus of original German literature allow us to investigate whether certain features of Oeser's style as a translator and post-editor are particularly indicative of his work, or whether they are also typical of German letters in general.

Of course in such a study, or rather series of studies, there are hundreds of formal features – involving varying levels of abstraction, quantification and qualitative analysis – that could be investigated (see Herrmann, van Dalen-Oskam and Schöch 2015: 45). For reasons elaborated upon in Winters and Kenny (2024), we start our investigative journey by tracing keywords in the various data sets. The analysis is thus, in the first instance, a lexical one and focuses on those lexical items that distinguish Oeser's post-edited text from the DeepL MT output. The findings reported in Winters and Kenny (2024) suggest that there is hardly a keyword in Oeser's post-edited version that cannot be explained with reference to his style. By way of illustration, and due to space constraints already alluded to, we give just a single example here.

The word form ranked fifteenth in the keywords list for Oeser's post-edited text is '*weshalb*' (why).[14] It occurs nineteen times in Oeser's post-edited text and not at all in the raw MT output.[15] Inspection of the two texts conducted

Table 3.1 Absolute frequency and frequency per 100,000 tokens of '*weshalb*' in Oeser's post-edited text, *Oeser 12* and *German Original Literature*

Keyword '*weshalb*'	Oeser's post-edited text	*Oeser 12*	*German Original Literature*
Absolute frequency	19	168	257
Frequency per 100,000 tokens	18.27	23.97	7.15

using a parallel concordancer shows that in every case, Oeser inserts '*weshalb*' to replace '*warum*' (a more common word also meaning why) in the MT output.[16] A comparison of the relative frequency of '*weshalb*' in Oeser's post-edited text, the reference corpus of his translations (*Oeser 12*) and the *German Original Literature* reference corpus, as depicted in Table 3.1, suggests that '*weshalb*' is a word characteristically used by Oeser.

It occurs nearly 24 times in 100,000 words (tokens) in *Oeser 12*. It is far less commonly used by writers of original prose in German, however, occurring slightly more than 7 times in 100,000 words (tokens) in the *German Original Literature* corpus. Interestingly the use of '*weshalb*' is mentioned by Oeser in his comments on the post-editing work we reported on in Kenny and Winters (2020). In that source, Oeser characterized its use as one of his 'quirks': 'Auch so eine Marotte von mir: lieber "weshalb" als "warum" zu verwenden' (Another one of my quirks: I prefer using '*weshalb*' to '*warum*') (quoted in Kenny and Winters 2020:143). The presence of so many instances of '*weshalb*' in his post-edited text thus brings that text more into line with what is known about Oeser's style, as observed in a corpus of his translation work and as revealed in Oeser's own commentary on his previous work.

Conclusion

In this chapter we have reflected on current strands in research into literary MT and have pointed up the relative lack of conceptualization regarding style in the field. We have revisited the concept of style in the light of recent thinking in literary studies and made the case for adopting Herrmann, van Dalen-Oskam and Schöch's (2015) broad and digitally accommodating definition in studies of translator style, whether or not those translators are working with MT output. We have exemplified our approach with one simple example, using just one analytical technique, and focusing on just a single translator and a

single post-edited text. In ongoing work (Winters and Kenny 2024; Kenny and Winters 2022) we account for far more of our data and apply additional techniques. We hope here, nevertheless, to have offered 'proof of concept' that a literary translator's post-editing activity can be regarded as downstream translator-specific personalization of MT output.

Notes

1 By 'MT research' we mean research conducted primarily by computer scientists specializing in MT.

2 Toral and Way's (2018) early experiments with NMT, for one, were based on training data composed entirely of novels translated from English into Catalan and a monolingual corpus of novels in Catalan.

3 'Generic' engines are understood here as MT engines that are trained on sentences taken from texts covering 'a wide range of topics, styles and genres, and not specialized in any particular domain' (Ramírez-Sánchez 2022: 165). These are typically contrasted with 'custom', 'domain-adapted' or 'bespoke' engines, which are designed for use with specific genres in specific domains (e.g. legislation, patents, novels, etc.). The distinction between 'systems' and 'engines', meanwhile, roughly follows that made in Kenny (2022: 45): an MT system is understood as an MT product or service that is made available by a single supplier or developer. Familiar examples include Google Translate and Microsoft Translator. Many such systems often come in a free online version, yielding the acronym FOMT. An MT engine, on the other hand, is an MT 'program (or even a "model"') that has been trained to deal with a particular language pair and, often, domain or genre' (2022: 45). A single system provider may thus offer several different generic and/or custom engines. In cases where it is not necessary to make the distinction between system and engine or model, the term 'system' serves as a useful superordinate.

4 A discussion of how domain adaptation is achieved in practice (e.g. through fine-tuning and data interpolation) is beyond the scope of this chapter. The reader is referred instead to Koehn (2020: 239–61).

5 An engine trained on a parallel corpus of legislation, for example, should be well adapted to the task of translating previously unseen legislative texts in the language pair in question.

6 The impressive corpora used in this study have also been used in subsequent studies, including Moorkens et al. (2018) and Guerberof-Arenas and Toral (2020, 2022), none of which, however, problematize the notions of domain or style.

7 Stylistic differences are, in turn, assumed to exist between these corpora (Korotkova et al. 2018: 3), and despite acknowledged intra-corpus heterogeneity in

OpenSubtitles, many 'stylistic traits' are said to be 'similar' even within this corpus, 'which means that these common traits can be learned as a single style' (Korotkova et al. 2018: 3).

8 Note here also that, in line with other natural language processing studies in 'style transfer', Korotkova et al. (2018) distinguish between 'semantic content' and 'style', while acknowledging that in practice the two concepts are sometimes dealt with together.

9 For example, the speaker token is added at the start of the target sentence for each speaker in Michel and Neubig (2018).

10 Herrmann et al. (2015: 30) point out that their contribution 'centers on style definitions that are valid at the textual level of usage' and not 'those approaches that deal explicitly with style at the level of the reader's processing, such as psychological theories of style processing and cognitive-linguistic theories in their pure form'. The 'eminently important' analytical distinction between 'the levels of reader response and textual structure' is, however, acknowledged, as is the role of cognition in subjective judgements about deviation and norms, for example (2015: 30).

11 Note that corpus linguists in general have also been criticized in the past for setting too much store by difference at the expense of recognizing similarity (see, e.g. Baker 2004; Taylor 2018). Note also that the dropping of a requirement of contrast with others from the definition of (translator) style does not mean that we do not use contrastive techniques in the exploration of translator style. On the contrary, Winters and Kenny (2024) make heavy use of (and defend) one such technique, namely keywords analysis.

12 Note here that a further difference between Winters and Kenny (2024) and Kenny and Winters (2020) is that the latter involved a three-way comparison (in the first instance) between a human translation, a machine translation and a post-edited machine translation. Winters and Kenny (2024), on the other hand, involves (in the first instance) a two-way comparison between a machine translation and a post-edited machine translation. Further comparisons with the larger translator-specific corpus and reference corpus of original German literary texts are then carried out.

13 As already indicated, by using a contemporaneous corpus to get a snapshot of Oeser's style at a particular point in time, we implicitly acknowledge that his style (or any translator's style) might change over time. Such change over time would be best captured using a diachronic corpus design. The 'synchronicity' of *Oeser 12* with his post-edited work could, however, be questioned given the longer timeframe over which the source texts behind *Oeser 12* were written: while the majority of those source texts were published between 2010 and 2020, some were originally published earlier, for example in the 1940s, 1950s and 1970s. We also acknowledge that both the time of publication of the source text and the time period of the fictional setting

might influence style in translation, motivating, for example, the use of elevated lexis. In the case of the word we focus on later in this chapter, however, we note that while there is much variation in the frequency of '*weshalb*' across the texts in *Oeser 12*, we have not found any evidence of a relation between its frequency and the publication date of the originals.

14 Keywords are computed using WordSmith Tools version 8.0 (Scott 2022) and using the raw MT output as the reference text. They are ranked using log likelihood scores in conjunction with Bayesian Information Criterion (BIC) scores, which measure effect size. Full details of the keyword analysis are given in Winters and Kenny (2024).

15 The log likelihood score for '*weshalb*' in this text is 14.33 (BIC = 26.58).

16 For ease of analysis the MT raw output and post-edited versions were aligned and searched using Tetrapla, developed by David Woolls (2008–22).

References

Alvstad, C., A. K. Greenall, H. Jansen, and K. Taivalkoski-Shilov (eds) (2017), *Textual and Contextual Voices of Translation*, Amsterdam and Philadelphia: John Benjamins.

Baker, M. (1993), 'Corpus Linguistics and Translation Studies', in M. Baker, G. Francis, and E. Tognini-Bonelli (eds), *Text and Technology*, 233–50, Amsterdam: Benjamins.

Baker, M. (2000), 'Towards a Methodology for Investigating the Style of a Literary Translator', *Target*, 12 (2): 241–66. https://doi.org/10.1075/target.12.2.04bak

Baker, P. (2004), 'Querying Keywords: Questions of Difference, Frequency and Sense in Keywords Analysis', *Journal of English Linguistics*, 32 (4): 346–59. https://doi.org/10.1177/0075424204269894

Bentivogli, L., A. Bisazza, M. Cettolo, and M. Federico (2016), 'Neural Versus Phrase-Based Machine Translation Quality: A Case Study', in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 257–67, Austin: Association for Computational Linguistics. Available online: https://aclanthology.org/D16-1025.pdf

Castilho, S. and N. Resende (2022), 'Post-Editese in Literary Translations', *Information* 13 (66). https://doi.org/10.3390/info13020066

Čulo, O. and J. Nitzke (2016), 'Patterns of Terminological Variation in Post-editing and of Cognate Use in Machine Translation in Contrast to Human Translation', *Baltic Journal of Modern Computing*, 4 (2): 106–14.

Farrell, M. (2018), 'Machine Translation Markers in Post-Edited Machine Translation Output', in *Proceedings of the 40th Conference Translating and the Computer*, 50–9, London, 15–16, November. Available online: https://www.transcreate.it/tempfolder/wp-content/uploads/2018/11/TC406.pdf

Fitzgerald, F. Scott (1922), *The Beautiful and Damned*, New York: Scribner's.

Fitzgerald, F. Scott (1998), *Die Schönen und Verdammten*, trans. H. -C. Oeser, Zürich: Diogenes.

Fonteyne, M., A. Tezcan, and L. Macken (2020), 'Literary Machine Translation under the Magnifying Glass: Assessing the Quality of an NMT-Translated Detective Novel on Document Level', in *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, 3790–8. Available online: https://aclanthology.org/2020.lrec-1.468.pdf

Guerberof-Arenas, A. and A. Toral (2020), 'The Impact of Post-editing and Machine Translation on Creativity and Reading Experience', *Translation Spaces*, 9 (2): 255–82. https://doi.org/10.1075/ts.20035.gue

Guerberof-Arenas, A. and A. Toral (2022), 'Creativity in Translation. Machine Translation as a Constraint for Literary Texts', *Translation Spaces*, 11 (2): 184–212. https://doi.org/10.1075/ts.21025.gue

Hansen, D. (2021), 'Les lettres et la machine: un état de l'art en traduction littéraire automatique', in *Actes de la 28ème Conférence sur le Traitement Automatique des Langues Naturelles. Volume 2: 23e REncontres jeunes Chercheurs en Informatique pour le TAL (RECITAL)*, 28–45, Lille: ATALA. Available online: https://aclanthology.org/2021.jeptalnrecital-recital.3.pdf

Hansen, D, E. Esperança-Rodier, H. Blanchon, and V. Bada (2022), 'La traduction littéraire automatique: Adapter la machine à la traduction humaine individualisée', *Journal of Data Mining and Digital Humanities*. https://doi.org/10.46298/jdmdh.9114

Herrmann, J. B., K. van Dalen-Oskam, and C. Schöch (2015), 'Revisiting Style, a Key Concept in Literary Studies', *Journal of Literary Theory*, 9 (1): 25–52. https://doi.org/10.1515/jlt-2015-0003

Hovy, D., F. Bianchi, and T. Fornaciari (2020), '"You Sound Just Like Your Father". Commercial Machine Translation Systems Include Stylistic Biases', in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 1686–90, Association for Computational Linguistics. Available online: https://aclanthology.org/2020.acl-main.154.pdf

Isherwood, Christopher (1954), *The World in the Evening*, London: Methuen.

Isherwood, Christopher (2019), *Die Welt am Abend*, trans. H. -C. Oeser, Hamburg: Hoffmann und Campe.

Jockers, M. L., F. Nascimento, and G. H. Taylor (2020), 'Judging Style: The Case of Bush Versus Gore', *Digital Scholarship in the Humanities* 35 (2): 319–27.

Kenny, D. (2001), *Lexis and Creativity in Translation. A Corpus-based Study*, Manchester: St. Jerome.

Kenny, D. (2018), 'Machine Translation', in P. Rawling and P. Wilson (eds), *The Routledge Handbook of Translation and Philosophy*, 428–45, London and New York: Routledge.

Kenny, D. (2020), 'Machine Translation', in M. Baker and G. Saldanha (eds), *The Routledge Encyclopedia of Translation Studies*, 3rd edn, 305–10, London: Routledge.

Kenny, D. (2022), 'Human and Machine Translation', in D. Kenny (ed.), *Machine Translation for Everyone: Empowering Users in the Age of Artificial Intelligence*, 23–49, Berlin: Language Science Press.

Kenny, D. and M. Winters (2020), 'Machine Translation, Ethics and Literary Translator's Voice', *Fair MT: Towards Ethical, Sustainable Machine Translation, Special Issue of Translation Spaces*, 9 (1): 123–49. https://doi.org/10.1075/ts.00024.ken

Kenny, D. and M. Winters (2022), 'A Text of One's Own: Post-editing and the Affirmation of Style', *Translation in Transition 6*, Prague: Charles University, Czech Republic, 22–23 September.

Koehn, P. (2020), *Neural Machine Translation*, Cambridge: Cambridge University Press.

Korotkova, E., M. Del, and M. Fishel (2018), 'Monolingual and Cross-lingual Zero-shot Style Transfer', *arXiv Preprint*, arXiv:1808.00179.

Kuzman, T., Š. Vintar, and M. Arčan (2019), 'Neural Machine Translation of Literary Texts from English to Slovene', in *Proceedings of the Qualities of Literary Machine Translation*, 1–9, Dublin, Ireland: European Association for Machine Translation. Available online: https://aclanthology.org/W19-7301.pdf

Lee, C. (2021), 'How Do Machine Translators Measure up to Human Literary Translators in Stylometric Tests?', *Digital Scholarship in the Humanities*. https://doi.org/10.1093/llc/fqab091

Lennon, B. (2014), 'Machine Translation: A Tale of Two Cultures', in S. Bermann and C. Porter (eds), *A Companion to Translation Studies*, 136–46, Oxford: Wiley.

Matusov. E. (2019), 'The Challenges of Using Neural Machine Translation for Literature', in J. Hadley, M. Popović, H. Afli, and A. Way (eds), *Proceedings of the Qualities of Literary Machine Translation*, 10–19, Dublin. Available online: https://www.aclweb.org/anthology/W19-7302.pdf.

Michel, P. and G. Neubig (2018), 'Extreme Adaptation for Personalized Neural Machine Translation', in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 312–18, Melbourne, Australia: Association for Computational Linguistics. Available online: https://aclanthology.org/P18-2050.pdf

Moorkens, J., A. Toral, S. Castilho, and A. Way (2018), 'Translators' Perceptions of Literary Post-editing Using Statistical and Neural Machine Translation', *Translation Spaces*, 7 (2): 240–62. https://doi.org/10.1075/ts.18014.moo

O'Brien, S. (2022), 'How to Deal with Errors in Machine Translation: Postediting', in D. Kenny (ed.), *Machine Translation for Everyone: Empowering Users in the Age of Artificial Intelligence*, 105–20, Berlin: Language Science Press.

O'Brien, S. and M. Simard (2014), 'Introduction to Special Issue on Post-editing', *Machine Translation*, 28: 159–64. https://doi.org/10.1007/s10590-014-9166-8

Oeser, H.-C. (2020), 'Duel with DeepL. Literary Translator Hans-Christian Oeser on Machine Translation and the Translator's Voice', *Counterpoint*, 4: 19–23. Available online: https://www.ceatl.eu/wp-content/uploads/2020/12/Counterpoint_2020_04_article_05.pdf

Ramírez-Sánchez, G. (2022), 'Custom Machine Translation', in D. Kenny (ed.), *Machine Translation for Everyone: Empowering Users in the Age of Artificial Intelligence*, 165–86, Berlin: Language Science Press.

Saldanha, G. (2005), *Style of Translation: An Exploration of Stylistic Patterns in the Translations of Margaret Jull Costa and Peter Bush*, PhD Thesis, Dublin City University.

Saldanha, G. (2011), 'Translator Style: Methodological Considerations', *The Translator*, 17 (1): 25–50. https://doi.org/10.1080/13556509.2011.10799478

Scott, M. (2022), 'How Key Words are Calculated', *WordSmith Tools Help*. Available online: https://lexically.net/downloads/version_64_8/HTML/keywords_calculate_info.html (accessed 31 March 2022).

Taivalkoski-Shilov, K. (2019a), 'Free Indirect Discourse: An Insurmountable Challenge for Literary MT Systems?', in J. Hadley, M. Popović, H. Afli, and A. Way (eds), *Proceedings of the Qualities of Literary Machine Translation*, 35–9, Dublin. Available online: https://aclanthology.org/W19-7305.pdf

Taylor, C. (2018), 'Similarity', in C. Taylor and A. Marchi (eds), *Corpus Approaches to Discourse: A Critical Review*, 19–37, London: Routledge.

Toral, A. (2019), 'Post-editese: An Exacerbated Translationese', in *Proceedings of Machine Translation Summit XVII: Research Track*, 273–81, Dublin, Ireland: European Association for Machine Translation. Available online: https://aclanthology.org/W19-6627.pdf

Toral, A. and A. Way (2015a), 'Translating Literary Text between Related Languages Using SMT', in *Proceedings of NAACL-HLT Fourth Workshop on Computational Linguistics for Literature*, 123–32, Denver, CO. Available online: https://aclanthology.org/W15-0714.pdf

Toral, A. and A. Way (2015b), 'Machine-assisted Translation of Literary Text: A Case Study', *Translation Spaces*, 4 (2): 240–67. https://doi.org/10.1075/ts.4.2.04tor

Toral, A. and A. Way (2018), 'What Level of Quality can Neural Machine Translation Attain on Literary Text?', in J. Moorkens, S. Castilho, F. Gaspari and S. Doherty (eds), *Translation Quality Assessment: From Principles to Practice*, 263–87, Heidelberg: Springer International Publishing.

Van de Cruys, T. (2020), 'Articuler ou trahir. La poésie des machines à traduire', *La traduction littéraire et SHS à la rencontre des technologies de la traduction: enjeux, pratiques et perspectives, Translating Europe Workshop*. Université Toulouse Jean Jaurès, 7 February. Available online: http://timvandecruys.be/media/presentations/pres_tradlit_070220.pdf

Vanmassenhove, E., D. Shterionov and M. Gwilliam (2021), 'Machine Translationese: Effects of Algorithmic Bias on Linguistic Complexity in Machine Translation', in *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 2203–13, Association for Computational Linguistics. Available online: https://aclanthology.org/2021.eacl-main.188.pdf

Voigt, R. and D. Jurafsky (2012), 'Towards a Literary Machine Translation: The Role of Referential Cohesion', in *Proceedings of the NAACL-HLT 2012 Workshop on*

Computational Linguistics for Literature, 18–25, Montréal, Canada: Association for Computational Linguistics. Available online: https://aclanthology.org/W12-2503.pdf

Wang, Y., C. Hoang, and M. Federico (2021), 'Towards Modeling the Style of Translators in Neural Machine Translation', in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1193–9, Association for Computational Linguistics. Available online: https://aclanthology.org/2021.naacl-main.94.pdf

Winters, M. (2007), 'F. Scott Fitzgerald's *Die Schönen und Verdammten*: A Corpus-based Study of Speech-act Report Verbs as a Feature of Translators' Style', *Meta: Journal des traducteurs/Translators' Journal*, 52 (3): 412–25. https://doi.org/10.7202/016728ar

Winters, M. (2009), 'Modal Particles Explained. How Modal Particles Creep into Translations and Reveal Translators' Styles', *Target: International Journal of Translation Studies*, 21 (1): 74–97. https://doi.org/10.1075/target.21.1.04win

Winters, M. (2013) 'German Modal Particles – From Lice in the Fur of Our Language to Manifestations of Translators' Styles', *Perspectives: Studies in Translatology*, 21 (3): 427–45. https://doi.org/10.1080/0907676X.2012.711842

Winters, M. (2015), 'Hans-Christian Oeser: Translation Strategies of a Literary Translator', in S. Egger (ed.), *Cultural/Literary Translators: Selected Irish-German Biographies II (Irish-German Studies, Vol. 9)*, 210–24, Trier: Wissenschaftlicher Verlag Trier.

Winters, M. and D. Kenny (2024), 'Mark my Keywords: A Translator-specific Exploration of Style in Literary Machine Translation', in A. Rothwell, A. Way, and R. Youdale (eds), *Computer-assisted Literary Translation: The State of the Art*, 69–87, London: Routledge.

Woolls, D. (2008–2022), *Tetrapla*, CFL Software Limited.

4

The Figure of the Literary Translator amidst New Technologies

Damien Hansen

The renewed (im)possibility of literary translation

Recent decades have witnessed the emergence of an ever-growing range of tools that aim to boost productivity and computerize a multitude of daily tasks, linguistic and otherwise, involved in the translation profession. Some of these have become so prevalent in a translation professional's day-to-day life that it is easy to take them for granted: grammar and spell checkers, browsers and search engines, email and invoicing software are just some of the many tools at the fingertips of translators today (Froeliger 2013: 10). While these have been widely and, for the most part, uncontroversially adopted, other innovations have proven to be revolutionary for the industry. Notable among these are computer-assisted translation (CAT) tools and, more recently, machine translation (MT) systems, which, despite their growing prominence, have traditionally been regarded as entirely inappropriate in the literary field (Youdale and Rothwell 2022; Guerberof-Arenas and Toral 2020). With an increasing number of events and publications exploring the convergence of new technologies and creative texts, however, the time seems ripe for a more nuanced view of the question and a closer inspection of where translators might benefit most, if indeed at all, from a greater overlap. In this context, the aim of this chapter is to present an ongoing research project that set out to examine the possibility of computer-assisted literary translation (CALT) and later included the subject of literary machine translation (LMT), resulting in a general study of new technologies in literary translation.

To this end, this first section will attempt to contextualize the discussion around the use of translation technologies in the literary field, both from a

historical point of view and from the perspective of translation studies. Following this introduction, the focus will turn to the practical aspects of devising a customized literary corpus, before illustrating its possible uses in computer-assisted and machine translation scenarios. In doing so, I will try to answer two key questions: Are translation memories of any use when translating novels; and can they also be used to train a personalized machine translation engine that would reflect individual translator style? Finally, some of the shortcomings of the current technology will be addressed, with the aim of anticipating potential developments in these programmes, the future direction of CALT studies and the foreseeable societal challenges for the profession. This holistic approach to the issue of technology and creative translation will ultimately show that there are fundamental similarities between the tools discussed in this particular context, the principles they rely on, and the conclusions that can be drawn for each piece of technology.

In fact, the literary field as a whole demonstrates recurring patterns in which literary translations have routinely sparked objections. Well before the consolidation of translation studies as a discipline, theorists have argued that the very nature of literary texts renders them impossible to translate, as summarized in the age-old saying *traduttore, traditore* (translator, traitor). Although there are multiple historical, political and theoretical grounds for this reasoning (Mounin 2016: 13–26), the cornerstone of these arguments relies on the idea that the poetic dimension of an original can never be transposed into another language (Ladmiral 1994b: 85–98). Yet, the actual literary translation work carried out over hundreds of years of practice is suffice to reject the supposed impossibility of the endeavour (Mounin 2016), just as it illustrated the sharp divide at the time between theorists who defended the dogma of untranslatability and translators in the field (Ladmiral 1994b: 88).

Even before the appearance of CAT tools, computers and word processors were accused of introducing a 'computerized' or 'mechanical' style into translated literary works (ATLAS 1988: 124). Some went so far as to say that translations should only be delivered on paper, while, conversely, those who had tried the tools were delighted to be working with them (143). In time, translators 'tamed the computer thingy', rendering it some sort of a small pet that most of them could not do without anymore (125). Although some of these arguments might raise a smile today – many professionals would arguably feel hard-pressed to imagine working without word processors these days – it is interesting to find that there has been very little change in the objections formulated against CALT or machine translation, when compared to those originally raised against

human literary translation itself or earlier computer tools. Proof of this is the recurrent contrast between expressive or creative texts, on the one hand, and pragmatic or technical texts, on the other hand, even though this dichotomy has repeatedly been called into question (Durieux 2000; Newmark 1981: 5–6; Hatim and Mason 1997: 1–2). In essence, translation technologies appear to be instigating a renewal of arguments around the theoretical impossibility of literary translation. As with human translation and as will be demonstrated in the following sections, these theories must be tested by putting them into practice. Another key observation that will be developed throughout this chapter is that CALT is a research area that extends beyond the traditional CAT tool. In each case, however, the common denominator is the use of corpora.

Custom-made corpora

Whatever their aim, translation technologies such as CAT, MT or Web concordancers rely on large datasets. Whether we call them corpora, translation memories (TM), bitexts or something else entirely, these data troves have changed the practice of translators, becoming a fundamental part of their daily routine. Commonly used tools, including Google's search engine or DeepL's Linguee, also make use of corpora in their singular, albeit less obvious, way, and the abundance of online or offline data has even proven beneficial for translation scholars and teachers alike (Loock 2016). In all of these cases, however, the results can vary significantly depending on the quality of the data as translation technologies work best with carefully curated in-domain datasets. But reliable parallel corpora of literary texts are hard to come by, or simply non-existent for some languages. Fortunately, literary translators naturally accumulate a substantial amount of valuable resources – that is, their own translations – over years of work. All that is needed to take advantage of this data in the literary field are direct and efficient ways of building, updating, converting and querying these databases, ideally with as few time and technical constraints as possible.

A prime example of a database is the translation memory; these have been used in conjunction with CAT tools for decades now, but recent advances in machine translation over the last years have also crucially depended on these parallel corpora. In both technologies, the datasets consist of a collection of texts – an original and its translation(s) – that have been segmented into chunks, in the same way they would appear in a bilingual concordance, for instance. The essential feature of these datasets is that the source text has to be aligned with

its target text(s). Ideally, the aligned segment should be as small as possible – usually at sentence-level – but they can also be expanded substantially to ensure that the content is equivalent across languages, sometimes reaching the size of entire paragraphs.

As a way of assessing the advantages of reusing these personal corpora for the translation of literary texts, I therefore set out to build a translation memory for use in a computer-assisted translation experiment that would eventually come to include MT. The initial motivation for this study was not only to challenge the idea that translation technologies and literature are entirely incompatible but also to imagine scenarios in which translators could use their own data without having to share it with third parties so as to evaluate the benefits of customized tools and determine what aspects could be improved, while hopefully empowering literary translators through technology. The experiment relied on a specific case study, namely the *Septimus Heap* saga, an English-language fantasy series for young adults written by Angie Sage (2005–13) and translated into French by Nathalie Serval (2005–13), a recognized translator with whom I enjoy an ongoing collaboration.

The case study began with the creation of a translation memory; this was done from scratch, although such datasets are generally accumulated over time in real-world contexts. The first six volumes of the series were compiled and automatically aligned in a dedicated software programme, and the output of the alignment – the bitext, or TM – was then manually reviewed and corrected. This also entailed the analysis of translator style through the TM, although I will argue that the optimal situation would be that of a professional translator working with their own texts.

There were several underlying motives behind this corpus-building and alignment process. First, it made it possible to artificially reproduce a realistic scenario where a translation professional would have assembled a short TM over time while using their CAT tool, as mentioned earlier. A second reason was that the manual revisions resulted in a high-quality dataset that could be used without fear of alignment issues. Lastly, it offered a chance to reflect on the issues and benefits that arose during this process, and to evaluate the creation of a TM from scratch as a means of offering insights to translators who have never relied on translation memories but would like to replicate this scenario.

From the first versions of AntPConc (Anthony 2014) and Sketch Engine (Kilgarriff et al. 2004) onwards, many tools have become available to help create and explore custom-made corpora. While each of these programmes offers distinct functionalities, all potentially useful and worthy of attention, the aim of

this approach was to create a parallel corpus, in a standard translation memory format, which could then be used with a variety of software. Once again, multiple tools can help in this regard, many of which are integrated in most CAT tools, but LogiTerm (Bédard 2016) was specifically selected as the user has the ability to create and edit TM files through a relatively intuitive interface.[1] Indeed, LogiTerm is a tool that provides a vast array of terminology-related features, not least the alignment editor.

In addition, opting for a multivolume series was an opportune way of seeing whether a relatively restricted but coherent corpus would be a useful dataset. As well as being representative of the popular heroic fantasy genre and an international bestseller, the *Septimus Heap* saga was also unique in that the seventh tome of the series was never published in French. Despite this, the first six volumes contained enough data to enable approximately 45,000 segments, or translation units, to be retrieved, which in turn could be used to inform my own translation of the last novel of the series.

Overall, the raw output from the automated alignment proved to be highly satisfactory, and the quality was all the more impressive given that the textual and syntactic structure of the books can differ considerably from one language to the other. Although the French translation often combined two or three source-text sentences into one, most of these instances were correctly merged by the programme to form a single valid translation unit. In extreme cases where multiple paragraphs were entirely missing from or added into the target version, the software was able to resolve the alignment mismatch within only four or five segments, instead of affecting and misaligning the thousands of consecutive segments. Fortunately, this was not a regular occurrence and the features offered by the tool made it easy to correct the occasional mistake.

Indeed, the majority of the problems encountered involved words or characters rather than the alignment itself. LogiTerm, for example, inverted some characters (e.g. 'sniff' was read by the software as 'snfif'), transformed some words into abbreviations (e.g. 'edge' became 'e.g.') and tended to merge words on either side of a single capital letter (e.g. 'WellIlike them').[2] A large number of errors also derived from the original files themselves (spelling mistakes, missing punctuation etc.) or were due to formatting (unusual breaks, use of images etc.). These errors would be problematic for TM matches in a CAT tool, but less so with the concordance search feature that will be discussed in the next section. Furthermore, some programmes enable users to see the textual context surrounding a translation unit, a feature that helps to alleviate most alignment issues, as well as the occasional case where a word or idea in the source sentence

is rendered not into the corresponding target sentence but into a neighbouring segment and would need to be merged manually otherwise.

Even then, the quality of the resulting TM indicated that it could have been used as a personal fit-for-purpose corpus and that this approach could be reproduced with either light or no editing at all, depending on the intended use. Literary translators could therefore create a relatively reliable corpus from scratch – in a matter of seconds or, if quality is an issue, within the short period of time needed to refine the output – and still derive many benefits from it, as will be illustrated later. Likewise, Loock (2016) specifically challenges the widely held view that becoming familiar with such tools and creating a TM is not only too time-consuming but also pointless, offering many counterexamples and suggestions for how to take advantage of DIY and even disposable corpora.[3]

It is important to note, however, that all these steps are unnecessary if a translator chooses to work directly within the CAT interface from the outset. While my aim for this exploratory work was to create a specific TM *ex nihilo*, it is worth pointing out that one of the main advantages of CAT tools is that they allow users to automatically build up and expand their own translation memories over time, without having to dedicate any additional time and energy to the process. They can subsequently explore, export and exploit the resulting TMs as they wish. Of course, as with any other domain, it can take time before these resources become noticeably useful.

Computer-assisted literary translation

With a newly created TM in hand, the translator is then ready to import the file into their tool of choice. And while CALT might cover a very large spectrum of tools and software, including machine translation, corpus-based analysis and data visualization, there is often a tendency to reduce this approach to CAT tools in their narrowest conception, that is, as translation memory systems. The first question that needs to be addressed, then, is how useful might this new translation memory be if it were used for the translation of a novel in a computer-assisted translation programme?

Indeed, this question has formed the crux of existing debates around the value of TMs in a literary context, fuelling controversy over the very notion of CALT by extension. The reasons for this are numerous, and the arguments that are put forward by its opponents can vary greatly, but comments on a single thread on Proz, a public translation forum, responding to the question, 'Do

you use Trados for literary texts, too?' (Original poster 2015) offer unsolicited, genuine and representative insights into the different responses to its feasibility and utility. In fact, many of these are reminiscent of the theoretical opposition between scientific and literary translation, as the following anonymized responses illustrate:

> I love to work with my CAT Tool which is memoQ. But for literary translation I found it totally useless, it is a different, sofisticated [*sic*] process where you need to have much more possibilities, choose and create the right one. It is art. (Contributor 1 2015)

> I've been translating books (novels, comics, essays and romances) for so many years and never used a CAT tool. It's a creative job. Of course I do translate even technical texts, but literature is rather different. (Contributor 2 2015)

Another observable pattern is the resurgence of Mounin's (2016) *objection préjudicielle* (prejudicial objection), as coined by Ladmiral (1994b), namely the presumption that literary translation – whether human or computer assisted – is impossible, even though such claims may not be founded on first-hand experience. As one user remarks:

> Funny how it seems that most of those who are saying a firm 'no' are probably those who have never used any CAT tools or never bothered to learn how to use them properly. I should know: not too long ago, I was one of them, and I am now the first person to admit that. (Contributor 3 2015)

To better capture the range of arguments commonly voiced against CALT, I devised a simple typology.[4] First, we can find practical objections relating to the idea that 'CAT tools are too tedious to use'. These generally revolve around the belief that such tools do not provide any added value, are overly complicated or, for example, that they might restrict the translator's freedom. In the second category, we can find more theoretical objections resting on the assumption that 'CAT tools offer no benefits', not due to the tool as such but because the nature of these texts makes them incompatible with the technology. These arguments usually reflect the view that literature, by definition, is not repetitive enough or is simply too sophisticated for computer software to be of any use.

To assess the first set of arguments, it is worth challenging two commonly held ideas from the start: first, that CAT tools serve exclusively as translation memory systems; and secondly, that they serve only to constrain the translation process. Indeed, the criticism levelled at CALT often involves the misconception that if TMs cannot be useful, the same is true for the programme as a whole.

Nevertheless, from a purely practical point of view, CAT tools offer a convenient and seamless interface which is better suited to translation than simple word processors, for instance, regardless of the field of expertise concerned. In addition to translation memories, other useful features include the bilingual display, termbases and access to various online resources and plugins. There are also passive advantages such as backups or having each version of the translation archived in the same project. Machine translation is also making more and more of an appearance in CAT environments, and having all of these functionalities within the same interface removes the hassle of switching constantly between windows, in addition to possibly freeing a monitor if the translator uses more than one.

For these reasons alone, an increasing number of people are turning to CALT, simply because they grew accustomed to working within a CAT interface. Moreover, rather than seeing it as a constraint, they have come to rely on the features that they feel are useful to them:

> There is nothing in these tools that prevents you from varying the sequence of words, sentences and paragraphs. There is nothing that imposes consistency or lack of it, curtails your production speed, lowers the writing quality, sterilises your work, or imposes a structure. What matters is how you use the tool. (Contributor 4 2015)

> Ever since I started using a CAT tool (MemoQ) for the first time a couple of years ago, I have used it whenever possible, even when dealing with creative texts (which is what I mostly do anyway). Personally, I find that having the text segmented into sentences helps, if for no other reasons than helping avoid accidental omissions. I also find it beneficial to have the source and target side by side in case I have to return to any completed segment again. When you are just overtyping in Word, you have to pull up the original file for that, and either toggle between them or divide the screen in two, which is an extra hassle. So no, it is not absolutely useless, at lest [*sic*] for me it isn't. Go figure, I used to be a convinced skeptic here before I tried that thing myself. (Contributor 3 2015)

This is consistent with other accounts of CALT, where users explain that they turn to it by force of habit and because they find the interface useful whatever the field in which they work. Mazoyer (2020), for instance, reports how the systematic use of CAT tools has become a natural and obvious reflex that has extended to the translation of essays in the human sciences. The author thus explains that, even without relying on any TM beyond the translation itself, he has found that it accelerated the translation process and made the revision

phase easier, all the while helping him ensure greater consistency and harmony. Rothwell (2020b) reaches a similar conclusion in a retranslation experiment, in which he observes that the CAT tool does help in some respects, such as revising, checking for additions and omissions, or avoiding data loss. In the specific case of a retranslation, the author also notes that it is particularly useful to have the previous translation in the same interface and that the process can reveal valuable information on translation strategies by previous translators, across other languages and in other time periods. On the whole, these accounts, together with the following forum post, indicate that the aforementioned practical objections against CALT do not necessarily hold true and that they are not sufficient to invalidate this approach:

> I have used it for fiction and for literary essays. In the literary essays it was very useful because it allowed me to maintain consistency in the translations I gave of the books being considered. In fiction, that consistency can be useful, too – sometimes characters' names need to be translated, and you can chuck them all into your termbase. But mostly I use it because I'm used to it. It's a part of my working process now. I find the segmenting into sentences helpful, because it makes me focus and just churn through the text, leaving the fine tuning till later. Of course, with literature the editing process outside Trados is longer and more intensive. But it's fine to use it if you like it. In the middle of the book of literary essays I did, I suddenly got fed up with using Trados – for some reason it seemed to be interfering with the flow of paragraphs. So I stopped using it for a few chapters, then later I went back to it. There's no need to be dogmatic either way! (Contributor 5 2015)

One element that has not been mentioned so far, however, is that becoming comfortable with such programmes does require considerable learning effort (O'Brien et al. 2017: 155). It implies differing work processes that change translators' relationship to the text and require a different skill set to navigate the technical complexity of these new tools, which explains why CALT appears to have attracted some users who were already familiar with them, and why new generations of translators hold more positive views on the question, as computer-assisted translation has become a formal component of their training. As the testimonial suggests, there is no need to feel any sense of dogmatic pressure: CALT is not compulsory, nor ought it be imposed and used consistently. It should only improve the translators' comfort. This is the double and sometimes paradoxical objective of translation technologies: improve both the user's productivity and their working conditions (Lavault-Olléon 2011: 6). While some aspects of CAT tools could be improved, as will be explored later, these are

not necessarily specific to the literary field. Of course, using these tools involves a significant change in the way translators work, but it also allows them, in time and with practice, to tweak the programme in a way that is better adapted to their own needs and workflow. Considering CALT as a tool, as simple as word processors, is a step towards dissolving the supposedly strict boundary between scientific texts on the one hand, deemed to be ideally suited to computer-assisted translation, and literary texts on the other hand. The lines blur even further if we acknowledge that both fields share many challenges, such as the terminology searches typically associated with scientific and technical texts or segments where form is more important than content.

This leaves us with the second type of objections levelled against CALT. These theoretical arguments are not unlike those that rely on the dichotomy between scientific and literary translation. A commonly found criticism, for instance, is that translation memories offer absolutely no benefit due to the very nature of literary texts. Multiple arguments are presented in a similar light, but in variously vague terms. However, the answer to these assumptions might be summarized in a singular question: Are CAT tools of any use when it comes to issues of style, form and creativity? In order to try to provide insight into this question, I will describe and reflect on my own experiences of using the TM I tailor-made from the *Septimus Heap* series to carry out my translation of the seventh novel within Trados Studio 2015 (SDL 2015). This following discussion therefore comes with a certain level of subjectivity, which is nevertheless necessary to bridge the gap between theory and practice, and to address criticism voiced against CALT.

CALT in practice

To begin, and despite the argument that literature is not repetitive enough to derive any benefit from translation memories, it was surprising to find a number, albeit limited, of automatic translations from the TM. It might be argued at this point that the fantasy genre facilitates these matches – this assumption will be examined more explicitly in the section on LMT – due in particular to its short dialogue segments. Yet, it should be noted that these are penalized more heavily: TM systems such as Trados retrieve segments by comparing words in the source and TM sentences and assigning a 'match score' based on the number of corresponding units, meaning that a single change in a segment comprising two words will be marked as 50 per cent dissimilar and will consequently not be shown by the software. Furthermore, these matches

in the TM did not only involve segments on the shorter side but also longer descriptive ones.

The 'pre-translation' and 'auto-propagation' features,[5] however, proved much less useful than the 'context' or 'concordance search' that allows users to quickly look up segments of varying length, as they would in a traditional concordancer.[6] In this respect, the translation memory represents a truly invaluable resource that offers particularly relevant results. Moreover, the fact that these texts are less repetitive actually made the results especially useful, as they offered additional reference material and translation suggestions to explore.

Without delving into a complex definition of stylistics, it can safely be said that style is inherently connected to the choices made at various levels of the text (see Taber 1972: 61; Baker 2000: 260–1; Boase-Beier 2014: 393–4; Youdale 2020: 2–5). That being so, the greatest advantage offered by TMs in the specific context of a literary translation is that they inherently reflect the author and translator's choices and style, and that consulting them supports and reinforces the decision-making process. Concordance searches allow the user to quickly navigate through the corpus and to take decisions based on observations of, for example, the name of a place, the translation of a term relating to the fictional world of the series, or how a specific linguistic issue had been previously resolved. Subsequently, it also becomes much easier to maintain consistency, either within one particular novel or across a series of novels, in terms of style and terminology. Indeed, picking up the translation from another translator made each of these decisions all the more noticeable, and the TMs proved to be especially useful in ensuring that my choices were in line with those of the six preceding novels, which, in this case, was one of the translation goals.

Among the other advantages of using TMs and concordance searches in literary translation is that they make it particularly convenient for the user to execute fast lookups for simple issues: searching for the translation of common terms, finding synonyms or merely recovering a translation solution enacted a few thousand words earlier. Humans are not machines that always work at full capacity, but this is precisely where CAT tools can offer their technical and linguistic support. They provide valuable and welcome assistance on long-term projects such as literary translation, without having to leave the software interface and consistently switch between various windows. Arguments to the contrary might highlight the risks of creating a more repetitive text, but the opposite is also true, as concordance searches can also afford a quick insight into previous solutions that the translator can then consciously avoid during the selection process. This strategy proved especially useful given that French is less tolerant of repetitions than English.

Thus, TM lookups make it easy to find alternative solutions, and ultimately help translators to reflect on their own production and style.

In turn, Loock (2016) and Youdale (2020) both demonstrate how the use of translation technologies and corpus exploration/visualization tools can be beneficial, not only for researchers 'as a form of computer-assisted translation studies', but also for professional translators 'as computer-assisted translation tools' in the broadest sense. In the specific case of the *Septimus Heap* novels, for example, they helped to identify variations in the translation that diverged from the source. Those variations could be as simple as a character having a different hair colour; or more complex, when ascertaining which characters used regional or old-fashioned variant of French; or even subtler, such as discerning that the interactions between two specific characters were systematically rendered in a higher register. Once again, the tools proved particularly advantageous when following in the footsteps of another translator, although this experiment showed that the same holds true when using personal translations.

It can generally be concluded that translation memories have the potential to be as effective in the literary field as in any other. This is nonetheless dependent on the TMs containing sufficient and relevant data, and multivolume series lend themselves particularly well to this scenario. As it transpires, termbases and terminology searches can be equally useful to literary and scientific translators as a way of increasing consistency, just as TMs can serve to refine style, regardless of the type of text, notably by entering queries as would be the case with a reference corpus. Despite these similarities, a difference perhaps still persists in that literary texts rely more on manual research using the concordance tool. The ability to compare and draw inspiration from various options was indeed much more beneficial than a single and precise perfect match. This is nevertheless in keeping with recent studies indicating that the use of this concordance feature seems to be underestimated and is much more significant than the typical discourse or previous research on CAT suggests (Bundgaard and Christensen 2019). From this perspective, relying on a custom, trustworthy corpus of personal translations can offer great advantages. Translators could similarly engage in their own exploration by compiling a reference corpus of their favourite authors and translators, while remaining alert to potential infringement issues.

On the whole, my own experience revealed that CAT tools ensure greater quality and creativity in a number of ways. Far from being useless, TMs help maintain coherence and consistency throughout the translation. By providing suggestions that the user can either adopt or steer away from, they serve as a semantic, lexical and stylistic foundation on which to build a new translation.

These numerous and diverse solutions shown in context also greatly improve the decision-making process and allow for a stylistically better-informed translation. At the same time, CALT facilitates the analysis of the source and previous translations by allowing a 'stylistically-aware reading' of these texts (Boase-Beier 2014: 401). In general terms, the conclusion from my experience with CAT tools is that they also speed up the process as a whole and give the translator more time to focus on what is important and where the human expertise is more crucially needed. This process has been shown to increase creativity in other fields, such as those of the European institutions, where it would normally not be considered as important (Strandvik 2001: 3).

These findings echo a growing number of researchers and translators who have, likewise, experimented with CALT. Among these, Rothwell (2020b) illustrates how the tools can reveal unseen problems or ambiguities in the source text, confirm or modify local points of interpretation, suggest unforeseen translation solutions, and provide access to other inscriptions of a particular text. Regarding CAT as a corpus exploration and textual analysis tool thus has the potential to improve the quality of the output, because it can bring out stylistic patterns which might otherwise be invisible to the naked eye. Similarly, corpus-based studies have long shown that taking advantage of corpora can facilitate the translation and revision process, but also the understanding of the source text, information retrieval and terminology research (Loock 2016: 175–6). Youdale (2020) illustrates this through his 'close and distant reading' approach to literary translation. The author perfectly summarizes this method by describing it as 'a way in which translators can interact with technology as part of the translation process – a way which neither dilutes nor deskills the art of translation, but actually enhances it by revealing information about a text which even close reading is unlikely either to measure accurately or to detect at all' (1). This ties in with this chapter's view of CALT as a mere set of tools that have the potential to support human translation, depending on how they are used, by offering additional information about a text and guiding translators' decisions. Even so, while there is an apparent shift towards a more positive attitude on the use of CAT tools in the literary field, the same cannot be said of the emergent interest in another key topic: literary machine translation.

Literary machine translation

In recent years, machine translation has become another increasingly common feature within computer-assisted translation programmes, further blurring

the boundaries between CAT, MT and the broader conception of CALT. And although the idea of automated creative writing can be traced back to the early computers, with Braffort's 1975 computerized version of Queneau's *Cent mille milliards de poèmes* (A Hundred Thousand Billion Poems) for instance, LMT is a more recent issue in comparison to CAT tools.[7] The subject seems to have received special attention since 2019 and the advances brought about by neural machine translation (NMT),[8] yet remains an early research avenue. Furthermore, if LMT appears to be facing the same theories of untranslatability that have been used against human and computer-assisted literary translation, this particular scenario is all the more complex. Machine translation as a whole is characterized by an unprecedented presence not just in translation circles but also in the general media, and by a polarizing and often exaggerated discourse on both sides of the debate on neural MT. It is therefore important that more nuanced arguments are brought to the debate, as well as concrete and substantiated views of what NMT can or cannot do, and what its performance in the literary domain would be if it was specifically trained on literary data.

Despite improvements in the performance of NMT, and to Transformer models in particular (Vaswani et al. 2017), it is unfair to judge the feasibility of LMT solely on publicly available systems, as they were not trained with this particular use in mind. Indeed, domain robustness still remains a key challenge for machine translation (Müller, Rios and Sennrich 2020: 162), which underlines the importance of training these systems on data taken from the same domain in which they will be used. Through extensive research, Toral and colleagues have already shown that it is possible to get good results with systems trained exclusively on literary data (Toral, Oliver and Ribas-Bellestín 2020). Nevertheless, one of the main obstacles to training such systems is that it remains difficult to find good quality literary data that is publicly available. Accordingly, the chapter will now turn its attention to exploring the feasibility of a scenario whereby translators can train their own systems, as is made possible by the OPUS-CAT project for instance (Nieminen 2021). In similar low-resource settings, researchers such as Matusov (2019) have tried to 'fine-tune' generic MT systems on a smaller literary corpus. This approach is in keeping with the general principle that NMT needs very large training data to expand its vocabulary and improve the syntax, as well as specialized in-domain data, so that it can reproduce terminology and style. Nonetheless, these attempts were met with mixed results, except for one of the settings presented in the work of Kuzman, Vintar and Arčan (2019), who showed that training on another novel by the same author and translator yielded better results. So, in addition to providing the first new experiment on LMT on the

English–French pair since Besacier's (2014) work on the automated translation and post-editing of a fiction short story, the research detailed hereafter also aims to assess whether the translation memory I created and discussed earlier could also be used to train a system on one literary series, with one author and one translator, and whether the MT system would be able to learn and reproduce the stylistic patterns displayed by the French translator.[9]

To this end, I first trained a generic system with the open-source OpenNMT toolkit (Klein et al. 2017), using the following corpora from the OPUS repository: Books, Europarl, GlobalVoices, News and TED (Tiedemann 2012). To these, I also added a personal corpus of video game translations (Hansen and Houlmont 2022) to increase the diversity in an otherwise relatively homogeneous set. The resulting model, trained on approximately four million sentences, was then fine-tuned on the six novels (approximately forty-five thousand sentences), with the default parameters of the base Transformer architecture at both stages.[10] This domain adaptation process showed promising results, as the BLEU score (Papineni 2002) of the generic system increased from 9.93 to 18.56. Table 4.1 provides three metrics[11] used to evaluate the resemblance of this output with regard to the reference produced by the original translator, as well as the scores for DeepL and Google Translate. These were calculated with sacreBLEU (Post 2018).[12] For the first two metrics, improved performance is marked by a higher score, whereas it should be ideally as low as possible for TER.

The results demonstrate that even though this customized system is trained on very limited in- and out-of-domain data – a training corpus of six million sentences is considered 'frugal' by today's standards (Blin 2021) – relying on training data from the same author and translator proved highly beneficial (+9 BLEU compared to the generic system). This is especially true if this improvement is compared to the experiments carried out by Kuzman, Vintar and Arčan (2019) and Matusov (2019), where the score increased between 1 and 3 BLEU points or decreased in some cases. These scores are lower in comparison to the performances obtained by Toral, Oliver and Ribas-Bellestín (2020), however, who have suggested in previous work a minimum reference point of 20 BLEU

Table 4.1 Automatic evaluation of the systems on a fantasy novel

	BLEU ⇧	**chrF2++ ⇧**	**TER ⇩**
Google Translate	10.79	35.20	91.08
DeepL	10.04	34.88	92.81
Customized system	**18.56**	**40.43**	**76.06**

Source: Adapted from Hansen and Esperança-Rodier (2023).

to mark useful post-editing quality (Toral and Way 2015). It should be noted that not all scores are comparable, since they are obtained on different systems, languages and testing material. Nevertheless, the improvement reported in Table 4.1 is all the more substantial given that the novel used for this test case appears to be particularly challenging for MT. This is indicated first by the low scores displayed by both public systems, but also by various measures performed in a previous study to evaluate the degree of freedom and lexical diversity in the translations compared to other literary works (Hansen et al. 2022).

This complexity did not come as a surprise, considering that the translator opted for a very free translation of the original novel as a general strategy, as well as a more formal register overall, with significant variations between certain characters. The initial low score of the generic and public systems can also be attributed to the use of regionalisms and archaisms in French, as well as the presence of dialogues adapted from Old French. Such terms are not likely to have been present in these systems' training data, nor are the many terms, or irrealia (Loponen 2009) that are specific to the universe developed in the saga. Consequently, even if BLEU is not a good indicator of quality, this measure can contribute to refuting the common idea that fantasy works and literature for young adults is easier to translate with MT.[13] Although this will subsequently be the subject of a more thorough investigation, early observations reveal that the customized system was able to retain all of the novel-specific terms, to opt for solutions in line with the translator's style, and, most surprisingly, to merge sentences where the translator also chose to do so. Notwithstanding these positive findings, the system conversely suffered from general shortcomings of MT that are not specific to the literary field, such as poor choices on the lexical level and a strong tendency to translate literally.

Among the advantages of using such technology would be an increase not only in productivity but also in creativity and quality as well. As with the use of CAT tools, it could also reduce cognitive load and make the entire process more enjoyable (Taivalkoski-Shilov 2019). My interpretation of this issue is that it can only happen in an ergonomic and adapted interface, such as a CAT or other interactive environment. If it were to be adequately integrated in the translator's workflow, with personal engines controlled by the user, the effects on quality and creativity could be much more positive than the raw post-editing of a pre-translated text in a word processor. In that case, MT could possibly speed up the translation process as a whole, leaving time to focus on more challenging sections of the text, on research and on revision, or it could support the translator by offering alternative solutions for the segment being translated. These form

only preliminary assumptions in the growing body of research in the field. In the future, new generations of translators may turn to LMT out of habit, as seems to be happening now with CAT tools. These developments, however, raise new questions and challenges for the field.

The way forward

Some of the outcomes from these experiments reveal that the technologies explored in this chapter have their limitations and that there is a need for better tools across the profession as a whole (O'Brien et al. 2017), but especially if we want to improve their usefulness to literary translators, as likewise pointed out by Rothwell (2020b). This study hopes to establish that existing software can already be useful, but also that CAT tools can be used differently for literary texts. More importantly, it shows that we could imagine programmes even better suited to the translation of creative texts.

To move forwards, some of the current hurdles to using CAT tools will need to be addressed. In addition to being expensive and requiring a steep learning curve, CAT tools have search options that can sometimes feel limited in comparison with traditional corpus tools. Thus, Teixeira and O'Brien (2017: 98) put forward the call for increased search functionalities, to which could be added other visualization aids offering more context for instance, as it is sometimes necessary to use an additional programme for this purpose.[14] Segmenting the text into sentences can also pose a problem for some users (O'Brien et al. 2017), so I would argue that taking paragraphs as the basic unit of segmentation could be extremely useful to literary translators, giving them more freedom over the structure of the text. Expanding the boundary of segments would also resolve most of the alignment errors made by CAT tools and facilitate the fully automatic creation of custom-made literary corpora. Some programmes already allow this, but they prevent any automatic retrieval from TMs, highlighting the need for a better handling of this segmentation option. Nevertheless, the lack of this pre-translation feature is also in keeping with the account of translators who say that they prefer turning off the automatic suggestions completely, as they want to carry out all their research manually carried out with the concordance tool.

For translation technology to move ahead, developers should ideally reach out to translators to see how best to adapt the tools to their needs. But, according to Lavault-Oléon (2011: 10), this is not often, if ever, the case; an observation

that still stands today. Yet, research shows that simple changes to aspects as trivial as the display and text presentation of CAT tools, which have been taken for granted for decades, can lead to significant discrepancies in translator performance and satisfaction (Läubli et al. 2022). In the specific case of CALT, the prevalent image of the noble literary endeavour as completely opposed to the technicist use of machines is so prevalent that it has deterred developers from even trying to develop adapted tools (Lacour 2019). So far, the only attempts have been made by volunteers and scholars through open-source initiatives that I will briefly outline later, as they provide interesting insights into the computer-assisted translation of creative texts.

Indeed, the concept of CAT covers more than systems such as Trados (Loock 2016: 103), and it is easily possible to think of new tools that could facilitate the translation of literary texts. Among these, software dedicated to specific challenges of creative language can be found, including the wordplays that are at the core of PunCAT's design (Miller 2019). TraduXio is another and is illustrative in that it places the emphasis on concordance searches between literary works and their many translations, by comparing multiple versions of a text (Goncharova and Lacour 2011). This approach is strongly reminiscent of the 'stereoscopic translation' in Rothwell (2020a: 178), presented by the author as a translation from slightly different but complementary angles, or visions, of the same text. Furthermore, the TraduXio initiative also aims to foster collaboration between professionals, which resonates with the findings that translators would welcome tools that allow them to connect with their peers (Ruffo 2018: 130). This is also in line with renowned German literary translator Tophoven's promotion of collegiality, which eventually resulted in the establishment of the European College of Translators as well as the Réseau Européen des Centres Internationaux de Traducteurs littéraires. Tophoven, credited by Berman (1995: 9) as 'the first to foresee the wonderful possibilities that computers could offer to literary translators', was also an early adopter of these machines, which supported his approach of 'transparent translation' (see Cordingley 2020). Likewise, this chapter has discussed how keeping track of translation choices could be beneficial for the literary translator. Features facilitating this process or allowing edits to be saved would also be a key functionality, in a similar fashion to what is offered by the software Smartcat (Youdale and Rothwell 2022).

With these initiatives and positive aspects in mind, it remains nonetheless important to study how all of these programmes are presented to translators and how they might affect the workflow, given that CAT tools, and resource

consultation tasks in particular, can be critical in either reducing cognitive load or inducing cognitive friction (Taivalkoski-Shilov 2019; Teixeira and O'Brien 2017). This type of 'ecological study' of translation technologies becomes all the more important if MT is to join the current range of tools available to literary translators (Kenny and Winters 2020). To minimize priming effects and the detrimental impact on quality and creativity, I would argue that the implementation of LMT should move away from the traditional post-editing (PE) model, where the entire text is pre-translated and the user only has to revise the output. In an ideal scenario, translators should be able to toggle MT suggestions on and off in an interactive way or display the output outside of the immediate translation pane, alongside TM results, for instance. Moreover, it should be possible to show more than one MT solution if translators have access to multiple engines, as suggested in the setup of Besacier's (2014) experiment, or if the system is set to produce multiple hypotheses. The main concern in all of these recommendations is to enable the user to remain in control of the process as much as possible. For the same reason, paragraph segmentation might be best suited for the task, as is the case with CAT tools and as pointed out in Moorkens et al. (2018). Advances in this direction could prove particularly meaningful in the literary field, notably since the custom-made MT system in this study has the capacity to reorganize sentences in the same way as the translator had. Unfortunately, the handling of paragraphs is not yet optimal, even though document-level translation is currently an important research avenue for MT. Lastly, studies evaluating the impact of LMT on creativity and style, as in Guerberof-Arenas and Toral (2020) or Kenny and Winters (2020), will be of special significance.

So far, I have depicted an optimistic and ideal use case scenario for these technologies, which would hopefully empower translators. However, translation technologies also raise concerns that make it vital to turn our attention to CALT and LMT to anticipate possible changes. If these tools can have a favourable impact all along the translation chain, notably in the case of LMT (Besacier 2014; Hansen 2021), unthinking use of such technology could also lead to more effort than time savings, and its use in the name of productivity may arguably have a drastic impact on quality. In turn, such an impact could affect the recognition of the translator and author's work, as well as the reader's experience and language learning process. Foreseeable issues include the selling of unedited MT by mercenary publishers, a stronger temptation to hire non-professionals, as well as negative effects on remuneration, deadlines or intellectual copyright and ownership. Although not unique to the literary field, these issues may

well be exacerbated in the literary domain due to global market struggles, the tendency of publishing houses to reduce costs by whatever means, and the already precarious situation of literary translators (Taivalkoski-Shilov 2019). The availability and use of corpora have undoubtedly changed the industry, and language service providers have been quick to react to the trend. Indeed, reports in 2020–1 – for example, at Nimdzi (Akhulkova, Hickey and Agulló García 2021) or TAUS (Aslan 2021) – undeniably mark a shift in the language industry from traditional translation tasks to data mining, annotation and provision services, with providers also leveraging and transforming their language data into 'business assets'. The TAUS Data Marketplace (TAUS 2020) perfectly illustrates this trend, as the platform allows companies and language providers – including publishing companies – to sell datasets with a price per word that is reminiscent of and sometimes higher than regular market translation rates. These trends further confirm that the data produced by translators has become 'the new oil' of the language industry and highlight the need for new regulations, which associations of literary translators could help to establish.

Conclusion

While there is still a degree of uncertainty surrounding the changes to come, this uncertainty is precisely why it is important to keep an interest in new technologies and how these can affect the working conditions of translators, as they ultimately have a direct effect on the quality of translations (Lavault-Olléon 2011: 7). What we know for certain, however, is that literary translators have always made extensive use of corpora, in one form or another. By investigating this avenue as a corpus-based approach to literary translation, this chapter hopes therefore to promote a reasoned approach to CALT.

While there has been a traditional assumption that technology and literature should be entirely incompatible, few attempts have been made to adapt translation technologies or use them with literary data. Just as human translators need to train and specialize in certain fields, it is important to feed the tools, whether CAT or MT, with relevant resources so that they reflect the specific challenges of literary translation. Furthermore, relying on custom corpora allows professionals to create personal datasets that are tailored to their own style, voice and choices. They are then free to explore, draw on and reuse these resources advantageously in different scenarios. These may lead to a closer reading of the source text (that the alignment process or the use of other 'inscriptions' of the

text encourages), while others might help with the translation process in its entirety (as CAT tools do) or facilitate the production of the target text (which is the main objective of MT).

As an early proponent of CALT, Tophoven drew attention to the multitude of tools available on computers to assist the translator along the various phases of the translation process. As Cordingley notes:

> Rather than being anxious that translators would be superseded by technology, [Tophoven] saw the potential for computers to register and communicate the complexity of translation, to retain and share each translator's professional knowledge and possibly to increase efficiency by allowing each to build upon the work of their peers. (2020)

Today, CALT should still be viewed as an array of tools that largely surpass the automatic retrieval of segments from translation memories, and as a working environment that is better suited to the work of the translator than word-processing software. Musy (1989) also shared this notion, asserting, more than thirty years ago, that literary translators had, at the time, barely been able to capitalize on the benefits brought about by the advent of computers. It was for this reason that he exhorted professionals to familiarize themselves with these emerging tools, in the hope that new technologies could be adapted to serve them better. Unfortunately, it would appear that the situation has barely changed since then with respect to software development.

Considering that current technologies rely primarily on corpora that are produced by human translators, crucial issues of copyright and data ownership have arisen and are becoming increasingly complex and urgent with MT and the recent arrival of large language model applications (Koponen, Nyqvist and Taivalkoski-Shilov 2023). Although investigation into this ethical dimension is envisaged as a continuation of this study, the framework depicted throughout this chapter also offers a possible solution to the challenge of intellectual property, as it suggests a way for translators to use these tools on their own data, without any intermediary, and retain ownership over the entire process. In any case, relying on personal translations as literary corpora allows for interesting avenues that could potentially improve the comfort of translators, the creative process and the quality of the target text. Instead of restricting the translator's freedom, TMs can support the decision-making process and inspire new solutions, all the while maintaining style and reinforcing consistency. Consulting TMs to consciously and voluntarily deviate from a solution is a conversely surprising use of the technology that illustrates this freedom. Moreover, machine translation could

add to these suggestions and further help the translation process, particularly if it can be trained on the production of individual translators.

Evidently, CALT differs from working in a simple word-processing environment, and familiarity or formal training can play an important role in overcoming the constraints of these technologies. What should be stressed is that the human-machine interaction is a complementary one, and that it can actually reinforce the human aspects of translation by providing different perspectives on a text that would otherwise be invisible to the translator's eye. Nevertheless, despite these advantages, machines cannot truly be creative in themselves. By definition, they can only copy patterns learned from human productions and reproduce them in predictable but occasionally surprising ways. Thus, human translators are not destined to disappear in the foreseeable future – not until a major paradigm shift changes the status quo. Their ability to understand the actual meaning of a text and come up with creative solutions is still far beyond the reach of machines. Notwithstanding these current drawbacks, the use of MT and CAT tools can, on the other hand, support translators in multiple ways during the complex set of tasks that comprises the lengthy process of translation. Ultimately, the choice mainly rests on the ability, willingness, personal habits, professional outlook and, indeed, the training of the individual translator.

Overall, literary translators enjoy a privileged position in the field of translation, in that the sole motivation for them to turn to CALT is arguably their own personal curiosity. This should not prevent us from anticipating changes that are likely to come about in the near future, however, especially when considering that the majority of these issues still apply across the profession: no ownership over the data, decreased translation rates, shorter deadlines, ergonomic concerns, visibility, quality and so on (Doherty 2016). Even so, as observed by Cartano in what now seems to be the perennial story of (literary) translation, this is not a cautionary tale against computers themselves, but against the possibility that the benefits of a translator's personal equipment – and of their investment in their capacity to use technology with confidence – notably the financial ones, only flow to someone else, in our case the publishers. On the contrary, we should try to better equip translators, who are indeed asking for tools tailored to their activity (Ruffo 2018). Training is equally essential, as opportunities are scarce in this field, and many professionals may not be up to date with the latest developments in digital tools (Slessor 2020: 249). In the current climate, then, we should always bear in mind that work should be adapted to humans rather than the reverse (Lavault-Olléon 2011: 6) and strive for tools that more adequately reflect the challenges of the literary translator.

Notes

1 A web-based version, YouAlign, also allows users to create bitexts for free: https://youalign.com/.
2 It should be noted that this process was first conducted in 2015 and that the performances of the software could have changed since then. Automatic alignment is an ongoing and improving research area, and there are multiple other programmes available.
3 'DIY corpora' refer to sets of texts that are compiled by a user for their own use, while 'disposable corpora' designate documents that will be used only once, therefore requiring barely any processing. In his work, Loock explains that the compilation of such datasets does not require considerable time or computer knowledge and is easily compensated by the relevance and variety of queries that they allow.
4 A survey conducted by Braga (2021) reveals that almost all of the arguments levelled against CALT fall into these categories, and that they tend to underline the confusion about what CAT tools really are as well as how they work. Braga, however, puts another reason forward, namely the price of the software, which represents a substantial investment and a strong disincentive to experiment with these tools. On that note, while free alternatives are available and sometimes offer ingenious features that could be of particular interest for creative texts, they can also lack the nifty functionalities that make their paid counterparts useful for CALT.
5 These features refer to the automatic retrieval and display in the translation pane of the most similar TM segment, and to the propagation of a newly translated segment to similar ones throughout the rest of the document, respectively.
6 Contrary to the two previous automated features, context searches take place in a dedicated pane and allow the user to manually search for terms, phrases or segments within the TM and then view multiple translations in context for these queries.
7 Readers interested in a more thorough survey of the literature on this topic, in French, can refer to Hansen (2021).
8 Although there are experiments before and after the successful applications of neural models of Sutskever, Vinyals and Le (2014) and Bahdanau, Cho and Bengio (2015), I posit that the workshop on *The Qualities of Literary Machine Translation* at the *Machine Translation Summit XVII* in August 2019 (EAMT, Dublin City University) represents a turning point in the research on LMT.
9 For more details on the training process, results and analysis of this experiment, see Hansen et al. (2022).
10 Although it continues to receive minor updates, the Transformer architecture is still the state of the art for machine translation and most natural language processing

applications, including large language models such as GPT, that takes its name from this architecture. In this experiment, I used the same parameters as those given in the original paper, so as to determine the capabilities of this base model without having to take into consideration questions of computational resources or careful optimization.

11 These automated metrics are commonly used in machine translation tasks to compare how different systems perform on a given task. BLEU estimates the similarity between a machine-translated text and a reference text by comparing strings of words. The second measure also estimates similarity on the basis of strings of characters, while TER calculates the number of edits to the machine output that are needed. BLEU is the most widely used metric by far, but the reliability of these tests has also been questioned repeatedly (Berg-Kirkpatrick, Burkett and Klein 2012).

12 Publicly available systems tested on 25 December 2020. See Hansen and Esperança-Rodier (2023) for metric signatures and additional measures.

13 As a point of comparison, the translation into French of the English novels contained in the books corpus produced BLEU scores comprised between 13 and 31 with the same public systems.

14 During the CALT experiment, I often relied on Xbench (ApSIC 2011) to carry out queries.

References

Akhulkova, Y., S. Hickey, and B. Agulló García (2021), 'Nimdzi Language Technology Atlas 2021', *Nimdzi*, 5 August. Available online: www.nimdzi.com/nimdzi-language-technology-atlas-2021 (accessed 18 May 2023).

Anthony, L. (2014), *AntConc* (Version 1.0). Tokyo: Waseda University. Available online: https://www.laurenceanthony.net/software

Aslan, Ş. (2021), 'Data and AI Trends in 2022', *TAUS*, 1 December. Available online: www.taus.net/resources/blog/data-and-ai-trends-in-2022 (accessed 18 May 2023).

ATLAS (1988), *Actes des quatrièmes assises de la traduction littéraire*, Arles: Actes Sud.

Bahdanau, D., K. Cho, and Y. Bengio (2015), 'Neural Machine Translation by Jointly Learning to Align and Translate', in Y. Bengio and Y. LeCun (eds), *3rd International Conference on Learning Representations: Conference Track Proceedings*, 1–15, San Diego, CA.

Baker, M. (2000), 'Towards a Methodology for Investigating the Style of a Literary Translator', *Target* 12 (2): 241–66, doi:10.1075/target.12.2.04bak.

Bédard, C. (2016), *LogiTerm*. Montreal: Terminotix.

Berg-Kirkpatrick, T., D. Burkett, and D. Klein (2012), 'An Empirical Investigation of Statistical Significance in NLP', in J. Tsujii, J. Henderson, and M. Paşca (eds), *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language*

Processing and Computational Natural Language Learning, 995–1005, Jeju Island: ACL.

Berman, A. (1995), 'In Memoriam Elmar Tophoven', *TTR*, 2 (1): 9, doi:10.7202/037029ar.

Besacier, L. (2014), 'Traduction automatisée d'une œuvre littéraire: Une étude pilote', in P. Blache, F. Béchet, and B. Bigi (eds), *Actes de la 21ème conférence sur le Traitement Automatique des Langues Naturelles*, 389–94, Paris: ATALA.

Blin, R. (2021), 'Neural Machine Translation, Corpus and Frugality', *arXiv*, arXiv:2101.10650.

Boase-Beier, J. (2014), 'Stylistics and Translation', in Michael Burke (ed.), *The Routledge Handbook of Stylistics*, 393–407, Abingdon: Routledge.

Braga, G. (2021), 'Translating Literature from Scratch and the Computer: CAT Tools for Literary Translators', [Presentation] *Séminaire TRACT 2021–2022*, Paris, 9 December.

Bundgaard, K. and T. P. Christensen (2019), 'Is the Concordance Feature the New Black? A Workplace Study of Translators' Interaction with Translation Resources While Post-Editing TM and MT Matches', *JoSTRans*, 31: 14–37.

Contributor 1 (2015), 'Do You Use Trados for Literary Texts, Too?', *Proz*, 23 March–3 November. Available online: www.proz.com/forum/literature_poetry/283550-do_you_use_trados_for_literary_texts_too.html (accessed 18 May 2023).

Contributor 2 (2015), 'Do You Use Trados for Literary Texts, Too?', *Proz*, 23 March–3 November. Available online: www.proz.com/forum/literature_poetry/283550-do_you_use_trados_for_literary_texts_too.html (accessed 18 May 2023).

Contributor 3 (2015), 'Do You Use Trados for Literary Texts, Too?', *Proz*, 23 March–3 November. Available online: www.proz.com/forum/literature_poetry/283550-do_you_use_trados_for_literary_texts_too.html (accessed 18 May 2023).

Contributor 4 (2015), 'Do You Use Trados for Literary Texts, Too?', *Proz*, 23 March–3 November. Available online: www.proz.com/forum/literature_poetry/283550-do_you_use_trados_for_literary_texts_too.html (accessed 18 May 2023).

Contributor 5 (2015), 'Do You Use Trados for Literary Texts, Too?', *Proz*, 23 March–3 November. Available online: www.proz.com/forum/literature_poetry/283550-do_you_use_trados_for_literary_texts_too.html (accessed 18 May 2023).

Cordingley, A. (2020), 'Tophoven's Dream: A Prototype for Genetic Translation Studies', *Palimpsestes*, 34: 152–68, doi:10.4000/palimpsestes.6102.

Doherty, S. (2016), 'The Impact of Translation Technologies on the Process and Product of Translation', *International Journal of Communication*, 10: 947–69.

Durieux, C. (2000), 'Traduction littéraire et traduction technique : Même démarche', *Revue des lettres et de traduction*, 6: 13–25.

Froeliger, N. (2013), *Les Noces de l'analogique et du numérique*, Paris: Les Belles Lettres.

Goncharova, Y. and P. Lacour (2011), 'TraduXio: Nouvelle expérience en traduction littéraire', *Traduire*, 225: 86–100, doi:10.4000/traduire.94.

Guerberof-Arenas, A. and A. Toral (2020), 'The Impact of Post-Editing and Machine Translation on Creativity and Reading Experience', *Translation Spaces*, 9 (2): 255–82, doi:10.1075/ts.20035.gue.

Hansen, D. (2021), 'Les lettres et la machine: Un état de l'art en traduction littéraire automatique', in P. Denis, N. Grabar, A. Fraisse, R. Cardon, B. Jacquemin, E. Kergosien, and A. Balvet (eds), *Actes de la 28ème Conférence sur le Traitement Automatique des Langues Naturelles*, 2: 29–45, Paris: ATALA.

Hansen, D. and E. Esperança-Rodier (2023), 'Human-Adapted MT for Literary Texts: Reality or Fantasy?', in S. Castilho, R. Caro Quintana, M. Stasimioti, and V. Sosoni (eds), *Proceedings of the New Trends in Translation and Technology Conference - NeTTT 2022*, 178–90, Rhodes: NeTTT.

Hansen, D., E. Esperança-Rodier, H. Blanchon, and V. Bada (2022), 'La traduction littéraire automatique: Adapter la machine au travail de l'humain', *Journal of Data Mining & Digital Humanities, Vers une robotique du traduire*, 1–19, doi:10.46298/jdmdh.9114.

Hansen, D. and P.-Y. Houlmont (2022). 'A Snapshot into the Possibility of Video Game Machine Translation', in J. Campbell, S. Larocca, J. Marciano, K. Savenkov, and A. Yanishevsky (eds), *Proceedings of the 15th Biennial Conference of the Association for Machine Translation in the Americas*, 2:257–69, Orlando: AMTA.

Hatim, B. and I. Mason (1997), *The Translator as Communicator*, London: Routledge.

Kenny, D. and M. Winters (2020), 'Machine Translation, Ethics and the Literary Translator's Voice', *Translation Spaces*, 9 (1): 123–49, doi:10.1075/ts.00024.ken.

Kilgarriff, A., P. Rychlý, P. Smrž, and D. Tugwell (2004), 'The Sketch Engine', in *Proceedings of the 11th EURALEX International Congress*, 105–16. Available from: http://www.sketchengine.eu/

Klein, G., Y. Kim, Y. Deng, J. Senellart, and A. Rush (2017), 'OpenNMT: Open-Source Toolkit for Neural Machine Translation', in M. Bansal and H. Ji (eds), *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 67–72, Vancouver: ACL.

Koponen, M., S. Nyqvist, and K. Taivalkoski-Shilov (2023), 'Translating with Technology: How Digitalisation Affects Authorship and Copyright of Literary Texts', in J. L. Hadley, K. Taivalkoski-Shilov, C. S. C. Teixeira, and A. Toral (eds), *Using Technologies for Creative-Text Translation*, 180–98, New York: Routledge, doi:10.4324/9781003094159-9.

Kuzman, T., Š. Vintar, and M. Arčan (2019), 'Neural Machine Translation of Literary Texts from English to Slovene', in J. Hadley, M. Popović, H. Afli, and A. Way (eds), *Proceedings of the Qualities of Literary Machine Translation*, 1–9, Geneva: EAMT.

Lacour, P. (2019), 'TraduXio: Un environnement numérique pour la traduction collaborative et multilingue de textes culturels', [Presentation] *Atelier 2019 Digit-Hum*, Paris, 17 October.

Ladmiral, J.-R. (1994a), 'Le traducteur et l'ordinateur', *Langages*, 116: 5–19, doi:10.3406/lgge.1994.1691.

Ladmiral, J.-R. (1994b), *Traduire : Théorèmes pour la traduction*, 2nd edn, Paris: Gallimard.

Läubli, S., P. Simianer, J. Wuebker, G. Kovacs, R. Sennrich, and S. Green (2022), 'The Impact of Text Presentation on Translator Performance', *Target*, 34 (2): 309–42, doi:10.1075/target.20006.lau.

Lavault-Olléon, É. (2011), 'L'ergonomie, nouveau paradigme pour la traductologie', *ILCEA*, 14: 1–16, doi:10.4000/ilcea.1078.

Loock, R. (2016), *La traductologie de corpus*, Villeneuve-d'Ascq: Presses universitaires du Septentrion.

Loponen, M. (2009), 'Translating Irrealia: Creating a Semiotic Framework for the Translation of Fictional Cultures', *Chinese Semiotic Studies*, 2 (1): 165–75, doi:10.1515/css-2009-0117.

Matusov, E. (2019), 'The Challenges of Using Neural Machine Translation for Literature', in J. Hadley, M. Popović, H. Afli, and A. Way (eds), *Proceedings of the Qualities of Literary Machine Translation*, 10–19, Geneva: EAMT.

Mazoyer, R. (2020), 'Traduction d'essai et TAO. Le racisme est un problème de Blancs de Reni Eddo-Lodge: Une étude de cas', [Presentation] *La traduction littéraire et SHS à la rencontre des technologies de la traduction: enjeux, pratiques et perspectives*, Toulouse, 7 February.

Miller, T. (2019), 'The Punster's Amanuensis: The Proper Place of Humans and Machines in the Translation of Wordplay', in I. Temnikova, C. Orăsan, G. Corpas Pastor, and R. Mitkov (eds), *Proceedings of the Second Workshop on Human-Informed Translation and Interpreting Technology (HiT-IT 2019)*, 57–64, Varna: Incoma Ltd, doi:10.26615/issn.2683-0078.2019_007.

Moorkens, J., A. Toral, S. Castilho and A. Way (2018), 'Translators' Perceptions of Literary Post-editing using Statistical and Neural Machine Translation', *Translation Spaces*, 7 (2): 240–62, doi:10.1075/ts.18014.moo.

Mounin, G. (2016), *Les Belles Infidèles*, 3rd edn, Villeneuve-d'Ascq: Presses Universitaires du Septentrion.

Müller, M., A. Rios, and R. Sennrich (2020), 'Domain Robustness in Neural Machine Translation', in M. Denkowski and C. Federmann (eds), *Proceedings of the 14th Conference of the Association for Machine Translation in the Americas*, 1: 151–64, Washington: AMTA.

Musy, G. (1989), 'La traduction littéraire "assistée par ordinateur"', *Études de lettres*, 221: 85–8.

Newmark, P. (1981), *Approaches to Translation*, Oxford: Pergamon Press.

Nieminen, T. (2021), 'OPUS-CAT: Desktop NMT with CAT Integration and Local Fine-Tuning', in D. Gkatzia and D. Seddah (eds), *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, 288–94, Kyiv: ACL, doi:10.18653/v1/2021.eacl-demos.34.

O'Brien, S., M. Ehrensberger-Dow, M. Hasler, and M. Connolly (2017), 'Irritating CAT Tool Features that Matter to Translators', *Hermes – Journal of Language and Communication in Business*, 56: 145–62, doi:10.7146/hjlcb.v0i56.97229.

Original Poster (2015) 'Do You Use Trados for Literary Texts, Too?', *Proz*, 23 March–3 November. Available online: www.proz.com/forum/literature_poetry/283550-do_you_use_trados_for_literary_texts_too.html (accessed 18 May 2023).

Papineni, K., S. Roukos, T. Ward, and W.-J. Zhu (2002), 'BLEU: A Method for Automatic Evaluation of Machine Translation', in P. Isabelle, E. Charniak, and D. Lin (eds), *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 311–18, Stroudsburg: ACL, doi:10.3115/1073083.1073135.

Post, M. (2018), 'A Call for Clarity in Reporting BLEU Scores', in O. Bojar, R. Chatterjee, C. Federmann, M. Fishel, Y. Graham, B. Haddow, M. Huck, A. Jimeno Yepes, P. Koehn, C. Monz, M. Negri, A. Névéol, M. Neves, M. Post, L. Specia, M. Turchi, and K. Verspoor (eds), *Proceedings of the Third Conference on Machine Translation: Research Papers*, 186–91, Stroudsburg: ACL, doi:10.18653/v1/W18-6319.

Rothwell, A. (2020a), 'CAT Tools and Creativity', in M. De Meo and E. Di Martino (eds), *Translation and Creativity: Readers, Writers, Processes*, 169–81, Canterano: Aracne editrice.

Rothwell, A. (2020b), 'Retranslating Zola in Stereo with CAT Tools', [Presentation] *La traduction littéraire et SHS à la rencontre des technologies de la traduction: enjeux, pratiques et perspectives*, Toulouse, 7 February.

Ruffo, P. (2018), 'Human-Computer Interaction in Translation: Literary Translators on Technology and Their Roles', in D. Chambers, J. Drugan, J. Esteves-Ferreira, J. M. Macan, R. Mitkov, and O.-M. Stefanov (eds), *Proceedings of the 40th Conference Translating and the Computer*, 127–31, Geneva: Éditions Tradulex.

Sage, A. (2005–13), *Septimus Heap*, 7 vols, London: Bloomsbury Publishing.

SDL Trados (2015), Maidenhead: SDL plc.

Serval, N. (2005–13), *Magyk*, ed. Angie Sage, 6 vols, Paris: Albin Michel.

Slessor, S. (2020), 'Tenacious Technophobes or Nascent Technophiles? A Survey of the Technological Practices and Needs of Literary Translators', *Perspectives*, 28 (2): 238–52, doi:10.1080/0907676X.2019.1645189.

Strandvik, I. (2001), 'Las memorias de traducción en la Comisión Europea', *Tradumática*, 1–4.

Sutskever, I., O. Vinyals, and Q. V. Le (2014), 'Sequence to Sequence Learning with Neural Networks', in Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger (eds), *NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems*, 2: 3104–12, Montréal: MIT Press.

Taber, C. R. (1972), 'Traduire le sens, traduire le style', *Langages*, 28: 55–63, doi:10.3406/lgge.1972.2098.

Taivalkoski-Shilov, K. (2019), 'Ethical Issues Regarding Machine(-Assisted) Translation of Literary Texts', *Perspectives: Studies in Translation Theory and Practice*, 27 (5): 689–703, doi:10.1080/0907676X.2018.1520907.

TAUS (2020), *Data Marketplace*, 5 November. Available online: https://datamarketplace.taus.net (accessed 18 May 2023).

Teixeira, C. S. C. and S. O'Brien (2017), 'Investigating the Cognitive Ergonomic Aspects of Translation Tools in a Workplace Setting', *Translation Spaces*, 6 (1): 79–103, doi:10.1075/ts.6.1.05tei.

Tiedemann, J. (2012), 'Parallel Data, Tools and Interfaces in OPUS', in N. Calzolari, K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, and S. Piperidis (eds), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, 2214–18, Luxembourg: ELRA.

Toral, A., A. Oliver, and P. Ribas-Bellestín (2020), 'Machine Translation of Novels in the Age of Transformer', in J. Porsiel (ed.), *Maschinelle Übersetzung für Übersetzungsprofis*, 276–95, Berlin: BDÜ Fachverlag.

Toral, A. and A. Way (2015), 'Machine-Assisted Translation of Literary Text: A Case Study', *Translation Spaces*, 4 (2): 240–67, doi:10.1075/ts.4.2.04tor.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin (2017), 'Attention Is All You Need', in U. von Luxburg, I. Guyon, S. Bengio, H. Wallach, and R. Fergus (eds), *NIPS'17*: *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 6000–10, Long Beach: Curran Associates.

Youdale, R. (2020), *Using Computers in the Translation of Literary Style: Challenges and Opportunities*, Abingdon: Routledge.

Youdale, R. and A. Rothwell (2022), 'Computer-Assisted Translation (CAT) Tools, Translation Memory and Literary Translation', in S. Deane-Cox and A. Spiessens (eds), *The Routledge Handbook of Translation and Memory*, 381–402, Abingdon: Routledge.

Part II

Shifting Methods and Models

5

Risk Management for Content Delivery via Raw Machine Translation

Maarit Koponen and Mary Nurminen

Introduction

Technological developments in machine translation (MT) have led to an ever-increasing use of MT as an aid in communicating across languages. In content delivery, MT is most commonly used in a post-editing (PE) process, where a professional translator checks and edits the MT (see, e.g., Guerberof-Arenas 2020). Parallel to this, MT can also be used in its raw, unedited form to help people interact with either content or other people across language barriers. In addition to people using free online MT tools, companies and public organizations may also provide machine-translated content. An organization might choose to deliver information as raw MT so that they can offer more content in more languages and reach new audiences while saving on time and costs. For example, some technology companies have aimed to broaden the range of languages in which they offer product support information by delivering the content in some languages as raw MT (Thicke 2013). In another example, the field of intellectual property rights (IPR) has relied on raw MT for more than a decade (Nurminen 2019) to enable patent professionals to identify relevant patent documents in languages they do not speak themselves.

When MT is used for delivering content, it is important to manage the risks involved. Recent literature (Vieira, O'Hagan and O'Sullivan 2020; Scott and O'Shea 2021) has highlighted the consequences associated with the use of raw MT. Other studies have discussed risk management, although mainly with a focus on scenarios where MT is post-edited (Nitzke, Hansen-Schirra and Canfora 2019; Canfora and Ottmann 2020). Scenarios where raw MT is used, however, entail specific risk considerations. Despite the recently reported

improvements of neural MT systems, the possibility of inaccurate translations still exists, and translation errors may entail high risks. At the same time, the use of MT should be balanced against the potential consequences of not translating the information at all, as argued by O'Brien (2020: 313). In this way, delivering information via raw MT can be considered not only a risk but also a mitigation of the higher-level risk of people receiving no information. To better understand the contexts where delivering information via raw MT would be effective, and to manage the potential risks and mitigating factors, a more nuanced understanding would be needed about how people react to and consume raw MT. However, research and discussions of best practices in this area are currently lacking.

This chapter aims to investigate how risk management principles can be applied to scenarios where information is delivered to end-users of products and services as raw MT. We examine risk factors for assessing the use of raw MT, as well as options for managing and mitigating the risks. We contend that a simple checklist of rules that would lead to minimal or no-risk delivery of raw MT content is not possible, and therefore do not aim to propose one. Rather, our goals are to provide general considerations for assessing and managing the risks involved in content delivery via raw MT and to contribute a new viewpoint to existing risk management models in translation. Besides contributing to academic discussions, this can offer a concrete tool for organizations considering using raw MT to communicate across languages. The chapter first provides an overview of risk management principles based on International Standard ISO 31000:2018 and related work on risk management in the context of translation. We then apply the risk management framework to the delivery of content as raw MT, before illustrating these considerations through a concrete use case for raw MT in the patent world. The final section presents the conclusions of this analysis.

Related work: Risk management

Key concepts and framework for risk management

A commonly used risk management framework is the International Standard ISO 31000 (ISO 2018), which defines principles, key concepts and processes. The central concept of risk is defined as the 'effect of uncertainty on objectives' (ISO 2018, term 3.1). The uncertainty arises from a given risk source (term 3.4)

and can lead to positive or negative consequences (term 3.6). Risk management according to the ISO standard involves a systematic, iterative process integral to an organization's operation. This process should be customized to each organization's internal context, which relates to the values, strategies and culture of the organization itself, and the external context, which involves social, regulatory, technological and economic factors among others (ISO 2018: 10). The standard also emphasizes the need to account for stakeholder needs and perspectives through active communication and consultation. The overall risk management process of ISO 31000:2018 is illustrated in Figure 5.1.

Within the risk management process, the ISO standard defines risk assessment to encompass risk identification, where the organization examines both tangible and intangible risk sources and consequences, followed by risk analysis to understand the nature of a risk and the likelihood, nature and severity of consequences. Risk evaluation then compares the analysis results to risk criteria, and, based on this evaluation, the organization can decide to take the risk, to avoid the risk by not starting or continuing an activity that causes risk, or to employ some risk treatment measures (see ISO 2018). The standard emphasizes that decisions should not be made on an economic basis alone but should also consider organizational obligations and the diverse stakeholder perspectives.

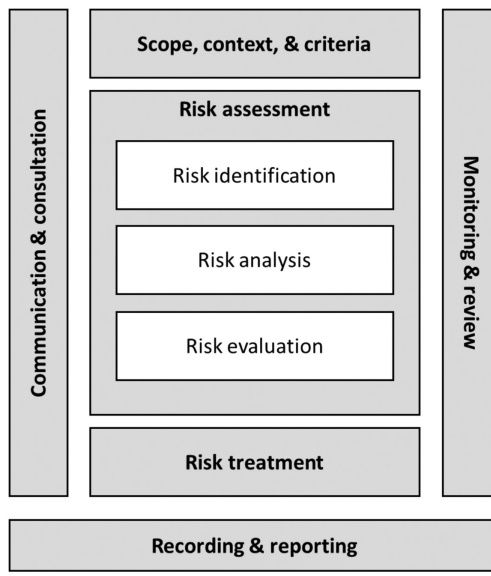

Figure 5.1 Risk management process according to ISO 31000:2018, figure 4.

Reporting and monitoring risk treatment is also important to dynamically address new and emerging risks, and to continually improve the process.

Translation, MT and risk

Relatively little research has been conducted on risk management in the context of (human) translation, but in recent years this topic has received increasing interest. Risk management has been applied at different levels and stages, from an individual translator's management of uncertainty and risk mitigation (see Pym 2021 for an overview) to risk management in terms of business considerations and project management (e.g. Dunne 2013; Zaveckaite and Ulbinaite 2018).

A common view of risk is as something negative, and although risk-taking can also lead to positive outcomes, risk management tends to focus on potential negative consequences and mitigation (see Zaveckaite and Ulbinaite 2018: 1325–6). Translation errors are a central source of risk that can affect 'the author, translator, translation agency, client or end user of the translation' (Canfora and Ottmann 2018: 168), and previous work has examined risks and potential consequences of translation errors in safety-critical situations (Canfora and Ottmann 2018; 2019) or in conjunction with commercial, legislative, normative and political texts (Byrne 2007; Cismas 2010). Potential consequences are discussed further in the 'Risk analysis' section.

Recently, risk related to the use of MT has received increased attention. Canfora and Ottmann (2020: 59) note that although many of the risks involved are similar to risks in human translation, some are more specific to the use of MT and entail different considerations. These include potential liability and cyber risks, such as the potential exposure of confidential information. As with Canfora and Ottmann's focus on translation in safety-critical contexts (2018, 2019), several studies on MT suggest that a greater consideration of risk is necessary when the texts involved in MT are from high-risk areas. Way (2013) describes the use of raw MT as risky for translating stock reports because of the high financial risk. Both Nurminen and Koponen (2020) and Vieira, O'Hagan and O'Sullivan (2020) cite public health and medicine as high-risk contexts, and Haddow, Birch and Heafield (2021) argue that post-editing of MT is always needed in these settings. Risk in legal contexts has been discussed by Vieira, O'Hagan and O'Sullivan (2020), Scott and O'Shea (2021) and Guerberof-Arenas and Moorkens (2023).

On the other hand, researchers have also pointed out that some environments in which MT could be used might be somewhat resistant to risk (e.g. Way

2013). Nurminen (2021) lists risk-tolerant environments as a contextual factor that positively influences if and how MT gisting is used, specifically citing the patenting process as one such an environment. She describes the paradox of risk tolerance in the environment: since patenting has a wide variety of inherent risks, affordances for mitigating those risks have been built into processes. Those same affordances help to provide an environment which is tolerant to the risks involved with relying on raw MT (2021: 94–5). This is discussed further in the 'Risk management in action: The case of patent professionals' section.

A few models have been proposed for classifying different use cases as more or less appropriate for raw MT. Guerberof-Arenas and Moorkens (2023) propose a model for classification that relies on two factors: the shelf life of the content to be translated and its riskiness. They offer a number of suggested use cases that fit into one of four categories: short term/low risk, short term/high risk, long term/low risk, and long term/high risk. However, shelf life and risk seem to be separate factors that are simply put into the same model rather than factors that influence each other. A more comprehensive model for risk assessment is offered by Nitzke, Hansen-Schirra and Canfora (2019: 240), who argue that such a model can be employed for both strategic decisions (whether to use MT at all) and operative ones (whether to use MT for a specific purpose or with a specific text). Their model includes factors related to technology, the sensitivity of the text to be translated, the number of people who might read the translation, the level of quality required of the translation and the number of translators and time available for human translation.

A few studies discuss risk treatment strategies, such as introducing guidelines and legislation (Canfora and Ottmann 2020; Martindale 2020), developing technology that enables risk mitigation (Canfora and Ottmann 2020; Martindale 2020) and promoting MT literacy among users (e.g. Bowker and Buitrago Ciro 2019; Martindale 2020). Besides these articles, however, past research on risk treatment tends to emphasize mitigating risk through post-editing and quality control by human translators.

A different approach to the question of risk and MT gisting is taken by Nurminen (2021: 140–1) in her proposal that the practice of MT gisting could be conceptualized as an exercise in risk management. She maps the risk assessment and treatment processes discerned in her study of patent professionals to ISO 31000 (see previous section) to demonstrate how the processes conform to the standard's description of risk management. She then proposes that conceptualizing MT gisting as risk management offers us a framework for

analysing people's use of raw MT, an idea that was first proposed by Pym (2021: 453): 'The categories of risk management invite studies that assess the strategies of all participants in a translation event and use that matrix to try to explain translator decisions.' Nurminen concludes that MT literacy programmes could benefit from including risk management in their syllabi. It would encourage end-users of raw MT not only to acknowledge the risk but also to realize that the risk can be managed through concrete actions.

Risk management considerations for raw machine translation

In this section, we apply the ISO 31000 risk management framework to content delivery via raw MT. Risk is always present in this context: there is no guarantee that a text will convey the meaning intended. The same naturally applies to human translation, but the revision and review controls commonly used to minimize risk in human translation workflows are not available with raw MT. Thorough analysis is therefore required if the use of raw MT is considered, and this section aims to support that analysis. Our discussion focuses mostly on the 'heart' of the process: scope, context and criteria; risk assessment; and risk treatment. It is here that issues that are specific to the delivery of information using raw MT are likely to arise. The other processes – communication, monitoring and reporting – are equally important. In particular, it would be important to implement processes to collect feedback from end-users or internals who notice issues themselves or encounter them in customer comments. For example, Canfora and Ottmann (2018) propose a process for monitoring and analysing incidents related to human translation, and a similar process could be applied to the raw MT situation. Other aspects of risk communication, monitoring and reporting are likely already covered by existing risk management processes in the organization.

We make certain background assumptions in our discussion. First, the risk management considerations are appropriate for any type of organization, including private businesses, public administration or non-profits, so we use the generic organization to refer to them collectively. Second, the considerations presented are not intended to account for decisions regarding which MT tool to select. Instead, we assume that the organization has already evaluated and identified appropriate MT tools or will do so separately. Information on this process is available both in academic literature (e.g. Canfora and Ottmann 2020), and from commercial technology providers and consultants. We also assume that the MT tool selected would be trained on the organization's own

material and would be in the organization's private use. This would offer better quality than generic MT solutions and ensure data protection and privacy.

Communication and consultation

A key part of the risk management framework is communication and consultation with both internal and external stakeholders. This entails involving people with different backgrounds and expertise in order to build a comprehensive and nuanced understanding of different types of risk, the situations where they can occur, and the potential consequences. Needs and expectations with regard to information may vary between stakeholders, as do understanding of MT technology and attitudes towards it. Furthermore, different stakeholder groups may have differing perspectives on risk levels and the acceptability of specific risk treatment options.

Internally, people or groups responsible for content creation and translation would logically be involved in risk management processes focused on MT and information delivery. Other relevant internal stakeholders include managers, people responsible for language technology, and groups who interact with end-users. Consultation can also involve external technology or language service providers or independent consultants, as suggested by Martindale (2020).

A vital external stakeholder group is of course the people to whom the organization intends to offer raw machine-translated information. The importance of this end-user perspective should be recognized, and the organization should seek to proactively involve them in the early stages of planning to start delivering information as raw MT, as well as in later processes like monitoring. Forms of involvement could include user studies and focus groups, such as those described in Bowker and Buitrago Ciro (2015) or Haddow, Birch and Heafield (2021).

Scope, context and criteria

One of the first steps in implementing risk management is to establish the scope of the exercise. First, the organization needs to determine what unit of content is assessed when considering delivery via raw MT. Assessing the suitability of raw MT for each document produced by the organization would probably be a cumbersome and fractured process. In contrast, different information types involve varying types and levels of risk, so assessing risk for all content collectively is not feasible. Therefore, we assume that the basic unit for risk assessment is a specific content set or information type, such as product support materials or internal announcements. Similarly, the decision model for MT use

by Nitzke, Hansen-Schirra and Canfora (2019) uses text type as the assumed unit of analysis. The scope definition should also involve the number of intended target languages and the estimated number of people who would receive the information via raw MT. As noted by Canfora and Ottmann (2018: 172), the wider a translation is circulated and the more target languages are involved, the greater the likelihood of a serious incident (see also Nitzke, Hansen-Schirra and Canfora 2019).

To establish the context, the organization should evaluate external and internal, organization-specific factors that affect the identified risks. External factors include the organization's operating environment or industry; social, regulatory and technological factors; and general perceptions of information and risk. Awareness of and attitudes towards automation, MT and the implications of using raw MT are also important considerations. In some environments, information delivery via raw MT could be viewed negatively, while other target audiences might consider it a sign of innovation. Internal factors include the organization's operational culture and 'risk appetite' (see Nitzke, Hansen-Schirra and Canfora 2019: 241), specifically, its willingness to take risks in information delivery. Such factors can change over time, so analyses should also be updated periodically.

These various factors will inform the criteria for evaluating risks, determining risk likelihood and severity and deciding which risks can be accepted and under what conditions. An important consideration for risk criteria is identifying potential benefits. Being able to deliver more information via MT to more people, especially to those who currently are not offered information in their own languages, can mitigate risks posed by not providing information at all (see O'Brien 2020). The increased reach could also provide a better understanding of what content and languages end-users access the most (Dillinger and Gerber 2009: 11) or increase customer satisfaction and reduce support calls (Thicke 2013: 50). When the output is provided by the organization's own domain-specific MT solution, it can also offer higher-quality information and mitigate the risks posed by stakeholders' use of free online tools (see Canfora and Ottmann 2020). However, potential benefits must be carefully weighed against the risks (Nitzke, Hansen-Schirra and Canfora 2019: 246; Vieira, O'Hagan and O'Sullivan 2020: 12).

Risk identification

One of the most important phases of risk management is risk identification. However, this phase can be challenging because the practice of delivering

content as raw MT is not yet widespread and best practices are lacking. In our viewpoint, the most important principle is to carry out the most comprehensive analysis possible. This can be achieved by systematically analysing potential risks through distinct categories. Besides providing a more comprehensive analysis, the approach also makes it easier to identify specific risks and allows people with diverse expertise to participate in analysing areas that are familiar to them without needing to participate in the entire exercise. The need for a holistic approach to risk analysis has been noted in past research as well. For example, Canfora and Ottmann (2018: 171) point out that translation is a complex activity and that serious issues are not usually caused by an individual translation error but result from a series of failures. We propose that the analysis focus on the categories of technical, content and context of use.

As discussed briefly earlier, we also propose that a diverse group of people should be involved in analysing these four phases, and that carrying out four distinct analyses allows the appropriate parties to participate in the areas they are most familiar with. In an exercise focused on information delivery in different languages, it is logical that people or groups responsible for content creation and translation would be involved in all phases. In addition, the analysis of organizational or business risks could include managers or other people responsible for the organization's strategy, people responsible for risk management, and representatives from marketing or sales. The analysis of risks related to technology might include people who are responsible for language technology, the organization's other technology and external participants such as the provider of the organization's MT solution, language service providers or, as suggested by Martindale (2020), independent consultants.

The analysis of the content could include those responsible for language technology and representatives from internal groups that rely on the organization's content, for example, those who train customers and other stakeholders. It might also include external participants such as language service providers or external consultants. Finally, the phase devoted to identifying risks in the context of use should include people who study or interact with end-users, for example, people who train customers and other stakeholders, people involved in marketing and sales and members of the user interface team.

A key point in contexts where unedited MT is used is that MT output is not error-free, and although truly misleading translations are rare (Martindale and Carpuat 2018: 16; Martindale 2020: 31), MT will produce errors and predicting where they occur and how they impact the output is difficult. Therefore, errors

are not risks but rather risk sources, and actual risk arises when someone relies on faulty information when taking actions and those actions lead to consequences.

The following sections examine possible risk factors related to technology, content and context of use in more detail. Based on prior research, we provide examples that are intended as guidance for identifying risks, not as definitive lists of all risk factors. The risks relevant to each organization are specific to its context, information and target audience and may involve factors differing from the examples given here.

Technological risk factors

Examples of technological risk factors include the quality of the MT output, the language pairs involved and cyber risks. The quality of MT output depends on several factors, such as the text types and languages involved (see, e.g., Barrault et al. 2020 for an evaluation of multiple language pairs). Even after MT technology has been chosen and implemented, it would be important to analyse the quality of the MT for the specific information type and language pair an organization is planning to use for raw MT delivery.

A second issue with the quality of MT output is that it is unpredictable, even within one text: it can be good, it can contain clear and frequent errors that lead to confusing or incomprehensible translations, or it can be fluent and easily understandable but nevertheless inaccurate, particularly in the case of the currently predominant neural MT technology. In particular, 'believable output' that is both fluent and plausible can lead the reader to judge the information as credible (Martindale 2020: 25). Problems can then arise when inaccurate translations are trusted by readers. MT output quality also varies depending on the language pair. In addition to a potentially higher likelihood of risk due to more frequent errors, Vieira, O'Hagan and O'Sullivan (2020: 7) suggest that this variability could lead to the unequal treatment of the target audience, or at least the perception of inequality. Byrne (2007: 7) discusses problems caused by faulty English-to-Chinese human translation in a US election process and the ensuing perception that it was 'proof of the indifference of the federal government to the [Chinese American] community'.

The use of MT also involves cyber risks (e.g. Canfora and Ottmann 2020; Nitzke, Hansen-Schirra and Canfora 2019) which centre around the use of external systems that are not secure and may expose the organization's non-public information to others. This cyber risk is lessened if the organization

employs a secure MT solution and the machine-translated information offered to external parties is already publicly available.

Content risk factors

Another set of risk factors involve the content or information type that an organization is considering delivering as raw MT. The genre, subject matter and content creation processes should be examined.

The first concern is whether the subject matter involves safety-critical information. Information meant to be used in situations where the safety of the end-user or others may be endangered is particularly risky due to the potentially severe consequences (see Canfora and Ottmann 2018; Vieira, O'Hagan and O'Sullivan 2020). As an example, Byrne (2007: 17) discusses a case where a translation error in a product manual led to unsafe behaviour when the product malfunctioned. Safety-critical information types would therefore be unsuitable for delivery via raw MT (Nitzke, Hansen-Schirra and Canfora 2019; Canfora and Ottmann 2020). Other content types less suited for MT would include information with potential financial, legal or political ramifications (see Byrne 2007; Cismas 2010; Vieira, O'Hagan and O'Sullivan 2020).

A second factor concerns whether the genre is conducive to producing high-quality MT output. Although evidence on which genres lead to better MT output is somewhat unclear, creative texts such as literature, marketing and advertising are commonly suggested as poor candidates for MT (see Nitzke, Hansen-Schirra and Canfora 2019 or, for an industry perspective, Densmer 2020). Conversely, content types in which translations are required to follow the source text strictly would be more suitable for MT.

A related factor is the content creation process. In general, content that is created in a standard way is considered to be better suited for MT than content that varies greatly (Nitzke, Hansen-Schirra and Canfora 2019). In contrast, using raw MT to deliver content authored by a variety of people using non-standard methods is riskier. Nitzke, Hansen-Schirra and Canfora (2019: 243) outline a general rule that texts (and processes) suited for translation memories are also good for MT if they do not contain other risks.

Contextual risk factors

To date, little attention has been devoted to the contexts in which MT is used, although research suggests that context influences how people use and receive raw MT (see Nurminen 2021). Relevant contextual factors include the size of the

intended group of end-users; their qualities, such as familiarity with the subject matter and MT; their goals and processes when using raw MT; and factors related to the environment in which MT would likely be used, including access to possible auxiliary tools and materials.

The user's knowledge of the subject matter can positively impact the reception of raw MT (Nurminen 2021), whereas a low level of familiarity can make MT gisting more challenging. For example, Nitzke, Hansen-Schirra and Canfora (2019: 245) suggest that customer service agents could have sufficient background knowledge to understand a question posed via raw MT messages, but the customers who lack the same knowledge might have difficulty understanding the agents' machine-translated replies.

Another factor is the users' MT literacy, meaning their knowledge regarding how MT works, how it can be used and what the implications of using it are (Bowker and Buitrago Ciro 2019). A low level of MT literacy could make the users more prone to trusting MT uncritically, while a better understanding of the technology can help them recognize potential errors. On the other hand, a low level of MT literacy could be associated with a negative perception of MT and cause the users to not accept MT use at all.

Contextual factors like the user's goals are also important. If the user of the raw MT will act on the information, for example, install an electronic device, then defective translations in the user manual can lead to unsafe situations (see Cismas 2010: 493; Byrne 2007: 17). Contextual factors may offer additional considerations for safety-critical content. For example, a text of medium safety risk might be considered less risky if the target audience has a high level of expertise in the subject matter. Caution is required in such considerations, however, because the audience and their level of expertise may not be fully known to the organization (cf. Pym 2021: 451–2). The users' goals also impact how acceptable they find raw MT. For example, users may be more accepting if the purpose is to find practical information about services than if the information somehow impacts the rights of a language minority (see Bowker and Buitrago Ciro 2015).

Environmental factors such as access to other resources also affect the users' means and motivation to verify raw MT output when they encounter an unclear passage or suspected error (see Martindale 2020). The easy availability of auxiliary language resources like glossaries or alternate translation versions – or a colleague who speaks the source language – could help verify unclear passages. Higher risk is involved if the end-users of the raw MT are assumed to work in an environment without such verification aids.

Risk analysis

Each identified risk is then analysed in terms of likelihood, potential consequences and their severity. As noted, the main risk is that someone may use incorrect information delivered via raw MT and that may lead to consequences. To define likelihood levels and types of consequences, the organization could apply a risk matrix like the one outlined by Canfora and Ottmann (2019).

Likelihood of occurrence

The likelihood of a risk (in terms of an incident leading to some consequences) can be categorized using different levels. For example, Canfora and Ottmann (2019: 83) use the likelihood levels of very likely, probable, occasional, remote and improbable. Estimating the likelihood of translation-related risks is challenging because, beyond anecdotal reports, clear and reliable evidence of faulty translations leading to severe consequences are difficult to find (Canfora and Ottmann 2018; Byrne 2007: 3). Canfora and Ottmann (2018: 169) note that this lack of information can be dangerous due to a human tendency to underestimate the likelihood of (negative) incidents and ignore incidents with low likelihood but potentially serious consequences.

As discussed earlier, likelihood of risk may be higher for some texts and information types. Additionally, even within one text, certain parts might carry higher risk than other parts (see Pym 2015). Another factor is the MT output quality. Frequent MT errors increase the likelihood of faulty translations leading to consequences. Conversely, fluent and believable MT output may increase the likelihood that the user places unwarranted trust in the output. Finally, the exposure of the translation needs to be considered, as the size of the audience and number of languages affect the likelihood of an event.

Possible consequences

As in the case of human translations, consequences arising from faulty MT output can involve physical harm like injury and death, legal consequences, property damage, financial damage, embarrassment, damage to reputation or trust and impaired or prevented communication (see Byrne 2007; Cismas 2010; Canfora and Ottmann 2018: 171; Canfora and Ottmann 2019: 81–2; Vieira, O'Hagan and O'Sullivan 2020: 10; Scott and O'Shea 2021: 26–7). Various consequences are also tied to each other in that the same incident can cause more than one consequence, and one consequence can lead to others.

Impaired or prevented communication with the target audience could be caused, on the one hand, by MT output that is difficult to understand or contains clear errors. Conversely, impaired communication can be caused by output that is easy to understand in that it is fluent, but it nevertheless contains errors. Neural machine translation, which is currently the predominant MT technology in use, can produce output that sounds fluent but that contains errors (see Martindale and Carpuat 2018). Both Martindale and Carpuat (2018) and Rossetti, O'Brien and Cadwell (2020) found that fluent MT output engendered trust among MT users. Martindale (2020: 25) identifies 'believable output' as a risk and defines believability as a combination of fluency, plausibility and human judgement of credibility. Problems can arise when texts contain translation errors, yet they are trusted by readers.

When someone uses information that is incorrect due to translation errors, they may incur property damage, financial damage or physical harm, which in turn can result in questions of liability and legal consequences for the organization that provided the information. As noted by several researchers (e.g. Canfora and Ottmann 2020; Vieira, O'Hagan and O'Sullivan 2020) currently legislation is just starting to react to liability questions concerning MT, leading to uncertainty in predicting and planning for potential consequences. Some laws and regulations concerning human translation do exist (for a more detailed discussion, see Byrne 2007), but it is unclear whether or which of these might also apply in cases in which information is translated by a machine.

The final example risk could also be viewed as the consequence of all the risks listed above. Impaired communication, damages to property or body, and questions of liability can cause embarrassment and loss of reputation of the organization. As illustrated by Zaveckaite and Ulbinaite (2018: 1336), although language and translation are often seen as less important support features to the main product or service, they have direct impact on how the user perceives that product or service. If the end-user finds an error in the translation (see Pym 2015: 78), loss of credibility of the translation can potentially also affect the credibility of the organization offering the information in the raw MT scenario (see Vieira, O'Hagan and O'Sullivan 2020). In addition to a loss of reputation due to concrete issues that end-users notice, an organization's reputation might also suffer from a general negative attitude towards MT. For example, Baumgarten and Cornellà-Detrell (2019: 18) discuss a case where translation errors drew negative attention to the use of MT for translating subtitles in Catalonia, leading to backlash from the viewers towards the TV channel. Also, simply the knowledge that MT has been used may lead the users to perceive the content more negatively (Asscher and Glikson 2021).

Severity of consequences

Death and major physical harm are generally viewed as the most severe potential consequences. Martindale (2020: 22) labels the most severe consequences 'dangerous catastrophic failures' while Google designates high-risk content as YMYL or Your Money or Your Life (Google 2022: 26). Canfora and Ottmann (2018: 176) deem fatal or lifelong injuries as most severe, followed by other injuries, and then material or financial damage, legal consequences or damage to reputation. If the same incident leads to multiple consequences, severity should be assessed according to the most severe (see Canfora and Ottmann 2019). However, such severity categories are not straightforward and depend on situational factors. Impaired communication is often considered the lowest in severity, but in the voter information case discussed by Byrne (2007: 6), impaired communication had rather important political ramifications.

How severe the consequences will be is affected by the various factors. As noted previously, consequences can be particularly severe for safety-critical content (see Canfora and Ottmann 2018). Even for a given content type, consequences can vary from negligible to very severe depending on situational factors (see Cismas 2010: 493). According to Martindale (2020), potentially dangerous situations are characterized by 'believable' MT output, lack of means or motivation to verify the output on the part of the users, and use cases where the user will act on the information provided.

Risk evaluation

The risk analysis results are then evaluated against the risk criteria set earlier (see section above on scope, context and criteria). Based on the risk likelihood, potential consequences and their severity, as well as the organization's context and risk appetite, the organization can decide that some risks may be tolerated while others are not. Using a risk matrix (see Canfora and Ottmann 2019: 82–3), the organization might, for example, deem occasional risk with low-impact consequences like impaired communication to be acceptable, but even improbable risks with potentially catastrophic consequences to be unacceptable. Pym (2021: 446) likens such evaluation to distinguishing 'kittens' (risks with low likelihood and lesser impact) from 'tigers' (risks with high likelihood and severe consequences): one can live with kittens, but protective measures are needed with tigers.

Factors for this evaluation again include information type and content as well as contextual factors. As noted, raw MT is better suited to low-risk texts

(Nitzke, Hansen-Schirra and Canfora 2019: 242), and not advisable for safety-critical information, at least until a clearer understanding can be formed of its use and users. Evaluation might additionally include factors like resources and time available for translation and expected life span of the translation (see Nitzke, Hansen-Schirra and Canfora 2019: 244–5). However, decisions should not be based on time or financial aspects alone. Finally, evaluation should also balance the risks of providing raw MT against the risk that no translated information is provided and the risks posed by customers turning to generic online systems.

Risk treatment

Based on the risk evaluation, the organization can decide to avoid the risk, take the risk, remove the risk source, change the likelihood or consequences of risk or share the risk with other stakeholders. When risks are deemed unacceptable, the primary way to avoid risk is to not use raw MT for a specific text type or purpose. Conversely, if the risk level of raw MT is low enough to accept (cf. 'kittens' in Pym 2021: 446; see also Canfora and Ottmann 2019), the organization can decide to take the risk without additional measures. Removing the risk source could mean changing the process so that the machine translation is checked by humans, which differs from the use case in this chapter (see Nitzke, Hansen-Schirra and Canfora 2019 for details).

The organization can also decide that the risk involved in raw MT use is acceptable with some further treatment. The following sub-sections provide examples of measures categorized as changing the likelihood or the consequences of a risk or sharing the risk with other stakeholders. Effective risk management generally requires combining multiple measures at different levels (Canfora and Ottmann 2018: 179–80) and should account for user behaviour, such as the tendency to use default options (see Canfora and Ottmann 2020: 65). Many of the proposed measures could be implemented without a large investment of resources.

Change the likelihood or consequences of the risk

The likelihood of a risk can be reduced by helping users achieve the best possible understanding of raw MT content, and outcomes can be changed by aiming to ensure faulty information will not lead to severe consequences. To accomplish this, users could be encouraged to check an auxiliary source to verify a raw MT text, or in critical situations, even be required to do so

(see Martindale 2020: 38). What follows are examples of treatment options organizations could consider.

Increase MT literacy among users. MT literacy can help raw MT users understand risks, recognize common errors and follow best practices (Bowker and Buitrago Ciro 2019; Nurminen 2019; Vieira, O'Hagan and O'Sullivan 2019; Martindale 2020). Organizations could embed information, tips and hints in the interfaces through which users access raw MT, or even offer training in MT use. Guidelines on using raw MT can also raise MT literacy and promote best practices.

Show the source document. People using MT often translate from languages in which they have some proficiency, and they tend to look at the source text when reading raw MT. This can contribute to their ability to understand raw MT (Nurminen 2019, 2020). Especially if the organization publishes content in a language in which the target audience might have some proficiency (e.g. English), users could benefit from having that language version easily viewable, either side by side with the raw MT output, or by hovering the mouse over the corresponding target passage.[1] Pictures and other multimodal elements of the source text also provide important information when using raw MT (Nurminen 2019, 2020). Some MT solutions translate full documents, showing the original multimodal elements in place (e.g. Microsoft; see Doss Mohan 2021). Alternatively, users should be given other ways to quickly refer to the source document.

Provide easy access to multiple MT outputs. MT users sometimes translate the same text with different tools to improve their understanding of it (Nurminen 2019, 2020; Martindale 2020), and Gao et al. (2015) found that providing MT users with two outputs is an effective way to improve understanding. Access to multiple MT outputs could be offered by embedding a second MT tool directly into the interfaces used to access raw MT (see an example in Nurminen 2019: 37). Another option is automatically generating output with multiple engines and showing all outputs to end-users (Martindale 2020).

Provide access to other language resources. Verifying the correctness (or incorrectness) of specific words or passages can be helpful in MT gisting. The organization can encourage such verification by providing easy access to resources. For example, the organization's own glossaries could be translated (by humans) into the languages in which raw MT is delivered and opened to end-users. Technology could also be used to highlight terms or offer links between source and target versions of terms. Integrating access to freely available resources such as online dictionaries can also enable easy verification of raw MT content (Martindale 2020: 41).

Help users recognize errors. Some technological solutions can flag output with potential errors. For example, MT confidence estimation (or quality estimation) produces an automatically estimated score for the quality of a translated piece of text (see Guerberof 2020: 347). Highlighting passages with low scores can prompt MT users to approach them with caution and verify the information (Martindale 2020: 42).

Offer an alternative path to information. Whenever possible, MT users should be offered a second way to get information in case of MT failure. Commonly, people are offered the option of contacting support services, and this should continue to be an option even after information is delivered in users' own languages via MT.

Share the risk

The organization can also decide to share the risk with other stakeholders, most importantly, with the people reading the raw MT content. Sharing the risk means ensuring that the end-user is aware of the risk and obtaining confirmation of that awareness for reasons of liability. The following are examples of treatment options organizations could consider.

Make MT transparent. People should always be aware that they are reading raw MT output. If users request MT themselves, they are naturally aware of it, but if they are offered pre-translated content, it may not be obvious to them. Therefore, clear labels and disclaimers are needed. People might also save machine-translated documents for future use or share them with others (Nurminen 2020). Therefore, raw MT should be properly labelled as such. Disclaimers, explanations and warnings placed in critical points of the end-user's process of accessing raw MT are also important for increasing MT literacy among users (see previous section).

Confirm that end-users are aware of risks. Asking end-users to confirm that they understand the risks involved in using raw MT might be appropriate in addition to warnings and disclaimers. For example, users might be required to check a box confirming their understanding before proceeding to raw machine-translated content.

Change the external environment. Sharing risks could also mean joint risk management initiatives with other organizations that deliver content as raw MT. For example, organizations in an industry might join forces to implement industry-wide guidelines and best practices for raw MT publishing, develop MT literacy programmes, sponsor sessions on MT gisting in conferences and user group meetings, or even introduce legislation if needed (Canfora and Ottmann 2020).

Monitoring and review, recording and reporting

An important part of the risk management framework in ISO 31000 is the ongoing monitoring, review and reporting of risks and risk treatment. Explicit and transparent communication about risks is important (see Zaveckaite and Ulbinaite 2018: 1338) for monitoring. People within the organization should be encouraged to report any MT-related problems or errors they notice themselves or encounter in customer comments. For human translation, translators themselves are in key positions to report translation errors, including near-misses that were caught during revision (see Canfora and Ottmann 2018: 178), but no such revision stage exists for raw MT use. For this reason, it is vital to proactively collect feedback and experiences from the end-users of the organization's machine-translated texts, for example, in the form of surveys. The organization should also implement easy procedures for end-users to report MT errors and encourage them to use this option regardless of whether major consequences occurred.

It is also important to establish clear procedures for identifying incidents involving raw MT use. This will provide the organization valuable information on risks and consequences, and will facilitate the assessment of the effectiveness of risk treatment. Canfora and Ottmann (2018) propose a process for monitoring and analysing incidents related to human translation, and a similar process could be applied to the raw MT situation. In addition to major incidents where more or less serious consequences have occurred, Canfora and Ottmann (2018: 181) emphasize the importance of analysing 'near-miss' incidents where a translation error could have caused a potentially hazardous situation even if no consequences occurred. For example, a customer might report on a translation mistake they encountered in the raw MT that would have led to severe consequences if not noticed. Observing and analysing such cases is important because near-misses and major incidents share the same root causes, and addressing the causes helps reduce the likelihood of more serious events (Canfora and Ottmann 2018: 172–3).

Risk management in action: The case of patent professionals

To illustrate the principles of risk management, this section examines the use of raw MT for information in a real-life use case involving patent professionals in the intellectual property rights (IPR) field. This group has actively relied on

raw MT to understand patent documents needed in their work for more than a decade (Nurminen 2019: 32). The background for our discussion of risk management is formed by a study conducted in 2018–19 by Nurminen, who interviewed nine Scandinavian patent professionals about how raw MT is used in their working environments (for more detailed results, see Nurminen 2019, 2020; 2021).

In an IPR process such as patenting an invention or defending a patent, one of the most important tasks of patent professionals is to review all potentially relevant patent documentation. This review can involve hundreds of documents in various languages. Documents written in languages the patent professional does not speak are read as raw MT, and the goal is to understand each document well enough to evaluate whether it is relevant to the IPR case or not. If raw MT is not sufficient for making this decision, patent professionals can request a human translation. However, time and costs limit the number of documents that can be translated by humans. On the other hand, overlooking a relevant document can lead to negative consequences and is considered one of the main risks patent professionals face. These conflicting needs and risks are weighed against each other when deciding whether to rely on raw MT or choose human translation.

Nurminen (2019) concluded that this decision-making was a process of risk assessment. Although the process was described as more informal and manifested through patent professionals' own thought processes and discussions with their colleagues, the way risks are evaluated and weighed against benefits nevertheless reflects more formalized risk management principles. In what follows, we examine the use of raw MT by patent professionals through the lens of the risk management principles presented in this chapter.

To begin with, the patent professionals displayed a good knowledge of the scope and context in which they evaluated risk in decisions concerning translation. They discussed their own organizations' views on risk and were aware of the criteria that determined if risks were acceptable. They also displayed a familiarity with the external context of IPR work and its various stakeholders and processes.

When faced with the need to understand a potentially relevant patent document written in a foreign language, patent professionals described identifying the risks of using raw MT. For example, they pinpointed risk factors related to the IPR process in which the document would be used and risks related to their own initial understanding of the machine-translated version. They then analysed the identified risks, considering likelihood and potential consequences, as in this example that was voiced by informant #4 in Nurminen (2019: 36):

> If we make the wrong decision and allow a product to the market which does not have freedom to operate, there is a risk of using time and money and goodwill in a court case and potentially being responsible to cover the damages of a client.

Finally, they evaluated the identified risks against their organizations' risk criteria and decided whether to rely on the machine-translated document or to incur further time and cost by ordering human translations.

If the treatment decision was to take the risk by relying on raw MT, patent professionals might continue to try and achieve a better understanding of the raw MT document in question, for example, through discussions with technical experts or accessing further material such as multimodal components of the original patent document (see Nurminen 2019: 37). In the IPR context, the risk of using raw MT is always shared through full transparency. All stakeholders are kept fully aware of the general use of raw MT and of which texts are authored by humans and which are raw MT.

Conclusion

This chapter set out to examine risks and risk management in situations where an organization intends to deliver information to end-users as raw MT. Such practices are still relatively new but will likely increase as technology advances and people adjust to new ways of communicating and receiving information. Although such practices offer the benefit of making more information available in more languages to more people, those benefits must nevertheless be weighed against risks. As illustrated in this chapter, applying the risk management principles and framework defined in ISO 31000:2018 can help organizations identify and analyse potential risks, evaluate whether the risks are acceptable, and determine how to treat the risks. Examples of risk factors, analysis and evaluation criteria, and proposed treatment options are provided in various sections of the chapter and illustrated with the real-life use case of MT gisting by patent professionals. Nevertheless, we emphasize that risk management requires analysis of the organization's specific operating context and active communication with a diverse group of stakeholders.

Although risk management can offer a lens through which translation processes and actions can be analysed, as suggested by Pym (2021: 447) and Nurminen (2021), risk management in the specific context of content delivery using raw MT has to date gone unexplored. This chapter contributes a new viewpoint to considerations of risk in various types of translation and contributes

to practical risk management planning and decision-making in organizations that use raw MT to deliver content or are considering doing so.

The nascent nature of research on MT gisting and its inherent risks means that our approach focuses more on initiating a conversation than offering final solutions, and we look forward to further discussions and contributions on the topic. Future research could concentrate on examining specific areas of risk management in greater detail. Additionally, empirical investigations into how risk is conceptualized in actual environments in which content is delivered as raw MT would be welcome.

Note

1 For example, the European Patent Office's MT tool, Patent Translate, has this function.

References

Asscher, O. and E. Glikson (2021), 'Human Evaluations of Machine Translation in an Ethically Charged Situation', *New Media and Society*, 25 (5): 1087–107, https://doi.org/10.1177/14614448211018833.

Barrault, L., M. Biesialska, O. Bojar, M. R Costa-Jussà, C. Federmann, Y. Graham, R. Grundkiewicz, B. Haddow, M. Huck, E. Joanis, T. Kocmi, P. P. Koehn, C. Lo, N. Ljubešić, C. Monz, M. Morishita, M. Nagata, T. Nakazawa, S. Pal, M. Post, and M. Zampieri (2020), 'Findings of the 2020 Conference on Machine Translation (WMT20)', *Proceedings of the Fifth Conference on Machine Translation*, 1–55. Available online: https://www.aclweb.org/anthology/2020.wmt-1.1.

Baumgarten, S. and J. Cornellà-Detrell (2019), 'Translation and the Economies of Power', in S. Baumgarten and J. Cornellà-Detrell (eds), *Translation and the Global Spaces of Power*, 11–26, Bristol: Multilingual Matters.

Bowker, L. and J. Buitrago Ciro (2015), 'Investigating the Usefulness of Machine Translation for Newcomers at the Public Library', *Translation and Interpreting Studies*, 10 (2): 165–86, https://doi.org/10.1075/tis.10.2.01bow.

Bowker, L. and J. Buitrago Ciro (2019), *Machine Translation and Global Research: Towards Improved Machine Translation Literacy in the Scholarly Community*, Bingley: Emerald Publishing.

Byrne, J. (2007), 'Caveat Translator: Understanding the Legal Consequences of Errors in Professional Translation', *The Journal of Specialised Translation*, 29 (7): 2–24, https://doi.org/10.1037/023079.

Canfora, C. and A. Ottmann (2018), 'Of Ostriches, Pyramids, and Swiss Cheese: Risks in Safety-Critical Translations', *Translation Spaces*, 7 (2): 167–201.

Canfora, C. and A. Ottmann (2019), 'Who's Afraid of Translation Risks?', in H. E. Jüngs, L. Link, K. Schubert, and C. Zehrer (eds), *Challenging Boundaries: New Approaches to Specialized Communication*, 73–92, Berlin: Frank & Timme GmbH.

Canfora, C. and A. Ottmann (2020), 'Risks in Neural Machine Translation', *Translation Spaces*, 9 (1): 58–77, https://doi.org/10.1075/ts.00021.can.

Cismas, S. C. (2010), 'Risk Assessment and Mitigation in Drafting Translation Contracts', *Proceedings of the International Conference on Risk Management, Assessment and Mitigation*, 491–5.

Densmer, L. (2020), 'Machine Translation and the Scale of Emotional Weight', *RWS*, 4 February. Available online: https://www.rws.com/blog/machine-translation-and-the-scale-of-emotional-weight/

Dillinger, M. and L. Gerber (2009), 'Success with Machine Translation: Automating Knowledge-Base Translation, Part 2', *Client Side News*, February: 9–11.

Doss Mohan, K. (2021), 'Translator Announces Document Translation (Preview)', *Microsoft*, 17 February. Available online: https://techcommunity.microsoft.com/t5/ai-cognitive-services-blog/translator-announces-document-translation-preview/ba-p/2144185 (accessed 23 January 2023).

Dunne, E. S. (2013), *Project Risk Management: Developing a Risk Framework for Translation Projects*, Doctoral dissertation, Kent State University. Available online: https://etd.ohiolink.edu/apexprod/rws_etd/send_file/send?accession=kent1368700402&disposition=inline (accessed 23 January 2023).

Gao, G., B. Xu, D. Hau, Z. Yao, D. Cosley, and S. R. Fussell (2015), 'Two is Better than One: Improving Multilingual Collaboration by Giving Two Machine Translation Outputs', *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 852–63. Available online: https://dl.acm.org/doi/10.1145/2675133.2675197

Google (2022), 'Search Quality Rater Guidelines: An Overview'. Available online: https://services.google.com/fh/files/misc/hsw-sqrg.pdf

Guerberof-Arenas, A. (2020), 'Pre-Editing and Post-Editing', in E. Angelone, M. Ehrensberger-Dow, and G. Massey (eds), *The Bloomsbury Companion to Translation Industry Studies*, 333–60, London: Bloomsbury.

Guerberof-Arenas, A. and J. Moorkens (2023), 'Ethics and Machine Translation: The End User Perspective', in H. Moniz and C. Parra Escartín (eds), *Towards Responsible Machine Translation: Ethical and Legal Considerations in Machine Translation*, 113–33, Cham: Springer.

Haddow, B., A. Birch, and K. Heafield (2021), 'Machine Translation in Healthcare', in Ş. Susam-Saraeva and E. Spišiaková (eds), *The Routledge Handbook of Translation and Health*, 1st edn, 108–29, Abingdon, Oxon; New York: Routledge.

ISO (2018), *ISO 31000 Risk Management*, Guidelines, Geneva: International Organization for Standardization.

Martindale, M. J. (2020), 'Responsible Gist MT Use in the Age of Neural MT', in *Proceedings of the 14th Conference of the Association for Machine Translation in the Americas, Workshop on the Impact of Machine Translation*, 18–44. Available online: https://www.aclweb.org/anthology/2020.amta-impact.2/

Martindale, M. J. and M. Carpuat (2018), 'Fluency Over Adequacy: A Pilot Study in Measuring User Trust in Imperfect MT', *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Papers)*, 13–25.

Nitzke, J., S. Hansen-Schirra, and C. Canfora (2019), 'Risk Management and Post-Editing Competence', *Journal of Specialised Translation*, 31: 239–59.

Nurminen, M. (2019), 'Decision-making, Risk, and Gist Machine Translation in the Work of Patent Professionals', *Proceedings of the 8th Workshop on Patent and Scientific Literature Translation*, 32–42. Available online: https://www.aclweb.org/anthology/W19-7204/

Nurminen, M. (2020), 'Raw Machine Translation use by Patent Professionals: A Case of Distributed Cognition', *Translation, Cognition & Behavior*, 3 (1): 100–21, https://doi.org/10.1075/ts.00025.nur

Nurminen, M. (2021), *Raw Machine Translation in Context: How Context Influences the Use and Reception of Raw Machine Translation*, Doctoral dissertation, Tampere University, Finland.

Nurminen, M. and M. Koponen (2020), 'Machine Translation and Fair Access to Information', *Translation Spaces*, 9 (1): 150–69.

O'Brien, S. (2020), 'Translation Technology and Disaster Management', in M. O'Hagan (ed.), *The Routledge Handbook of Translation and Technology*, 304–18, Abingdon: Routledge.

Pym, A. (2015), 'Translating as Risk Management', *Journal of Pragmatics*, 85: 67–80, https://doi.org/10.1016/j.pragma.2015.06.010.

Pym, A. (2021), 'Translation, Risk Management and Cognition', in F. Alves and A. Jakobson, *The Routledge Handbook of Translation and Cognition*, 445–58, Abingdon: Routledge.

Rossetti, A., S. O'Brien and P. Cadwell (2020), 'Comprehension and Trust in Crises: Investigating the Impact of Machine Translation and Post-Editing', *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, 2: 9–18.

Scott, J. and J. O'Shea (2021), 'How Legal Documents Translated Outside Institutions Affect Lives, Businesses and the Economy', *International Journal for the Semiotics of Law*, 1–43, https://doi.org/10.1007/s11196-020-09815-5

Thicke, L. (2013), 'Revolutionizing Customer Support through MT', *MultiLingual*, 49–52.

Vieira, L. N., M. O'Hagan, and C. O'Sullivan (2020), 'Understanding the Societal Impacts of Machine Translation: A Critical Review of the Literature on Medical and Legal Use Cases', *Information, Communication & Society*, 24 (11): 1515–32, https://doi.org/10.1080/1369118X.2020.1776370

Way, A. (2013), 'Traditional and Emerging Use-Cases for Machine Translation', *Proceedings of Translating and the Computer*, 35, London: Aslib.

Zaveckaite, A. and A. Ulbinaite (2018), 'Assessment Criteria of Project Risk Management in Language Translation Service Companies', *Technological and Economic Development of Economy*, 24 (4): 1323–43, https://doi.org/10.3846/20294913.2017.1295287

6

Machine Translation in the Legal Context

A Spanish-to-English Comparative Product Study of Statistical vs. Neural MT Output

Jeffrey Killman

Introduction

Machine translation (MT) is increasingly common in translation workflows and has, to a certain extent, shed some of the negative connotations that it originally had (Koponen 2016). The explosion of content has required language service providers and users to investigate and embrace new approaches to translation technology use, even in contexts where it previously would have seemed out of place. While statistical MT (SMT) and rule-based MT (RBMT) have been effectively implemented in some environments, the advent of neural MT (NMT) represents a shift in the translation technology landscape. Though NMT has shown improvements according to automatic metrics or public benchmarks (e.g. Bahdanau, Cho and Bengio 2015; Bojar et al. 2016; Sennrich, Haddow and Birch 2016; Toral and Sánchez-Cartagena 2017) or has been quickly adopted in commercial deployments (Crego et al. 2016; Junczys-Dowmunt, Dwojak and Hoang 2016; Wu et al. 2016), results are not as straightforward in more specialized domains (Koehn and Knowles 2017). And while studies employing human evaluations have also shown NMT improvements (e.g. Bentivogli et al. 2016; Bojar et al. 2016; Castilho et al. 2017a, 2017b, 2018; Toral and Sánchez-Cartagena 2017; Popović 2018; Van Brussel, Tezcan and Macken 2018; Stasimioti et al. 2020), coverage of specialized domains has not specifically included legal content. In any event, NMT improvements have not been entirely consistent, and additional work should continue to compare results in more specialized areas of translation.

This study presents a comparative evaluation of NMT and SMT output in the context of Spanish-to-English legal translation, focusing on results at the terminological level. Data are drawn from Google Translate (GT) at two points in time: in 2013 when it was a phrase-based statistical system and in 2019 after the provider had transitioned at the end of 2016 to a neural-net-based system. The English output in both cases was generated for a text of civil law judgement summaries (12,000+ words) produced by the Supreme Court of Spain. In a previous study (Killman 2014), accuracy of the SMT output was measured for a sample of characteristic terms and phrases in the text (600+ items) that had been identified as challenging. Results of the previous study indicated that in well over half of the cases, the SMT renditions were adequate, though in the majority of inadequately rendered cases context needed to be considered to a greater extent. That is, terminology presented aspects of lexical ambiguity or translation solutions needed to be written in a contextually variable way. This study asks whether the neural iteration of GT can maintain a level of legal terminological accuracy similar to that of its statistical predecessor, especially in cases where context is crucial. According to Alcaraz and Hughes (2002: 16), '[p]robably the greatest single difficulty encountered initially by legal translators is the unfamiliarity of the vocabulary characteristic of this type of discourse.' With translators now increasingly incorporating MT into their workflows, it is especially important to assess the ability of these technological resources to produce legal terminological accuracy. Concept designations having various legal and non-legal meanings or different levels of potential lexical ambiguity may be especially challenging for natural language processing, as well as prioritization of certain translation renditions over others when the meanings or functions they convey may be drafted in different ways depending on the situation.

To situate this study, this chapter first describes important aspects of SMT and NMT, as well as related MT studies with SMT and NMT, general-purpose online MT and legal text. A description of the criteria used to classify and characterize the terminological sample under study is presented before the methodology and results. Finally, the chapter wraps up with some conclusions and possible future avenues.

SMT, NMT and related research

To better understand the comparative nature of this study, it is important to set out certain aspects of the underlying architecture of SMT and NMT, both of which

are corpus-based approaches. As such, they draw on large bilingual corpora where a significant amount of content comes from international organizations whose documents often concern legal matters, including the European Union and the United Nations (e.g. Koehn 2005; Koehn 2010, 53; Crego et al. 2016; Junczys-Dowmunt, Dwojak and Hoang 2016; Koehn and Knowles 2017). The corpora have been segmented into hundreds of thousands or even millions of sentence pairs, and in the case of both types of system 'a target sentence is a translation of a source sentence with a certain probability of likelihood' (Forcada 2017: 300).

Nevertheless, the computational approaches that each type of MT system applies to determine this probability differ. Neural MT relies on artificial neural networks which may represent knowledge in multidimensional or 'deep' ways in the form of neural networks 'composed of thousands of artificial units that resemble neurons in that their output or activation [. . .] depends on the stimuli they receive from other neurons and the strength of the connections along which these stimuli are passed' (Forcada 2017: 292). Activations of different sets of neural units are combined in layers comprising hundreds of neural units, while the units in one layer are connected by weights with the units in the following layer so that connections range in the thousands (Forcada 2017: 295). Activation states 'are trained to build distributed representations of words and their contexts, both in the context of the source sentence being processed and in the context of the target sentence being produced' (Forcada 2017: 293–4). In the case of SMT, however, translations are built by chunking source sentences into subsegments that may be multiword depending on the availability of reliable statistics. Statistical systems may also run cross-checks against corpora in the target language to further determine whether multiword segment translations are desirably or naturally worded. As Forcada (2017: 301) aptly points out, in neural MT, unlike in statistical MT, 'the identification of subsegments and their translations is not straightforward: the raw translation is produced word by word taking the whole source segment into account'. In other words, neural MT attempts a more holistic approach to context, whereas statistical MT builds translations in a piecemeal fashion based on subsegments in the sentence being translated.

Neural MT has often been described as making gains in sentence fluency (Bentivogli et al. 2016; Bojar et al. 2016; Castilho et al. 2017b; Forcada 2017; Toral and Sánchez-Cartagena 2017; Moorkens 2018; Van Brussel, Tezcan and Macken 2018; Stasimioti et al. 2020), sometimes at the expense of semantic accuracy (Castilho et al. 2017a), especially when it comes to rare words or terminology

(Sennrich, Haddow and Birch 2016; Wu et al. 2016). Using bilingual evaluation understudy (BLEU) (Papineni et al. 2002), a popular automatic metric for evaluating the quality of machine-translated text, Koehn and Knowles (2017: 31), however, find that NMT BLEU scores are better than SMT BLEU scores in German-English when it comes to highly infrequent words thanks to byte-pair encoding, though they find that both systems still have difficulty translating certain infrequent words. In any event, studies also reveal NMT gains in semantic accuracy to varying extents (Bentivogli et al. 2016; Castilho et al. 2017b; Van Brussel, Tezcan and Macken 2018; Stasimioti et al. 2020) if we interpret the reduction in 'lexical' errors in Bentivogli et al. (2016) as contributing to semantic faithfulness to the source text or if we consider the improvements in 'adequacy' reported in Castilho et al. (2017b), Van Brussel, Tezcan and Macken (2018) and Stasimioti et al. (2020). Given the results of the latter two of these comparative studies (Van Brussel, Tezcan and Macken 2018; Stasimioti et al. 2020) and the fact they involve SMT and NMT output from GT, it is reasonable to expect that NMT would demonstrate an improvement in additional studies with the same language pairs or others, including English-Spanish which has not yet been addressed. Nevertheless, it remains an open question whether the specialized nature of legal terminology, at least in the civil law context, might give rise to similar or different levels of quality when comparing NMT and SMT output from a general-purpose online provider such as GT from Spanish to English.

There are several studies involving legal texts and GT, during both its SMT phase and its NMT phase (Gotti et al. 2008; García 2010, 2011; Şahin and Dungan 2014; Wiesmann 2019). In Şahin and Dungan (2014: 77) quality was a bit higher when the participants in their study translated legal texts from English to Turkish using just internet resources, as opposed to post-editing GT output. In García (2011: 226), however, post-editing resulted in a slight quality gain in a similar test with English-Chinese translation, but post-editing slowed participants down slightly (2011: 223). Gotti et al. (2008), for their part, compare GT and a system they designed, TransLI (Translation of Legal Information), to assist federal courts in Canada with their English-French translation needs. According to various automatic metrics, TransLI outperformed GT in the translation of judgements thanks to its being trained on corpora originating from the same courts. In the German-Italian language pair, Wiesmann (2019) also compares GT and different systems in the legal context: DeepL, a more recent online general-purpose NMT provider, and MateCat, a translator workbench drawing on a combination of DeepL, GT (already neural) and Microsoft Translator (then still a statistical system). According to Wiesmann's assessment, DeepL generally

performed better. In turn, Heiss and Soffritti (2018) and Dik (2020) regard DeepL as having potential in the legal context, though they do not compare this system with any others in their respective English-Italian and English-Dutch studies. Roiss (2021), who is not optimistic about DeepL in her German-Spanish evaluation in the legal domain (2021), also does not compare this system with any others.

Outside of the context of free online systems, Koehn and Knowles (2017) compare outputs from different SMT and NMT systems that they train using corpora from five different domains, including the legal domain (Acquis), recognizing that 'a crucial step in developing machine translation systems targeted at a specific use case is domain adaptation' (Koehn and Knowles 2017: 29). The language pair tested is German-English, and while they find similar BLEU scores for in-domain SMT and NMT systems (i.e. scores from systems trained on data from which the test sets were subsampled; these subsampled tests were excluded from the training of the systems for test validity purposes), SMT was superior in the legal, medical and Quran domains, while NMT was better for IT and subtitles. In terms of out-of-domain performance (i.e. results obtained using test sets from data on which the system was not trained), the NMT systems were 'worse in almost all cases, sometimes dramatically so' (Koehn and Knowles 2017: 29). In the case of the individual NMT and SMT systems trained on all five corpora, SMT performed better in the cases of law and IT alone, NMT performed better in the medical and subtitling cases, and both systems were equal in the case of the Quran. It is interesting to note that the legal BLEU score of the SMT system trained on all five corpora was somewhat higher than that of the legal in-domain trained NMT system. Given the legal domain results in these tests, similar English-Spanish results in the case of GT might be possible, especially since it is trained with corpora from a variety of domains. Moreover, Spanish is part of GT's first stage of supported languages that could be translated to and from English and, in the case of GT's neural transition, it is part of the first eight languages for which the neural system was enabled in 2016. That Spanish has been supported with training and resources since the beginning likely renders it a language in which results might be higher than in less supported or lower-resource languages. Wu et al. (2016), for instance, compare errors produced by GT's neural and statistical systems in the context of Wikipedia and news website sentences translated to English from Chinese, French and Spanish and from English to these languages. Though error reductions in all these language pairs and directions were noted in the case of neural GT, English-to-Spanish error reductions were the most substantial

overall, and Spanish experienced the greatest reduction in the to-English results (Wu et al. 2016: 19).

The present study seeks to respond to the area of specialized translation in the legal domain for which NMT/SMT comparative studies are limited in general (Koehn and Knowles 2017; Wiesmann 2019), if not non-existent in English-Spanish. Moreover, the study focuses on offering a human comparison of SMT and NMT output at the terminological level, taking into consideration the fact that terminology is often regarded the most challenging aspect of legal translation, as illustrated later by the various authentic examples from a translation task that required extensive research. Results from MT comparative studies in general have varied to different extents, with NMT showing some advantages with increasing consistency in less specialized or other domains. According to the results from Koehn and Knowles (2017), however, SMT in the legal domain fared better. Wiesmann (2019) rated DeepL more favourably than MateCat in her study with legal texts, but it is not possible to determine if or to what extent her assessment was affected by each of the three systems from which MateCat drew output in her study: GT, DeepL or Microsoft Translator, the only SMT system in her study. For this reason, it is not clear whether Microsoft Translator contributed to or detracted from quality. Based on the results from these two studies and other NMT/SMT comparative studies reviewed, clear consensus has yet to be established, and it is hoped that the current study might help shed some necessary light on what might be considered the most important determinant of legal translation quality: terminology. This area is urgent in that it is often subject to a number of potentially complex contextual parameters. Nevertheless, NMT attempts a more holistic approach to text whether systems combine encoder-decoder architectures (Sutskever, Vinyals and Le 2014) with attention models (Bahdanau, Cho and Bengio 2015) or rely on Transformer architecture (Vaswani et al. 2017).

Legal terminology: Classification criteria and context

The terminological sample in this study comprises 621 source text terms and phrases that warranted research when the author was involved in the translation of the civil law judgement summaries more than a decade ago, a translation that was completed without MT. These 621 terms and phrases can be characterized as symbolic items in all but seven instances; in other words, 614 of them are 'things or ideas found in the world of reality, physical or mental'. The remaining

seven instances can be classified as functional items, or 'grammatical words or phrases that have no direct referents either in reality or in the universe of concepts, but which serve to bind together and order those that do' (Alcaraz and Hughes 2002: 16). The latter appear as conjunctions and prepositional phrases of varying complexity that are either frequent in legal texts or somewhat peculiar to them, such as *al régimen de* (under), *en el que se discute* (concerning) or *ex artículo* (referred to in article).

The 614 symbolic items can be classified as terms in 421 cases and as phrases in the other 193 cases, following the criteria that a term is a 'designation of a defined concept in a special language by a linguistic expression' and a phrase 'combines more than one concept in a lexicalized fashion to express complex situations' (TTT.org 2001). While a term may consist of one or more words, '[t]he distinguishing characteristic of a term is that it is assigned to a single concept, as opposed to a phrase' (Alcaraz and Hughes 2002: 16). A phrase example is *iniciar actuaciones judiciales contra* (open proceedings against), whereas *actuaciones judiciales* (proceedings) is a term forming part of this phrase.

These 614 items were considered 'legal' according to the following criteria. In 575 cases they could be classified, according to Alcaraz and Hughes (2002), as technical terminology (331), semi-technical terminology (126), or as everyday terminology frequently found in legal texts (118). The remaining thirty-nine cases fell under the label of official terminology, which includes the names of specific laws, conventions, titles of legal professions or documents. The technical, semi-technical and official cases (496) were all related to specific areas of law: that is, procedural (148), civil (144), commercial (123), constitutional (18), family (12), criminal (10), tax (8), European Union (7), international (7), administrative (6), inheritance (5), insurance (5), employment (2) and United Nations (1).

A final parameter used to further categorize the 621 total items was context. Recognizing that context has not been easily or consistently defined in translation studies (e.g. Alcaraz 1996; Baker 2006; Hatim 2009; Melby and Foster 2010), this study sees context as operating in different degrees of source-text lexical ambiguity and in terms of specific formulation patterns or drafting in the target texts. In 371 cases (60 per cent), the items were considered contextually sensitive. Legal terminology may be especially prone to different forms of ambiguity (e.g. Alcaraz and Hughes 2002; Chromá 2011; Duro Moreno 2012; Glanert 2014; Prieto Ramos 2014; Simonneaes 2016). Legal phraseology, for its part, typically follows certain patterns and may especially depend on co-text, as well as extratextual context (Kjaer 1990; Vanallemeersch and Kockaert 2010).

There are two categories in which all the items exemplify context sensitivity in these ways: semi-technical terminology (126) and functional phrases (7). Semi-technical items have at least one meaning in non-legal or everyday contexts and another meaning in a legal domain. For example, *sujeto pasivo* in tax law is a 'taxable person', but in the everyday sense the term could be a 'passive subject'. The seven functional items feature co-occurrents that must be taken as a whole and are non-compositional (see functional examples in the first paragraph of this section). That is, their co-text may act more as a single unit rather than as the sum of its parts.

The composition of contextually sensitive terminology in the remaining categories varies. The 331 technical items are contextually sensitive in 164 cases (almost 50 per cent), as single-word items may have more than one legal meaning or multiword terms or phrases may contain words with more than one meaning, for example, *competencia desleal* (unfair competition) vs. *competencia judicial* (jurisdiction). The 118 everyday items, 'which are easier to understand than to translate, precisely because they tend to be contextually bound' (Alcaraz and Hughes 2002, 162), are contextually sensitive in 51 cases (43 per cent) because of polysemy, non-compositional co-text and because they may need to be worded a certain way. *Situaciones consolidadas* is translated more acceptably as 'previous situations' than 'consolidated situations', for example. Finally, there are thirty-nine official terminological cases, of which twenty-three items are deemed contextually sensitive for the same reasons as the technical and everyday items. Cases involving EU or UN law, for example, require specific equivalents in the target language. It is worth noting that context is not only a challenge in legal translation practice but also remains a continuous MT obstacle (e.g. Bar-Hillel 1960; Arnold 2003; Forcada 2010; Killman 2015; Koehn 2020). Given the potential challenges of translating legal terminology both for humans and natural language processing, this study specifically assesses the contextual sensitivity of the terminological sample and the NMT and SMT output in each of 621 total cases.

Methodology

GT is the focus of the chapter given its ubiquity both online and in many commercial translation memory tools. It has also been selected for continuity purposes as it allows different MT iterations from a same provider to be tested. As Koehn and Knowles (2017) report, SMT may outperform NMT when trained

with corpora from one or more domains. This study tests if this might still hold true in the case of an online open-domain system that is widely available to many users. GT output was collected for the 621 terms and phrases during 2 separate periods: between 3 June and 2 August 2013, when GT was a SMT system, and on 10 January 2019, by which point GT had been an NMT system for a little over 2 years. On both occasions, GT was fed the entire source text of each judgement summary to ensure all surrounding co-text was available given its potential relevance to terminological items. The data were organized into tables where the source judgement summary was aligned with the outputs for analysis. Had the entire collection of judgement summaries been fed to the system at once, the context would not have been altered. Each judgement summary operates as its own independent text, and there are no cross references to other summaries or direct co-textual interrelatedness between them. Going one judgement summary at a time facilitated the organization of the data during both periods and is why during the 2013 period, data collection spanned a couple of months while the author simultaneously collected, analysed and annotated data. In 2019, the entire output was collected on one day and analysed and annotated subsequently.

The accuracy of the output provided by the 2 systems across the 621 terms and phrases was then assessed. These sample items were chosen because they were considered challenging and warranted research in order to be translated effectively when the judgement summaries had been previously (human) translated. Extensive use was made of specific, high-quality translation resources such as the IATE termbase, the Eur-Lex multilingual document repository, or lexicographical sources such as Alcaraz and Hughes' *Diccionario de Términos Jurídicos: Inglés-Español, Spanish-English* (2005) to assess possible translation equivalents. The solutions arrived at through this research served as a point of departure to help determine the accuracy of different MT solutions and to carry out additional research where necessary.

For a terminological or phraseological translation rendition to be deemed accurate, it had to be so in terms of both fluency (grammatical correctness and idiomaticity) and adequacy (semantic equivalence or appropriateness), or what has been referred to as 'a harsh correctness standard' (Koehn 2010: 218). Inaccurate renditions were considered either influent or inadequate in some way, or both. While such a standard may potentially lack granularity, and although multiple-range scoring could potentially increase results validity, disambiguating the extent to which grammar mistakes or inappropriate wording may bear on semantics can be challenging, especially in the case of terms and

phrases given their more limited word length and given the potentially delicate legal meanings that terms and phrases can carry in specialized legal contexts. Moreover, simplified scoring can help avoid the leniency variability that is more likely to occur in multiple-range scoring.

Terms and phrases appearing more than once in the text were counted once in the study. In cases where there were both accurate and inaccurate renditions for these items, the item was deemed as being translated accurately. This measure helped expedite data collection by allowing the researcher to focus on a wide selection of terms of phrases rather than focusing on a smaller selection of highly repetitive items. Repetitive items translated accurately or inaccurately in all instances were counted just once as either accurate or inaccurate depending on the case. Of course, inconsistent MT output with respect to repetitive terms and phrases represents a different type of user concern; however, analysis of these inconsistencies would likely amount to enough data worthy of a separate study.

Results

The results have been analysed in terms of general accuracy across the NMT and SMT output sets, as well as with respect to overlaps and differences between the systems. In addition, the results reveal how the two systems performed with contextually sensitive items. It follows that they provide more insight into the key question of whether the recent neural iteration of GT can maintain a similar level of legal terminological accuracy when compared to its previous SMT iteration, especially in cases where context is crucial.

General accuracy and accuracy consistency among the NMT and SMT output sets

In 61 per cent of cases (380 out of 621 terms and phrases), GT's NMT output was accurate, while in 39 per cent of cases (241), it was inaccurate. These results are slightly inferior to the previous SMT GT output sample, which was 64 per cent accurate (398) and 36 per cent inaccurate (223).

In sixty-seven cases NMT was accurate where SMT was inaccurate, and in eighty-five cases SMT was accurate where NMT was inaccurate. There were 313 cases where NMT and SMT were both accurate and 156 cases where they were both inaccurate. Based on these results (see Table 6.1), the 2 GT iterations

Table 6.1 Cases in which both systems were accurate or inaccurate or in which one or the other of the systems was accurate or inaccurate

469 Consistent Cases 76% of sample	Both Systems Accurate 313	Both Systems Inaccurate 156
152 Inconsistent Cases 24% of sample	NMT Accurate/SMT Inaccurate 67	NMT Inaccurate/SMT Accurate 85

performed with consistent accuracy in 76 per cent of the cases (469) and with inconsistent accuracy in 24 per cent of cases (152).

Because NMT and SMT were applied to the same text, the differences in results can be compared statistically using McNemar's test (Mellinger and Hanson 2017). The observed differences are not statistically significant ($^{2}[1] = 2.13$, $p = .144$). While the overall error rate is 24.5 per cent, the proportion of errors for the two methods is quite similar, suggesting that neither approach represents a substantial improvement over the other.

The remainder of this section presents accuracy data for various subsets. Given the lack of significance at the omnibus level, it would be inappropriate to test these subsets quantitatively. Instead, what follows is a description and qualitative examination of the contextual accuracy of NMT and SMT to provide a better understanding of the types of errors.

Accuracy of contextually sensitive items in NMT and SMT output

To determine the extent to which context may have been involved in the accuracy results, the consistent and inconsistent cases in Table 6.1 were checked for contextually sensitive items. Among the 156 items with consistent NMT/SMT inaccuracy, 135 (87 per cent) are contextually sensitive. Of the 313 items with consistent accuracy, 129 (41 per cent) are contextually sensitive. The eighty-five cases in which NMT was inaccurate and SMT was accurate involve sixty-six contextually sensitive items (78 per cent), and of the sixty-seven items where NMT was accurate and SMT was inaccurate, forty-one instances (61 per cent) are contextually sensitive. Especially in cases of consistent inaccuracy or inconsistent accuracy, there are higher concentrations of contextually sensitive items, though more so in the case of the former than in the latter. Something else worth pointing out in this regard is that the eighty-five exclusively accurate SMT cases involve more contextually sensitive items and a higher concentration thereof than do the sixty-seven exclusively accurate NMT cases.

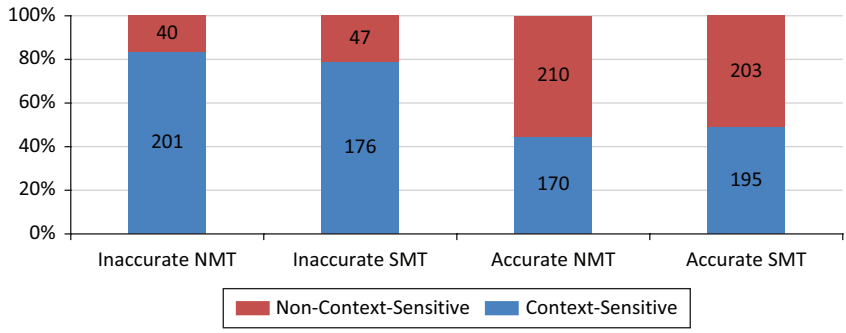

Figure 6.1 Contextually sensitive cases and non-contextually sensitive cases comprising inaccurate/accurate NMT and SMT.

Inaccurate NMT output has the highest concentration of contextually sensitive items at 83 per cent, followed by inaccurate SMT (79 per cent), accurate SMT (49 per cent) and accurate NMT (45 per cent) (see Figure 6.1).

Looking at these results from the standpoint of non-contextually sensitive cases, however, NMT demonstrated a slight accuracy advantage in seven cases, which is about a 3 per cent improvement over SMT in this area.

When observing the output accuracy of the entire sample of 371 contextually sensitive items in this study, NMT output was 46 per cent accurate (170 cases) and 54 per cent inaccurate (201 cases). This is a somewhat noteworthy departure from SMT, which was accurate in 53 per cent of the cases (195) and inaccurate in 47 per cent thereof (176). Simply put, NMT was accurate in a little under half of the contextually sensitive cases, while SMT was accurate in a little over half of such cases. Given these results, NMT appears to struggle more than SMT with contextually sensitive legal terms and phrases.

NMT/SMT output examples

Legal terminology can be particularly illustrative of the feats and failures of artificial intelligence approaches to language translation, given the often-multifaceted contextual nature of legal terms and phrases. As discussed in the previous subsection, the highest concentration of contextually sensitive items could be found in cases where both NMT and SMT were inaccurate (135 out of 156 items, or 87 per cent). It may also be said that these items are especially challenging, seeing that both systems were inaccurate. Interesting examples of such cases can be observed in Table 6.2.

Table 6.2 Contextually sensitive examples where both NMT and SMT were inaccurate

Source Item	Researched Solution	Inaccurate SMT	Inaccurate NMT
legitimación de	standing of	entitlement of	legitimization of
reitera la doctrina sentada en	upholds the case-law laid down in	reiterates the doctrine set forth in	reiterates the doctrine established in
retracto arrendaticio	right of first refusal	withdrawal rentable	retract lease
sociedad de gananciales	joint estate	community property	partnership society
transmisión de dominio	transfer of title	transmission domain	transmission of domain

These illustrative examples reveal how the ambiguity of source text and/or the desirable wording of target text were unattainable in either system. For example, the procedural law phrase '*reitera la doctrina sentada en*' should not be translated according to the most common interpretation of *doctrina* (doctrine), and the verb '*reiterar*' has a preferable collocation in this context (uphold) instead of its more common translation (reiterate).

Table 6.3 Contextually sensitive examples where SMT was accurate and NMT inaccurate

Source Item	Accurate SMT	Inaccurate NMT
daños y perjuicios	damages	damages and damages
efectos patrimoniales	property consequences	patrimonial effects
ejecución de una hipoteca	foreclosure	execution of the mortgage
omisión del deber de seguridad jurídica	failure to legal certainty	omission of the duty to legal security

Table 6.3 provides examples of instances of the second most contextually sensitive category discussed in the previous subsection, where NMT was inaccurate, but SMT was not (sixty-six out of eighty-five items, or 78 per cent). The examples in Table 6.3 display a more fragmented or word-based approach in the case of NMT in contrast to a more holistic or phraseological approach in that of SMT. The first example is a civil law doublet that SMT recognized, but NMT did not in its duplication of 'damages'. *Omisión del deber* is an everyday phrase SMT appropriately rendered by interpreting the phrase as a whole (failure to), while NMT provided a word-for-word unfortunately literal rendition (omission of the duty to).

Table 6.4 includes examples taken from the third most contextually sensitive group of cases, that is, where NMT was accurate and SMT, inaccurate (forty-one out of sixty-seven items, or 61 per cent).

Table 6.4 shows examples of how NMT produced either a more phraseological approach or a broader contextual view than SMT. The civil law term 'actor' highlights a wider contextual lens in the case of NMT in that this system may have taken better account of relevant surrounding phraseology: *reconocimiento del derecho de propiedad del actor sobre una finca reivindicada* (recognition of the plaintiff's right of ownership over the claimed property). Further evidence of NMT's holistic approach to this surrounding phraseology is how it also correctly translated *finca reivindicada*, unlike SMT with 'farm claimed', a more everyday/literal interpretation than NMT's 'claimed property'. The procedural law phrase *lo que obra probado en autos* very much exemplifies a phraseological and contextual advantage offered by NMT. The verbal phrase *obrar en* is particularly context driven and varies according to its immediate co-text surroundings (in this case the participle *probado*). *En autos* is an ambiguous procedural law phrase, as SMT exemplifies with 'cars', which would have been an acceptable translation were *autos* being used as an abbreviated form of *automóviles* (automobiles).

Table 6.4 Contextually sensitive examples where NMT was accurate and SMT inaccurate

Source Item	Accurate NMT	Inaccurate SMT
actor	plaintiff	actor
desestimación de la demanda	dismissal of the claim	estoppel
en el suplico	in the plea	in beg
improcedencia	inadmissibility	irrelevance
lo que obra probado en autos	which is proven in the case file	which works cars tested

Table 6.5 presents examples from the least contextually sensitive group, the category in which both NMT and SMT were accurate (129 out of 313 items, or 41 per cent). The examples highlight where both systems provided contextually appropriate solutions at the level of phrases and terms, taking co-occurrents as a whole and providing solutions that are not contextually misplaced. *En el caso de autos* is the unabbreviated form of *en autos* (noted earlier), and here both systems are able to disambiguate the procedural law phrase and express it appropriately, though in different ways.

Table 6.5 Contextually sensitive examples where NMT and SMT were both inaccurate

Source Item	Accurate NMT	Accurate SMT
administradores de sociedades	directors of companies	directors of companies
con fundamento en que	on the grounds that	on the ground that
culposo	negligent	negligent
el perjudicado	the injured party	the injured party
en el caso de autos	in the present case	in this case

Conclusion

This study conducts a human evaluation of what gains NMT has made in an area of translation known for being especially difficult at the terminological level and one that has received limited attention in the literature. Results indicate that, in many instances, NMT is comparable to SMT in terms of terminological accuracy, with the latter showing a slight accuracy advantage. Furthermore, context continues to play an important role in terminological and phraseological precision, and despite NMT advances, this is an area where NMT appears to struggle more than SMT and that still must continue to be addressed in tool development.

These results coincide on a general level with Koehn and Knowles (2017) in that these authors also found SMT performance slightly superior in the legal domain with multidomain training data. The results from the multidomain system in the current study may be considered as reaffirming. Moreover, the results, which highlight key areas of challenge facing artificial intelligence approaches to language translation, were achieved in a well-resourced language pair not covered by earlier studies. It is not that these findings definitively show SMT is better in the legal domain, but rather that it is not evident that NMT advances have been made in all aspects of legal translation, including that of terminology. The results of this study relate to previous findings of considerable NMT terminological errors in legal texts (Heiss and Soffritti 2018; Wiesmann 2019; Dik 2020; Roiss 2021). Outside of legal translation, the results of this study may very well coincide with previous findings such as inconsistent NMT-SMT comparative adequacy or semantic accuracy (Castilho et al. 2017a) or inadequate NMT renditions of rare words or terminology (Sennrich, Haddow and Birch 2016; and Wu et al. 2016).

As previously stated, this study contributes a human comparative evaluation of SMT and NMT, which has been regarded as an essential element of, and is being increasingly applied in, NMT-SMT comparative studies (e.g. Castilho et al. 2017b; Castilho et al. 2018; Toral et al. 2018; Dowling et al. 2020; Läubli et

al. 2020). Moreover, the sample of challenging terms and phrases was selected from a translator's perspective, that is, somebody actually tasked with translating this collection of judgement summaries in a previous commission without MT assistance, which to the author's knowledge has not been replicated, at least in these comparative studies.

As illustrated in this study, legal terms and phrases may be especially prone to different aspects of contextual sensitivity, in terms of different degrees of lexical ambiguity and in terms of requiring specific formulation patterns or styles in the target text. Contextually sensitive categorizations in this study indicate that inaccurate NMT output had the highest concentration of contextually sensitive terms and phrases and erred more often in this category. While NMT may appear to struggle more with the challenging contextual aspects of legal terminology, perhaps NMT-SMT comparative legal translation studies may reveal specific gains at the sentential level, where improvements have often been documented in general (Bentivogli et al. 2016; Bojar et al. 2016; Castilho et al. 2017b; Forcada 2017, 305; Toral and Sánchez-Cartagena 2017; Moorkens 2018; Van Brussel, Tezcan and Macken 2018; Stasimioti et al. 2020). Improvements at this level would be especially welcome in the legal domain, where a second source of major translation difficulty has been identified in the peculiarity of the morphology and syntax in legal texts (Alcaraz and Hughes 2002: 18). Nevertheless, legal genres such as judgement summaries, rulings or regulations can include particularly long or complex sentence structures. Sentence length is important because past a certain threshold, NMT sentence fluency may be inferior to SMT sentence fluency (Koehn and Knowles 2017), negative correlations with sentence length may be observed in NMT (Toral and Sánchez-Cartagena 2017), or quality may degrade faster with sentence length in the case of NMT (Bentivogli et al. 2016). Whatever the case may be, future studies could assess performance similarities and differences between SMT and NMT in specific areas of morphology and syntax in the judgement summaries in this study or in any other legal text where sentences may be considerably long and complex.

While the results of this study do not indicate progress in the case of GT, more studies could be carried out on MT performance in the legal domain in related or non-related areas to further understand what might be expected from MT systems in the legal context. Studies could include other legal text types, areas of law, language phenomena, language pairs and MT systems (e.g. eTranslation) to help answer this question and determine how MT might be integrated in different legal translation workflows and how translators might be trained to use these tools in legal translation contexts.

References

Alcaraz, E. (1996), 'Translation and Pragmatics', in R. Álvarez and A. Vidal Claramonte (eds), *Translation, Power, Subversion*, 99–115, Clevedon: Multilingual Matters.

Alcaraz, E. and B. Hughes (2002), *Legal Translation Explained*, Manchester: St. Jerome Publishing.

Alcaraz, E. and B. Hughes (2005). *Diccionario de términos jurídicos*, 8th edn, Barcelona: Editorial Ariel.

Arnold, D. (2003), 'Why Translation is Difficult for Computers', in H. Somers (ed.), *Computers and Translation: A Translator's Guide*, 119–42, Amsterdam: John Benjamins.

Bahdanau, D., K. Cho, and Y. Bengio (2015), 'Neural Machine Translation by Jointly Learning to Align and Translate', in *Proceedings of the International Conference on Learning Representations*, San Diego, CA. Available online: https://arxiv.org/pdf/1409.0473.pdf

Baker, M. (2006), 'Contextualization in Translator– and Interpreter–mediated Events', *Journal of Pragmatics*, 38 (3): 321–37.

Bar-Hillel, Y. (1960), 'The Present Status of Automatic Translation of Languages', *Advances in Computers*, 1: 91–163.

Bentivogli, L., A. Bisazza, M. Cettolo, and M. Federico (2016), 'Neural Versus Phrase-Based Machine Translation Quality: A Case Study', *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 257–67, Austin: Association for Computational Linguistics (ACL). Available online: https://aclanthology.org/D16-1025.pdf

Bojar, O., R. Chatterjee, C. Federmann, Y. Graham, B. Haddow, M. Huck, A. Jimeno Yepes, P. Koehn, V. Logacheva, C. Monz, M. Negri, A. Neveol, M. Neves, M. Popel, M. Post, R. Rubino, C. Scarton, L. Specia, M. Turchi, K. Verspoor, and M. Zampieri (2016), 'Findings of the 2016 Conference on Machine Translation', *Proceedings of the First Conference on Machine Translation of the Association for Computational Linguistics*, 131–98, Berlin: Association for Computational Linguistics. Available online: https://aclanthology.org/W16-2301.pdf

Castilho, S., S. Doherty, F. Gaspari, and J. Moorkens (2018), 'Approaches to Human and Machine Translation Quality Assessment', in J. Moorkens, S. Castilho and S. Doherty (eds), *Translation Quality Assessment: From Principles to Practice*, 9–38, Cham: Springer International Publishing.

Castilho, S., J. Moorkens, F. Gaspari, I. Calixto, J. Tinsley, and A. Way (2017a), 'Is Neural Machine Translation the New State of the Art?', *The Prague Bulletin of Mathematical Linguistics*, 108 (1): 109–20.

Castilho, S., J. Moorkens, F. Gaspari, R. Sennrich, V. Sosoni, P. Georgakopoulou, P. Lohar, A. Way, A. V. Miceli Barone, and M. Gialama (2017b), 'A Comparative Quality Evaluation of PBSMT and NMT Using Professional Translators', in *Proceedings of MT Summit XVI. Volume 1: Research Track*, 116–31, Nagoya, Japan. Available online: https://aclanthology.org/2017.mtsummit-papers.10.pdf

Chromá, M. (2011) 'Synonymy and Polysemy in Legal Terminology and Their Application to Bilingual and Bijural Translation', *Research in Language*, 9 (1): 31–50.

Crego, J. M., J. Kim, G. Klein, A. Rebollo, K. Yang, J. Senellart, E. Akhanov, P. Brunelle, A. Coquard, Y. Deng, S. Enoue, C. Geiss, J. Johanson, A. Khalsa, R. Khiari, B. Ko, C. Kobus, J. Lorieux, L. Martins, D. C. Nguyen, A. Priori, T. Riccardi, N. Segal, C. Servan, C. Tiquet, B. Wang, J. Yang, D. Zhang, J. Zhou, and P. Zoldan (2016). 'Systran's Pure Neural Machine Translation Systems', *CoRR*, abs/1610.05540.

Dik, H. (2020), *CTRL+V for Verdict: An Analysis of Dutch to English Legal Machine Translation*, Unpublished MA Thesis, Leiden University.

Dowling, M., S. Castilho, J. Moorkens, T. Lynn, and A. Way (2020), 'A Human Evaluation of English-Irish Statistical and Neural Machine Translation', in *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, 431–40, Lisbon, Portugal. Available online: https://aclanthology.org/2020.eamt-1.46.pdf

Duro Moreno, M. (2012), 'El modelo de los entornos de la traducción: Aplicaciones didácticas a la traducción jurídica', in E. Ortega Arjonilla, C. Balliu, E. Alarcón Navío, and A. B. Martínez López (eds), *À propos de l'enseignement de la traduction et l'interprétation en Europe/Sobre la enseñanza de la traducción y la interpretación en Europa*, 279–88, Granada: Comares.

Forcada, M. L. (2010), 'Machine Translation Today', in Y. Gambier and L. V. Doorslaer (eds), *Handbook of Translation Studies Volume 1*, 215–23, Amsterdam: John Benjamins.

Forcada, M. L. (2017), 'Making Sense of Neural Machine Translation', *Translation Spaces*, 6 (2): 291–309.

García, I. (2010), 'Is Machine Translation Ready Yet?', *Target*, 22 (1): 7–21.

García, I. (2011), 'Translating by Post-Editing: Is it the Way Forward?', *Machine Translation* 25: 217–37.

Glanert, S. (2014), 'Law-in-Translation: An Assemblage in Motion', *The Translator* 20 (3): 255–72.

Gotti, F., A. Farzindar, G. Lapalme, and E. Macklovitch (2008), 'Automatic Translation of Court Judgments', in *Proceedings of The Eighth Conference of the Association for Machine Translation in the Americas*, 1–10, Waikiki, USA. Available online: https://aclanthology.org/2008.amta-govandcom.11.pdf

Hatim, B. (2009), 'Translating Text in Context', in J. Munday (ed), *The Routledge Companion to Translation Studies*, 36–53, London: Routledge.

Heiss, C. and M. Soffritti (2018), 'DeepL Traduttore e didattica della traduzione dall'italiano in tedesco. Alcune valutazioni preliminari', *InTRAlinea*. Special Issue: *Translation and Interpreting for Language Learners*.

Junczys-Dowmunt, M., T. Dwojak, and H. Hoang (2016), 'Is Neural Machine Translation Ready for Deployment? A Case Study on 30 Translation Directions', in

Proceedings of the International Workshop on Spoken Language Translation (IWSLT). Available online: https://arxiv.org/pdf/1610.01108.pdf

Killman, J. (2014), 'Vocabulary Accuracy of Statistical Machine Translation in the Legal Context', in S. O'Brien, M. Simard, and L. Specia (eds), *Proceedings of the Third Workshop on Post-Editing Technology and Practice (WPTP-3), The 11th Conference of the Association for Machine Translation in the Americas, 22–26 October 2014, Vancouver, BC, Canada*, 85–98, Vancouver, Canada. Available online: https://www.amtaweb.org/AMTA2014Proceedings/AMTA2014Proceedings_PEWorkshop_final.pdf

Killman, J. (2015), 'Context as Achilles' Heel of Translation Technologies: Major Implications for End-Users', *Translation and Interpreting Studies*, 10 (2): 203–22.

Kjær, A. L. (1990), 'Context-Conditioned Word Combinations in Legal Language', *Journal of the International Institute for Terminology Research*, 1 (1/2): 21–32.

Koehn, P. (2005), 'Europarl: A Parallel Corpus for Statistical Machine Translation', in *Machine Translation Summit X*, 79–86, Phuket, Thailand. Available online: https://aclanthology.org/2005.mtsummit-papers.11.pdf

Koehn, P. (2010), *Statistical Machine Translation*, Cambridge: Cambridge University Press.

Koehn, P. (2020), *Neural Machine Translation*, Cambridge: Cambridge University Press.

Koehn, P. and R. Knowles (2017), 'Six Challenges for Neural Machine Translation', in *Proceedings of the First Workshop on Neural Machine Translation*, 28–39, Vancouver, Canada. Available online: https://arxiv.org/pdf/1706.03872.pdf

Koponen, M. (2016), 'Is Machine Translation Post-editing Worth the Effort? A Survey of Research into Post-editing and Effort', *The Journal of Specialised Translation*, 25: 131–48.

Läubli, S., S. Castilho, G. Neubig, R. Sennrich, Q. Shen, and A. Toral (2020), 'A Set of Recommendations for Assessing Human–Machine Parity in Language Translation', *Journal of Artificial Intelligence Research*, 67: 653–72.

Melby, A. and C. Foster (2010), 'Context in Translation: Definition, Access and Teamwork', *Translation & Interpreting*, 2 (2): 1–15.

Mellinger, C. D. and T. A. Hanson (2017), *Quantitative Research Methods in Translation and Interpreting Studies*, New York: Routledge.

Papineni, K., S. Roukos, T. Ward, and W. J. Zhu (2002). 'BLEU: A Method for Automatic Evaluation of Machine Translation', in *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 311–18, Philadelphia, USA. Available online: https://aclanthology.org/P02-1040.pdf

Popović, M. (2018), 'Language-Related Issues for NMT and PBMT for English–German and English–Serbian', *Machine Translation*, 32 (3): 252–73.

Prieto Ramos, F. (2014), 'International and Supranational Law in Translation: From Multilingual Lawmaking to Adjudication', *The Translator*, 20 (3): 313–31.

Roiss, S. (2021), 'Y las máquinas rompieron a traducir… Consideraciones didácticas en relación con la traducción automática de referencias culturales en el ámbito jurídico', *TRANS. Revista de Traductología*, 25: 491–505.

Şahin, M. and N. Dungan (2014), 'Translation Testing and Evaluation: A Study on Methods and Needs', *Translation & Interpreting*, 6 (2): 67–90.

Sennrich, R., B. Haddow, and A. Birch (2016), 'Edinburgh Neural Machine Translation Systems for WMT 16', in *Proceedings of the First Conference on Machine Translation (WMT16)*, Berlin, Germany. Available online: https://aclanthology.org/W16-2323.pdf

Simonneaes, I. (2016), 'Legal Language – Pragmatic Approaches to Its Interconnectivity with Legal Interpretation and Legal Translation', *Meta*, 61 (2): 421–38.

Stasimioti, M., V. Sosoni, D. Mouratidis, and K. Kermanidis (2020), 'Machine Translation Quality: A Comparative Evaluation of SMT, NMT and Tailored-NMT Outputs', in *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, 441–50, Lisboa, Portugal. Available online: https://aclanthology.org/2020.eamt-1.47.pdf

Sutskever, I., O. Vinyals, and Q. V. Le (2014), 'Sequence to Sequence Learning with Neural Networks', in Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger (eds), *Advances in Neural Information Processing Systems*, 3104–112, Montréal: NIPS.

Toral, A., S. Castilho, K. Hu, and A. Way (2018), 'Attaining the Unattainable? Reassessing Claims of Human Parity in Neural Machine Translation', in *Proceedings of the Third Conference on Machine Translation*, 113–23, Brussels, Belgium. Available online: https://aclanthology.org/W18-6312.pdf

Toral, A. and V. M. Sánchez-Cartagena (2017), 'A Multifaceted Evaluation of Neural Versus Phrase-Based Machine Translation for 9 Language Directions', in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, 1063–73, Valencia, Spain. Available online: https://aclanthology.org/E17-1100.pdf

TTT.org (2001), 'CLS Framework: ISO 12620 Data Categories Section 01', *ttt.org-CSL Framework*. Available online: https://www.ttt.org/clsframe/datcats01.html (accessed 19 January 2023).

Vanallemeersch, T. and H. Kockaert (2010), 'Automatic Detection of Inconsistent Phraseology Translation', *Southern African Linguistics and Applied Language Studies*, 28 (3): 283–90.

Van Brussel, L., A. Tezcan, and L. Macken (2018), 'A Fine-Grained Error Analysis of NMT, SMT and RBMT Output for English-to-Dutch', in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*, 3799–804, Miyazaki, Japan. Available online: https://aclanthology.org/L18-1600.pdf

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin (2017), 'Attention is All You Need', in *Advances in Neural Information*

Processing Systems, 5998–6008, Long Beach, USA. Available online: https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Wiesmann, E. (2019), 'Machine Translation in the Field of the Law: A Study of the Translation of Italian Legal Texts into German', *Comparative Legilinguistics*, 37: 117–53.

Wu, Y., M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, J. Klingner, A. Shah, M. Johnson, X. Liu, L. Kaiser, S. Gouws, Y. Kato, T. Kudo, H. Kazawa, K. Stevens, G. Kurian, N. Patil, W. Wang, C. Young, J. Smith, J. Riesa, A. Rudnick, O. Vinyals, G. Corrado, M. Hughes, and J. Dean (2016), 'Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation', *CoRR*, abs/1609.08144.

Part III

Shifting Translation and Interpreting Pedagogies

7

Open-source Statistical Machine Technology in Translator Training

From Machine Translation Users to Machine Translation Creators

Khetam Al Sharou

Introduction

With the growing importance of machine translation (MT) and the recent move from statistical machine translation (SMT) to neural machine translation (NMT), alongside the diversification of technological demands in industry, there is a pressing need to include MT authoring within the syllabus of translator training programmes, especially in higher education contexts where technological training has often been understated. In response, this chapter proposes a training module based on open-source software, the Linux-based Moses SMT, as a training platform to encourage making MT part of translation training programmes in low-resource contexts where access to resources and commercial translation technology is limited. Developed primarily at the University of Edinburgh and with the support of European Union funding, Moses provides a free SMT toolkit that 'also includes a wide variety of tools for training, tuning and applying the system to many translation tasks' (Koehn et al. 2007). The software and user manual are available, alongside other information, from the Moses website (Moses 2018). Teaching an open-source MT platform to English-Arabic and Arabic-English translation students so that they learn how to use MT and build their own customized MT engines is a largely unexplored solution. The biggest challenge is that these students have often lacked technological proficiency, and the platform requires a full conceptual understanding of the tasks the user has to perform to obtain a translation output. The proposed module was tested in

educational settings on two groups of translation students in Oman and Jordan to explore its effectiveness in teaching students about MT and its applicability in the classroom. This work does not simply champion the use of one single system (the chosen system is a tool in the process, rather than the primary focus); instead, it aims to examine whether the skills required to build working customized MT engines using Moses can be attained by translation students and whether they can develop the confidence to work with the tool autonomously. It also explores the type of learning and challenges students may experience when trained on Moses for the first time.

This chapter advocates the notion that translation technology training should empower translators, so they are not limited to 'after-the-event' roles such as post-editing MT outputs (Doherty, Kenny and Way 2012: 2). It further supports translators' calls for more flexibility to be built into the tools they are using (Moorkens and O'Brien 2017), and which they themselves can adapt to their own needs by using their skills, rather than passively accepting what is on offer. By proposing a module based on Moses as a training platform, this work offers alternative ways of introducing MT into translator training beyond what is typically taught in translation studies; the aim being to expand learners' capabilities in translation technologies and, in so doing, ascertain how training in translation technology can shift translators' role from mere users to actual creators (or at least adapters) of MT engines. Training on a Linux-based MT system builds not only translation and MT skills but also IT skills, and allows the acquisition of knowledge that would enable them to transition from SMT to NMT and future MT paradigms, while also acquiring a deeper understanding of computational linguistics and natural language processing (NLP).

Translation technologies have grown in significance and have been used in different ways along the various phases of the translation process, contributing to (and probably forcing) the reshaping of these phases, and affecting how translators work and interact with technology. They have changed from mere visions of potential integrated devices (Kay 1997) to now truly integrated and adaptable tools available to both professional and general users. From tools that streamline information mining, word searches and terminological issues, to tools for drafting, translating, revising, reviewing and proofing (as defined in the ISO 17100:2015+A1:2017), there are now technologies to cover the full spectrum of needs (often integrated into one package). Human translators benefit from these tools devised to aid them in the process of their work and attain much higher productivity and quality, but this still comes at the price of accepting specific modi operandi. From their introduction in the 1990s

(O'Brien and Kenny 2001) to the NMT announcements of 2016, many practices have been transformed (see Olohan 2017a, 2017b). The work of language professionals, especially when dealing with high-volume repetitive texts, has changed to revising and editing other translators' work, due to the introduction of translation memory tools (TMs) and their wide uptake by the translation industry (O'Brien 2012; O'Hagan 2013). TM software was designed to allow translators to reuse previously translated segments, with the aim of increasing translators' productivity (O'Hagan 2013). The editing component of the task has increased further with the adoption of MT by language service providers and the increased use of free online MT systems by freelancers. This acceptance reflects the market pressures within the industry to streamline the translation process and reduce the time and cost associated with human translation (O'Hagan 2016). Advances in MT – availability in a variety of language pairs, improved quality and the accessibility of free online MT systems – has also led to new forms of interaction with technology. Many computer-assisted translation (CAT) tools now integrate advanced MT systems, making CAT-based translation even more of a form of post-editing. Instead of having to work with a completely machine-translated text which they need to then post-edit, a new work environment has emerged called interactive translation prediction (ITP) that allows translators to use the MT output as an 'auto-complete feature' (Knowles, Sanchez-Torron and Koehn 2019: 136). This mode purports to grant translators more control over their work in that it allows greater interaction with the technology, giving translators the option to accept a translation suggestion or reject it and then the MT system provides new suggestions as alternatives (Knowles, Sanchez-Torron and Koehn 2019: 136).

Nevertheless, despite all these developments to the translator's workflow, there remains resistance to the widespread adoption of MT, which continues to be seen as a threat to the profession in the long term. Paradoxically, there is still scepticism as to whether MT quality can be improved (Läubli and Green 2019), with recommendations for further investigation into the influence of MT on translators' textual voice and style in the post-editing work, especially in literary-adapted MT, being urged (Kenny and Winters 2020). The new technologies and practices have changed the profile of translators, obliging professionals and trainees to develop a more comprehensive skill set so as to secure additional professional opportunities (Gaspari, Almaghout and Doherty 2015). The development of such a profile can clearly be fostered by training institutions, whose current trainees will become future professional translators only if there is pro-active attention to diversifying their translation technology competencies.

This chapter first reflects briefly on evolving didactic and educational concerns surrounding technology training for translators. In the second section, it outlines the steady, yet at times, cautious changes to translation curricula to include more extensive didactic approaches covering translation technologies, including computer-aided translation tools and machine translation. It then reflects on the merits and challenges of hands-on technology training, presenting the case study that tested Moses as a training platform.

Translation technologies, market needs and curricula

Integrating more technology-based tasks and syllabi into translator training has been a constant request (Doherty, Kenny and Way 2012; Bowker 2014; Guo 2020). From the point of view of educators and trainers, technologies can no longer take a marginal place in professional translators' activities and workspaces, especially with the increased tendency to automate the translation workflow and its development towards more post-editing MT output and changing the translator's skill set as a result (Pym 2012). The modules on post-editing and automation developed for some translation training programmes need to go further to equip their students for the new market requirements, where proficient and competent users of translation technologies are more competitive and de facto more employable as translators (see Cronin 2012).

Translator training degrees, their technological content and their relation to employability and market expectations have been debated over the years. Li (2007) observes a discord between the general consensus that translation teaching cannot exist in a vacuum, detached from the market forces (such as post-editing, quality assurance checking, multilingual/multimedia communication engineering, terminology management, data cleaning) that shape the translation profession, and the different opinions on how and to what extent training programmes should reflect the actual industry in which professionals operate. For example, Pym (1993) cautions against any direct relation between translation training programmes and market demands. According to him, any lessons that could be learned from the market take too long to emerge to be applicable. However, by the turn of the century, Li (2000) strongly advocates for the incorporation of market needs into translation curriculum design to make the transition into the professional market smoother for new graduates. An entirely market-driven education, however, poses its own risks. For example, relegating the role of universities to merely providing a training course in a specific company with

bespoke tasks and workflows limits the graduate to work in a specific context, and allows companies and business approaches that are dominant at the time of training to consolidate their position, which can lead to the monopolization of the market, and be a barrier to the creativity and diversification of both the market and the graduates (see discussion in Pym 1993; Mossop 2003).

With the technological revolution arguably being led by large public institutions (including the EU or the UN) and by the global market, discussions around pedagogical concerns led to a more cautious approach to the adoption of translation technologies in some areas of academic training. For example, the alleged needs of the industry to increase translation productivity does not necessarily justify the training of users in just one popular software platform alone which can limit the graduate to work only in a certain context, as explained above. Similarly, in a world characterized by innovation and fast-changing technology, responding to immediate market demands can result in training programmes that quickly become unsustainable or obsolete as they fall out of step with evolutions in industry. These are legitimate concerns which relate to the major role of academic translator training in allowing students to become familiar with current practices, while preparing them to embrace, adopt and tailor emerging and new practices and technologies to serve their own needs as future translation professionals. Furthermore, while language technology in general has been seen as a facilitator of translation processes, studies focused on the evaluation of translation tools have shown the limited role end-users play in their design and emphasize the importance of their involvement (O'Hagan 2013; Läubli and Green 2019). The lack of the involvement of end-users in the development of the technology explains the late adoption of new strategies by trainers, the reticence in students to engage with the technology, and complaints about the ergonomic constraints of the translator's modern work environment (Cadwell, O'Brien and Teixeira 2018).

Translation technology teaching in translator training

The topic of translation technology and its teaching has become a mainstay in all translation and translator training-related activities. This is borne out by the substantial amount of literature devoted to the teaching and assessment of translation technologies that signals the increased interest in curriculum development and course syllabi design (see, e.g. O'Brien 2002; Bowker 2014; Kenny and Doherty 2014; Moorkens 2018). The 'technological turn' of the

discipline (O'Hagan 2013), including research methods and ontologies, is reflected in the changed landscape of training at university level, where the adoption of multiple forms of technology, especially CAT tools and MT systems, is the main point of focus for many educators and scholars.

Training courses on translation technology started to include TMs around the mid-1990s as it became a requirement in industry (O'Brien and Kenny 2001). CAT tools such as SDL Trados, Déjà Vu and MemoQ, whose TMs are their most significant feature, have been commonly adopted in academic training programmes world-wide over the last two decades (see Kelly 2005; Starlander and Morado Vazquez 2013; Alotaibi 2014). Since their introduction, CAT tools have witnessed major developments in their features, and new versions of these tools with the ability to integrate online MT have been introduced and adopted in both academia and industry. In recent years, the tendency to integrate MT into TM platforms has grown with the increased use of cloud and browser-based translation systems such as Phrase TMS (formerly Memsource) and Smartcat. Tools have been introduced that integrate multiple MT engines (e.g. MateCat), while others, such as Lilt, offer an interactive and adaptive 'MT-first' CAT environment (including support for the English-Arabic pair), where users can work with MT while having the freedom to accept or reject a translation on the go (O'Brien 2012; Knowles, Sanchez-Torron and Koehn 2019; see also Mitchell-Schuitevoerder 2020). Although these developments offer more flexibility, they still limit the role of translators to post-editing. They do not attempt to enable them to build their own tools to meet their translation needs, a lacuna that drives this research to establish a form of training that can empower translators and ensure they can move beyond the role of post-editors.

It is apparent from reviewing the published content of selected translation technology modules embedded in translation training programmes, especially in Arabic-speaking countries, that MT has not been well addressed and is only briefly included in the broader context of translation teaching technology, being taught mainly theoretically. Proposals for the incorporation of MT training into translation programmes have mainly focused on post-editing (PE) (O'Brien 2002; Flanagan and Christensen 2014; Guerberof and Moorkens 2019). The rise of PE courses in HE institutions has served to familiarize students with PE as a different form of translation and prepare them for a future where they might, or will probably have to, work as post-editors, as dictated by the aforementioned market forces (Mellinger 2017). However, limiting the role of future translators in the MT workflows to that of post-editors only has been challenged by translation studies (TS) researchers with the availability of free and open-source

MT technology such as Moses that offers users the flexibility and independence to build customized MT systems (O'Hagan 2013). Furthermore, given the reliance of MT on human-generated data, the interdependency between human translations and MT has increased. To bring about better MT systems, MT developers need to understand the translation processes underlying the raw data production, while linguistic experts (translators) need to understand how MT systems work and how the training data are used so that they might contribute to MT developments (see discussion in Hearne and Way 2011). It has, therefore, become important for all the parties involved to show greater interest in this topic. Accordingly, new trends in the field of translation pedagogy have highlighted the need to train translation students in how to build customized MT engines using cloud-based MT platforms designed for such a purpose. Only a few proposals on teaching SMT have been developed and tested so far (i.e. Doherty, Kenny and Way 2012; Kenny and Doherty 2014). These proposals introduced translation students to a ready-made cloud-based MT platform (SmartMATE), initially based on Moses SMT technology, where their intervention was limited to adding more training data to improve the performance of their systems (Kenny and Doherty 2014). With the advent of NMT, new teaching proposals have emerged, but they are still limited in their objectives. For example, Moorkens (2018) presents a practical in-class translation evaluation exercise as part of a CAT module in which translation students compare SMT and NMT output; and Olkhovska and Frolova (2020) study the impact of using NMT engines on translation students' performance in the classroom.

The proposal presented in this chapter both reiterates and expands upon these previous attempts. The goal is to create competent users of MT, equipped with the skills and experience that will allow them to move beyond being mere users to becoming actual creators of their MT engines. Translators can then become independent users of MT who are able to customize MT engines to suit their own needs, rather than passively accepting what is on offer. As professional translators, training on open-source MT technology allows them to understand how such technology can affect the translation process and when and how to use it in their work. This chapter further explores the didactic implications of teaching open-source MT technology to translation students, resolving teaching and learning issues associated with training students on translation technology (in terms of the content and teaching approaches, technical support and licencing, and student's attitudes towards the tested tool), and contributing to the growing body of research that focuses on teaching translation technologies to translation students.

Critical discussion: Existing issues

Integration of technology: A challenging task

Setting up a framework that allows technologies to be embedded into and across translator education programmes is a challenging task, as highlighted in previous research (e.g. Bowker 2014). A set of must-ask questions were considered by the course designer, before, during and after the introduction of MT into translation training programmes. The first question was when to introduce it. According to Bowker (2014), there is no general agreement regarding this, but 'common sense suggests that it should be possible to introduce more general tools earlier in the translator training process, while reserving some of the more complex tools for later integration' (2014: 98). The second was how it should be introduced. Core courses devoted to translation technologies are valuable in providing students with a deep knowledge of, and extensive skills in, using the tools. However, teaching these tools in isolation deprives them of learning about how they work, including how they can be integrated into other tools and how this integration impacts the translation workflow (Bowker 2014). For these reasons, Mellinger (2017) favours integrating translation tools into practical translation courses. Finally, a decision had to be taken around what should be introduced. Pym (2012) argues that a syllabus on translation technologies should include the teaching of a variety of translation tools. However, according to Rodríguez-Castro (2018), teaching multiple systems might make students feel overwhelmed and under-prepared. With these questions driving the design and content of the MT training programme, many of the answers have proven to be highly dependent on the context in which the training is taking place, not least in terms of technical support and licencing, as explained in what follows.

Hands-on technology training: Benefits and challenges

Hands-on experience is central to any practical course on technology. As a form of 'learning by doing', it helps to deepen the relevant knowledge acquired by students (Shuk Man 2015). Task-based, hands-on training has been tested in several environments for CAT tools (for a comprehensive, albeit outdated, discussion, see Mitchell-Schuitevoerder 2020). Its benefits have been underlined by several TS educators and scholars. For example, Shuk Man (2015) states that hands-on training allows students to gain practical experience in using translation tools, and for Doherty and Kenny (2014), it enhances student confidence. At the same time,

though, it creates issues that may or may not be easily solved. Practical lessons with translation technologies require many contact hours in the already busy schedules of tutors. Additionally, technical restrictions affect the integration of translation technology, such as institutional inflexibility where technical support may not be available for certain platforms, not having enough memory space for students, and disruptions to lab sessions caused by software maintenance (Doherty and Moorkens 2013). The budgetary and technical restrictions can be more keenly felt in non-Western institutions, and so the use of open software in translator training, possibly employing a non-proprietary operating system (OS), offers an alternative solution, especially where licences and technical support are free. This approach was put to the test using an open-source SMT system on Linux as a training platform and the results are discussed in the next section.

Emerging opportunities: Open-source Moses SMT as a hands-on technology training platform – a case study

In 2016, two groups of mixed postgraduate and undergraduate translation students, one in Oman and the other in Jordan, working from English into Arabic and Arabic into English, were offered a task-based training course to develop their MT engines in this language combination as an extracurricular component (for full details, see Al Sharou 2019). As translation technology training continues to lag behind in Arabic-speaking countries (see discussions in Alotaibi 2014; Al-Batineh and Bilali 2017), and because the Arabic language system presents specific challenges for newly introduced technologies (making open-source technology a valid choice because it provides a way to use, test and develop a tool to make it suitable for a specific language combination at a low cost), the proposed training course was tested as an opportunity to expand translation training programmes by integrating MT training in Arabic-speaking countries. This work replicates some details reported in previous co-authored work (Federici and Al Sharou 2018) because it is based on the same study design. However, while Federici and Al Sharou (2018) reflect on the Oman experiment, this work considers data collected from both Oman and Jordan.

Open-source Moses SMT as an educational platform

The infrastructure of Moses underpins most available SMT systems; yet, understanding Moses can be a significant challenge for students without

a computer science background. The payoff is that it represents a way of understanding all the MT features that are being integrated into CAT tools, as well as a direct, experiential route to understanding alignment, information mining and other technology-related concepts (e.g. tokenization, translation and language models, phrase-tables, pre-reordering). Moses allows users to build and customize their SMT engines with considerable freedom and enables translation and language models to be trained for any language combination using the corpora provided by the user (Koehn et al. 2007). Then, the decoder searches quickly through a huge number of possible translations to find those with the highest probability according to these models (Koehn et al. 2007). The Moses team, who still maintains its source code and runs an active mailing list, gives extensive help, provided that the Moses SMT is installed on a Linux operating system.

Even though focus has recently shifted towards NMT technology, teaching SMT is still useful and relevant. First, SMT is more easily accessible in a classroom environment, with minimum technical facilities needed for training MT engines, making it useable, especially for the targeted contexts. More importantly, SMT achieves better quality performance than NMT systems with small parallel corpora, that is, less than one million sentence pairs, as they are considered too small for NMT systems to train robust MT models (Liu et al. 2018; Sen et al. 2021).

Moses can be used to teach students aspects of NLP. Students are expected to gain knowledge of the foundations of MT and acquire a deep understanding of the fundamental principles of probabilistic corpus-based translation techniques. While the underlying architecture of NMT is different (NMT uses neural network-based techniques), both are data-driven approaches and require the same set of technical skills and knowledge acquired by using Moses. To clarify further, Moses is an indicative open-source MT system, hence understanding and mastering the workings of Moses will develop transformative and transferable skills in the users that will allow them to also work with other MT systems. Moses equips users with (1) IT skills to operate in the Linux environment and interact via the command line, which is considered essential for integrating training and test programmes; (2) skills to use various pre-processing and post-processing NLP techniques which form a vital part of the basic skills required to carry out tasks needed to develop MT engines (e.g. tokenization, pre-reordering); and (3) skills for corpus preparation of data. These skills also act as the basis for learning any programming languages, which will enable students to make a positive contribution to the MT community through carrying out tasks including pre-

processing and automated data cleaning. Overall, it can be said that, while NMT is the latest technology in MT, the use of an SMT in the classroom is more than justified, and it is still a better option for the targeted educational context and for working with the Arabic–English language pair.

Participant profile

Twenty-one participants joined the training that consisted of both lectures and practical sessions (or 'labs'). The participants had similar levels of proficiency in English as their second language and were enrolled at the Department of English Language and Literature at Sultan Qaboos University, Oman, and the Translation Department, Faculty of Foreign Languages at the University of Jordan, Jordan. The majority of the participants were students (six participants were master's students, and eleven were undergraduates). Three participants were graduates in the Omani group and one member of staff at the University of Jordan. None of the participants had received training on MT before.

Course design and content

Given that the course design is explained in detail in Federici and Al Sharou (2018) and Al Sharou (2019), this work will, therefore, only cover the main points:

- A task-based approach (TB) to teaching translation technologies was used to design the syllabus of the Moses training course (for discussions on task-based approaches, see Hurtado Albir 1999; González Davies 2004).
- The syllabus was divided into eight lessons, delivered as one theoretical session and seven hands-on sessions (see Table 7.1) within the framework of a two-hour session per day over fifteen consecutive days.
- The syllabus included tasks, interlinked to measurable learning objectives, whose overall goal could only be achieved once learners had incrementally acquired all the skills introduced in the lessons. The aim was to enable learners, guided by the tutor, to develop SMT engines working from English into Arabic and vice versa.
- The freely available online UN corpora was used to build SMT engines during the training.
- At the end of each session, learners were requested to complete a task-based assessment, which were also the stages of creating an SMT engine for this

Table 7.1 Syllabus, adapted from Federici and Al Sharou (2018)

Lesson	Content
1	**Overview of MT, history, approaches and linguistic issues. Statistical machine translation in action (SMT), i.e. Moses SMT**
2	**Introduction to basic Linux commands**
3	**Installation of Moses from source**
4	**Installation of essential external tools needed to set up Moses**: tools for word alignment
5	**Sourcing and preparing data to create the MT system**: including looking for appropriate parallel data freely available online; checking its quality; and processing the data by dividing it into training, tuning and testing sets
6	**Running Moses**: including editing the configuration file; running the system; examining the status of the system; and fixing issues when needed
7	**Translating with Moses**: translating sentences interactively in the Linux shell or from a text file
8	**Quality evaluation of the MT engines**: including introduction to both automatic and manual evaluating measurements where students learn how to check the quality of the MT engine in terms of BLEU scores; and then how to assess the quality of the MT output using error analysis using an error classification adapted for the English-Arabic language combination

language combination, aimed at evaluating their acquired skills rather than their theoretical knowledge. These independent activities replicated the classroom tasks that enabled learners to work autonomously to evaluate their learning at their own pace. Task repetition proved to be a useful technique, supporting its use in cognate fields such as language learning (see, e.g. Bygate 1996).

- The course considered learner levels of self-efficacy as useful ways of assessing the potential of using this platform for training. Issues of adoption and rejection of the technology were also explored and expanded upon as part of learners' personal reflections on whether they were good at working with the technology. The interest in participants' perception of learning and the link with self-efficacy principles were highly influenced by Doherty and Kenny's study (2014).
- During the instructor-led hands-on sessions, learners were given handouts and a reading list of online sources as a follow-up to the classroom tasks, with recommended tasks for independent practice (see Federici and Al Sharou 2018).

Data collection

To assess the viability of the MT training, learner engagement was monitored by collecting qualitative data on their experience. This included two questionnaires, one administered before and the other after the completion of the training. The two questionnaires had two sections: (1) The first section had questions selected from Gaspari (2001) that similarly aimed to collect information about the participants' habitual computer use and their knowledge and experience of using MT before and after they had taken a course on MT. Only eleven questions in the first section in the preliminary questionnaire and fourteen questions in the end-of-module questionnaire were used from Gaspari's fifty-question survey. Unlike Gaspari's study which aimed to assess the improvement of students' theoretical knowledge and opinions and impressions of MT before and after they had taken a course on MT only, the focus of this study was more on building and testing students' technical skills through other methods as explained later. (2) The second section of the questionnaires included a set of questions to test students' self-efficacy based on Compeau and Higgins' (1995) ten-question computer self-efficacy scale. These questions were revised to gauge participants' self-efficacy before and after they had been trained in using Moses for three weeks. The two questionnaires also included questions to elicit essential data on the participants' profiles (e.g. native and second languages, professional and extracurricular experience related to translation) (for further details on the questionnaires, see Federici and Al Sharou 2018). In addition, learners kept a learning log to reflect on what had been achieved and what needed remediation, and the tutor kept systematic classroom observation notes as well as a reflective teaching log.

Technical settings

Linux, the main operating system for running Moses, had to be installed using a virtual machine. At the University of Jordan, the installation of Linux on the university's computers was completed by the instructor with the help of the computer lab's supervisor. At Sultan Qaboos University, however, learners were taught how to install it on their own machines as there was no technical support available. It can be said that the implementation of the training course is dependent on the infrastructure of the translation department and the university. Microsoft made a Linux emulator available within Windows (Windows Subsystem for Linux) which could change the type of IT support offered to staff with open-source software. Furthermore, from a financial perspective, the running costs

of the module would only need to cover staffing and class space, especially if students install the system on their own laptops.

General findings

At the conclusion of each course in Oman and Jordan, overall, the students evaluated their knowledge of MT as having improved from 'very poor' to 'fairly high', while signalling a change of attitude towards MT in that fewer learners now considered MT as a threat (see Federici and Al Sharou 2018 for findings from the Omani experiment). Learners developed their self-efficacy and confidence which corresponds to the development of their skills, as well as their ability to work with Moses and build their own MT systems. At the end of the training, the qualitative data from learners indicate that this type of training in an unfamiliar OS represents a potentially significant platform towards increasing the effectiveness of translation technology, especially as the two iterations showed similar outcomes with minor differences that are specifically noticed in their assessment of some tasks (see Al Sharou 2019).

Students' reflections on their experience

Learners thought that the hands-on training was effective in enabling them to build and create an SMT engine using Moses, and most of them liked the idea of the proposed module, understood the need to master MT technology, and recognized its importance in their career as professional translators, as stated by a Jordanian learner: 'I found it very useful. I had not thought about learning about MT before, but now I am convinced it is important.' The unfamiliar learning environment, therefore, became a source of discovery and curiosity. Nevertheless, the experience was still 'challenging' due to their limited hands-on training in using translation technologies and the need to operate in the unfamiliar Linux command line, which was initially daunting and very demanding.

Learning techniques and procedures through practice

Task-based and routine learning

During the training in Oman and Jordan, there was always an emphasis on the fact that regular and sustained practice enhances and consolidates learning. In the TB syllabus, repeated practice in controlled procedures and a requirement

for the rigorous (and patient) implementation of specific commands allowed participants to develop a sense of incremental learning. They were able to improve their technical skills and gain control over their learning by repeating step-by-step tasks and mastering the system. Through repetition, they acquired familiarity with the task and grew in confidence, and the system subsequently became easier to work with. This finding is in line with previous work on language learning that proved repetition as a useful technique (see Bygate 1996).

Error acceptance

Learners gradually recognized that making mistakes was almost inevitable and natural, and they became more accepting of errors as part of the learning curve. Indeed, new confidence emerged with the realization that solving these mistakes was just a step forward in their learning. This particular realization is of enormous significance in terms of developing learner flexibility and adaptability to switch across technologies and making them more than users of just one or a limited number of translation packages. To resolve any problems encountered, the students were encouraged to find solutions by themselves first, before seeking external help via resources they had access to, including their fellow participants, the internet and the provided handouts. They were also encouraged to collaborate and help their fellow learners who were struggling. This collaboration and knowledge exchange became a form of active learning; the students became teachers to their peers when they had attained a degree of confidence in their own learning, thereby further consolidating their skill set. In case of major errors, the instructor provided elaborative and scaffolding feedback to enhance their learning and build their confidence.

Task-based assessment

The Jordanian group's task assessment differs somewhat from that of the Omani group. As they could not install Moses on their laptop, their access to the Linux lab was limited to classroom time only which meant they could not practice what they had learned and test Moses outside the supervised sessions. Their self-study was limited to reading the handouts and the user guide. This lack of self-study is clear in their self-assessment of their ability to carry out some steps (e.g. installing Moses and the word alignment tool, as well as data preparation steps) which they found to be challenging and confusing, unlike the Omani group

who were more comfortable with these tasks as they had access to Moses on their laptops. However, the Jordanian group described editing the configuration file task as easy and interesting, unlike the Omani group who described it as demanding. This can be attributed to the adjustments made to this task following the prior recommendations of two Omani respondents.

Instructor observations on factors influencing learning

The complexity of the tasks was not the main cause of issues while working with Moses; problems also occurred because learners sometimes did not maintain high levels of attention to detail or concentration on the task at hand, both of which are needed for successful learning in this new computer-based environment. Learners who struggled with (re)occurring errors would sometimes mistakenly underestimate their ability, attributing those errors to a lack of skill. In turn, this misconception could dent confidence and adversely affect the learning process.

Scaffolding and learner motivation

The instructor was presented with a mixed-level class where learners had different backgrounds, abilities, motivations and interests. This meant that more focus needed to be paid to learners who showed low levels of motivation and improvement. For this reason, a step-by-step approach to training was supported by scaffolding, a teaching method proposed by Kiraly (2000), to increase the level of engagement among learners. Forms of support used included explaining unclear or complex tasks by providing examples and activities (Schwartz and Bransford 1998), increasing student expectancy for success (Belland, Kim and Hannafin 2013) and increasing their curiosity by assigning them new tasks (Oudeyer, Gottlieband and Lopes 2016).

Conclusion

The case study demonstrates that integrating open-source MT into translator training programmes is a viable option and is also much needed in Arabic-speaking countries where students have fewer chances of having first-hand contact with a range of translation technologies. The proposed syllabus, with its practical element, created competent and independent users of a freely available tool with no previous experience of working with MT. It transformed

them into creators of their MT engines, thus expanding their capabilities in terms of translation technologies. Moreover, considering the increased prevalence of MT in professional translation workflows, further integration of MT training into existing translator training programmes should, or perhaps must, become an educational priority. Learners need to be able to select the appropriate tools they need to use in their work. Having this awareness of how to match technologies with the needs and briefs of clients is of paramount importance for their future career progression. Furthermore, understanding the need for technical skills also encourages learners to ensure they have a critical grasp of their intellectual and professional value in the translation process given the steady gains that are being made by MT and integrated technologies. The proposed syllabus can serve as a starting point for testing the validity of such hands-on training on open-source MT technology for other language combinations under different educational contexts (including adapting it to teach NMT technology). It can also be used for training trainers or as self-taught training material, all of which are equally valid approaches to explore.

The discussion presented in this chapter, although still ongoing, leaves no doubt that training programmes need to be up to date and reflect in some way the requirements imposed by industry if they are to support translation graduates in their future employability. This employability will depend on the lifelong ability of those graduates to engage with yet-to-be released or designed support technologies. The question is how to realize this engagement. First, translation students need basic competencies and intellectual flexibility to assess, select and monitor the quality of the output of translations produced using technological tools (Korošec 2011). Furthermore, they need to be empowered with more advanced skills and given the opportunity to learn as much as they can about any translation technology that could be useful in their future careers (González Davies 2004), thereby allowing them to take a more advanced role in the development of the technology. That means translation technology teaching should no longer be in the periphery, but rather a central pillar in translation programmes. Practical classes need to be delivered with a focus on CAT tools and MT translations, in addition to human-only translation components. Otherwise, learners are at risk of ignoring intrinsic aspects of their future professional role, a blind spot that will considerably influence their expectations of what the 'job' will look like and their presentation to prospective employers as 'translators' should they choose to enter the profession.

References

Al-Batineh, M. and L. Bilali (2017), 'Translator Training in the Arab World: Are Curricula Aligned with the Language Industry?', *The Interpreter and Translator Trainer*, 11 (2–3): 187–203. https://doi.org/10.1080/1750399X.2017.1350900

Alotaibi, H. M. (2014), 'Teaching CAT Tools to Translation Students: An Examination of Their Expectations and Attitudes', *Arab World English Journal*, 3: 65–74.

Al Sharou, K. (2019), *Training Translators on Open-source MT Technology: An Empirical Assessment of Learning Using a Task-based Syllabus*, Unpublished PhD Thesis, University College London.

Belland, B., Kim, C., and M. Hannafin (2013), 'A Framework for Designing Scaffolds that Improve Motivation and Cognition', *Educational Psychologist*, 48 (4): 243–70. https://doi.org/10.1080/00461520.2013.838920

Bowker, L. (2014), 'Computer-Aided Translation: Translator Training', in S. W. Chan (ed.), *Routledge Encyclopedia of Translation Technology*, 88–104, Abingdon and New York: Routledge.

Bygate, M. (1996), 'Effects of Task Repetition: Appraising the Development Language of Learners', in J. Willis and D. Willis (eds), *Challenge and Change in Language Teaching*, 136–46, Oxford: Heinemann.

Cadwell, P., S. O'Brien, and C. S. C. Teixeira (2018), 'Resistance and Accommodation: Factors for the (Non-)Adoption of Machine Translation Among Professional Translators', *Perspectives*, 26 (3): 301–21. https://doi.org/10.1080/0907676X.2017.1337210

Compeau, D. R. and C. A. Higgins (1995), 'Computer Self-efficacy: Development of a Measure and Initial Test', *MIS Quarterly*, 19 (2): 189–211. https://doi.org/10.2307/249688

Cronin, M. (2012), 'The Translation Age: Translation, Technology and the New Instrumentalism', in L. Venuti (ed.), *The Translation Studies Reader*, 3rd edn, 469–82, Abingdon and New York: Routledge.

Doherty, S. and D. Kenny (2014), 'The Design and Evaluation of a Statistical Machine Translation Syllabus for Translation Students', *The Interpreter and Translator Trainer*, 8 (2): 295–315. https://doi.org/10.1080/1750399X.2014.937571

Doherty, S., D. Kenny, and A. Way (2012), 'Taking Statistical Machine Translation to the Student Translator', *AMTA, the Tenth Biennial Conference of the Association for Machine Translation in the Americas, San Diego, 28 October–1 November*. Available online: http://doras.dcu.ie/17669/1/AMTA-2012-Doherty-1.pdf

Doherty, S. and J. Moorkens (2013), 'Investigating the Experience of Translation Technology Labs: Pedagogical Implications', *The Journal of Specialised Translation*, 19: 122–36.

Federici, F. M. and K. Al Sharou (2018), 'Moses, Time, and Crisis Translation', *Translation and Interpreting Studies*, 13 (3): 486–508. https://doi.org/10.1075/tis.00026.fed

Flanagan, M. and Christensen, T. P. (2014), 'Testing Post-Editing Guidelines: How Translation Trainees Interpret Them and How to Tailor Them for Translator Training Purposes', *The Interpreter and Translator Trainer*, 8 (2): 257–75. https://doi.org/10.1080/1750399X.2014.936111

Gaspari, F. (2001), 'Teaching Machine Translation to Trainee Translators: A Survey of Their Knowledge and Opinions', in *MT Summit VIII Workshop on Teaching Machine Translation*, Santiago de Compostela, Spain, 18–22 September. Available online: https://aclanthology.org/2001.mtsummit-teach.5.pdf

Gaspari, F., H. Almaghout, and S. Doherty (2015), 'A Survey of Machine Translation Competencies: Insights for Translation Technology Educators and Practitioners', *Perspectives: Studies in Translatology*, 23 (3): 333–58. https://doi.org/10.1080/0907676X.2014.979842

González Davies, M. (2004), *Multiple Voices in the Translation Classroom: Activities, Tasks and Projects*, Amsterdam and Philadelphia: John Benjamins.

Guerberof, A. and J. Moorkens (2019), 'Machine Translation and Post-Editing Training as Part of a Master's Programme', *The Journal of Specialised Translation*, 31: 217–38.

Guo, Y. (2020), 'Prospects for the Teaching of Translation Majors in the New Era', *Babel*, 66 (4–5): 867–81. https://doi.org/10.1075/babel.00179.guo

Hearne, M. and A. Way (2011), 'Statistical Machine Translation: A Guide for Linguists and Translators', *Language and Linguistics Compass*, 10: 1–21. https://doi.org/10.1111/j.1749-818X.2011.00274.x

Hurtado Albir, A. (1999), 'La competencia traductora y su adquisición. Un modelo holístico y dinámico', *Perspectives: Studies in Translatology*, 7 (2): 177–88. https://doi.org/10.1080/0907676X.1999.9961356

Kay, M. (1997), 'The Proper Place of Men and Machines in Language Translation', *Machine Translation*, 12 (1–2): 3–23. (Original work published 1980). https://doi.org/10.1023/A:1007911416676

Kelly, D. (2005), *A Handbook for Translator Trainers: A Guide to Reflective Practice*, Manchester: St. Jerome Publishing.

Kenny, D. and S. Doherty (2014), 'Statistical Machine Translation in the Translation Curriculum: Overcoming Obstacles and Empowering Translators', *The Interpreter and Translator Trainer*, 8 (2): 276–94. https://doi.org/10.1080/1750399X.2014.936112

Kenny, D. and M. Winters (2020), 'Machine Translation, Ethics and the Literary Translator's Voice', *Translation Spaces*, 9 (1): 123–49. https://doi.org/10.1075/ts.00024.ken

Kiraly, D. (2000), *A Social Constructivist Approach to Translator Education: Empowerment from Theory to Practice*, Manchester: St. Jerome Publishing.

Knowles, R., M. Sanchez-Torron, and P. Koehn (2019), 'A User Study of Neural Interactive Translation Prediction', *Machine Translation*, 33 (1): 135–54. https://doi.org/10.1007/s10590-019-09235-8

Koehn, P., H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin, and E. Herbst (2007), 'Moses: Open Source Toolkit for Statistical Machine Translation', in *Proceedings of the ACL 2007 Demo and Poster Sessions, Association for Computational Linguistics*, 177–80, Prague, Czech Republic: Association for Computational Linguistics. Available online: https://aclanthology.org/P07-2045.pdf

Korošec, M. K. (2011), 'Applicability and Challenges of Using Machine Translation in Translator Training', *English Language Overseas Perspectives and Enquiries*, VIII (Autumn): 8–18. https://doi.org/10.4312/elope.8.2.7-18

Läubli, S. and S. Green (2019), 'Translation Technology Research and Human-Computer Interaction (HCI)', in M. O'Hagan (ed.), *The Routledge Handbook of Translation and Technology*, 370–83, Abingdon and New York: Routledge.

Li, D. (2000), 'Tailoring Translation Programs to Social Needs: A Survey of Professional Translators', *Target*, 12 (1), 127–49. https://doi.org/10.1075/target.12.1.07li

Li, D. (2007), 'Translation Curriculum and Pedagogy Views of Administrators of Translation Services', *Target*, 19 (1): 105–33. https://doi.org/10.1075/target.19.1.07li

Liu, C. H., C. C. Silva, L. Wang, and A. Way (2018), 'Pivot Machine Translation Using Chinese as Pivot Language', in J. Chen and J. Zhang (eds), *Machine Translation. CWMT 2018. Communications in Computer and Information Science*, 954, Singapore: Springer. https://doi.org/10.1007/978-981-13-3083-4_7

Mellinger, C. D. (2017), 'Translators and Machine Translation: Knowledge and Skills Gaps in Translator Pedagogy', *The Interpreter and Translator Trainer*, 11 (4): 280–93. https://doi.org/10.1080/1750399X.2017.1359760

Mitchell-Schuitevoerder, R. (2020), *A Project-based Approach to Translation Technology*. Abingdon and New York: Routledge.

Moorkens, J. (2018), 'What to Expect from Neural Machine Translation: A Practical In-Class Translation Evaluation Exercise', *The Interpreter and Translator Trainer*, 22: 1–13. https://doi.org/10.1080/1750399X.2018.1501639

Moorkens, J. and S. O'Brien (2017), 'Assessing User Interface Needs of Post-Editors of Machine Translation', in D. Kenny (ed.), *Human Issues in Translation Technology*, 109–30, Abingdon and New York: Routledge.

Moses (2018), 'Welcome to Moses!', *Moses*, 24 June. Available online: http://www2.statmt.org/moses/ (accessed 17 August 2023).

Mossop, B. (2003), 'What Should be Taught at Translation School?', in A. Pym, C. Fallada, J. R. Biau, and J. Orenstein (eds), *Innovation and E-Learning in Translator Training*, 20–2, Tarragona: Universitat Rovira i Virgili.

O'Brien, S. (2002), 'Teaching Post-Editing: A Proposal for Course Content', in *Proceedings of the 6th EAMT Workshop on Teaching Machine Translation, EAMT/BCS, UMST*, 99–106, Manchester, UK. Available online: http://mt-archive.info/EAMT-2002-OBrien.pdf

O'Brien, S. (2012), 'Translation as Human-computer Interaction', *Translation Spaces*, 1(1): 101–22. https://doi.org/10.1075/ts.1.05obr

O'Brien, S. and D. Kenny (2001), 'In Dublin's Fair City: Teaching Translation Technology at Dublin City University', *Language International*, 13 (5): 20–3.

O'Hagan, M. (2013), 'The Impact of New Technologies on Translation Studies: A Technological Turn?', in C. Millán and F. Bartrina (eds), *The Routledge Handbook of Translation Studies*, 503–18, Abingdon and New York: Routledge.

O'Hagan, M. (2016), 'Massively Open Translation: Unpacking the Relationship between Technology and Translation in the 21st Century', *International Journal of Communication*, 10: 929–46.

Olkhovska, A. and I. Frolova (2020), 'Using Machine Translation Engines in the Classroom: A Survey of Translation Students' Performance', *Advanced Education*, 15: 47–55.

Olohan, M. (2017a), 'Knowing in Translation Practice: A Practice-theoretical Perspective', *Translation Spaces*, 6 (1): 159–80. https://doi.org/10.1075/ts.6.1.08olo

Olohan, M. (2017b), 'Technology, Translation and Society: A Constructivist, Critical Theory Approach', *Target. International Journal of Translation Studies*, 29 (2): 264–83. https://doi.org/10.1075/target.29.2.04olo

Oudeyer, P. Y., J. Gottlieband, and M. Lopes (2016), 'Intrinsic Motivation, Curiosity, and Learning: Theory and Applications in Educational Technologies', *Progress In Brain Research*, 229: 257–84. https://doi.org/10.1016/bs.pbr.2016.05.005

Pym, A. (1993), 'On the Market as a Factor in the Training of Translators', *Koiné*, 3: 109–21.

Pym, A. (2012), 'Translation Skill-sets in a Machine-Translation Age', in *16th International Symposium on Interpretation and Translation Teaching, Graduate Institute of Cross-Cultural Studies*, Taiwan: Fu Jen Catholic University, 14–15 January. Available online: https://usuaris.tinet.cat/apym/on-line/training/2012_competence_pym.pdf

Rodríguez-Castro, M. (2018), 'An Integrated Curricular Design for Computer-Assisted Translation Tools: Developing Technical Expertise', *The Interpreter and Translator Trainer*, 12 (4): 355–74. https://doi.org/10.1080/1750399X.2018.1502007

Sen, S., M. Hasanuzzaman, A. Ekbal, P. Bhattacharyya, and A. Way (2021), 'Neural Machine Translation of Low-Resource Languages Using SMT Phrase Pair Injection', *Natural Language Engineering*, 27 (3): 271–92. https://doi.org/10.1017/S1351324920000303

Schwartz, D. L. and J. D. Bransford (1998), 'A Time for Telling', *Cognition and Instruction*, 16: 475–522. https://doi.org/10.1207/s1532690xci1604_4

Shuk Man, C. W. (2015), 'The Teaching of Machine Translation. The Chinese University of Hong Kong as a Case Study', in S. W. Chan (ed.), *Routledge Encyclopedia of Translation Technology*, 237–51, Abingdon and New York: Routledge.

Starlander, M. and L. Morado Vazquez (2013), 'Training Translation Students to Evaluate CAT Tools Using Eagles: A Case Study', in *Aslib: Translating and the Computer 35*, London. Available online: https://aclanthology.org/2013.tc-1.1.pdf

8

Teaching Machine Translation Literacy to Non-translation Students

A Case Study at a Canadian University

Lynne Bowker

Introduction

Technology occupies a significant space in our society, and the effects of the speed and breadth of technological change have been felt deeply in areas relating to language and communication, and certainly in multilingual communication. This includes professions such as translation, but it reaches far beyond professional translators. Machine translation is one technology that has undergone a major paradigm shift with regard to its underlying approach, resulting in improved translation quality. Subsequently, many users outside the translation profession have adopted this technology; however, these users, who have no formal education in translation, need support to make good use of machine translation tools. Translation educators are one group who have the knowledge and experience to help students within and beyond translator education programmes to develop relevant skills.

Translation technology courses have long been integrated into translator education programmes; however, instruction in the use of machine translation is not yet common outside of translation and language classes. Nevertheless, translation educators currently have an opportunity to assist non-language professionals to become more critical and informed users of machine translation tools, such as Google Translate or DeepL Translator, which are freely available online to anyone with an internet connection. Machine translation tools are easy to access and use, but this does not mean that users who have no background in translation will instinctively know how to use them effectively or responsibly.

Rather, such users need to develop machine translation literacy skills (Bowker and Buitrago Ciro 2019).

In this chapter, I present a case study in which machine translation literacy has been actively taught as part of a translation course for non-translation students at a large Canadian university. Though data were collected only from a single institution, this experience serves as a useful case study whose findings may inspire a similar undertaking at other institutions. The results are applicable for universities around the globe given that machine translation is widely used by students in many countries (e.g. Alhaisoni and Alhaysony 2017; Mundt and Groves 2016; Lee 2020; Delorme Benites et al. 2021; Loock and Léchauguette 2021; Dorst, Valdez and Bouman 2022), and that the student body at most universities primarily comprises students studying subjects other than translation.

Following this introduction, I briefly introduce machine translation and the concept of machine translation literacy, explaining some of the factors that have contributed to the need for non-translation students to develop this skillset. Next, I describe the institution and the course on translation for non-translation students in which a machine translation literacy module was embedded, along with the demographic profile of the students in the course, and the general content of the machine translation literacy module as well as some of the pedagogical practices used to deliver and evaluate it. This is followed by a discussion of some high-level student feedback on the module that was collected using an anonymous online survey. Finally, I share some concluding remarks and suggestions for possible next steps.

Machine translation: An evolving landscape

The context of machine translation use has evolved considerably since the first tools were developed in the early 1950s. This includes changes in the way that the tools work, in the types of users, and in the nature of the education needed to ensure critical use.

Serious interest in machine translation was sparked by the now famous memorandum issued by Warren Weaver (1949). Following its release, machine translation research programmes were launched in a number of countries, and the overall approach adopted by researchers is now described as rule-based machine translation (RBMT) (Hutchins and Somers 1992). Developing RBMT systems essentially involved programming computers to look up words in

large bilingual dictionaries and to apply grammar rules, in an approximation of how it was thought that people process language. However, after about a decade of hard work, developers became frustrated after running up against the so-called semantic barrier and realizing that people often disambiguate language by drawing upon their real-world knowledge, whereas computers could not do this (Bar-Hillel 1960). Although it remained the dominant paradigm for nearly half a century, RBMT had limited success. In general, the quality of the output was usually quite poor, and in most cases the machine-translated texts needed to be extensively post-edited by a person who had a strong background in translation. Often, this post-editing process was more labour-intensive than translating from scratch, and so RBMT systems were not widely used.

In the following decades, computer technology itself became faster and more powerful, and the rapid growth of the internet in the late 1990s and early 2000s was accompanied by increasing online linguistic diversity. During this period, researchers began to experiment with new approaches to machine translation that drew on the strengths of computers, such as number crunching and pattern matching. One of the most promising data-driven approaches that emerged became known as statistical machine translation (SMT) (Koehn 2010). In SMT, translation is seen as a problem in which the system has to select the most probable translation for a given source text by referring both to a probabilistic model of translation that it has learned from an aligned bilingual corpus, and to a probabilistic model of the target language that it has learned from a monolingual target language corpus. The corpora needed to train SMT systems are extremely large, meaning that data-driven approaches tend to be more successful for widely used languages and language pairs, and somewhat less successful for languages and pairs that are less widely used. Overall, however, the translations produced by SMT systems were of a considerably higher quality than those generated by RBMT systems. In addition, the SMT era witnessed the launch of the first free online machine translation systems, including the high-profile Google Translate in 2006, which was quickly followed by a host of similar products. According to Turovsky (2016), Google Translate went from supporting two languages and hundreds of users in 2006 to supporting 103 languages and hundreds of millions of users in 2016, and both the number of languages and users look set to continue expanding. Since the introduction of Google Translate, we can describe machine translation tools as being well and truly in the wild, a commonly used expression in information technology which means that a tool has passed through the research and development stage and has become a publicly used tool.

Even more recently, since around late 2016, there has been another paradigm shift in the underlying approach to machine translation. Researchers have applied artificial intelligence techniques, such as machine learning, in an approach that is referred to as neural machine translation (NMT) (Forcada 2017; Koehn 2020). This approach, which is also data-driven and so relies on providing the computer with enormous collections of previously translated texts, incorporates an artificial neural network that can be trained to learn from these examples. Whereas the original RBMT systems, and to a somewhat lesser extent the later SMT systems, produced translations that often contained errors and sounded awkward, the translations produced by NMT systems are considerably more accurate and natural sounding (Castilho et al. 2017). Though still not perfect, the output of NMT tools is more viable as a starting point than was the output of the older systems (ISO 2017), and since the results may be usable for some purposes (e.g. basic knowledge acquisition), or may at least provide a reasonable first draft, more and more people are beginning to use machine translation, especially the free online tools. Adding to their appeal is the fact that, from a user interface perspective, machine translation tools are very straightforward to use because they require users to do little more than copy/paste a text and select the language for translation, or sometimes just click a button.

In summary, over the course of the history of machine translation, we have seen these tools become easier to access, easier to use and able to produce a higher quality of output that is usable for some purposes. Nevertheless, users must still take care and show good judgement because the natural sounding text may lull users into a false sense of security and result in them overlooking errors. Yet, despite this need to exercise caution, users do not instinctively know how to optimize NMT or even how to use it wisely in a given context. Unlike language professionals, who have both training and experience in knowing how, when or whether to use translation technology, as well as how to recognize and compensate for its limitations, people outside the language professions do not typically receive such training. Moreover, for users without a formal background in translation, the popular media can shape people's perceptions of machine translation.

Vieira (2020) investigated the way that the international English-language press presents translation technologies to the general public. After examining 284 articles published in six different countries between 1986 and 2019, Vieira (2020: 13) found that the coverage of machine translation in the written press tends to emphasize the positive aspects of the technology, while at times it 'inflated the capabilities of the technology by comparing it to human translators

or by implying that MT had the power to make users speak and understand any language without the prospect of encountering any issues'. Meanwhile, articles that adopted a negative view were more narrowly focused on the poor quality of the output and used a sensationalist tone to draw particular attention to errors that were deemed shocking. Overall, this investigation suggests that the reporting lacks nuance and tends either to overstate this technology's capabilities or to position it as being highly problematic. The truth falls somewhere between these extremes, and users with no formal background in translation may need guidance to develop their machine translation literacy. In other words, they need to increase their understanding of how machine translation works, to assess when and where it can be usefully employed, and to learn how to work with it more effectively to get better results. If users do not have a good level of machine translation literacy, the speed and convenience of machine translation tools may cause them to simultaneously underestimate the intricacies of translation and overestimate the capabilities of these tools, which could in turn result in their misuse.

On a more positive note, the scholarly literature contains examples of machine translation being used outside the translation profession with varying degrees of success. For instance, Nurminen (2020) presents a case study explaining how patent professionals use this technology to search for international patents (see also Nurminen and Koponen, this volume), while Cadwell, O'Brien and DeLuca (2019) explore how machine translation may be used by humanitarian aid workers to facilitate communication in a crisis situation, and Anazawa, Ishikawa and Takahiro (2013) describe how practicing nurses in Japan use machine translation to stay on top of the latest developments in international nursing literature. However, in each case, the researchers emphasize that some form of training is needed to allow users to make better decisions about employing the technology and to optimize its use. I believe that offering machine translation literacy instruction to students who are not studying translation can be a way for translation educators to empower a greater number of machine translation users to become informed users. In the next section, I describe a machine translation literacy module offered as part of a first-year university course on translation for non-translation students.

Institutional profile, course description and student profile

The University of Ottawa is a large Canadian university with over 45,000 students. It is a fully bilingual (English-French) institution with a mission to ensure that

French-language higher education opportunities are available to Canadians. The university is also committed to internationalization and to supporting Canada's Indigenous communities, as outlined in its strategic plan *Transformation 2030* (University of Ottawa 2019). Overall, language has an important profile at this institution, which hosts several language-oriented academic units, including a School of Translation and Interpretation. Until recently, the undergraduate provision at this school focused exclusively on delivering professional programmes for students wishing to work as translators or terminologists, and the courses were open only to students majoring in translation. The sole exception was a single first-year course called 'Introduction to translation', which was intended to act as a feeder that would attract students with an as-yet-undeclared major into the translation programme. However, beginning in the year 2019, the school began reforming its undergraduate programmes, and in that context, I began specifically to rethink the nature of the 'Introduction to translation' course that I had been teaching.

Over the past decade, popular media have reported widely on what is often described as a 'language crisis'. It has been noted that universities in countries such as the United States (e.g. Foderaro 2010; Johnson 2019), the United Kingdom (e.g. Boffey 2013; Bowler 2020; Parsons 2020) and Australia (e.g. Mason and Hajek 2019; Lees et al. 2020) have responded to the need for belt tightening by cutting language programmes. This also happened at the University of Ottawa, where the Department of Modern Languages and Literatures has undergone major restructuring. It is perhaps not surprising to see this happening in English-speaking countries given that English seems entrenched as the world's lingua franca for international business (e.g. Neeley 2012), scientific research (e.g. Montgomery 2013) and higher education (Wächter and Maiworm 2014). At the same time, as noted previously, machine translation quality has vastly improved – even to the extent that the media often portray these tools as having near-magical powers that obviate the need for learning other languages. In the face of such phenomena, cutting language programmes must appear to be an easy win for higher education institutes in times of fiscal austerity.

In response, arguments have been mounted which underscore the value of maintaining language courses regardless of the status of English as a lingua franca or the possibilities offered by technology for facilitating multilingual communication (e.g. Reisberg 2017; Ben-Ghiat 2019). Nevertheless, language departments at many universities have already been shuttered or shrunk (Boffey 2013; Johnson 2019). In contrast, translation programmes are thriving to a greater extent, for the present at least. Translation is a highly specialized skill that

often requires working with specialized subjects (e.g. legal, medical, technical), or requires knowledge of specialized tools (e.g. for audiovisual translation, localization). Therefore, it cannot usually be tackled successfully by machine translation alone or by a bilingual person with no training in translation. Translator training does not necessarily include language teaching per se but often assumes that students will enter the programme with a solid knowledge of two or more languages, which they will continue to hone through the acquisition of specialized translation skills. The European Master's in Translation (EMT) is a network of master's level programmes in translation which counts more than eighty members across Europe (European Commission 2021), suggesting that interest in translator training remains strong. Meanwhile, translation programmes are faring well in other regions of the world too. For instance, in Canada, more than a dozen universities are members of the Canadian Association of Schools of Translation (Mareschal 2005), while in China, the number of translator training programmes has grown to include nearly 400 bachelor's and over 250 master's programmes (Zhong 2020).

While professional translator training programmes focus on preparing individuals to work as professional translators, and these courses are not typically open to the broader student population, students in other disciplines may nonetheless be interested in translation, and instructors who teach on translator training programmes are well placed to deliver such a course. In the following section, I report on my experience of designing and delivering a course about translation for non-translation students, paying particular attention to the module on machine translation literacy, which is highly relevant for this group.

Benefits of offering a translation course to non-translation students

As noted, many language programmes are being scaled back at a time when campuses are becoming more linguistically diverse. Not only are universities working hard to attract international students (e.g. Redden 2018), but migration means that domestic students may speak a heritage language if they or their parents have moved from another country (Duff and Li 2009). In addition, some countries have more than one official language and may also have one or more Indigenous languages (Lewington 2018). As a result, many students have an interest in or experience of studying, working or living in a multilingual context. Even if it is not their primary focus of study, such students may be motivated to learn more about translation since it already features in their life in some way. In other cases, students may have plans to integrate an international component to

their future studies or career. If they are unable to take a language course, they may be interested in feeding their desire for knowledge about other languages and cultures through a course about translation.

Another reason to offer a course about translation to non-translation students is to try to dispel some misunderstandings. It is well known that many members of the general public do not know the difference between translation and interpretation, and that many falsely believe that bilingualism is the only criterion required for being a translator. Given that the popular media tend to present translation in a way that lacks nuance, such as by either over- or understating the capabilities of machine translation, a course about translation can help to sensitize students to translation and give them a greater appreciation for what is involved.

While a single course about translation will not prepare students to work as professional translators, it can nonetheless make them aware of this profession and of what is involved, and in this way, it might attract some of the students to pursue translation, even at a later stage in life. Since much translation deals with specialized material, and since the most successful specialized translators are those with a sound subject field knowledge, students who are studying engineering, science or administration and who take a course about translation may later be inspired to pursue a professional master's in translation and work as a specialized translator.

Finally, while a course about translation cannot replace language courses, it could demonstrate to decision-makers in governments or university administration that students – even students of other disciplines – *do* have a strong interest in language and culture. In this way, a course about translation may provide evidence in favour of re-introducing or expanding the offer of language and culture courses in the future.

Course content

The idea for such a course is that it is principally *about* translation, rather than learning to *do* translation. In other words, it is not a translation practice course but a high-level exploration of the field. The course need not be specific to any language pair and can even encourage students to incorporate their varied linguistic and cultural backgrounds into the discussions and activities. Inspiration for content can come from ways in which the lives of ordinary people intersect with some aspect of the translation profession. For example, many people have watched foreign-language films or series and so have encountered

subtitling or dubbing. Similarly, many people have used a machine translation tool such as Google Translate, and many have acted informally as an interpreter for a friend or family member.

At the University of Ottawa, a teaching term consists of twelve weeks, with the majority of courses having three contact hours per week. This means that a first-year course about translation can be divided into twelve one-week modules, the possible high-level themes of which are summarized in Table 8.1. The remainder of this chapter will focus on the module on machine translation literacy; however, a more detailed overview of the entire course can be found in Bowker (2023).

General student profile

At the University of Ottawa, the translation course for non-translation students is offered as a first-year course with no pre-requisites; however, students could take it at a later point in their studies. In addition, the course is not required for or tied to any specific programme but is offered as an elective open to students in all disciplines. The course was offered for the first time in the Fall 2020 term, with a target of having at least thirty-five registered students. As illustrated in Table 8.2, it exceeded this target and has continued to do so in subsequent terms. Based on the number of registrations during these first two academic years of running the course, the school has committed to continue offering it regularly for the foreseeable future.

Table 8.1 Themes covered in the course

Module	Theme
1	Basic concepts and terms used in the language professions
2	Brief history of translation and interpretation
3	The language industry today
4	Lexicography and terminology basics
5	Tools and resources (e.g. term banks, electronic dictionaries, online concordancers)
6	Machine translation literacy
7	Localization
8	Adaptation and transcreation
9	Summarization and interlingual multimodal communication (e.g. sight translation, records of meetings)
10	Audiovisual translation (subtitling and dubbing)
11	Introduction to interpreting
12	Wrap-up

Table 8.2 Number of students registered

Term	Number of registered students
Fall 2020	42
Winter 2021	45
Summer 2021	38
Fall 2021	47
Winter 2022	49
Summer 2022	35
TOTAL	256

Up to this point, the translation course has attracted students from forty-seven different degree programmes across the broad spectrum of both humanities and sciences. Among the programmes represented, it is not too surprising to see a large cluster of ninety-seven students emerging from programmes that are concerned with other aspects of language and communication, such as communication, second-language teaching, linguistics and various programmes dealing with different modern languages and literatures. From the social sciences, the course attracted twenty-two students majoring in law, political science, public administration, international affairs, conflict studies and human rights and social work – all of which are professions that could entail regular interactions with people from other linguistic and cultural backgrounds. While the sciences were less well represented than the humanities overall, the course nonetheless attracted twenty-five students majoring in a science (e.g. biology, environmental science, mathematics), twenty-eight students from programmes in the health sciences (e.g. human kinetics, nursing, psychology) and twenty-one students from different types of computer and engineering programmes. Finally, seventeen students were majoring in a programme called interdisciplinary studies, in which they are encouraged to sample courses from a wide range of fields. The remaining programmes represent a long tail with five or fewer students each (e.g. economics, history, music, gender studies).

Overall, the diversity of programmes represented suggests that there is interest in, and a market for, a course about translation for students not planning to become professional translators. Moreover, when asked to share on the course discussion forum what motivated them to choose this course, numerous students indicated that, while they were happy with their major and did not intend to change it, they nonetheless hoped to integrate an international component into their career or future studies and so they were interested in learning about interlingual and intercultural communication. Meanwhile, others cited personal interest, with a considerable number noting that they had already engaged in

some informal translation activity (e.g. interpreting for a family member or employer).

At the University of Ottawa, courses on the professional translator training programme are limited to English, French and Spanish, but the range of languages used by the students in the course suggests that the interest in translation at this institution extends to speakers of other languages too. With regard to the languages spoken by the students, thirty different native or dominant languages were declared. Unsurprisingly, Canada's two official languages (English and French) were the best represented languages with ninety-two and sixty-one native speakers respectively. Meanwhile, other languages that are commonly used by heritage speakers or by international students in the course include Chinese (thirty-two students), Arabic (twenty-five students) and Spanish (fifteen students). This breakdown is in line with data recently released by Citizenship and Immigration Canada, which reports that Canada issued a record-breaking number of permits for international students in 2021, with China and Mexico both figuring in the top ten countries for which Canadian study permits were issued (El-Assal 2022). In addition, because the University of Ottawa in particular offers many programmes taught in French, it is a popular choice for international students who may have French (rather than English) as their second language, such as Arabic-speaking students from the Maghreb; indeed, as reported by El-Assal (2022), residents of Morocco are eligible for an expedited permit to study in Canada. Once again, the remaining languages form a long tail with just a few speakers each; however, in many cases, these languages are used in countries that were formerly colonized by France or Belgium, meaning that these students may be attracted by the opportunity to study on French-taught programmes. For instance, various students in the course identified themselves as native speakers of Haitian Creole, Vietnamese, Kirundi (spoken in Burundi), Kinyarwanda (spoken in Rwanda, Burundi and the Congo) and Swahili (spoken in Rwanda, Burundi and the Congo, among other places).

Machine translation literacy module

The module on machine translation literacy was integrated into the course at the halfway point, meaning that the students had already acquired some basic translation concepts before tackling machine translation. This module comprised three contact hours; however, students were expected to do some advance preparation (e.g. consulting introductory texts and videos on how

machine translation systems work), as well as some homework (e.g. exercises on revising machine translation input and output, quiz). As noted previously, machine translation tools are quite straightforward to use, meaning that the module is less about 'how to' use machine translation (i.e. which buttons to press) and more about whether, when and why to use it. The focus is more on critical thinking tasks, such as evaluating a text's suitability for translation by machine or weighing the benefits and risks of using machine translation against other translation solutions. Owing to space constraints, it is not possible to provide a comprehensive description of the contents of the module; however, the general and specific learning outcomes, key content and general pedagogical approaches that were employed are briefly summarized.

The general objective of the module is to encourage students to approach machine translation use in a critical way. Specifically, students are expected to be able to do the following things once they have completed the module:

- explain the general concept of machine learning and the overall neural approach to machine translation;
- appreciate the type, quantity and quality of data required for data-driven NMT, identify ways in which machine translation systems can be sensitive to data and articulate potential consequences of data insufficiency;
- describe the need for transparency around machine translation use;
- conduct basic risk assessment regarding machine translation use;
- compare and evaluate the results produced by a selection of free online machine translation systems;
- modify input texts to reduce ambiguity and improve the quality of the machine translation output;
- apply basic post-editing techniques to improve machine translation output according to fit-for-purpose principles.

To enable students to achieve these learning outcomes, the module was divided into four main parts, as outlined in the following.

Data-driven approaches to machine translation

Using introductory videos and readings (consulted before class), and a short lecture from the professor, students learned the essentials of how NMT systems operate. This knowledge allows students to understand the strengths and limitations of these tools. For instance, an understanding of sensitivity to the volume and type of training data can help users to realize why machine translation

systems could be more or less useful for different language combinations, domains or text types.

An understanding of the fact that each machine translation system is trained using a different corpus also alerts students to the fact that each system is likely to produce different results. Most students have tried Google Translate, but not necessarily other products. Realizing that different tools exist and that these have been trained on different corpora can encourage students to try others. In addition, students come to understand that NMT systems are always learning, but the new training data (and other factors, such as sentence length) may affect the quality of the MT output in different ways. Therefore, students should not dismiss a tool altogether because of one poor performance, but rather they should be aware that unpredicted fluctuations in quality may occur in any system.

Finally, knowledge about sensitivity to training data alerts students to the potential for algorithmic bias, such as gender or racial bias, if the training data is not well selected (Monti 2020).

Transparency

Students discuss mini-case studies to learn the ethical importance of transparency in regard to machine translation use. For example, students contemplate whether using machine translation for coursework is acceptable or contrary to the learning objectives of a language course vs a course in another domain. Students also consider fair use of machine translation and sustainability, which includes encouraging users not to use machine translation in a way that might harm the language industries. For instance, the module emphasizes importance of clearly labelling machine-translated texts since this allows readers to take this information into account when deciding whether to trust the content.

Risk assessment

Mini-case studies also allow students to develop their judgement when assessing the benefits and risks of using machine translation for a particular task. Pragmatically, students learn not to enter sensitive information into free online machine translation systems because the companies that own the tools can keep the data and use it for other purposes. However, risk assessment extends to determining whether a translation task is a high- or a low-stakes task. For example, using machine translation to read a friend's social media post or a manga comic book is low-stakes because translation errors are unlikely to

have serious consequences. In contrast, using machine translation to translate texts in a legal or healthcare context is a high-stakes scenario because a poor translation could have serious consequences (Vieira, O'Hagan and O'Sullivan 2021).

A related concept is understanding that translation can be undertaken for different purposes (e.g. comprehension vs production) and that, while raw machine translation may be useful for helping a reader to understand a text in a familiar field, it may be less appropriate for a purpose such as publishing. Likewise, external factors such as time or budget may come into play when deciding the extent to which machine translation could be a good choice.

Interacting with machine translation

Finally, students do some text editing exercises to learn how they can interact with machine translation systems to improve their output. While most students are already somewhat familiar with the concept of post-editing, few have considered that a key way to improve machine translation output is to ensure that the input text is written in plain language with little to no ambiguity (e.g. using the active voice, short sentences, consistent terminology).

Although the errors made by a machine translation system will likely differ from one system, language, domain or text type to the next, it is worth learning some basic techniques to spot and fix errors. In particular, NMT systems are known to produce texts that sound very plausible, even though they may not be correct (Way 2020), and users may easily overlook such errors. By practicing both pre- and post-editing and comparing the results of different machine translation systems, students become more comfortable with manipulating texts to achieve a desired level of quality.

Evaluation

The methods used to evaluate the students' knowledge of machine translation literacy were varied and included the following:

- crossword puzzle to test students' recognition of definitions of key notions;
- quiz with true/false and multiple choice-style questions (which can be auto-corrected by the learning management system) to evaluate students' understanding of basic concepts;

- a debate (in-person or on a discussion forum) where students argued for or against a decision to use machine translation for a given task;
- posts to a discussion forum in response to prompts (e.g. what is the most surprising thing that you learned about machine translation that you did not know before this module?);
- pre- and post-editing exercises.

Examples of some of these activities can be found on the Machine Translation Literacy Project website (https://sites.google.com/view/machinetranslationliteracy/). Note that not all activities need to be graded; some can simply serve to allow students to demonstrate their understanding of the material. As work on machine translation literacy becomes more established, we will undoubtedly see more resources being developed to support this type of instruction. In addition to the Machine Translation Literacy Project website and the textbook that we have recently published (Bowker 2023), instructional resources for machine translation literacy are also being developed by other groups, such as the Erasmus+ strategic partnership MultiTraiNMT—Machine Translation Training for Multilingual Citizens (Ramírez-Sánchez et al. 2021; Delorme Benites et al. 2021; Dorst, Valdez and Bouman 2022; Kenny 2022).

High-level feedback

At the end of the course, I asked students to respond to a voluntary and anonymous survey to gather feedback with a view to improving the course for the future. The following data presented were gathered from the students who completed the course in the Fall 2020, Winter/Summer/Fall 2021 and Winter 2022 terms. Of the 221 students who completed the course, 193 provided complete responses to the survey, which corresponds to an 87 per cent response rate.

One question on the survey asked the students to select their favourite module on the course (i.e. the one that they enjoyed the most or found to be the most interesting). In another question, students were asked to indicate which module they found to be the most useful (e.g. which gave information that they had already applied or could see themselves applying in the future). As indicated in Tables 8.3 and 8.4, the machine translation literacy module was the only one to be ranked among the top three in both cases, suggesting that non-translation students find the topic to be both interesting and helpful.

Table 8.3 Top three most enjoyable modules

Module	Ranking	Number/percentage of respondents who selected this module
Adaptation and transcreation	1	70 (36%)
Audiovisual translation	2	52 (27%)
Machine translation literacy	3	37 (19%)

Table 8.4 Top three most useful modules

Module	Ranking	Number/percentage of respondents who selected this module
Machine translation literacy	1	101 (52%)
Tools and resources	2	60 (31%)
Summarization and multimodal communication	3	23 (12%)

Table 8.5 Importance of machine translation literacy for non-language professionals

Response to the question 'For people who are not language professionals, I think that machine translation literacy is . . .'	Number/percentage of respondents
Essential	38 (20%)
Very important	96 (50%)
Moderately important	53 (27%)
Not very important	6 (3%)
Not at all important	0
TOTAL	**193 (100%)**

Meanwhile, in answer to the question about how important they perceive machine translation literacy to be for people who are not professional translators, the majority of respondents (96/193 or 50 per cent) identified this skillset as being 'very important', while an additional 38/193 (20 per cent) went so far as to categorize it as 'essential' (see Table 8.5). No respondents felt that machine translation literacy was 'not at all important'.

Finally, as summarized in Table 8.6, in answer to the question about whether machine translation literacy instruction should be offered to all students at the University of Ottawa, 82/193 (42 per cent) responded 'definitely', while an additional 71/193 (37 per cent) answered 'probably'. Just two respondents said 'probably not' and none selected 'definitely not'.

Table 8.6 Offer of machine translation literacy training

Response to the question 'Do you think that the University of Ottawa should make machine translation literacy training available to all students?'	**Number/percentage of respondents**
Definitely	82 (42%)
Probably	71 (37%)
Maybe	38 (20%)
Probably not	2 (1%)
Definitely not	0
TOTAL	**193 (100%)**

This feedback is very high level rather than detailed, and it does not speak to the specific content of the module on machine translation literacy, nor to the way in which it was taught or assessed; it nonetheless offers some general insights into how relevant students beyond translator training programmes perceive machine translation literacy instruction to be. In future iterations of the course, I hope to be able to capture some more specific feedback on the teaching and learning aspects of this machine translation literacy module.

Concluding remarks

As machine translation technology has become more pervasive in our increasingly digital world, machine translation literacy is emerging as a valuable skillset for people outside the language professions. At many universities, translation courses are delivered mainly to students on professional translator training programmes, and to some extent to students pursuing modern languages degrees. However, my experience developing and delivering a course about translation for non-translation students suggests that students in a wide range of disciplines, and with a broad range of languages, are interested in learning about translation. Moreover, when asked to identify the modules that were the most enjoyable and the most useful, students in the course ranked the module on machine translation literacy in the top three in both cases. In addition, when asked if they perceive that machine translation literacy is useful for non-translation students the majority (134/193 or 70 per cent) identified it as being either essential or very important. Likewise, when asked whether machine translation literacy instruction should be made available to non-translation students, the majority

(153/193 or 79 per cent) indicated that this type of instruction should definitely or probably be offered to all students.

Clearly, mounting a complete course about translation for non-translation students represents a significant undertaking for a university; however, the strong and sustained uptake at the University of Ottawa suggests that it could be worthwhile. In addition, there are options for providing machine translation literacy instruction in particular that require a lower level of investment. Bowker (2020, 2021a, 2021b) describes several options, such as offering a workshop in collaboration with the university library, integrating machine translation literacy into more general information literacy courses that are taught at many universities, or teaming up with English-as-a-second language instructors (or indeed with instructors of other languages) on campus. Translation educators are well placed to help students – both on and beyond translation programmes – to increase their machine translation literacy, and taking steps to do so will help them to fulfil their social responsibility.

References

Alhaisoni, E. and M. Alhaysony (2017), 'An Investigation of Saudi EFL University Students' Attitudes Towards the Use of Google Translate', *International Journal of English Language Education*, 5 (1): 72–82.

Anazawa, R., H. Ishikawa, and K. Takahiro (2013), 'Use of Online Machine Translation for Nursing Literature: A Questionnaire-based Survey', *Open Nursing Journal*, 7 (1): 22–8.

Bar-Hillel, Y. (1960), 'The Present Status of Automatic Translation of Languages'. *Advances in Computers*, 1: 91–163.

Ben-Ghiat, R. (2019), 'Gutting US Foreign Language Education Will Cost Us for Generations', *CNN*, 29 January. Available online: https://www.cnn.com/2019/01/29/opinions/duke-professor-mla-report-foreign-language-departments-ben-ghiat/index.html (accessed 19 October 2021).

Boffey, D. (2013), 'Language Teaching Crisis as 40% of University Departments Face Closure', *The Guardian*, 17 August. Available online: https://www.theguardian.com/education/2013/aug/17/language-teaching-crisis-universities-closure (accessed 19 October 2021).

Bowker, L. (2020), 'Machine Translation Literacy Instruction for International Business Students and Business English Instructors', *Journal of Business and Finance Librarianship*, 25 (1–2): 25–43.

Bowker, L. (2021a), 'Promoting Linguistic Diversity and Inclusion: Incorporating Machine Translation Literacy into Information Literacy Instruction for

Undergraduate Students', *The International Journal of Information, Diversity and Inclusion*, 5 (3): 127–51.

Bowker, L. (2021b), 'Machine Translation Use Outside the Language Industries: A Comparison of Five Delivery Formats for Machine Translation Literacy Instruction', in R. Mitkov, V. Sosoni, J. C. Giguère, E. Murgolo, and E. Deysel (eds), *TRITON 2021: Proceedings of the Conference*, 25–36. Available online: http://triton-conference.org/proceedings/ (accessed 19 October 2021).

Bowker, L. (2023), *Introducing Translation to Non-translators*, London: Routledge.

Bowker, L. and J. Buitrago Ciro (2019), *Machine Translation and Global Research*, Bingley: Emerald.

Bowler, M. (2020), *A Languages Crisis? Higher Education Policy Institute Report 123*, Higher Education Policy Institute. Available online: https://www.hepi.ac.uk/wp-content/uploads/2020/01/HEPI_A-Languages-Crisis_Report-123-FINAL.pdf (accessed 30 January 2023).

Cadwell, P., S. O'Brien, and E. DeLuca (2019), 'More than Tweets: A Critical Reflection on Developing and Testing Crisis Machine Translation Technology', *Translation Spaces*, 8 (2): 300–33.

Castilho, S., J. Moorkens, F. Gaspari, R. Sennrich, V. Sosoni, Y. Georgakopoulou, P. Lohar, A. Way, A. Miceli Barone, and M. Gialama (2017), 'A Comparative Quality Evaluation of PBSMT and NMT Using Professional Translators', *Proceedings of Machine Translation Summit XVI*, Vol. 1, 116–31, Nagoya, September. Available online: https://aclanthology.org/2017.mtsummit-papers.10.pdf

Delorme Benites, A., S. Cotelli Kureth, C. Lehr, and E. Steele (2021). 'Machine Translation Literacy: A Panorama of Practices at Swiss Universities and Implications for Language Teaching', *Proceedings of EUROCALL 2021: CALL & Professionalisation*, 80–7. Available online: https://doi.org/10.14705/rpnet.2021.54.1313

Dorst, A. G., S. Valdez, and H. Bouman (2022), 'Machine Translation in the Multilingual Classroom: How, When and Why do Humanities Students at a Dutch University Use Machine Translation?', *Translation and Translanguaging in Multilingual Contexts*, 8 (1): 49–66.

Duff, P. and D. Li (2009), 'Indigenous, Minority and Heritage Language Education in Canada: Policies, Contexts and Issues', *The Canadian Modern Language Review*, 66 (1): 1–8.

El-Assal, K. (2022), 'Canada Welcomed 450,000 New International Students in 2021, an All-time Record', *Citizenship and Immigration Canada News*, 24 March. Available online: https://www.cicnews.com/2022/03/canada-welcomed-450000-new-international-students-in-2021-an-all-time-record-0323762.html (accessed 30 March 2022).

European Commission (2021), 'European Master's in Translation (EMT)', *European Commission*. Available online: https://ec.europa.eu/info/resources-partners/european-masters-translation-emt_en (accessed 8 December 2022).

Foderaro, L. W. (2010), 'Budget-Cutting Colleges Bid Some Languages Adieu', *The New York Times*, 3 December. Available online: https://www.nytimes.com/2010/12/05/education/05languages.html (accessed 19 October 2021).

Forcada, M. L. (2017), 'Making Sense of Neural Machine Translation', *Translation Spaces*, 6 (2): 291–309.

Hutchins, W. J. and H. L. Somers (1992), *An Introduction to Machine Translation*, London: Academic Press.

International Organization for Standardization (ISO) Technical Committee 37 (2017), *ISO 18587:2017 Translation Services – Post-editing of Machine Translation Output – Requirements*, Geneva: International Organization for Standardization.

Johnson, S. (2019), 'Colleges Lose a "Stunning" 651 Foreign-language Programs in 3 Years', *The Chronicle of Higher Education*, 22 January. Available online: https://www.chronicle.com/article/colleges-lose-a-stunning-651-foreign-language-programs-in-3-years/ (accessed 19 October 2021).

Kenny, D. (ed.) (2022), *Machine Translation for Everyone: Empowering Users in the age of Artificial Intelligence (Translation and Multilingual Natural Language Processing 18)*, Berlin: Language Science Press.

Koehn, P. (2010), *Statistical Machine Translation*, Cambridge: Cambridge University Press.

Koehn, P. (2020), *Neural Machine Translation*, Cambridge: Cambridge University Press.

Lee, S. (2020), 'The Impact of Using Machine Translation on EFL Students' Writing', *Computer Assisted Language Learning*, 33 (3): 157–75.

Lees, O., S. Wijaya,and E. Renaldi (2020), 'Axing Asian language courses from schools and universities may leave Australia underprepared for the future', *ABC News*, 4 December. Available online: https://www.abc.net.au/news/2020-12-05/asian-language-programs-fights-to-stay-hindi-indonesian/12934096 (accessed 11 December 2023).

Lewington, J. (2018), 'How Canadian Universities Help Fight to Save Indigenous Languages', *Maclean's*, 12 October. Available online: https://www.macleans.ca/education/how-canadian-universities-help-fight-to-save-indigenous-languages/ (accessed 19 October 2021).

Loock, R. and S. Léchauguette (2021), 'Machine Translation Literacy and Undergraduate Students in Applied Languages: Report on an Exploratory Study', *Revista Tradumàtica*, 19: 204–25.

Mareschal, G. (2005), 'L'enseignement de la Traduction au Canada', *Meta*, 50 (1): 250–62.

Mason, S. and J. Hajek (2019), 'Representing Language Education in Australian Universities: An Analysis of Press Reporting (2007–2016)', *Language Learning in Higher Education*, 9 (1): 179–202.

Montgomery, S. L. (2013), *Does Science Need a Global Language? English and the Future of Research*, Chicago: University of Chicago Press.

Monti, J. (2020), 'Gender Issues in Machine Translation: An Unsolved Problem?', in L. von Flotow and H. Kamal (eds.), *Routledge Handbook of Translation, Feminism and Gender*, 457–68, London: Routledge.

Mundt, K. and M. Groves (2016), 'A Double-edged Sword: The Merits and the Policy Implications of Google Translate in Higher Education', *European Journal of Higher Education*, 6 (4): 387–401.

Neeley, T. (2012), 'Global Business Speaks English', *Harvard Business Review*, May. Available online: https://hbr.org/2012/05/global-business-speaks-english (accessed 19 October 2021).

Nurminen, M. (2020), 'Raw Machine Translation Use by Patent Professionals: A Case of Distributed Cognition', *Translation, Cognition & Behavior*, 3 (1): 100–21.

Parsons, S. (2020), 'Edinburgh Napier is Third University in Scotland to Cut their Foreign Language Program', *Edinburgh Evening News*, October 5. Available online: https://www.edinburghnews.scotsman.com/education/edinburgh-napier-third-university-scotland-cut-their-foreign-language-programme-2993516 (accessed 19 October 2021).

Ramírez-Sánchez, G., J. A. Pérez-Ortiz, F. Sánchez-Martínez, C. Rossi, D. Kenny, R. Superbo, P. Sánchez-Gijón, and O. Torres-Hostench (2021), 'MultiTraiNMT: Training Materials to Approach Neural Machine Translation from Scratch', in R. Mitkov, V. Sosoni, J. C. Giguère, E. Murgolo, and E. Deysel (eds), *TRITON 2021: Proceedings of the Conference*, 181–6. Available online: http://triton-conference.org/proceedings/ (accessed 19 October 2021).

Redden, E. (2018), 'For International Students, Shifting Choices of Where to Study', *Inside Higher Ed*, 24 August. Available online: https://www.insidehighered.com/news/2018/08/24/international-enrollments-slowing-or-declining-some-top-destination-countries-look (accessed 19 October 2021).

Reisberg, L. (2017), 'Foreign Language Study Should be Mandatory', *Inside Higher Ed*, 14 March. Available online: https://www.insidehighered.com/blogs/world-view/foreign-language-study-should-be-mandatory (accessed 19 October 2021).

Turovsky, B. (2016), 'Ten Years of Google Translate', *Google Blog: The Keyword*. Available online: https://blog.google/products/translate/ten-years-of-google-translate/ (accessed 19 October 2021).

University of Ottawa (2019), *Transformation 2030*, University of Ottawa. Available online: https://transformation2030.uottawa.ca/ (accessed 19 October 2021).

Vieira, L. N. (2020), 'Machine Translation in the News: A Framing Analysis of the Written Press', *Translation Spaces*, 9 (1): 98–122.

Vieira, L. N., M. O'Hagan, and C. O'Sullivan (2021), 'Understanding the Societal Impacts of Machine Translation: A Critical Review of the Literature on Medical and Legal Use Cases', *Information, Communication & Society*, 24 (11): 1515–32.

Wächter, B. and F. Maiworm, eds (2014), *English-Taught Programmes in European Higher Education: The State of Play in 2014*, Bonn: Lemmens. Available online: https://www.lemmens.de/dateien/medien/buecher-ebooks/aca/2014_english_taught.pdf (accessed 19 October 2021).

Way, A. (2020), 'Machine Translation: Where Are We at Today?', in E. Angelone, M. Ehrensberger-Dow, and G. Massey (eds), *The Bloomsbury Companion to Language Industry Studies*, 311–32. London: Bloomsbury.

Weaver, W. (1949), 'Translation'. Available online: www.mt-archive.info/Weaver-1949.pdf (accessed 19 October 2021).

Zhong, W. (2020), 'Recent Developments and Future Directions of Translation and Interpreting Education in China's Mainland', in J. Zhao, D. Li, and L. Tian (eds), *Translation Education: A Tribute to the Establishment of World Interpreter and Translator Training Association (WITTA)*, 83–97. Berlin: Springer.

Index