

ROUTLEDGE



ROUTLEDGE  
HANDBOOKS



# The Routledge Handbook of Translation and Technology

Edited by Minako O'Hagan

# The Routledge Handbook of Translation and Technology

*The Routledge Handbook of Translation and Technology* provides a comprehensive and accessible overview of the dynamically evolving relationship between translation and technology.

Divided into five parts, with an editor's introduction, this volume presents the perspectives of users of translation technologies, and of researchers concerned with issues arising from the increasing interdependency between translation and technology. The chapters in this *Handbook* tackle the advent of technologization at both a technical and a philosophical level, based on industry practice and academic research.

Containing over 30 authoritative, cutting-edge chapters, this is an essential reference and resource for those studying and researching translation and technology. The volume will also be valuable for translators, computational linguists and developers of translation tools.

**Minako O'Hagan**, PhD, is the Discipline Convenor for Translation Studies at the School of Cultures, Languages and Linguistics at the University of Auckland in New Zealand. She specializes in applied translation studies with a technology focus, including game localization and non-professional translation. Her publications include the co-authored *Game Localization* (2013). Her current research interest lies in exploring the nexus of human and machine in translation.

# Routledge Handbooks in Translation and Interpreting Studies

**Routledge Handbooks in Translation and Interpreting Studies** provide comprehensive overviews of the key topics in translation and interpreting studies. All entries for the handbooks are specially commissioned and written by leading scholars in the field. Clear, accessible and carefully edited, *Routledge Handbooks in Translation and Interpreting Studies* are the ideal resource for both advanced undergraduates and postgraduate students.

For a full list of titles in this series, please visit [www.routledge.com/Routledge-Handbooks-in-Translation-and-Interpreting-Studies/book-series/RHTI](http://www.routledge.com/Routledge-Handbooks-in-Translation-and-Interpreting-Studies/book-series/RHTI).

## **The Routledge Handbook of Audiovisual Translation**

*Edited by Luis Pérez-González*

## **The Routledge Handbook of Translation and Philosophy**

*Edited by Piers Rawling and Philip Wilson*

## **The Routledge Handbook of Literary Translation**

*Edited by Kelly Washbourne and Ben Van Wyke*

## **The Routledge Handbook of Translation and Politics**

*Edited by Fruela Fernández and Jonathan Evans*

## **The Routledge Handbook of Translation and Culture**

*Edited by Sue-Ann Harding and Ovidi Carbonell Cortés*

## **The Routledge Handbook of Translation Studies and Linguistics**

*Edited by Kirsten Malmkjaer*

## **The Routledge Handbook of Translation and Pragmatics**

*Edited by Rebecca Tipton and Louisa Desilla*

## **The Routledge Handbook of Translation and Technology**

*Edited by Minako O'Hagan*

---

# The Routledge Handbook of Translation and Technology

*Edited by Minako O'Hagan*

First published 2020  
by Routledge  
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge  
52 Vanderbilt Avenue, New York, NY 10017

*Routledge is an imprint of the Taylor & Francis Group, an informa business*

© 2020 selection and editorial matter, Minako O'Hagan; individual chapters, the contributors

The right of Minako O'Hagan to be identified as the author of the editorial material, and of the authors for their individual chapters, has been asserted in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

*Trademark notice:* Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

*British Library Cataloguing-in-Publication Data*

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

*Library of Congress Cataloging-in-Publication Data*

Names: O'Hagan, Minako, 1959– editor.

Title: The Routledge handbook of translation and technology / edited by Minako O'Hagan.

Description: Abingdon, Oxon ; New York, NY : Routledge, 2020. |

Series: Routledge handbooks in translation and interpreting studies |

Includes bibliographic references and index. | Summary: "The Routledge

Handbook of Translation and Technology provides a comprehensive and

accessible overview of the dynamically evolving relationship between translation

and technology. Divided into five parts, with an editor's introduction, this volume

presents the perspectives of users of translation technologies, and of researchers

concerned with issues arising from the increasing interdependency between translation

and technology. The chapters in this handbook tackle the advent of technologization

at both a technical and a philosophical level, based on industry practice and academic

research. Containing over thirty authoritative, cutting-edge chapters, this is an essential

reference and resource for those studying and researching translation and technology.

The volume will also be valuable for translators, computational linguists and developers

of translation tools"–Provided by publisher.

Identifiers: LCCN 2019014720 | ISBN 9781138232846 (hardcover) | ISBN 9781315311258 (ebook)

Subjects: LCSH: Translating and interpreting–Technological innovations. |

Translating and interpreting. | Machine translating.

Classification: LCC P306.97.T73 .R69 2020 | DDC 418/.020285–dc23

LC record available at <https://lcn.loc.gov/2019014720>

ISBN: 978-1-138-23284-6 (hbk)

ISBN: 978-1-315-31125-8 (ebk)

Typeset in Times New Roman

by Newgen Publishing UK

*The book is dedicated to Professor Emeritus John Tiffin  
and the late Dr. Peter Toma, who chose New Zealand to test their visions  
for the future of education and translation.*



**Taylor & Francis**

Taylor & Francis Group

<http://taylorandfrancis.com>

# Contents

---

<i>Illustrations</i>	<i>xi</i>
<i>Contributors</i>	<i>xiii</i>
<i>Acknowledgements</i>	<i>xix</i>
1 Introduction: Translation and technology: disruptive entanglement of human and machine <i>Minako O'Hagan</i>	1
<b>PART I</b>	
<b>Translation and technology: defining underlying technology – present and future</b>	<b>19</b>
2 Standards for the language, translation and localization industry <i>Sue Ellen Wright</i>	21
3 XML for translation technology <i>Johann Roturier</i>	45
4 Terminology extraction and management <i>Kyo Kageura and Elizabeth Marshman</i>	61
5 Building and using parallel text for translation <i>Michel Simard</i>	78
6 Speech recognition and synthesis technologies in the translation workflow <i>Dragoş Ciobanu and Alina Secară</i>	91
<b>PART II</b>	
<b>Translation and technology: users' perspectives</b>	<b>107</b>
7 Multinational language service provider as user <i>Bert Esselink</i>	109
	vii

## Contents

8	Applications of technology in the Patent Cooperation Treaty (PCT) Translation Division of the World Intellectual Property Organization (WIPO) <i>Colm Caffrey and Cristina Valentini</i>	127
9	Small and medium-sized enterprise (SME) translation service provider as technology user: translation in New Zealand <i>Patrick King</i>	148
10	Freelance translators' perspectives <i>Jost Zetzsche</i>	166
11	Language learners and non-professional translators as users <i>Masaru Yamada</i>	183
<b>PART III</b>		
<b>Translation and technology: application in a specific context – shaping practice</b>		<b>201</b>
12	Technology, technical translation and localization <i>Debbie Folaron</i>	203
13	Technology and game localization: translation behind the screens <i>Nathan Altice</i>	220
14	Technology and non-professional translation <i>Miguel A. Jiménez-Crespo</i>	239
15	Technological advances in audiovisual translation <i>Jorge Díaz Cintas and Serenella Massidda</i>	255
16	Technology and interpreting <i>Sabine Braun</i>	271
17	Technology and sign language interpreting <i>Peter Llewellyn-Jones</i>	289
18	Translation technology and disaster management <i>Sharon O'Brien</i>	304

19	Post-editing of machine translation <i>Lucas Nunes Vieira</i>	319
<b>PART IV</b>		
	<b>Translation and technology: research foci and methodologies</b>	<b>337</b>
20	Translation technology evaluation research <i>Stephen Doherty</i>	339
21	Translation workplace-based research <i>Maureen Ehrensberger-Dow and Gary Massey</i>	354
22	Translation technology research and human–computer interaction (HCI) <i>Samuel Lüubli and Spence Green</i>	370
23	Sociological approaches to translation technology <i>Maeve Olohan</i>	384
24	Translation technology research with eye tracking <i>Arnt Lykke Jakobsen</i>	398
<b>PART V</b>		
	<b>Translation and technology: overarching issues</b>	<b>417</b>
25	Future of machine translation: musings on Weaver’s memo <i>Alan K. Melby</i>	419
26	Quality <i>Anthony Pym</i>	437
27	Fit-for-purpose translation <i>Lynne Bowker</i>	453
28	Copyright and the re-use of translation as data <i>Joss Moorkens and David Lewis</i>	469
29	Media accessibility and accessible design <i>Aline Remael and Nina Reviers</i>	482

Contents

30 Technology and translator training 498  
*Dorothy Kenny*

31 Translation, technology and climate change 516  
*Michael Cronin*

*Index* 531

# Illustrations

---

## Figures

7.1	TMS environment and workflow	112
8.1	Proportion of PCT Publications by Application Language (2007, 2017)	128
8.2a	Subject field browser in WIPO Pearl	140
8.2b	Concept map display in WIPO Pearl	140
8.3	WIPO Pearl portal design	141
8.4	Combined keyword search in PATENTSCOPE	142
8.5	Result of side-by-side comparison of PBSMT and NMT-translated sentences for English–Japanese	143
8.6	BLEU score comparison between WIPO Translate NMT and GNMT	143
9.1	Timeline of translation technology changes in New Zealand	153
11.1	Error distributions: human translator vs neural machine translation	195
13.1	A detail of FCEUX’s PPU Viewer displays the background (left) and sprite (right) tiles loaded from CHR-ROM during <i>Dragon Quest’s</i> opening throne room scene	227
13.2	<i>Dragon Quest’s</i> title screen (left) and its accompanying background tiles (right) loaded in CHR-ROM	228
13.3	The <i>Dragon Quest</i> battle screen	229
13.4	Many Famicom games incorporated <i>kanji</i> in their title screens, but ファザナドゥ (aka Faxanadu) was one of the few titles that intermixed <i>kanji</i> , <i>katakana</i> and <i>hiragana</i> in-game	231
13.5	The contents of <i>Dragon Warrior’s</i> tile memory during overworld sections show that the localization’s tile differences are primarily linguistic	232
13.6	Background tiles for the official <i>Dragon Warrior</i> localization, followed by its Swedish, Russian and Arabic fan translations	234
14.1	Example of approaches to ‘unit granularity selection’ (Morera, Filip and Collins 2014). A summary of how NPT participants access the text to translation during collaborative translation processes	248
16.1	Modalities of distance interpreting	272

16.2	Participant distribution in distance interpreting	273
19.1	Spectrum of human agency in the post-editing process	328
22.1	Automatic suggestions in IMT systems. The underlying MT engine predicts words (in grey) that match the words already translated by the user (in black). The suggestions are continuously updated as the user enters new words	372
22.2	Common UI layouts in segment-level translation editors	372
22.3	Information visualization through highlighting	373
22.4	Methods for assessing user experience	376
25.1	Diagram of the Vauquois triangle	424
27.1	The Iron Triangle illustrating the triple constraint of time, cost and quality	459

**Tables**

4.1	Basic patterns of terms in English	63
4.2	Contingency table for calculating unithood of complex terms with two elements	63
8.1	Training Data for WIPO Translate NMT	129
8.2	Functions of abstracts and common phrasing	131
8.3	Example of exact (top three) and fuzzy (bottom two) matches in title of the invention	132
8.4	Example of manual sub-segmentation of Chinese abstract body	133
8.5	Translations for ‘sealing device’	135
8.6	Classification of terms extracted in IPC Class B62	137
8.7	Search result in an English/French parallel corpus	138
8.8	Term equivalents and subject fields	138
8.9	Example of how to create multilingual records	139
8.10	HTER for WIPO Translate PBSMT and NMT systems for English to French post-editing	144
9.1	Assessment chart of MT suitability based on Gerber (2009)	157
9.2	Results of survey of NMT usage at NZTC, based on 34 responses	161
20.1	Overview of methods in translation technology evaluation research	342
25.1	Summary of the relationship between Weaver’s seeds, the enabling developments and the three paradigms	429

# Contributors

---

**Nathan Altice** is Teaching Professor of Computational Media at the University of California, Santa Cruz, where he teaches game design, game development and digital audio. His research focuses on computational platforms, the history of games and Japanese board games. His first book, *I AM ERROR*, was published by MIT Press in 2015.

**Lynne Bowker** holds a PhD in Language Engineering from the University of Manchester Institute for Science and Technology. She is Full Professor at the University of Ottawa's School of Translation and Interpretation where she teaches and conducts research in areas relating to specialized translation, language for special purposes and translation technologies. She is the author of *Computer-Aided Translation Technologies* (University of Ottawa Press, 2002) and *Machine Translation & Global Research* (Emerald, 2019).

**Sabine Braun** is Full Professor of Translation Studies and Director of the Centre for Translation Studies at the University of Surrey (UK). She is a member of the German Institute for Standardization Working Group on Interpreting Services and Technologies. Her research focuses on new methods, modalities and socio-technological practices of translation and interpreting. She has led several multinational research projects relating to video-mediated interpreting.

**Colm Caffrey** is a Translator-Reviser in the PCT Translation Division of the World Intellectual Property Organization, where his tasks involve the translation and revision of patent related documents, the training of fellows and the external liaison with translation providers. He holds a PhD in Translation Studies from Dublin City University. His current research interests include translator training and the integration of CAT tools into the workplace.

**Dragoş Ciobanu** is an Associate Professor in Translation Studies in the University of Leeds Centre for Translation Studies, specializing in Computer-Assisted Translation (CAT) tools, collaborative translation and training practices, as well as speech technologies in the translation workflow. He collaborates with Language Service Providers from around the world and trains heads of translation units and senior linguists from EU and UN Institutions to maximize the use of language and project management technologies.

**Michael Cronin** is 1776 Professor of French at Trinity College Dublin and Director of the Trinity Centre for Literary and Cultural Translation. He is a Member of the Royal Irish Academy, the Academia Europaea and an Honorary Member of the Irish Translators

and Interpreters Association. He is series editor of the Routledge New Perspectives in Translation and Interpreting Studies and he is Editor-in-Chief of the translation journal *MTM*.

**Jorge Díaz Cintas** is Professor of Translation and founder–director of the Centre for Translation Studies (CenTraS) at University College London. He is the Chief Editor of the series *New Trends in Translation Studies* and is a member of the Language Industry Expert Group of the European Commission. He is the recipient of the Jan Ivarsson Award (2014) and the Xènia Martínez Award (2015).

**Stephen Doherty**, PhD, is a Senior Lecturer in Linguistics, Interpreting and Translation in the School of Humanities & Languages at the University of New South Wales, Sydney, Australia. His research is based in language, cognition and technology with a focus on the psychology of language and technology. His recent publications include a co-edited volume *Translation Quality Assessment: From Principles to Practice* (Springer, 2018).

**Maureen Ehrensberger-Dow**, PhD, is Professor of Translation Studies in the ZHAW Institute of Translation and Interpreting. She is currently co-investigator of the project Cognitive Load in Interpreting and Translation, funded by the Swiss National Science Foundation. Her publications on workplace research include ‘Challenges of translation process research at the workplace’ (MonTI, 2014) and ‘Socio-technical issues in professional translation practice’ (with Gary Massey, *Translation Spaces*, 2017).

**Bert Esselink** graduated from the Dutch School of Translation and Interpreting in Maastricht and has been involved in multilingual software and content projects for over two decades. He worked as a localization specialist, focusing on project management, language technology, multilingual marketing and global content management. His book *A Practical Guide to Localization* is used throughout the localization and translation industry. After working for many years with Lionbridge in Amsterdam, he joined SDL as Strategic Account Director in 2017.

**Debbie Folaron** holds a PhD in Translation Theory/Comparative Literature from Binghamton University. She worked as Telecommunications Manager of a global shipping company in Barcelona and as Language Technology Manager at Eriksen Translations in NYC. She is Associate Professor at Concordia University in Montreal where she teaches localization, project management, translation technologies and translation theories. Her research focuses on translation in the digital age, with special emphasis on under-represented language-cultures.

**Spence Green** graduated from Stanford with a PhD in computer science under the direction of Chris Manning and Jeff Heer. He has published papers on machine translation, language parsing and mixed-initiative systems and has given talks on translator productivity. Spence is CEO of Lilt.

**Arnt Lykke Jakobsen** is Professor Emeritus of Translation and Translation Technology at Copenhagen Business School. In 1995 he invented *Translog*, and in 2005 he established the CBS Centre for Research and Innovation in Translation and Translation Technology

(CRITT), which he directed until his retirement at the end of 2013. From 2006 to 2009 he was a principal investigator in the EU Eye-to-IT project.

**Miguel A. Jiménez-Crespo** holds a PhD in Translation and Interpreting Studies from the University of Granada, Spain. He is a Professor in the Department of Spanish and Portuguese, Rutgers University, and directs the MA and BA programmes in Spanish–English Translation and Interpreting. His publications include *Crowdsourcing and Online Collaborative Translations: Expanding the Limits of Translation Studies* (John Benjamins Publishing, 2017) and *Translation and Web Localization* (Routledge, 2013).

**Kyo Kageura**, PhD, is a professor at the Graduate School of Education/Interdisciplinary Initiative in Information Studies, University of Tokyo. He serves as an editor of the journal *Terminology* and has published extensively across different fields, including the monograph *The Quantitative Analysis of the Dynamics and Structure of Terminologies* (John Benjamins Publishing, 2012). He runs Minna no Hon'yaku and Minna no Hon'yaku Jisshu – translation aid portals.

**Dorothy Kenny** is Full Professor of Translation Studies in the School of Applied Language and Intercultural Studies and a member of the Centre for Translation and Textual Studies at Dublin City University. Her publications include *Lexis and Creativity in Translation: A Corpus-Based Study* (St. Jerome, 2001) and the edited volume *Human Issues in Translation Technology* (Routledge, 2017).

**Patrick King** is the chief editor and a founding director of the New Zealand Translation Centre Ltd. (NZTC). With an MA in German and Russian, he worked as a translator and editor at the government Translation Service until 1985, when he and three colleagues founded NZTC. He is a founding member and past president of the New Zealand Society of Translators and Interpreters.

**Samuel Läubli** holds an MSc in Artificial Intelligence from the University of Edinburgh. He is a PhD Candidate in Interactive Machine Translation at the University of Zurich, supervised by Martin Volk and Rico Sennrich.

**Dave Lewis** is an associate professor of Computer Science at Trinity College Dublin and an Associate Director of the ADAPT Centre with responsibility for its collaborative research with industry and the Centre's role in the EDGE Marie-Curie Postdoctoral Fellowship programme. He is an ADAPT Principal Investigator, leading its research on Transparent Data Governance. With over 160 peer-reviewed publications, his research addresses the use of semantic models to manage the data protection and data ethics issues raised by digital content processing.

**Peter Llewellyn-Jones** qualified as a British Sign Language-English interpreter in 1972 and worked in both conference and Public Service domains. He has developed and taught sign interpreting courses for three UK universities and was the Programme Director of the MA in Interpreting and Translation: BSL-English at the University of Leeds from 2003 to 2012. His extensive publications include the co-authored *Redefining the Role of the Community Interpreter: The Concept of Role Space* (SLI Press, 2014).

**Elizabeth Marshman** holds a PhD in Translation (Terminology option) from the Université de Montréal. She is Associate Professor at the University of Ottawa's School of Translation and Interpretation, where she teaches primarily in areas relating to terminology and translation technologies. Her main research interests include corpus-based and computer-assisted terminology research, user-centred evaluation of terminology resources and translation technologies, and the place of technologies in translator training.

**Gary Massey**, PhD, is Professor and Director of the ZHAW Institute of Translation and Interpreting. He was co-investigator on the interdisciplinary project *Cognitive and Physical Ergonomics of Translation* and the preceding *Capturing Translation Processes* project. Currently co-editing the *Bloomsbury Companion to Language Industry Studies* (with Erik Angelone and Maureen Ehrensberger-Dow), his publications on workplace research include 'Socio-technical issues in professional translation practice' (with Maureen Ehrensberger-Dow, *Translation Spaces*, 2017).

**Serenella Massidda** is a Lecturer in Audiovisual Translation at Roehampton University and Honorary Research Associate at CenTraS at UCL. She holds a European Doctorate in Audiovisual Translation by the University of Sassari. She is the author of *Audiovisual Translation in the Digital Age – The Italian Fansubbing Phenomenon* (Palgrave MacMillan, 2015). She is a professional translator and subtitler, and a member of ESIST (European Association for Studies in Screen Translation).

**Alan K. Melby** is Professor Emeritus of Linguistics at Brigham Young University, where he serves as Associate Director of the Translation Research Group. He is a certified French-to-English translator, Vice-President of the International Federation of Translators and President of the non-profit LTAC Global. He began working on machine translation in 1970 with extensive publications, including *The Possibility of Language* (John Benjamins Publishing, 1995), in collaboration with C. Terry Warner.

**Joss Moorkens** is an Assistant Professor at the School of Applied Language and Intercultural Studies at Dublin City University and a researcher affiliated with the ADAPT Centre and the Centre for Translation and Textual Studies. He has authored articles and chapters on translation technology, MT post-editing, translation evaluation, translator precarity and translation technology standards. He co-edited the book *Translation Quality Assessment: From Principles to Practice* (Springer, 2018).

**Sharon O'Brien** has a PhD in the topic of controlled authoring, machine translation and post-editing. She is Professor of Translation Studies in the School of Applied Language and Intercultural Studies at Dublin City University. Her research centres on translator-computer interaction and crisis translation with numerous publications in translation process research. She is the Principal Investigator of a H2020 EU-funded RISE project – The International Network in Crisis Translation (INTERACT).

**Minako O'Hagan**, PhD, is the Discipline Convenor for Translation Studies at the School of Cultures, Languages and Linguistics at the University of Auckland. She specializes in applied translation studies with a technology focus, including game localization and non-professional translation. Her publications include the co-authored *Game Localization*

(John Benjamins Publishing, 2013). Her current research interest lies in exploring the nexus of human and machine in translation.

**Maevae Olohan** is Co-Director of the Centre for Translation and Intercultural Studies, University of Manchester, UK. Her research centres on the socio-materiality of translation practices, past and present. She is author of *Scientific and Technical Translation* (2016) and *Introducing Corpora in Translation Studies* (2004). Her forthcoming monograph on practice theory and translation will be published in Routledge's *Translation Theories Explored* series.

**Anthony Pym** holds a PhD in Sociology from the École des Hautes Études en Sciences Sociales, Paris. He currently teaches at the University of Melbourne. He is Distinguished Professor of Translation and Intercultural Studies at the Rovira i Virgili in Tarragona in Spain and Extraordinary Professor at Stellenbosch University in South Africa. He is the author of some 11 books in the general area of translation and intercultural communication.

**Aline Remael** (TricS research group) is Professor Emeritus at the Department of Applied Linguistics/Translators and Interpreters of the University of Antwerp where she is the co-founder of the new Expertise Centre for Accessible Media and Culture. Her main research interests and publications are in AVT/Media Accessibility. She is currently a partner in two European accessibility projects (ADLAB PRO and ILSA). In 2018 she received the ESIST Jan Ivarsson Award.

**Nina Reviere** (TricS research group) holds a PhD in Translation Studies from the University of Antwerp. She is a practitioner and researcher in the field of Media Accessibility. She is the coordinator of the newly founded Expertise Centre for Accessible Media and Culture in the University of Antwerp. She is an Editorial Board Member of the *Journal of Audiovisual Translation* and Steering Committee member for *Languages and the Media*.

**Johann Roturier** holds an MA and PhD in Translation Studies from Dublin City University. He is a Senior Principal Research Engineer in Symantec Research Labs where his research is conducted at the intersection of natural language processing, localization and usable security and privacy. He is the author of *Localizing Apps* (Routledge, 2015).

**Alina Secară** is an Associate Professor in Translation Studies in the University of Leeds Centre for Translation Studies, where she teaches subtitling, captioning and computer-assisted translation. She is also a freelance Stagertext accredited theatre captioner, working with theatres across the UK to create captions for the deaf and hard-of-hearing. She is currently co-managing the Leeds contribution to the EU-funded DigiLing Project (2016–2019) to create multilingual, multimedia e-learning resources for digital linguists.

**Michel Simard** holds an MSc in Computer Science from McGill University and a PhD in Computer Science from the Université de Montréal. He is a senior research officer at the National Research Council Canada, where he conducts research in areas relating to machine translation, computer-assisted translation and multilingual text processing.

## Contributors

**Cristina Valentini** is Head of the Terminology Unit in the PCT Translation Division of the World Intellectual Property Organization. Her tasks involve the design and development of terminology databases and guidelines, and terminology workflow management. She holds a PhD in Languages, Cultures and Intercultural Communication from the University of Bologna where she worked as a researcher on topics including occupational health and safety terminology, terminology databases design and multimedia corpora design.

**Lucas Nunes Vieira** is a Lecturer in Translation Studies with Technology at the University of Bristol, UK. He researches the use of machine translation in human translation workflows, with a focus on the cognitive aspects that surround the use of technology in translation tasks and how this affects processes, products and attitudes. Lucas holds a PhD on machine translation post-editing from Newcastle University (UK).

**Sue Ellen Wright** is an emerita professor from Kent State University, where she has taught computer applications for translators and German to English scientific and technical translation. She is Chair of the US Technical Advisory Group to ANSI for ISO TC 37, *Language and Terminology*. Her extensive publications on computer assisted terminology management include the *Handbook of Terminology Management* with Professor Gerhard Budin. She holds her doctorate from Washington University.

**Masaru Yamada**, PhD, is a Professor in the Faculty of Foreign Language Studies at Kansai University. He specializes in Translation and Interpreting Studies with a focus on translation process research (TPR), including machine translation and post-editing, and translation in language teaching (TILT).

**Jost Zetsche** is a translation technology consultant, an author and speaker on various aspects of translation and an American Translators Association (ATA)-certified English-to-German technical translator. He holds a PhD in Chinese translation history and linguistics from the University of Hamburg. In 2018 he was awarded an honorary membership of ATA. He is on the editorial board for *MultiLingual*, the *ATA Chronicle* and the *Journal of Internationalization and Localization*.

# Acknowledgements

---

The editor wishes to express her gratitude first and foremost to the contributors across a large geographical span who entrusted me with the fruits of their intellectual labour through this project. I am also grateful for peer-reviewers, among them particularly to Professor Dorothy Kenny. Special thanks go to the Routledge team, especially to Louisa Semlyen whose encouragement and perseverance led to this publication. I am grateful to the School of Cultures, Languages and Linguistics at the University of Auckland for its assistance towards this project. I am grateful to Kiri O'Hagan for her copy-editing assistance.

I would also like to acknowledge the initial contribution made by Professor Tony Hartley.



**Taylor & Francis**

Taylor & Francis Group

<http://taylorandfrancis.com>

# Introduction

## Translation and technology: disruptive entanglement of human and machine

Minako O'Hagan

---

### Background

This book builds on the increasing evidence of the impact of technology on contemporary translation, which serves diverse communicative situations across languages, cultures and modalities. The 2018 European Language Industry Association (ELIA) survey of over 1,200 respondents across 55 countries highlighted 2018 'as the year in which more than 50% of both the companies and the individual language professionals reported as using MT' (ELIA 2018: n.p.). Although the ELIA report is cautious not to overstate the penetration of MT, concluding that the use of MT in the translation industry is not yet mainstream, it is clear that technology has already profoundly affected the way translation is produced. Similarly, the wider public is exposed to machine translated texts of varying quality in different scenarios, including user-generated content (e.g., social media postings) and information gisting for personal use (e.g., hotel reviews). Furthermore, portions of the increased production and circulation of translations are attributable to the work of fans, volunteers or activists who have different backgrounds and motivations, yet are operating in parallel to their professional counterparts. The increased visibility of *non-professional translation* (NPT) can be traced to the availability of technology-supported social and collaborative platforms, on which NPT typically operates (see Chapter 14 by Jiménez-Crespo). In this way, technology has contributed to translation of diverse types and quality, accompanied by an increasing awareness in society at large of translation and the role played by technologies in the translation process. More recently, the newest MT paradigm, neural MT (NMT) is making inroads into translation practice and adding to substantial research interests in Translation Studies (TS), as demonstrated in this volume. The influence of technology, ranging from translation-specific technologies such as MT to more general-purpose speech technologies and cloud computing, is far-reaching and calls into question some of the assumptions about who should translate, how and to what level of quality.

Commercially viable translation today is all *computer-aided (or -assisted) translation* (CAT) and has been for some time. This is a term which comes across as somewhat redundant, given the ubiquitous use of computers in text production practices in

general, except that the extent and the nature of the computer aid is constantly shifting. Another frequently used term in the translation industry is *translation environment tools* (TEnts), which conveys an image of translators' work surroundings being enveloped by technology. Among the newer terms coming into use is *augmented translation* (AT), introduced by Common Sense Advisory (Lommel 2018). AT puts the human translator in the centre (Kenny 2018), supported by an advanced suite of technologies, including automated content enrichment (ACE). This allows automatic searches of relevant information associated with the source content and informs the translator and MT to generate better translation (Lommel *ibid.*). AT and ACE concepts align with AI-supported medicine, which augments human expert judgement with rapid access to vast and relevant key information (see Susskind and Susskind 2015). Such complex technological infrastructure shaping macro and micro translation environments in turn relies on ongoing behind-the-scenes standardization work (see Chapters 2 and 3 by Wright and Roturier respectively) to ensure that all technological elements meet required standards and can therefore interoperate. However, the technology-driven *modus operandi* and technology-based infrastructure on which translation increasingly rests adds to quality concerns (see Pym in Chapter 26). For example, according to nearly 2,800 respondents to the *SDL Translation Technology Insight Survey* (SDL 2016), quality is currently of the utmost concern for the translation industry.

These snapshots highlight that the human–machine relationship is in a state of flux, with uncharted paths ahead. While human translation shapes and is shaped by technologies, we do not know exactly how this process will unfold. This contributes to a sense of uncertainty among professional translators, which Vieira (2018), following Akst (2013), calls 'automation anxiety' (also see Kenny in Chapter 30). In the midst of ongoing technological transformation, this collected volume is not about translation technology *per se*. Rather, it is about understanding the dynamic relationship being formed between translation and technology from a range of perspectives. In doing so, it aims to increase our awareness of how contemporary translation is evolving and what it means to be a translator, as the co-existence of human and machine could be qualitatively different in the near future. Such a theme has become a major agenda of the 21st century across different types of work, particularly with AI beginning to affect areas previously considered only fit for humans (Susskind and Susskind 2015, also see Chapter 30 by Kenny). This volume attempts to tackle the topic both at a technical and a philosophical level, based on industry practice and academic research, to present a balanced perspective with TS contributions to a dialogue of global importance.

## **Historical contexts of research on the nexus of human and machine in translation**

For translation, the explicit connection with 'the machine' started in earnest in the 1950s, with research and development (R&D) of MT as a new field for the non-numerical application of computers instigated by the Weaver memo (Weaver 1949) (see Melby in Chapter 25). However, as is well known, the 1966 Automatic Language Processing Advisory Committee (ALPAC) report put an abrupt end to MT R&D, especially in the US, for nearly a decade. Despite this, the frequent references to the ALPAC report in this volume and elsewhere are arguably evidence of its continuing legacy, which is perhaps not all short-sighted and misguided. For example, its support for 'machine-aided translation' has become mainstream in the translation industry under the banner of CAT. Martin

Kay's *translator's amanuensis* (Kay 1980/1997) envisioned an incremental adaptive electronic aid for the human translator. Similarly, Alan K. Melby's work on the *translator's workstation* (Melby 1981) embodied a workbench integrating discrete levels of machine aid. Reviewing these pioneers' concepts, Hutchins (1998: 11) highlighted how, in both cases, the human translator had been placed in control as someone who would use such tools in ways s/he 'personally found most efficient'. The questioning of this centrality of human translators in today's transforming translation workflow (Kenny 2018), further validates the aim of this volume to investigate the relationship between human and machine and its ramifications.

Initially CAT tended to be distinguished from MT on the assumption that in the former, it is the human who translates (e.g., Bowker 2002, Somers 2003), whereas MT is automatic computer translation without human intervention. However, this division has become blurred as MT is increasingly integrated into CAT environments (see Kenny in Chapter 30) where the human translator is presented with translation proposals from (human produced) translation memory (TM) matches, together with MT outputs. Similarly, the increasing practice of post-editing of MT (PEMT) is reflected in a growing body of research which has rapidly reached a critical mass especially in translation process research (see collected volumes such as O'Brien 2014, Carl, Bangalore and Schaeffer 2016).

There has been considerable progress made to address the earlier disconnect between MT research and research in TS, although the tendency to exclude professional human translators is still observable 'in certain quarters of MT research' (Kenny 2018: 439). Initially MT research focused on the application of computers to human language, with computer scientists and engineers 'knowingly or unknowingly' attempting to 'simplify the translation process' or 'downplay the nuances of human language' (Giammarresi and Lapalme 2016: 218). But the lack of cross-fertilization can also be blamed on the TS camp, with too few scholars interested in translation technology to widen the scope of translation theory, so that it could consider the increasing integration of technology into the translation process (O'Hagan 2013, Jakobsen and Misa-Lao 2017). In fact, the connection between translation research and MT research can be traced to the 1960s when the idea of equivalence relationships between source and target texts was explored by linguists such as Catford (1965). In particular, Catford's idea of a translation rule as 'an extrapolation of the probability values of textual translation equivalents' (1965: 31) is of direct relevance to subsequent data-driven approaches to MT (Kenny forthcoming), which are based on the use of parallel texts (or bi-texts) (see Simard in Chapter 5). In the 1960s, when Chomsky's linguistic theory (Generative Grammar) was exerting its influence, including on MT, Eugene Nida was among the few early translation theorists cognizant of MT research, and related to it in his foundation work *Toward a Science of Translating* (Nida 1964). In his endeavour to bring theorizing about translation into the scientific arena, Nida applied Chomskian linguistics and the information theory approach to communication (Nida 1964, Nida and Taber 1969). It is relevant to recall the fact that MT R&D precede the development of TS; it was only in 1972 that James Holmes (1972/1988) named the discipline as 'Translation Studies' (abbreviated as TS in this article) and laid the foundations for theorizing translation to 'explain and predict' translation with 'description' as the first step. In the 1980s TS was shifting away from a linguistic focus to a consideration of broader contexts through functionalism. Attention moved from the source to the target text and translation as action, before the cultural turn in the 1990s moved human translation largely outside the scope of interest of MT circles.

Into the 1990s and 2000s technologies played a key role in empirical TS research by providing research tools, including some for corpus analysis. Other tools, such as keyboard logging (e.g., Translog originally developed by Arnt Jakobsen at the Copenhagen Business School in the late 1990s) and eye tracking (see Jakobsen in Chapter 24), were also introduced more widely into TS, and these have been used to better understand translator behaviours and the behaviours of translation users in the context of translation reception; for example, in audiovisual translation (AVT) (see Kruger 2018). In particular, these research tools contributed to the further development of cognitive translation studies as a specialized field of research (see Schwieter and Ferreira 2017), one which is now set to probe neural representation with non-invasive neuroimaging techniques, such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) (see Shreve and Diamond 2016: 155).

This brief look back at the trajectory of the connection between translation and technology shows increasing 'border crossings' (Gambier and van Doorslaer 2016) to neighbouring disciplines such as computer science, computational linguistics and now neuroscience.

## **Aim and scope of the publication**

The spread of computers across global markets gave rise to new areas of practice and research in TS, such as localization (see Folaron in Chapter 12). This saw TS scholars engaging more fully in theorizing about technologies by tapping into sociological, cultural or philosophical aspects (see Chapters 23 and 31 by Olohan and Cronin respectively), on the one hand, and cognitive or usability/ergonomic dimensions on the other (see Chapters 21 and 24 by Ehrensberger-Dow and Murphy; and Jakobsen respectively). There is also a large body of knowledge being accumulated in translator training and education focused on technology (see Kenny in Chapter 30). Furthermore, as a result of technological advances, research-led practices are becoming more common in fields such as accessibility and universal design (see Remael and Revijs in Chapter 29). In this way, technology more than anything else started to bring together the interests of academy and industry. Technological dimensions continue to present fresh scope to bridge the gap between translation theory and practice, ideally to respond to ever-present translator suspicions as to the usefulness of theory in actual translation practice – a topic earlier addressed in Chesterman and Wagner (2002) and more recently in Polizzotti (2018). As demonstrated in this volume, the exploration of the relationship between technology and translation is leading to a fresh examination of contemporary translation benefitting not only translators as users of technologies but also those who develop and research translation technology. It is hoped that this volume contributes critical insight into the complex symbiosis between humans and machines so that translation (and interpreting, which is covered to a limited extent in this volume) can serve increasingly diverse communication needs in the best and most sustainable way.

With the above overall goal of the publication, the *Handbook* has a number of specific features. First, it is designed to represent the interests of different stakeholders in the translation industry. The fragmented nature of the translation industry is recognized as it affects the level of implementation and the types of technologies used in translation. The translation industry consists of a large population of freelance translators (see Zetzsche in Chapter 10) and language service providers (LSPs) which range from small-and-medium-sized (see King in Chapter 9) to multinational vendors (see Esselink

in Chapter 7). In addition, often well-resourced public international organizations (see Caffrey and Valentini in Chapter 8) play an important role as early adopters of new technologies. Although not officially part of the industry, non-professional translation is also contributing to translation production, forming part of a participatory culture (Chapters 13 and 14 by Altice and Jiménez-Crespo, respectively). Similarly, the use of translation technology in (second) language learning is part of the picture in the technology and translation alliance (see Chapter 11 by Yamada). The volume therefore reflects different settings for technology uses according to the different segments of the industry as users of translation technology, encompassing contributors who reside outside academia. Secondly, this publication attempts to make sense of the current position of technology from diachronic perspectives. What is considered new technology often had a prior incarnation as a rudimentary prototype or an embryonic concept which needed further maturing, perhaps requiring relevant surrounding technologies and conditions. While historical approaches are well explored in TS research in general, their application in the context of translation technology research has not been traversed to the same extent. In the context of MT, John Hutchins was the first to demonstrate the merit of a historical approach with his comprehensively chronicled *Machine Translation: past, present, future* (Hutchins 1986). The *Routledge Encyclopedia of Translation Technology* (Chan 2015) is a more recent example also with regional foci. Among the many chapters in the present volume which provide a historical trajectory, historical perspectives are more applicable and prominent in certain chapters. For example, Sue-Ellen Wright in her chapter on Standards follows periodization, drawing on Galinski (2004 cited in Wright) to cast a spotlight on key phases of the evolution of approaches and applications of standardization across language, translation and the localization industry. Similarly, Debbie Folaron (Chapter 12), in discussing technical translation as an established practice and localization as a relatively new addition within TS, traces their historical trajectories. The historical approach contextualizes and recontextualizes the development of specialized translation practices in dynamic interaction with technology. Such an approach allows Folaron to present a critical discourse on the links between technology and localization as well as technical translation, enabling the author to systematize the epistemology of the field. In turn, Sabine Braun (see Chapter 16 on technology and interpreting) tracks technological developments in telecommunications which have shaped varied modes of distance interpreting and configurations of technical settings. This richly traces the new demands on professional interpreters to serve different technological constructs as their working environments. Thirdly, this volume addresses a number of substantive matters under Part V as overarching issues that challenge translation practice and research concerned with technology, ranging from quality to ecology. This part, along with the research foci and methodologies addressed in Part IV, aims to provide scholars and industry players with key topics, future avenues for research and analysis and insight into the implications of technologies for translation. Finally, the volume takes into account readers who may not be familiar with the topics addressed by some chapters and provides additional information: a list of relevant standards in Chapter 2, a glossary of terms in game localization in Chapter 13, an explanation of eye tracking technology in Chapter 24 and a list of recent major funded projects relevant to accessibility research in Chapter 29.

In terms of the macro-structure, Part I addresses key underlying frameworks and related technologies as relevant across different modes and areas of translation. Part II examines the adoption of technologies by different user groups. Part III considers the impact of technologies on each of the distinctive areas of translation (and interpretation)

practice. Part IV discusses research settings and methodological issues for selected research areas particularly relevant to the emerging relationships with technology. Part V explores the overarching issues in TS resulting from the increasing influence of technologies. The micro-structure of each chapter has certain key elements that are common across all chapters, yet is not uniform, as the final decision on the key content was left to the liberty of the chapter authors. The cross-referencing to other chapters was mostly added by the editor.

The next section provides an overview of how each contributor addresses their specific topic.

## **Part I: Translation and technology: defining underlying technologies – present and future**

Part I consists of five chapters which explain the fundamental building blocks and related general-purpose technologies key to understanding translation and technology at present and in their emerging guises. In Chapter 2, 'Standards for the language, translation and localization industry', **Sue Ellen Wright** provides a historical overview of how and why standards have developed over time, concerning technology applications in sectors spanning the translation, language and localization industry. Various standards for processes, products and services in today's complex technological world play a key role, including generating the basis for a 'feedback-rich information life cycle' beyond individual documents which may be chunked, repurposed and retrieved. Drawing on Briggs (2004 cited in Wright), Wright stresses, '[s]tandards transform inventions into commercial markets'. This is why international cooperation and expert consensus in a given field are critical in setting standards. Wright uses a historical approach to illustrate the role of standards and connections among them, without which today's technologically interlinked world through the Internet and use of tools in collaborative modes would not have been possible. A closely linked theme is taken up in Chapter 3, 'XML for translation technology', by **Johann Roturier**. Roturier shows how XML forms key backbone file exchange standards to ensure interoperability between translation tools and also the offline portability of tools in different user settings. The significance of XML can be illustrated in the statement, as quoted by Roturier, that 'over 90% of data for translation is generated with XML' (Zydroń 2014 cited in Roturier). Nevertheless, as the chapter explains, dynamic changes including the emergence of non-proprietary open source formats mean that this is a constantly developing area. Translators working especially in areas such as localization face the issues associated with these underlying factors in dealing with translation tools and files. Chapter 4, 'Terminology extraction and management', by **Kyo Kageura** and **Elizabeth Marshman** addresses terminology extraction and management as particularly pertinent in specialized translation (see Chapter 8 by Caffrey and Valentini). Terminology was one of the earliest areas within the translation workflow to have exploited electronic, as opposed to manual, processing, yet efficient terminology management within real-life translation practice remains a challenge. The chapter explains in some detail the different methods used in automatic term extraction (ATE), which is a critical upstream process, but is a computationally complex task to perform. The authors see ATE as a challenge especially in terms of quality, as is the case with collaborative terminology management. Finally, the role of terminology in connection with data-driven MT, including NMT, is briefly discussed, highlighting the importance of terminology quality in the training data. Here the need for human judgement at critical junctures within terminology management

is stressed. Related to the theme of electronic processing of linguistic resources, the following chapter focuses on linguistic data as a bi-product of, and an ingredient for, translation technology. In Chapter 5, ‘Building and using parallel text for translation’, **Michel Simard** explains the key techniques behind collection, structure, alignment and management involved in parallel text (consisting of an aligned source text and target text pair). These issues gained great importance with the widespread adoption of TM and data-driven MT, which use parallel text as training data. In reference to the more recent word alignment process in NMT, Simard refers to a ‘soft’ alignment mechanism known as ‘attention’. The anthropomorphic use of ‘attention’ in reference to a computational operation highlights its human-like function, albeit one not always achieved successfully. In turn, the lack of trust by human translators towards MT outputs, as alluded to by Simard, aligns with the findings elsewhere in TS literature (see Chapter 19 by Vieira). The last point signals some fundamental questions that arise when thinking about human–machine cooperation in translation. Further probing the cooperative dimension, the next chapter turns the focus to general-purpose technologies whose relevance to translation is increasing. In Chapter 6, ‘Speech recognition and synthesis technologies in the translation workflow’, **Dragoş Ciobanu** and **Alina Secară** examine the development and deployment of speech technologies i.e. speech-to-text and text-to-speech and their emerging uses in the translation workflow. While the authors find actual use cases of speech technologies in CAT scenarios are currently limited they point to the way in which speech recognition systems are integrated into live subtitling in ‘respeaking’ mode (also see Remail and Reviere in Chapter 29). The chapter reports recent empirical research conducted to test productivity gains and quality issues when combining automatic speech recognition systems in the process of translating as well as other tasks, such as revision and PEMT. The results highlight productivity gains as well as accuracy and stylistic issues while also pointing to the need for improvement in achieving a smoother integration of such technologies into CAT tools, together with consideration of task types.

## Part II: Translation and technology: users’ perspectives

Consisting of five chapters, this section addresses the perspectives of different translation technology users. The chapters represent different sectors of the translation industry. It ranges from large-scale language service providers (LSPs) and public institutions to freelance translators as well as language learners and translation practitioners who are not professional translators, but who benefit from translation technologies. Chapter 7, ‘Multinational language service provider as user’ by **Bert Esselink** looks into large LSPs for their use of technologies centred on translation management systems (TMS) which are divided into: Process Management and Automation, Project Management and Administration, Customer Management and Commerce, and Translation and Quality Management. The detailed description of the features and functionalities of TMS gives insight into how technologies are used to deliver an optimum translation service to customers by large LSPs. The chapter signals the increasing presence of AI and its likely significant impact in future, including in the area of project management, with implications for substantial change to the current human-based model. In Chapter 8, ‘Application of technology in the Patent Cooperation Treaty (PCT) Translation Division of the World Intellectual Property Organization (WIPO)’ **Colm Caffrey** and **Cristina Valentini** provide the perspective of a large public institution as a technology user. Patents form one of the most targeted fields of specialized translation heavily facilitated by technology. Caffrey and Valentini describe how TM and terminology

management systems are used in the PCT Translation Division, with its concerted efforts to provide translators with sophisticated terminological support via their terminology portal WIPO Pearl. Such terminological resources are a result of the integration of corpora, MT and machine learning algorithms, which may not be achievable by smaller organizations, let alone freelance translators. The authors further report on WIPO NMT which has been used since 2017 for all of the Division's nine languages, benefiting from a large body of in-domain training data (i.e. parallel corpora) available in-house. However, the authors suggest that the integration of NMT into the workflow means a change in the way translators deal with particular characteristics of NMT output which may be fluent yet contain terminological issues. This in turn implies different ways of using the terminological resources independently according to the need of the translator. Compared to large organizations, smaller translation operators have different settings and contexts in which to consider technologies, as described by **Patrick King** in Chapter 9, 'Small and medium-sized enterprise translation service provider as technology user: translation in New Zealand'. Drawing on his experience as a translator, editor and translation company operator, King explains how a medium-sized LSP in New Zealand is implementing technologies to achieve a productivity gain while maintaining translation quality. In particular, he shares translators' perspectives on new technologies, showing evidence of the openness of (some) translators to using technology, and that of NMT in particular. At the same time, King advises that technology should be assessed 'on its own merit', not simply because it introduces some improvements on the previous version. These days, most LSPs and freelance translators alike operate internationally, yet local contexts are still significant, as in New Zealand where Māori and South Pacific languages have unique requirements. King reminds the reader of the reality of translation service operating requirements, for example, dealing with a range of languages with unequal levels of compatibility with machine-processing. The fragmented translation industry continues to be supported by a large number of freelance translators. In Chapter 10, 'Freelance translators' perspectives' **Jost Zetzsche** opens the discussion by defining what a freelancer is and then moves on to examine key issues which initially delayed the uptake of technologies by freelance technical translators. By tracing a historical trajectory since the 1990s when CAT tools first became widely available, Zetzsche shows why uptake was initially relatively low and how translators changed from careful crafters of text to recycling 'CAT operators' who 'fill-in-the-blanks'. He argues that, at least in certain contexts, some tools are found to be 'stifling instruments for the human sensitivities of the technical translator'. Among the high use general-purpose technologies, Zetzsche highlights freelance translators' use of social media platforms from relatively early on, such as various online translator forums as a means to stay in contact with peers rather than for finding clients. The author points out that freelance translators tend to see the value of technology investment for its immediate link to increased revenue and this is why terminology management is a constantly undervalued element. He observes that MT is more accepted by translators compared to CAT when it was first introduced. Into the future with the increasing use of AI, Zetzsche sees the ideal role of translators as providing support by guiding technology developers. Chapter 11, 'Language learners and non-professional translators as users' by **Masaru Yamada** shifts the focus from the role of technology in official translation service provision to that of second language learning. Yamada explores the link between translation technologies and TILT (Translation in Language Teaching), with language learners and also non-professional translators using such technologies to improve their second language competency. Based on recent research on TILT, the chapter highlights the benefit of using MT output as a 'bad model' to boost language learners' competency through post-editing (PE) tasks. Furthermore, Yamada

draws on research pointing to the benefit of human-like errors made by NMT, which incur a higher cognitive effort in PE compared to errors produced by SMT, which are generally easier (more obvious) to repair. The former are therefore more conducive to learning. The capacity of translation technologies to boost lesser skilled translators' abilities is seen as empowering in this chapter. Yamada suggests the use of translation technologies in TILT could logically link to Computer-aided Language Learning (CALL), providing further research avenues.

### **Part III: Translation and technology: application in a specific context – shaping practice**

The technologization of translation is affecting different translation practices but with specific implications for each specialized area. Part III looks into different translation practices divided into eight chapters. In Chapter 12, 'Technology, technical translation and localization', **Debbie Folaron** takes on technical translation and localization to deconstruct their relationship with technology, taking a historical, methodological and critical approach. Through such lenses the chapter highlights, for example, how the emergence of localization practice has cast this new practice in relation to globalization, as articulated in the industry framework of Globalization, Internationalization, Localization and Translation (GILT). Furthermore, the localization process, which cannot be completed without the use of a technological platform, led to the development of specialized tools, in turn contributing to the formation of localization ecosystem (also see Cronin in Chapter 31). Folaron demonstrates the relevance of a critical digital discourse in shedding light on such practices as localization which is intertwined with digital artefacts. She then calls for TS scholars to engage more with the field of digital studies, which provides scope for the critical analysis of translation practice in an increasingly digital world. In Chapter 13, 'Technology and game localization: translation behind the screens' **Nathan Altice** inadvertently responds to Folaron's call to engage with digital studies with his discussion on localization of video games, especially by fans as non-professional localizers. Focused on the technicity of game hardware and software, Altice identifies a core feature of game localization with the practice of ROM (Read Only Memory) hacking, which involves unauthorized access and modification of a game's ROM by game fans, including the modification of the original language of the product. Characterized by its subversive and highly technical nature, ROM hacking communities continue to be active and visible. Informed by platform studies perspectives within game studies, Altice shows how 'language' is encoded 'graphically, materially and procedurally' by design in both the console/platform (hardware) and the game (software). This topic then naturally links to the following chapter focused on the broader concept of non-professional translation (NPT), which has recently gained considerable research interest in TS. In Chapter 14, 'Technology and non-professional translation (NPT)' **Miguel A. Jiménez-Crespo** examines the phenomenon of NPT, exploring its symbiotic relationship with broader technological developments represented by Web 2.0. The chapter gives close scrutiny to the increasingly visible practices of NPT, such as translation crowdsourcing and online collaborative translation. NPT involves participants who are not 'classically' trained translators, operating as part of translation communities in diverse contexts from pursuing fandom to activism or humanitarian initiatives. The chapter highlights the close correlation between NPT and digital technologies. NPT is characterized by non-uniform uses of translation technologies compared to its professional counterpart. Consequently, human-machine interaction in NPT can often be different from that in professional

translation, adding to the complexity of such relationships in contemporary translation. NPT encroaches on a variety of research foci, ranging from audiovisual translation (AVT) to PEMT, as well as raising questions of quality and ethics, affording scholars multiple lenses of analysis.

Within TS literature, localization and AVT are considered to be the areas most affected by new technologies and as a result having the greatest influence on the theorization of translation (Munday 2016: 275). In Chapter 15, 'Technological advances in audiovisual translation' **Jorge Díaz Cintas** and **Serenella Massidda** reflect on some of the formidable transformations within the rapidly expanding field of AVT. The chapter surveys an increasing body of research on the application of TM and MT in AVT, although the authors point out the benefit of these technologies is currently relatively limited. Cloud subtitling is seen as a new way for professional translators from different geographical locations to work together on collaborative platforms. Cloud-based dubbing and voice-over as end-to-end managed services are shown as rapidly developing examples. The authors explain how the availability of a wide range of tools and platforms is having a democratizing impact on AVT, yet is also heating up the competition among industry participants and causing increased anxiety among professional translators. The authors observe the way technology is altering relationships between stakeholders, highlighting its deep-seated impact.

Translation technologies are seen to be closely associated with (written) translation, yet MT is also core to machine interpreting (MI) which combines MT with speech technologies. In Chapter 16, 'Technology and interpreting', **Sabine Braun** focuses on the field of interpreting, including the rising demand for 'distance interpreting' and the milestones in MI. The chapter provides a comprehensive survey of the historical development of technologies shaping distance and on-site computer-assisted interpreting by humans, introducing different terminology used for different technology application settings and configurations of participant locations. While MI currently cannot service situations requiring highly accurate professional interpreting, Braun suggests that ongoing research, especially into neural networks, provides scope for further development. Highlighting the increasing reports by remote interpreters of psychological and physiological problems, the author stresses that interpreting is a cognitively challenging task and any other distracting issues relating to the lack of physical presence can affect the interpreter's performance. At the same time Braun raises the question of the sustainability of the profession as an important consideration in light of implementing smart technologies. Overlapping with some of these concerns, in Chapter 17 'Technology and sign language interpreting', **Peter Llewellyn-Jones** addresses settings specifically for Deaf people. Beginning with how the invention of the telephone disadvantaged the Deaf community, the author charts the development of spoken-signed language interpreting services via telephone, computer and video links. Comparing the situation in the US to Europe, the UK and Australia, the chapter argues that services such as Video Relay Services (VRS), where all interlocutors are in different locations, or Video Remote Interpreting (VRI), where only the interpreter is in a separate location, should not be developed simply to exploit available technologies; they must be carefully thought through to adequately enable the highly complex cognitive task of sign interpreting. Drawing on the research literature, Llewellyn-Jones illuminates the serious consequences that can result from making decisions purely based on the cost efficiency seen to be achieved by the use of technologies.

As touched on in the earlier chapter by Jiménez-Crespo, technologies are increasingly used to facilitate volunteer translators' involvement in humanitarian causes. A tragic

reminder of the need for ‘crisis translation’ is the 2017 Grenfell Tower fire, in which the apartment block’s multi-ethnic occupants speaking a diverse range of languages were largely unable to receive accurate information in their language in a timely manner. In Chapter 18, ‘Translation technology and disaster management’, **Sharon O’Brien** homes in on the role of technologies in disaster management and translation, which constitutes a relatively new area of research in TS and elsewhere. O’Brien argues translation is a neglected aspect in disaster management literature and policy, yet its role can be significant. This chapter illustrates the function of translation with the use of technologies serving beyond the ‘response’ phase of disaster risk management to all of the ‘4Rs’: the pre-disaster phases of ‘reduction’ and ‘readiness’ and the stages of ‘response’ and ‘recovery’. However, despite the proven successes with translation technologies in disasters such as in the Haiti Earthquake, technology deployment can be challenging, given issues such as disrupted infrastructure. Additionally, information recipients of written texts may have differing levels of literacy, not to mention cultural and accessibility considerations. Above all, this field highlights the socially significant role of translation, with challenges ahead including ethical considerations, linking to translation ecology thinking (see Cronin in Chapter 31). During the second decade of the new millennium, the use of MT within professional translation has become highly visible, with a raised interest in post-editing, as discussed at the beginning of this introduction and also amply demonstrated by the contributors to this volume. In Chapter 19, ‘Post-editing of machine translation’, **Lucas Nunes Vieira** gives a comprehensive survey of the growing research interest and industry practice of post-editing of Machine Translation (PEMT). Vieira begins the chapter by reminding us that PE used to be a ‘machine-centric’ activity in a mode of ‘human-assisted machine translation’ but is now geared rather towards ‘machine-assisted human translation’ in CAT environments. Drawing on the literature, Vieira presents the evolution of post-editing as a spectrum from MT-centred (automatic PE) to human-centred PE (interactive/adaptive) (also see Chapter 22 by Läubli and Green). Vieira sees the future of PE as better integrated into the professional translation process, where PE is no longer a discrete task. His conclusion highlights the need for further research into human agency in relation to PE activities and wider CAT environments. Vieira then highlights the role of TS in providing evidence-based findings to temper the hyperbolic claims made by some NMT developers and enables well-informed assessments to be made about technology.

#### **Part IV: Translation and technology: research foci and methodologies**

This section consists of five chapters which address specific foci and methodologies adopted to answer research questions probing the relationship between translation and technology.

In Chapter 20, ‘Translation technology evaluation research’, **Stephen Doherty** highlights how translation technology evaluation has gained key importance due to the prevalent use of technologies in contemporary translation. In particular, MT and post-editing have provided a strong impetus for this research area with the continuing development of automatic evaluation methods (AEMs) to complement or as an alternative to human-oriented evaluation of MT. Technology evaluation affects different stakeholders who have diverse needs, including technology developers and suppliers as well as providers and buyers of translation products and services, end-users and translation researchers. Doherty argues that despite the often-highlighted differences in purpose and context between evaluation methods used in academia versus in industry settings, the evaluation process is

inherently the same in that the evaluator needs to align the evaluation purpose with the available resources and methods, and the desired format. While advances in technology evaluation research are providing increasingly sophisticated evaluation mechanisms, Doherty calls for further research focused on three areas: universalism and standardization, methodological limitations and education and training. These will allow more inclusive and standardized approaches to meet the needs of the different stakeholders. In Chapter 21, 'Translation workplace-based research' **Maureen Ehrensberger-Dow** and **Gary Massey** provide an up-to-date survey of workplace-based research, which has steadily gained importance in TS over the last decade. This is where research moves out of the translation classroom or laboratory into real life workplaces, filling the gap in the other research settings and providing ecological validity by capturing data from translators in situ. Ehrensberger-Dow and Massey show how increasing technologization has made it relevant to see expert activity as a manifestation of situated cognition, whereby human cognition is assumed to extend beyond individual minds to, for example, interaction with technological artefacts. The chapter articulates the way workplace-based research can highlight with empirical data how technologies can facilitate or disrupt the human translation process. The chapter calls for more transdisciplinary action research to ensure human translators are empowered by working with technologies and not undermined by their technological environments. In Chapter 22, 'Translation technology research and human-computer interaction (HCI)' **Samuel Läubli** and **Spence Green** address translation and technology from the perspective of professional translation as HCI. Focused on users of translation technology, they discuss 'interactive MT' (IMT) as opposed to the 'static' model (also see Vieira in Chapter 19), and examine factors behind the often-negative response of professional translators to PEMT tasks. The chapter draws on empirical evidence to highlight how seemingly trivial User Interface (UI) design issues, such as font size, lack of shortcuts, copy-paste functionality, etc. can hinder efficient human-computer interaction. Similarly, the authors point to the findings in the literature that user irritation relates, above all, to the repeated need to correct the same MT errors. The authors surmise the key challenge in HCI as the limitation of the machinery's ability to learn from (human) users, whereas humans can learn to use 'novel machinery'. Furthermore, 'making the state and effects of adaptation understandable to their users' is part of the challenge in creating adaptive systems. This in turn critically requires the iterative involvement of translators in the development process, a lesson still being learnt from the early MT projects that lacked translator participation. In Chapter 23, 'Sociological approaches to translation technology' **Maeve Olohan** examines the key research questions and methodologies in sociologically-oriented studies on translation technologies. The chapter traces the development of SCOT (social construction of technology) as a field of study to demonstrate how 'science and technology are socially constructed cultures' (Pinch and Bijker 1984: 404 cited in Olohan's chapter), accommodating both successful and failed technologies. In parallel with SCOT the author explains other sociological approaches applied in TS research. Despite the increasing use of sociological approaches in TS research to shed light on translation technologies, Olohan concludes that there is more to pursue in 'sociology of translation', both conceptually and empirically. For example, she argues that critical theory of technology can be fruitfully combined with constructivist approaches to reveal unequal power distributions, which often affect the adoption of technologies. Olohan suggests these lines of inquiry could lead to a further renewal of the traditional conceptualization of translation. Methodological innovations are part of the increasing sophistication of research in TS and eye tracking is one of the key

examples. In Chapter 24, ‘Translation technology research with eye tracking’, **Arnt Lykke Jakobsen** provides explanations about eye tracking technology and a detailed survey of this popular research tool now used in diverse areas of TS. This chapter shows how eye tracking software can trace with fine granularity the translation process and the dynamics of the translator’s interaction with translation technology, for example TM or MT, while performing translation or post-editing. Or it can capture the translation user’s response to dynamic text presentation modes, such as in subtitles. Translation is an ‘effortful cognitive activity’, yet to what extent technological tools add to or lessen such efforts is a question which calls for empirical evidence. Jakobsen suggests eye tracking could provide insight, for example, into reasons for ‘the global preference of multimodal, analogic communication’ compared to ‘unimodal, symbolic communication’ despite the assumption that the former is more effortful. While cautioning that not everything about visual attention and cognitive processing is fully explainable from eye tracking data, Jakobsen predicts there are likely to be widening avenues for eye tracking in future as part of mixed-methods research design used with ‘qualitative data and neuroscience technologies’.

## Part V: Overarching issues

The final section consists of seven chapters which focus on a number of critical overarching issues arising from significant uses of technology in translation. This section covers MT, quality, fit-for-purpose translation, accessibility, reuse of translation data, translator training and translation ecology. In Chapter 25, ‘Future of machine translation: musings on Weaver’s memo’, **Alan K. Melby** explores where the next paradigm of MT is headed, centring on the challenge arising from sub-symbolic deep learning (i.e. its inner workings are non-inspectable to humans) applied in the current generation of NMT. This issue of increased opacity in machine learning is noted by scholars as a cause for concern (see Kenny 2018). As a way to think about future developments of MT, Melby uses a detailed analysis of Warren Weaver’s 1949 Memorandum. The chapter treats the early pioneer’s concepts presented in the memo as ‘seeds’ behind the subsequent successive paradigms of MT, from RBMT (Rule-based MT) to SMT (Statistical MT) to the current evolving state of NMT. Melby then considers developments in the intervening times of enabling technologies and surrounding contexts to build his conjecture by mapping the seeds in Weaver to the MT paradigms. With sub-symbolic deep learning, Melby argues, even those who are modelling AI cannot seem to predict exactly what goes on inside the ‘black box’. The discussion leads Melby to the question of what it means to ‘understand’ the text in human and machine translation, which he believes is significant for the next phase of MT, i.e. seeing Weaver’s final seed – linking MT to the human brain – sprouting.

Not unrelated to the issue of ‘understanding’, quality is a key challenge for translators and the translation industry. Pym clarifies the two meanings of ‘quality’, the first being ‘properties’ a la Aristotle, if used in the plural, and, in the singular, meaning ‘the relative excellence of the thing’ for a given purpose. The two meanings are often related, as we are reminded by Pym, with changes of properties leading to changes in the status of excellence, as applicable to the quality of translation technologies. In Chapter 26, ‘Quality’, **Anthony Pym** addresses translation quality in the context of translation technologies by treating it as ‘relations’ based on Chesterman (2004 cited in his chapter); namely relations between the translation and the target text, comparative texts, purpose, industrial standards and the translator. For example, with the prevalence of TM and MT in the translation process, Pym highlights the critical need for greater quality control of such technologies

which are exerting 'unspoken forces behind the industrial standards'. In reference to the relation between translation and the translator, Pym argues that a likely consequence of technologization manifesting in more pre- and post-editing for translators, could still be made satisfying for them, if such work was presented as 'authorizing' the final translation. He suggests it is up to translators and their employers to ensure that the work is recognized and rewarded as such. Discussing these relations, the chapter teases out human elements in quality to remind the reader that evaluations of quality 'reside on human values' that are 'built on a fundamental indeterminacy'. As highlighted in his conclusion, Pym draws our attention to 'the human in the machine', so the quality debate is not overshadowed by the technology and the extreme ideological stances both for and against it.

This chapter is followed by the closely related topic of 'Fit-for-purpose translation' in Chapter 27, where the indeterminate nature of quality is explored. Here **Lynne Bowker** discusses translation as a balancing act between the 'triple constraint' used in the project management field of quality, cost and time. Furthermore, the author points to 'a perception problem' in reference to the negative associations of the use of translation tools. Bowker reminds the reader that 'translations can be commissioned for a diverse range of purposes', while a translator's job is to 'choose the strategy best suited to producing a target text that fits the specified purpose'. With the technologization of translation, Bowker highlights, translators need to be able to optimize technologies to best meet different translation purposes, as specified by the clients. This may result in different levels of quality in translation, in conflict with professional ethics, which currently do not provide adequate guidance to translators in respect of the use of technologies. As much as there is a need for client education, Bowker stresses the need for professional translator (re)education to recognize the challenges and not denigrate translators who are catering for 'bulk translation services'. The final thought offered by Bowker is indeed ironic as she suggests: if lesser quality translations produced for different purposes start to affect the quality of the training data for MT, in turn affecting MT quality, fit-for-purpose translation may inadvertently ensure the survival of human translators. Bowker's last point relates to the increasing harvesting of translation as data used for machine learning, as discussed next.

In Chapter 28, 'Copyright and the re-use of translation as data', **Joss Moorkens** and **Dave Lewis** address the increasing secondary use of translation currently treated as a cheap commodity. This is becoming an issue in the advent of data-driven MT and especially for NMT, due to its requirement for a significant amount of training data for machine learning. The authors highlight that the metaphor of 'oil' or 'gold' used for the translation as training data implies they are naturally occurring, which is untrue, giving rise to the question of translation rights. The issue is that this subsequent benefit generated by the original translation is not passed on to the translator who translated the text. In view of the 1889 Berne Convention, which codified the copyright of translation as a derivative work, the authors point out that the current reuse of translation as data was not envisaged in the Convention, nor was its potential liability in relation to NMT. They argue that current copyright laws are not equipped to deal with the situation of the reuse of translation data, while new proposals, such as digital commons with a range of rights, could potentially be applied through professional translation organizations. The authors suggest the latter is more conducive to ensuring the sustainability of the translation industry by improving the redistribution of equity within translation production networks. The authors suggest that this could be realized in collective agreements accruing royalties to the translators, as is already the case among some subtitlers in Nordic countries. However, the chapter

concludes that the forthcoming EU Directive on Copyright in the Digital Single Market is not likely to resolve the issue of translation copyright, which will remain as a key question requiring the attention of translators.

In Chapter 29, ‘Media accessibility and accessible design’, **Aline Remael** and **Nina Reviers** discuss media accessibility (MA), which has rapidly become integrated into research agendas in TS with practical implications for audiovisual translation (AVT) driven by digitization and globalization. The authors argue that in today’s ‘highly mediated society’, the accessibility of audiovisual content, and eventually accessible design, has become a central concern for society at large. They assert technology is making it ‘theoretically’ possible to cater for all types of media users, given the right policy and legislation. MA involves ‘translation’ of an ‘intersemiotic and multi-modal’ kind where aurally or visually conveyed information is converted into modes to suit the target audience’s needs. For example, subtitles for the Deaf and the hard-of-hearing (SDH) involve a conversion from aural to visual mode where the target audiences can ‘read’ the dialogue. SDH now includes live subtitling, which is commonly delivered in the form of respoken, whereby subtitles are generated synchronously through the use of speech recognition. The technology applications in this field are wide-ranging, from speech recognition and synthesis to MT as well as avatars used for sign interpreting. Initiatives on universal design for MA are well underway, with examples such as accessible filmmaking in which accessibility is foregrounded in the filmmaking process itself (Romero-Fresco 2018). Applying an actor-network theory framework, this chapter critically traces the developments taking place in media accessibility in which practice and research interact with technologies exerting considerable force as enablers. In the unpredictable technological milieu, the authors see ‘translation’ in its broad sense as playing a major role of a key ‘actant’ to progress this significant social issue of the modern age of technologies towards universal design.

In Chapter 30, ‘technology and translator training’, **Dorothy Kenny** addresses the issue of translator training in the advent of technologization, comprehensively drawing on the growing literature in the field. Kenny argues ‘a nuanced understanding of how technology and translation are intertwined should be a vital ingredient of any broad education in translation studies’. Kenny therefore advocates the view that technological competence need not remain merely ‘instrumental’ but can make ‘a significant contribution to the development of critical citizenship’. The chapter provides a critical analysis of contemporary thinking behind translator training and education, which is then extended to a key concern for the long-term health of the translation industry, including economic factors such as ‘technological unemployment’ in the advent of AI. In the author’s words the next challenge lies in ‘the integration of machine learning into translator training’, which would signify indeed a paradigm shift in translator education. Implicit in this chapter is ecological thinking, viewing translation and technology as an intrinsic part of the technologizing global world, which relates to the theme of the next final chapter.

In Chapter 31, ‘Translation, technology and climate change’, **Michael Cronin** interprets the agenda of translation and technology in terms of the big picture, employing ecological perspectives and proposing a new methodological approach based on eco-translation thinking. Cronin maintains that the fate of translation enterprise is inevitably implicated in what happens to technology which is, in turn, linked to accelerated climate change. This chapter constructs an argument through the notion of translation ecology, with the key concept of the ‘posthuman’ which provides an approach for understanding the deepening relationship developing between humans and digital technologies. Cronin insists on treating technology not as ‘an inert tool’ but as ‘an animated part of the human ecosystem,

a constituent element of the translator's transversal subjectivity'. His ecological thinking in turn gives rise to a renewed perspective on ethical issues, as Cronin asks: 'Is it...ethically responsible and professionally adequate to train translators using technologies that will not be sustainable in an environmentally compromised future?' This line of concern relates to crisis translation settings (O'Brien in Chapter 18), which may only allow low-tech solutions due to the destruction of the communications infrastructure. Also, it relates to the issue raised by Moorkens and Lewis (Chapter 28) in questioning the continuing secondary use of translation as if it is a bottomless resource to feed into MT until it is depleted – or until eventually the translation quality deteriorates as a consequence of fit-for-purpose translation (Bowker in Chapter 27). In this critical age of climate change and rapid technologization, Cronin directs our attention to planetary contexts as a productive way to locate translation through an eco-translation framework, as we grapple with the role of humans in relation to the role of technologies in translation research and practice. Joining Ehrensberger-Dow and Massey (Chapter 21), Cronin advocates for transdisciplinary approaches to be adopted by scholars. This could usefully lead to a re-evaluation of the role of translation and translators in the age of technologization through collaboration with community members and organizations. In this way, Cronin argues, Translation Studies can participate in the critical dialogue at a time of environmental crises brought about by the Anthropocene era.

### In summary

This volume sets out to discuss translation and technology as a growing yet disruptive relationship. Together the contributors paint a picture of a profession or an activity that is dynamic and plays important social and ecological roles, sustaining global communication needs for businesses and individuals in public and private spheres. The examples discussed in the volume span NMT, post-editing, ROM hacking, crisis translation in disaster settings, media accessibility and interpreting at a distance for the Deaf community, to name a few. The volume highlights the central position technologies are occupying in translation and in some interpreting practices while drawing the reader's attention to human agency. In all this, as already acknowledged by TS scholars, translation continues to defy an exact definition (Williams 2013: 5–9) and technological factors are only confirming the multiplicity of the practice and concept of translation. From a practising translator's perspective, Mark Polizzotti (2018: xi) describes in his *Sympathy for the traitor: a translation manifesto* the nature of translation work as ambivalent, 'skirt[ing] the boundaries between art and craft, originality and replication, altruism and commerce, genius and hack work'. His manifesto celebrates the variability of human translation and defends the oft-used analogy of a translator as a traitor in the sense that translation decisions are not always deducible from the words in the source text alone. The challenge for 'augmented translation' or any other advanced technology-mediated environment would therefore be to facilitate such a complex, ill-defined human decision-making process. The inquiry into the deepening connection between translation and technology, and also translation by the human and by the machine, will widen the scope for the future development of Translation Studies and the translation profession, as the contributors of this volume eloquently demonstrate. In the spirit of participatory culture, the more stakeholders who partake in the examination of what is happening with the human-machine unison or abrasion in contemporary translation, the more chance we have of grappling with the changes taking place. It is hoped that the diverse observations presented in this volume will provide a fresh impetus for theory building for

scholars, which will enable translators to better navigate increasingly technologized environments that are complex, unpredictable and fragile. This in turn will help us ensure the survival and sustainable evolution of translation, in the advent of algorithm-led intelligence. Finally, the reference to ‘entanglement’ in the title of this introduction is borrowed from quantum physics. Described by Einstein as ‘spooky action at a distance’, quantum entanglement refers to the phenomenon where particles separated in space and time are inextricably linked (de Ronde and Massuri 2018). This deep-seated correlation and the synched status of two entities evokes the inescapable bond being formed between the human and the machine. It could be the vision for the future of the refined, if inscrutable, art of translation with human and machine learning enriching each other. This is ultimately related to the question of what it is to be human and a translator in the technologizing age.

## References

- Akst, D. (2013) ‘Automation anxiety’, *Wilson Quarterly* Summer 2013, Available online: <http://archive.wilsonquarterly.com/essays/automation-anxiety> [last accessed 26 Feb. 2019].
- Bowker, L. (2002) *Computer-aided translation technology: a practical introduction*. Ottawa: University of Ottawa Press.
- Carl, M., S. Bangalore and M. Schaeffer (eds) (2016) *New directions in empirical translation process research: exploring the CRITT TPR-DB*. Cham: Springer.
- Catford, J. (1965) *A linguistic theory of translation*. Oxford: Oxford University Press.
- Chan, S-W. (ed.) (2015) *The Routledge encyclopedia of translation technology*. London & New York: Routledge.
- Chesterman, A. and E. Wagner (2002) *Can theory help translators?: a dialogue between the ivory tower and the wordface*. Manchester: St. Jerome Publishing.
- de Ronde, C. and C. Massuri (2018) A new objective definition of quantum entanglement as potential coding of intensive and effective relations. Available online: <https://arxiv.org/abs/1809.00202> [last accessed 31 Jan. 2019].
- ELIA (European Language Industry Association) (2018) *Expectations and Concerns of the European Language Industry*. Available online: [www.euatc.org/images/2018\\_Language\\_Industry\\_Survey\\_Report.pdf](http://www.euatc.org/images/2018_Language_Industry_Survey_Report.pdf) [last accessed 2 Feb. 2019].
- Gambier, Y. and L. V. Doorslaer (eds) (2016) *Border crossings: translation studies and other disciplines*. Amsterdam & Philadelphia: John Benjamins Publishing.
- Giammarresi, S. and G. Lapalme (2016) ‘Computer science and translation: natural languages and machine translation’ in Y. Gambier and L. V. Doorslaer (eds), *Border crossings: translation studies and other disciplines*. Amsterdam & Philadelphia: John Benjamins Publishing, 205–224.
- Holmes, J. (1972/1988) ‘The name and the nature of translation studies’ in *Translated!: papers on literary translation and translation studies*. Amsterdam: Rodopi, 67–80.
- Hutchins, J. (1986) *Machine translation: past, present, future*. New York: Halsted Press.
- Hutchins, J. (1998) ‘The origins of the translator’s workstation’, *Machine Translation* 13(4): 287–307.
- Jakobsen, A. and B. Mesa-Lao (eds) (2017) *Translation in transition: between cognition, computing and technology*. Amsterdam & Philadelphia: John Benjamins Publishing.
- Kay, M. (1980/1997) ‘The proper place of men and machines in language translation’, Technical Report CSL-80-11, Xerox Palo Alto Research Center (PARC). Reprinted in *Machine Translation* 12: 3–23.
- Kenny, D. (2018) ‘Machine translation’ in P. Rawling and P. Wilson (eds) *The Routledge handbook of translation and philosophy*. London & New York: Routledge, 428–445.
- Kenny, D. (forthcoming) ‘Machine translation’ in M. Baker and G. Saldanha (eds) (2019) *The Routledge encyclopedia of translation studies*, 3rd ed. London & New York: Routledge.
- Kruger, J-L. (2018) ‘Eye tracking in audiovisual translation research’ in L. Pérez-González (ed.), *The Routledge handbook of audiovisual translation*. Abingdon & New York: Routledge, 350–366.
- Lommel, A. (2018) ‘Augmented translation: a new approach to combining human and machine capabilities’, in J. Campbell, J. Doyon, A. Yanishevsky and D. Jones (eds) *Proceedings of Association*

- for Machine Translation for Americas (AMTA) 2018*, vol. 2: MT Users' Track. Boston: AMTA, 5–13. Available online: [www.aclweb.org/anthology/W18-1905](http://www.aclweb.org/anthology/W18-1905) [last accessed 23 Nov. 2018].
- Melby, A. K. (1981) 'Translators and machines – can they cooperate?' *Meta* 26(1): 23–34.
- Munday, J. (2016). *Introducing translation studies: theories and applications*, 4th ed. Abingdon & New York: Routledge.
- Nida, E. (1964) *Toward a science of translating: with special reference to principles and procedures involved in Bible translating*. Leiden: Brill.
- Nida, E. and C. Taber (1969) *The theory and practice of translation*. Leiden: Brill.
- O'Brien, S. (ed.) (2014) *Post-editing of machine translation: Processes and applications*. Newcastle upon Tyne, England: Cambridge Scholars Publishing.
- 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*. London & New York: Routledge, 503–518.
- Polizzotti, M. (2018) *Sympathy for the traitor: a translation manifesto*. Cambridge, Massachusetts: MIT Press.
- Romero-Fresco, P. (2018) 'Accessible filmmaking: translation and accessibility from production' in L. Pérez-González (ed.), *The Routledge handbook of audiovisual translation*. Abingdon & New York: Routledge, 498–515.
- Schwieter, J. W. and A. Ferreira (eds) (2017) *The handbook of translation and cognition*. Hoboken, NJ: Wiley-Blackwell.
- SDL Technology Insights (2016). Available online: [www.sdltrados.com/download/translation-technology-insights-executive-summary/103906/](http://www.sdltrados.com/download/translation-technology-insights-executive-summary/103906/) [last accessed 25 Jan. 2019].
- Shreve, G. M. and B. J. Diamond (2016) 'Cognitive neurosciences and cognitive translation studies: about the information processing paradigm' in Y. Gambier and L. V. Doorslaer (eds), *Border crossings: translation studies and other disciplines*. Amsterdam & Philadelphia: John Benjamins Publishing.
- Somers, H. L. (ed.) (2003) *Computers and translation: a translator's guide*. Amsterdam & Philadelphia: John Benjamins Publishing.
- Susskind, R. and D. Susskind (2015) *The future of the professions: how technology will transform the work of human experts*. Oxford: Oxford University Press.
- Vieira, L. N. (2018) 'Automation anxiety and translators' *Translation Studies*. Online First. DOI: 10.1080/14781700.2018.1543613.
- Weaver, W. (1949) Translation. Reprinted in: W. N. Locke and A. D. Booth (eds), *Machine translation of languages*. New York: The Technology Press of MIT, 15–23.
- Williams, J. (2013) *Theories of translation*. New York: Palgrave Macmillan.

## Part I

# Translation and technology: defining underlying technology – present and future

---



**Taylor & Francis**

Taylor & Francis Group

<http://taylorandfrancis.com>

# Standards for the language, translation and localization industry

*Sue Ellen Wright*

---

## Introduction

This chapter addresses the role of standards and standardization treating language, language resources of many kinds, and the global language enterprise. It traces their evolution from their inception as terminology standards to a 21st century environment where language and the content that it expresses are chunked and identified for efficient retrieval and reuse, with the goal of creating not just documents, but rather knowledge systems where relevant ontologies, terminology collections, source-target bitexts (e.g., translation memories), or other knowledge objects provide a coherent and holistic basis for the information life cycle rather than being limited to the one-off document instances common in the past.

Standards comprise documents that provide requirements, specifications, guidelines or characteristics that can be used consistently to ensure that materials, products, processes and services are fit for their purpose:

An International Standard provides rules, guidelines or characteristics for activities or for their results, aimed at achieving the optimum degree of order in a given context. It can take many forms. Apart from product standards, other examples include: test methods, codes of practice, guideline standards and management systems standards.  
*(ISO 2018b)*

This chapter will also treat standards and ancillary resources from beyond the ISO framework that nonetheless fit this profile.

Describing the environment around language standards starts with the methodology for creating industrial standards. Although approaches vary somewhat by region and domain, the fundamental principle involves collaboration among technical experts in committee working groups to achieve consensus on process, procedures, product specifications, and terminology used in creating and evaluating goods and services. Most standards are elaborated by national, regional, and international standards organizations under the aegis of the International Organization for Standardization (ISO, see below),

although some ‘industrial standards’ are created by industry-based representatives of corporate members of professional organizations outside the ISO context. In both models, establishing ‘consensus’ involves the proposal of draft specifications followed by revisions and refinements until objections are resolved and agreement is reached. The process is never easy, especially on the international scale, where ‘global relevance’ can be hard come by, and often takes three to five years.

The main section of this chapter treats both the historical development of the standards and literature documenting standards development as a continuum, inasmuch as the literature is primarily descriptive and reflects the state-of-the-art at any relevant point in time.

## Literature review and historical trajectory

### *General literature*

Literature will be reviewed in sync with standards development, bearing in mind that many years passed before anyone actually wrote about language standards as such. References abound on the web, but readers should be cautioned that all but the most up-to-date articles, book chapters, and webpages treating the language enterprise should be viewed with a certain scepticism because articles quickly become outdated, and industry resources often fail to refresh their information with any frequency.

Throughout this chapter, the term *language enterprise* refers to the totality of language documentation, manipulation, and exploitation wherever it occurs: in industry at large, in the language industry in particular, in government at all levels, in academia, and elsewhere in the public and private sectors. References will be made to specific ISO, CEN, ASTM, DIN,<sup>1</sup> and other standards, for which full title and publication dates can be found in the *Language Standards Organized by Topic* list at the end of this chapter.

### *Language-related standards for and as industrial standards*

At its core, most human activity involves language, be it spoken or written, actual ‘in-person’ communication or recorded in some medium – etched in stone, carved in clay, inscribed in ink on velum or paper, or digitized electronically. Initial efforts to normalize language *per se* featured valorization of specific language variants, e.g., in the European context, the evolution of dictionaries (English and German) and of national Academies of language (French and Spanish), resulting in conventionalized spellings and grammars. Milestones such as Samuel Johnson’s *Dictionary of the English Language* (1755) and François I’s *Ordonnance de Villers-Cotterêts* declaring that French would henceforth be the official language of the courts (1539) stand out on such a timeline. Interesting as these events may be to linguists, they concern general language, while this chapter will focus on language used in industrial standards, and then on the standardization of methods, processes and even products related to the emergence of the language enterprise itself as an industry.

Standards governing products and processes date from the beginnings of measurement systems: the Egyptian *ell* as a length measurement, for instance, or the calendar systems of the ancient world established consensus in specific areas in order to ensure effective human collaboration (ANSI 2018). Standardization of products, processes, test methods and materials specifications can be traced through the medieval guild system or even earlier, and the earliest formal international effort of significance probably involved

the International Bureau of Weights and Measures, which began in 1875 to establish standard measures and units, resulting in *The Treaty of the Metre* and eventually, today's International System of Units (SI; Page and Vigoureux 1975, BIPM 1875 to 2018<sup>2</sup>). While not a 'terminology standard' *per se*, this list of units and values did pave the way for dedicated standardization in response, not surprisingly, to industrial advances later in the 19th century: the introduction of uniform rail gauges in the US and Europe, for instance, followed in the histories by the elaboration of standards for fire-fighting equipment in the early 20th century (ANSI 2018, Wright 2006a, ASTM 2018).

Formed in 1898, the *International Association for Testing and Materials* became the *American Society for Testing and Materials* (now *ASTM International*), based on the principle of consensus standards. ASTM is now one of ca. 300 individual standards bodies under the umbrella of ANSI, the *American National Standards Institute*. Other national groups count their origins in the same general time period: *The Engineering Standards Committee* in London, 1901, became the *British Standards Institution* by 1931 (BSI 2018); the *Normenausschuß der deutschen Industrie*, founded in 1917, evolved into today's *Deutsche Institut für Normung* (DIN, Luxbacher 2017); and AFNOR, *l'Association Française de Normalization*, can be traced back to 1901 (AFNOR 2018). On the international level, the *International Electrotechnical Commission* (IEC 2018a) was founded in London in 1906 (IEC 2018c), followed by the *International Federation of the National Standardizing Associations* (ISA), which originated in 1926, was disbanded in 1942, and reorganized in 1946 after World War II as ISO, the *International Organization for Standardization*, occupying today the position of global umbrella organization for 161 national standardizing bodies such as those cited above. CEN, the European Committee for Standardization or *Centre Européen de Normalisation* came together in 1961 to unify the standards organizations in the European Union, although today ISO standards frequently supplant CEN (OEI 2018).

### *Foundation and build-up*

Early language standards defined terms used in product and process standards. The earliest documented terminology standard was ASTM D9, *Standard Terminology Relating to Wood and Wood-based Products*, published in 1907 (Ellis 1988). ASTM interest in standardizing terminology led to formation of a separate committee (E-8) in 1920 for *Nomenclature and Definitions*, followed in 1978 by the Committee on Terminology (CoT), which created a standardized practice for defining terms and a *Compilation of ASTM Standard Definitions*. The CoT was disbanded after its role was essentially absorbed into the US mirror committee for ISO TC 37.

On a global scale, the *International Electrotechnical Commission* (IEC), founded in 1906, published its *International Electrotechnical Vocabulary* (IEV) in 1938, whose direct descendent is today's *Electropedia* (IEC 2018a). Under the leadership of Eugen Wüster, the founding theorist of terminology studies, ISA/ISO established Technical Committee TC 37, which, despite a hiatus during the WWII years, continues today as ISO TC 37, *Language and Terminology* (Trojar 2017: 56, Campo 2013: 20 ff.), with some name changes on the way – *Terminology (principles and coordination)* (1951); *Terminology and other language resources* (2003); and *Terminology and other content and language resources* (2006). In his history of Infoterm, the ASI administrative entity governing TC 37, Galinski categorizes the period from 1951 to 1971 as the 'foundation phase', which focused primarily on the theory of terminology and terminology work (later called terminography),

the compilation of terms and definitions for the ‘vocabulary of terminology’, and guidance for preparing ‘classified vocabularies’. The audience for these standards was originally limited to standardizers identifying and defining terms used in their standards. Terminology documentation during this era involved the handwritten (or possibly typed) collection and documentation of terminology on paper *fiches* (Felber 1984: 162 ff., Wright 2001: 573). Through the mid-1980s, TC 37 focused on defining concepts and terms from an onomasiological, that is, concept-oriented, systematic standpoint. Then as now, the core standard, ISO 704, and its companion vocabulary, ISO 1087, reflected Wüster’s *General Theory of Terminology*, which is firmly grounded in a long tradition of linguistic theory culminating in the work of Saussure and Frege. ISO 639 had already originated in 1967 as a repository of ‘symbols’ for languages, the beginning of the two-letter language codes (Galinski 2004a & b, Martincic 1997).

### *Consolidation, outreach, and the digital turn*

Galinski calls his second phase (1972–1988) the *Consolidation Phase*, and focuses on outreach to other sub-disciplines, not only chronicling the maturation of TC 37 within ISO, but also mirroring a turn in the evolution of the overall language enterprise. By the late 1980s, work had been organized into three sub-committees: SC 1, *Principles and methods*; SC 2, *Terminology work*; and SC 3, *Computer applications in terminology*, with 704 and 1087 assigned to SC 1, and language codes and a new ISO 1951 for symbols used in lexicography assigned to SC 2. Sub-Committee 3, however, ushers in a paradigm shift in the role of language in standardization, or perhaps more accurately, of standardization with respect to language: in 1987 ISO 6156, *Magnetic Tape Exchange Format for Terminological / Lexicographical records (MATER)* appears. Writing almost two decades later, Briggs notes a phenomenon perhaps little understood in the early 1980s: ‘Standards transform inventions into commercial markets’ (Briggs 2004). Briggs points out that the invention of the radio and of flight did not actually trigger new industries until such time as the standards to regulate and actually enable these industries had been put in place. The MATER standard is the first in a growing flow of enabling standards (see Wright 2006b: 266 ff.), which is to say, standards that facilitate functions that would be impossible without the standard itself.

In the case of MATER, this task involves the exchange of terminological (and theoretically lexicographical) data, heralding the advent of large-scale main-frame-based terminology management systems such as the Siemens TEAM terminology management system. Subject to a long evolution over the course of the 1980s, this one document was the first in a series of standards, both inside and outside ISO TC 37 designed to ensure data compatibility and interchangeability in a nascent industry.

ISO 6156 itself did not enjoy a long-term or widespread reception – it was uniquely introduced to meet the needs of the TEAM system, whereby the parent database resided on a main-frame computer in Munich, but translators and terminologists working in far-flung Siemens centres around the globe provided data maintenance information and received updates either on tape or via daily overnight data transfers over dedicated phone lines. Nonetheless, the standard did not focus on a virtual data stream, but rather on data stored on magnetic tape.<sup>3</sup> The seminal ideas in the MATER serialization included: 1) the sharing, exchange, and interoperability of data, 2) that is configured in a series of records, each devoted to a single concept-oriented term entry, 3) consisting of discrete data fields identified by data category names (e.g., *English term*, *English source*, etc.) and 4) the use

of standardized encoding to mark up these data elements in an otherwise undifferentiated flat data flow. These criteria still apply, although the details of the successor standards have changed.

It is important to note some critical developments that occurred on the way to that 1987 MATER publication date. While standardizers were focused on mainframe-based terminological systems, events unfolding elsewhere were destined to doom MATER to a limited life span. The IBM PC appeared in 1981, signalling the beginning of the end of main-frame dominance and stabilizing the then fragmented personal computer market. The availability of an affordable, increasingly user-friendly platform for text authors, translators, and terminologists introduced a spate of word processing applications, as well as terminology management, and eventually translation memory, systems that inspired an interchange format for ‘micro’-computers – the so-called *MicroMATER* standard (Melby 1991).

Along these lines, the publication in 1986 by ISO/IEC JTC 1/SC 34 for *Document description and processing languages*, of a *Standard Generalized Markup Language (SGML)*, after an extended period of development, was also the harbinger of another significant paradigm shift that impacted knowledge and information representation and sharing, academic research environments, as well as commerce and industry, on a global scale. In fact, although MicroMATER did not begin as an SGML formalism, it parallels many features from SGML, which offered a powerful medium not just for marking up presentational features in text, but also for creating standardized encoding systems for managing corpora and other text-based data resources. This new medium inspired computational linguists to come together on an international scale, beginning in 1987, to establish the *Text Encoding Initiative* (TEI 2018a, 2018b).

### *Extension phase (1989–1996)*

Sticking with Galinski’s time frames, which justifiably feature Infoterm activities, this chapter also views a broader spectrum of standards from the period. It almost goes without saying that SGML begat HTML, and HTML begat XML, and even if it was not always good, the whole development was prolific, fostering the World Wide Web as we know it today. Much of this activity involves manipulating language and informational content in some form, thus extending the arena of language standardization far beyond TC 37.

The TEI project originally expanded the use of SGML markup into a wide variety of corpus-related approaches, in addition to lexicography and terminology. Indeed, the simmering MicroMATER bubbled over into TEI-Term (TEI 1994/1999) and eventually was re-published in 1999 as ISO 12200: *Computer applications in terminology – Machine-readable terminology interchange format (MARTIF) – Negotiated interchange* and its companion standard, ISO 12620: *Data categories*. Other TEI-related formats found their way into the ISO framework via the formation of TC 37/SC 4 in 2001/2002, focusing on *Language Resource Management*, specifically: feature structures, lexical markup, semantic annotation frameworks, and metadata infrastructures, among other topics (ISO 2018a).

### *Extending the language codes and beyond*

Identifying languages with short codes (treated as symbols in ISO) relies on a foundational family of language standards, which over time has engaged a growing set of

stakeholders, including standardizers, lexicographers and terminologists in TC 37; librarians and documentation specialists in TC 46 (*Information and Documentation*); Bible translators motivated to codify all human languages in their missionary work (*Summer Institute of Linguistics* or SIL); internet engineers and metadata experts associated with the Worldwide Web Consortium (W3C) desirous of stable, unchanging language-specific character code points; and most recently, ethnographically oriented field linguists, who have come to question an earlier lack of scholarly insight, coupled with conflicting stakeholder perspectives rather than more scientific linguistic, cultural and historical realities, particularly in the case of so-called ‘long-tail languages’.<sup>4</sup>

The initial ISO 639 developed by terminologists served as the progenitor to the current Part I two-letter codes for languages (*en, de, fr*). This approach arose out of the constrained paper fiche environment of terminographic practice before the digital age. The early PC, with its claustrophobic 360K floppy disks and limited RAM capacity, also boxed developers into a parsimonious language ID regime, reflected in the highly economical MicroMATER model. Unfortunately, however, the two-letter codes pose one insurmountable handicap: the mathematical permutation of the number two coupled with the 7- and 8-bit limitations of the character codes. Library scientists at the Library of Congress, also confined within the boundaries of paper fiches – the then common library catalogue cards – introduced three-letter codes for use in their MARC (MACHINE Readable Cataloguing) record (LoC 2009). The advantage of three letters is that they provide more satisfying mnemonics as well as enough code points to accommodate realistic projections of the world’s identified languages, but the library-oriented system is nonetheless administratively limited to languages that produce sufficient documents to be catalogued. SIL International, which has its origins in the *Summer Institute of Linguistics*, started with a primary focus on Bible translation, later branching out to support literacy in a wide range of language communities. As a consequence, the SIL three-letter codes (*eng, deu, fra*) have been motivated towards total inclusion, resulting in July 2018 in a total of 7097 languages (SIL 2018a & b).

Over time, the language codes have, however, played an integral role not only in identifying languages, but also in pinpointing locale-related information types. In parallel to the language codes, ISO 3166-1:2013, *Codes for the representation of names of countries and their subdivisions* is maintained by the ISO 3166 Maintenance Agency, which consists of a set of prominent ISO Participating Member bodies, along with a few international entities, including the United Nations Economic Commission of Europe, the Universal Postal Union, and the Internet Corporation for Assigned Names and Numbers (ICANN). The list provides alpha-2 and alpha-3 country code elements as well as numeric country codes assigned by the UN and comprises essentially those countries that are recognized as member bodies of the UN. The final piece in this particular encoding puzzle is ISO 15924, which provides four letter codes for the names of scripts and is maintained by an ISO Registration Authority under the aegis of UNICODE.

Where SGML (Standard Generalized Markup Language) and its web-enabling descendent HTML reported language identity as values of the encoding attribute *lang*, XML’s *xml:lang* attribute introduces a detailed sequence of codes based on a combination of the language symbols with country and script codes to report very specifically on the locale and character set of a document. Typical examples of combined locale codes might be *fr-CA* (French as spoken in Canada) or *zh-Hant-TW* (modern Chinese printed in traditional script as used in Taiwan). This composite form makes up a *language tag* (Ishida 2014, IETF BCP 47). In addition to their use in the language tags described here, the

country codes also form the basis for country-specific geographical Top-Level Domains (TLDs) on the web (Technopedia 2018).

### *Character encoding*

The digital turn cited above produced a broad spectrum of language-related standards that had been developing over the course of the mid-century. In the beginning, character representation in computer environments was sketchy at best. Some scripts did not even exist yet – a truly legible Arabic font, for instance, had to be designed in order to process the language with first-generation terminology applications, and a script encoding developed for one platform would in all likelihood be unintelligible if data were transferred to another environment. The ASCII code (American Standard Code for Information Interchange, sometimes referred to simply as ‘ANSI 1981’), notorious today for its 128-character limitations, nevertheless represented a step forward because it provided a compliant 7-bit realization of both lowercase and uppercase text for English (Brandel 1999). ASCII was extended first by 8-bit upper-ASCII, which accommodated most Latin-character languages, but was not truly implemented in much software until the mid-1980s. Augmented by the extensive ISO 8859 family of standards (8-bit single-byte coded graphic character sets, which Michael Everson has characterized, however, as a ‘font hack’ (NPR 2003) because they simply re-assign character fonts to the same few actual code points), they accommodated most requirements for Latin-character languages as well as mappings to other alphabetic scripts, such as Greek and Cyrillic, but with limitations – one could not use much more than one set of characters at a time because there were still only a limited number of usable code points. By the early 2000s, however, development in the 8859 space ceased (2004), followed by full web adoption of Unicode (ISO 10646) in 2007. Detailed description of these character encoding systems exceeds the scope of this chapter, but it is worth noting that the Basic Multilingual Plain of UNICODE, UTF-8, provides 55,029 characters (as opposed to 128/256) and accounts for so-called Unified Han, including Japanese Kanji and Simplified Chinese, followed in 2006 by introduction of the second plain, UTF-16, to accommodate the complete Chinese encoding, not to mention a seemingly endless proliferation of emojis. Unicode provides ‘a unique number for every character, no matter what the platform, no matter what the program, no matter what the language’ (Unicode 2018a). UTF-8 has become the character encoding scheme of choice on the web and in numerous other platforms, especially for the accommodation of multiple languages where non-English orthographic conventions require Latin letters with diacritics, additional punctuation marks, and non-Latin character sets.

### *POSIX and the CLDR (Unicode Common Locale Data Repository)*

Language and locale information provided by the extended *xml:lang* does not, however, fully cover all the cultural and regional information related to programming and web environments. The IEEE (ISO/IEC 9945) *Portable Operating System Interface (POSIX)*, begun in the late 1980s, supported compatibility between operating systems. By 2001–2004, POSIX had added locale definitions, including a ‘subset of a user’s environment that depends on language and culture conventions’, which accommodated character classification and conversion, collating order, monetary information, non-monetary number representation, date and time formats, and messaging formats (IEEE 2018).

Currently locale detail is documented using the Unicode *Common Locale Data Repository* (CLDR), including language tags, calendar style, currency, collation type, emoji presentation style, first day of week, hour cycle, line break style, word segmentation, number system, time zone, and POSIX legacy variants, in addition to transliteration guidelines (ICU; UNICODE 2018b). Language codes and CLDR notation are intimately embedded throughout the web and encoded in computer programs, resulting in an entailment that vigorously resists any trivial change, rendering the web community a powerful stakeholder that advocates for the status quo. Consequently, in the early 2010s, when socio-linguists proposed completely revamping the codes to eliminate legacy errors and to reflect modern scientific definitions of language and language variants, the UNICODE community, supported by IETF, the W3C, and the CLDR initiative, asserted the need for stability and continuity in the current system (Morey and Friedman 2013). The Maintenance Authority does, however, correct errors on a regular basis.

### *The language industry and the translation extension*

Galinski categorizes 1989 to 2003 as ‘extension’ years, citing SC 3’s move into computer-supported terminology management and the founding of SC 4 in 2002. This period marks a definite turn towards translation-oriented topics. Galinski’s terminal date of 2003 is artificial, however, because, although it marks a milestone for Infoterm, it does not map well to overall standards evolution. During this period, industry-oriented standards activity outside the framework of the national and international standards bodies began. It is tempting to select 2008/2009 as a critical intermediate juncture, because there is somewhat of a break in the documentation of standards development at this point, but this blip is more relevant to the global financial crisis than to the development of translation and localization industry standards. 2012 makes more sense because this year sees the creation of TC 37 SC 5 for *Translation, Interpreting and Related Technology* in 2012.

To understand these trends in language standards, it is important to look back at the creation of the Internet, followed by the appearance of the web, which was accompanied by the growth of multiple layers of enabling technologies supporting public, academic, and commercial computing environments. Over this time and in a number of venues, experts from a variety of major web technology companies came together to form organizations and associations to achieve consensus on best practices, particularly with regard to the translation and localization of information, computer programs, and web content. Often, their projects reflect pragmatic efforts to achieve consensus-based solutions more rapidly than is sometimes the norm in the formal ISO process.

### *Localization Industry Standards Association (LISA)*

LISA was founded in Switzerland in 1990, but only survived for a little over two decades, dissolving in 2011. LISA served as a professional and trade organization for companies, translators, language engineers, and other experts involved in the translation and adaptation of software and web content into a variety of languages. LISA hosted global conferences and promoted the interests of globalization and internationalization in the context of the localization industry, in addition to supporting the elaboration of a family of XML-based standards. For purposes of this discussion, localization involves the adaptation and transcreation involved in configuring a translated resource (computer program, web page, or even a concrete product like an automobile) to the expectations of a target

language and culture (see Folaron in this volume). At the time of its demise, the list of LISA standards included (LISA 2004, IEC 2015):

- TBX: Term Base eXchange (TBX);
- TBX: TBX-Basic (the first industry-supported public dialect of TBX);
- TMX: Translation Memory eXchange (TMX);
- SRX: Segmentation Rules eXchange (SRX);
- GMX: Global information management Metrics eXchange ;
- xml:tm: XML text memory for storing text and translation memory in XML documents;
- Term Link: A lightweight standard for linking XML documents to terminology resources.

Roturier in this volume provides a thorough discussion of these standards, with the exception of the last two, whose functions have essentially been subsumed in the OASIS-based XLIFF standard. As noted in the earlier chapter, the rights to these standards were turned over to ETSI, the *European Telecommunications Standards Institute*, where they were further elaborated until August of 2015, when ETSI discontinued the assigned Special Interest Group. This move probably reflects lack of buy-in from major industry players, who by then had made a significant commitment to OASIS. The TBX standard, together with its TBX-Basic dialect, had, however, already been ceded to ISO as ISO 30042, so it has continued development under the auspices of ISO TC 37/SC 3 and the TerminOrgs group, *Terminology for Large Organizations*, ‘a consortium of terminologists who promote terminology management as an essential part of corporate identity, content development, content management, and global communications in large organizations’ (TerminOrgs 2018). TMX, SRX, and GMX remain on the ETSI website for download, but are unfortunately not currently being updated. Efforts to resolve copyright issues between the OASIS/XLIFF group and ETSI have, at least for the present, failed, which may well have been a bad move on the part of ETSI, as all indicators point to a robust, comprehensive XLIFF as the standard of the future. LISA also developed a *Best Practice Guide and a Quality Assurance (QA) Model*, marketed as a computer utility, but also incorporated into some computer-assisted translation systems (Melby *et al.* 2001, LISA 2004).

## OASIS

Founded in 1993, OASIS (the *Organization for the Advancement of Structured Information Standards*) is an international association devoted to developing guides and standards for use in XML environments, both on the web and across networks. The emphasis of many OASIS standards is on text and information processing and interoperability. The great strengths of OASIS are the broad range of its standards and the active participation of a wide spectrum of major industry partners, not just from the web and computational environment, but also from manufacturing, government, and academia. Many collaboration projects also exist between OASIS committees and national, regional, and international standards bodies. OASIS is the home of DITA and of XLIFF (OASIS 2018), as recounted in Chapter 3.

### *Analysis for this phase*

Leading up to 2008, articles and reports began to appear treating standards in the language industry. Already in 2003/2004, in a now inaccessible web-based report, Wright

asserted that the digitization of text, accompanied by the ability to chunk and identify strings and segments, enabled the creation of data objects using embedded (or for that matter, stand-off) metadata and the automatic processing and manipulation of text by marking it for semantic content, thus facilitating the exploitation of language as both process and product. This trend transformed the traditional document production chain into what is understood today as a feedback-rich information life cycle, supporting data migration, conversion, re-importation and reutilization on a global scale. Wright's study identified XML and web-related standards, such as RDF, ISO 11179 (metadata registries), OMG's UML standard, UNICODE, and TC 46 activities surrounding controlled vocabularies (now ISO 25964), along with a number of W3C protocols, as a core stratum of standards supporting the information life cycle.

On this layer there resides a family of XML tools (many of which started out in the SGML environment) that was revolutionizing the authoring function, leading eventually to standards governing language mediation. OASIS' XLIFF (2003) already commanded attention in the translation area, as did LISA's TMX and segmentation standards, as well as controlled language solutions such as *AECMA Simplified English*. This list includes, in addition to LISA/OSCAR XML standards, W3C's OWL Web Ontology Language and a number of now defunct initiatives such as OLIF (Open Lexicon Interchange Format) for terminology interchange, along with Topic Maps and DAML-OIL. Under the heading of *Corpora*, Wright referenced the then active EAGLES project and TEI, precursors of TC 37/SC 4's mission. Toward the end of her report, Wright stated, '...formats like RDF, with its promise to revolutionize the Web in the direction of the visionary Semantic Web, have not yet met their full potential, but as more and more implementations are put in place, the power of this approach will increase momentum in this direction'. She predicted the implementation of full UNICODE compliance and 'semi-automated authoring environments', which she surmised would facilitate the embedding of semantic markers in text. She also expressed concern that TBX would be unlikely to gain wide acceptance until such time as it could achieve 'black-box turnkey status' without the user necessarily understanding how it works.

2004 was evidently the year of the private industry report, since Wright also created a second such report where she classified language standards according to a 'translation pyramid', a five-sided model with the translator at its base, supporting four elements: Text, Task, Tools and Training. Here she viewed the then prevalent normative trends as facets of incipient efforts to apply standardization approaches to translation and localization. Drawing on a background in manufacturing and engineering, she foreshadowed the later trend to apply the ISO 9000 Total Quality Management approach to the assessment of translation product and the evaluation of translators. This then relatively novel approach was melded with notions of equivalence and adequacy, which were surfacing in translation studies, stressing the affinity of the target-oriented translation brief to the ISO 9000 principle that quality is a function of client specifications. She visualized ISO 9000 as a top-level umbrella standard over-arching a level 2 (industry-specific certification standards, then reflected by ASTM F2575 and CEN 15038); level 3 (industry-specific metrics, reflected at the time in SAE J2540 and the ATA Framework for Error Marking); and level 4 (enterprise-specific customized procedures and metrics). This report also invokes the notion of Failure Mode and Effect Analysis (FMEA), an essential risk-management component of the ISO 9000 toolkit that is indeed being applied today as part of quality certification for Translation Service Providers. She drew an analogy between the FMEA Critical Items List (CIL) to quality metrics error categories, the

ATA Framework for Error Marking, and the LISA Quality Metrics project, pointing to existing or proposed industry standards related to her four major criteria (See also Wright 2013 and 2017).

In 2006, Lommel also focused on the role of digitization in the form of XML serialization used to transform human knowledge into ‘information as a business commodity’ (Lommel 2006: 225). Like Wright, he outlined a model whereby units of knowledge marked as units of text are abstracted into reusable and processable information units that can be repurposed on an industrial scale, which essentially describes information-centric environments where content is managed for a range of potential applications. In this context, he also described the then evolving LISA XML standards and stressed the advantage of vendor independence, collaborative environments, and data interchangeability in XML-empowered scenarios.

### *Translation standards*

Previous discussion has already set the stage for initial efforts to integrate language mediation activities into the environment of industrial standards.

### *Certification and service standards*

Work began in the late 1990s on ISO 12616: *Translation-oriented Terminology* (published in 2005), which may have been originally perceived to be a one-off extension of the SC 2 mandate to elaborate terminography-related standards. Along these lines, the translation industry did find a home in TC 37 in 2005, when a Working Group for *Translation and Interpreting Processes* was formed in SC 2. By 2012, activity migrated to a new SC 5, *Translation, Interpreting and Related Technology*. DIN issued a standard on translation contracts in the late 1990s, with ASTM following in 2003 and 2005 with standards on interpretation practice and requirements as well as customer-oriented guidelines for negotiating translation and localization specifications. CEN worked in parallel, publishing their EN 15038: *Translation services: Service requirements* in 2006.

Today, the most prominent of the translation standards, is ISO 17100:2015: *Translation services – Requirements for translation services/Amd 1:2017*. This standard has now become the lynchpin document for the certification of translators and translation service providers in the ISO 9000 environment, and it embodies the difficulty of establishing consensus on an international scale when confronted by contextual and linguistic differences across cultures. The original unamended standard published in 2015 stated that translators shall have earned ‘a **graduate qualification** in translation’ [emphasis for purposes of this discussion], which on the face of it may sound like a good idea, but in North America, this wording means 1) a Master of Arts degree at least, and 2) a degree that has ‘Translation’ in the title of the concentration. The intent of the drafters of the passage, however, was that translators should have graduated with an initial degree from a university, that is, with a baccalaureate degree. It took two years of arguing for the working group to accept the fact that it would be easier to change the wording in the standard than to change extensive usage by native speakers of the English language. The second problem was that although degrees ‘in Translation’ are common in Europe and in European languages in North America, they are not readily available in some countries – Japan, for instance, and for some languages. Hence the amended 2017 wording: ‘a) has obtained a degree in translation, linguistics or language studies or an equivalent degree that includes significant translation training, from a recognized institution of higher education; b) has

obtained a degree in any other field from a recognized institution of higher education and has the equivalent of two years of full-time professional experience in translating'. These requirements, along with an option for five years of full-time professional experience, pose problems, however, for speakers of long-tail languages who do not have access to formal curricula or who rarely have enough work to qualify as full-time translators. In its broader programme, SC 5 maintains an ongoing commitment to more specialized standards in translation and interpreting, specifically for community, legal and health-related interpreting. One major effort focuses on a set of standards for various types of interpreting equipment (see Braun in this volume), with an upcoming trend focusing on requirements for leading-edge mobile systems.

## Emerging issues and outlook

Important examples of new initiatives include the following. In SC 2, efforts are progressing to define criteria for linguistically motivated classification of language variants in ISO 21636: *Identification and description of language varieties*, which will define critical metadata values (data categories) for the exchange of secondary information as well as to support the re-usability of language resources. This activity addresses Morey's criticisms cited above (Morey and Friedman 2013) of inadequacies inherent in the language codes without disturbing the utility of the current system.

DIN in conjunction with German tekomp has submitted a work item on the *Vocabulary of Technical Communications*, which has led to the formation of a new TC 37 subcommittee for dealing with linguistic (as opposed to product-related) aspects of technical communication. Initial discussion of the standard places this terminology management activity firmly in the framework of the product information life cycle (Klump 2018).

A set of standards initiated by SC 4 features text markup for the creation of annotation schemes for linking semantic and syntactic elements in text to the generation of technical diagrams (ISO24627-3). The new standard is associated with existing SC 4 annotation models, but has the potential to interact with both SC 1 efforts to diagram concept structures as well as to integrate data-driven diagramming graphics into technical texts. Although the scope is currently limited, future directions might lead to standards for computer-generated scientific diagrams.

SC 3, partly with its ISO 12620 Data Category standard and the creation of the www.datcatinfo.net data category repository has long maintained a relationship with ISO/IEC JTC 1/SC 32 for metadata, patterning their work after standards for data element dictionaries. A new liaison with IEC SC3D: *Product properties and classes and their identification* expands on the definition of metadata elements in the context of the *International Electrotechnical Commission, IEC 61360 – Common Data Dictionary* (IEC 2018b)

These moves into technical communications, on the one hand, and metadata, on the other, reflect a holistic trend toward enterprise-wide document production and information management. Emboldened by some past success in predicting future directions for language standards, one might venture to suggest that this tendency will grow as standardizers try to support information management by integrating workflows and resources throughout enterprise processes. This trend can already be seen in the linkage between XLIFF, ITS and TBX, for instance. This kind of through-composed approach to technical writing, translation, localization and information management would ideally involve the coordination of standards cited here with other special

vocabulary standards, such as the SKOS (*Simple Knowledge Organization System*) from ISO Technical Committee 46, in order to create tools for linking data both within enterprises and across the web in support not only of translation and localization, but also of writing and information retrieval from massive bitext and multilingual text resources (SKOS 2018). Although building blocks are in place, future developments must of necessity lead in new directions to bind existing language standards with work going on in the W3C and elsewhere.

ASTM is elaborating a work item, provisionally called ASTM WK 46396, designed to standardize the top level of DFKI's EU-funded iteration of the old LISA Multidimensional Quality Metrics (MQM) (see Doherty and Pym's respective chapters in this volume, Lommel *et al.* 2015, Burchardt 2016, Görig 2017), leaving the rich detail of the subsequent layers to be presented as ancillary material in a flexible web-based environment (Lommel *et al.* 2015). In a parallel activity, ASTM is also developing a standard for holistic evaluation of published translations intended as a kind of triage activity designed to reveal when published translations are unacceptable to the target audience. A similar more recent ISO activity, ISO 21999, has the intent of introducing an as-yet-undefined quality metric. These standards add a critical product-oriented component to the process-oriented QA environment comprising the application ISO-9001 plus ISO 17100.

## Conclusion

This chapter chronicles a progression from engineer-linguists like Eugen Wüster, who inspired engineers standardizing vocabulary to serve the purpose of engineering-oriented standardization, to linguists applying computer engineering principles for the purpose of controlling and analysing language, and ultimately, of engineering knowledge, giving us today's notion of knowledge and language engineering. It begins with the efforts of terminologists, lexicographers, librarians and field linguists to normalize language names and word-related data categories, but then segues into standards that enable tools for a burgeoning translation/localization industry. It demonstrates the evolutionary trajectory from modest efforts to define small sets of highly specialized terms to culminate in information-oriented environments where standards-driven linguistic markup enables the transformation of surface-level word and character strings into semantically encoded terminological knowledge units populating massive data stores in the form of text and bitext corpora, supporting multilingual retrieval, text processing and further manipulation.

## Notes

- 1 See the abbreviations list in the References section for full forms of standards bodies and related acronyms.
- 2 Even as this article was being written, the core specifications of the treaty changed for the first time (and potentially last) time.
- 3 The reader should not imagine cassette tapes of any size: at end of day, enterprises would record entire data transactions for the day on a magnetic tape or tapes, often twice the size of a movie reel, and cart them off to some external storage site for safe keeping. In 1980, major institutions and companies had entire floors devoted to the storage of large reels of magnetic tape for data storage.
- 4 See <http://rosettaproject.org/blog/02010/jul/21/building-audio-collection-all-worlds-languages/>
- 5 \* = Under Development [SC3]

## *Standards, Bodies and Related Acronyms*

AECMA, European Association of Aerospace Industries  
AFNOR, Association française de normalisation  
ANSI, American National Standards Institute  
ASCII, American Standard Code for Information Exchange  
ASTM International [Originally: American Society for Testing and Materials]  
BIPM, International Bureau of Weights and Measures  
BSI, British Standards Institution  
CEN, Comité européen de normalisation  
CLDR, (Unicode) Common Locale Data Repository  
CoT, Committee on Terminology (ASTM)  
DFKI-MQM, Deutsches Forschungsinstitut für künstliche Intelligenz-Multidimensional  
Quality Metrics  
DIN, Deutsche Institut für Normung  
ETSI, European Telecommunications Standards Institute  
ICANN, Internet Corporation for Assigned Names and Numbers  
ICU, International Components for Unicode  
IEC, International Electrotechnical Commission  
IETF, Internet Engineering Taskforce  
ISO, International Organization for Standardization  
ISO/IEC JTC1, ISO/IEC Joint Technical Committee 1  
JTC, Joint Technical Committee  
LISA, Localization Industry Standards Association  
LoC, (US) Library of Congress  
MARC, Machine Readable Cataloguing  
OASIS, Organization for the Advancement of Structured Information Systems  
OMG, Object Management Group  
POSIX, Portable Operating System Interface  
SC, Sub-Committee  
SI Units, International System of Units  
SIL International, *originally* Summer Institute of Linguistics  
TC, Technical Committee  
TEI, Text Encoding Initiative  
UNICODE,  
UTF, Unicode Transformation Format  
W3C, Worldwide Web Consortium  
WG, Work Group, Working Group

## **Further reading**

- Filip, David (2019) 'Localization Standards Reader 4.0'. *Multilingual: language tech business*. #181 30 (1, January/February). Available online: <https://magazine.multilingual.com/> [last accessed 18 Jan. 2019]. | *Filip's in-depth article offers a highly technical, comprehensive overview of standards governing the localization process, including relevant Internet management standards, text encoding, translation and terminology-related management and interchange standards, quality metrics and standards designed for implementing specific locale-related details.*
- ISO 17100:2015/Amd. 2017. *Translation services – Requirements for translation services* | *Provides requirements for the core processes, resources and other aspects for the delivery of quality translation services according to specifications and supports demonstration of conformity, including capability of processes and resources to deliver translation services that meet the client specifications.*
- GALA Compendium, 2016. <https://web.archive.org/web/20160305195016/http://lsrp.galacrisp.org/#aboutthissite.html>. | *This comprehensive and informative web resource was originally published by GALA-CRISP, but has been taken down as of 08/04/2018. This web archive version is from March, 2016, so some pieces of information are outdated, but the scope of the information remains*

*exceptional. The current GALA webpages are better than most resources, and many items are downloadable:* [www.gala-global.org/translation-and-localization-industry-standards](http://www.gala-global.org/translation-and-localization-industry-standards).

## Related topics

- 3. XML for translation technology
- 12. Technology, technical translation and localization
- 16. Technology and interpreting
- 20. Translation technology evaluation research
- 26. Quality

## References

References to actual standards: Please see *Language Standards Organized by Topic* list at the end of this chapter and [www.ttt.org/standards-lists/](http://www.ttt.org/standards-lists/).

- AFNOR (2018) ‘Un peu d’histoire: AFNOR a 90 ans!’. Available online: <https://fr.linkedin.com/pulse/un-peu-dhistoire-afnor-90-ans-olivier-gibert> [last accessed 2 Jul. 2018].
- ANSI (2018) ‘Through history with standards’. Available online: [www.ansi.org](http://www.ansi.org) [last accessed 2 Jul. 2018].
- ASCII (2005) ‘ASCII Code – The Extended ASCII Table’, Available online: <https://ascii-code.com> [last accessed 2 Jul. 2018].
- ASTM (2018) ‘The history of ASTM International’. Available online: [www.astm.org/about/history\\_book\\_html#chapterOne](http://www.astm.org/about/history_book_html#chapterOne) [last accessed 2 Jul. 2018].
- BIPM (1875-Present) ‘Measurement Units: The SI’, Bureau International des Poids et Mesures. Available online: [www.bipm.org/en/measurement-units/](http://www.bipm.org/en/measurement-units/) [last accessed 9 Jul. 2018].
- Brandel, M. (1999) ‘1963: The Debut of ASCII’. Available online: [www.cnn.com/TECH/computing/9907/06/1963.idg/index.html](http://www.cnn.com/TECH/computing/9907/06/1963.idg/index.html) [last accessed 9 Aug. 2018].
- Briggs, B. (2004) ‘The importance of TMX for translation tool buyers’ [No longer available] [www.languagepartners.com](http://www.languagepartners.com)
- BSI (2018) ‘Our history’. Available online: [www.bsigroup.com/en-US/about-bsi/Our-history](http://www.bsigroup.com/en-US/about-bsi/Our-history) [last accessed 2 Jul. 2018].
- Burchardt, A. (2016) ‘*Harmonization of DQF and MQM sets the stage for a standard translation quality metrics*’. Available online: [www.slideshare.net/TAUS/harmonization-of-dqf-and-mqm-sets-the-stage-for-a-standard-translation-quality-metrics-aljoscha-burchardt-dfki](http://www.slideshare.net/TAUS/harmonization-of-dqf-and-mqm-sets-the-stage-for-a-standard-translation-quality-metrics-aljoscha-burchardt-dfki) [last accessed 24 Jul. 2018].
- Campo, A. (2013) *The reception of Eugen Wüster’s work and the development of terminology*. Département de linguistique et de traduction. Faculté des arts et des sciences, Université de Montréal.
- Ellis, W. (1988) ‘Management of technical terminology in standards organizations’, in R. Strehlow (ed.) *Standardization of terminology: principles and practices*. VII. STP991. Philadelphia: ASTM, 62–75.
- Felber, H. (1984) *Terminology manual*. Paris: Unesco, Infoterm.
- Galinski, C. (2004a) ‘30 Years of Infoterm’, Infoterm. Internal document.
- Galinski, C. (2004b) ‘50 Years ISO/TC 37 ‘Terminology and other language resources: A history of 65 years of standardization of terminological principles and methods’, Internal ISO TC 37 document.
- Görlig, A. (2017) ‘The 8 most used standards and metrics in translation quality evaluation’. Available online: <http://blog.taus.net/the-8-most-used-standards-and-metrics-for-translation-quality-evaluation> [last accessed 24 Jul. 2018].
- ICU International Components for UNICODE (2018) ICU 62 (The premier library for software internationalization). Available online: <http://site.icu-project.org/download/62#TOC-Release-Overview> [last accessed 12 Jul. 2018].
- IEC (2015) ‘Electro LISA: LISA OSCAR Standards’, *The GALA website*. Available online: [www.gala-global.org/lisa-oscar-standardsv](http://www.gala-global.org/lisa-oscar-standardsv) [last accessed 23 Jun. 2018].