MACHINE TRANSLATION AND GLOBAL RESEARCH

This page intentionally left blank

MACHINE TRANSLATION AND GLOBAL RESEARCH: TOWARDS IMPROVED MACHINE TRANSLATION LITERACY IN THE SCHOLARLY COMMUNITY

BY

LYNNE BOWKER
University of Ottawa, Canada

JAIRO BUITRAGO CIRO
University of Ottawa, Canada

United Kingdom – North America – Japan – India – Malaysia – China

Emerald Publishing Limited
Howard House, Wagon Lane, Bingley BD16 1WA, UK

First edition 2019

Copyright © 2019 Emerald Publishing Limited.

Reprints and permission service
Contact: permissions@emeraldinsight.com

No part of this book may be reproduced, stored in a retrieval system, transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without either the prior written permission of the publisher or a licence permitting restricted copying issued in the UK by The Copyright Licensing Agency and in the USA by The Copyright Clearance Center. Any opinions expressed in the chapters are those of the authors. Whilst Emerald makes every effort to ensure the quality and accuracy of its content, Emerald makes no representation implied or otherwise, as to the chapters' suitability and application and disclaims any warranties, express or implied, to their use.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-1-78756-722-1 (Print)
ISBN: 978-1-78756-721-4 (Online)
ISBN: 978-1-78756-723-8 (Epub)

ISOQAR certified Management System, awarded to Emerald for adherence to Environmental standard ISO 14001:2004.

Certificate Number 1985
ISO 14001

INVESTOR IN PEOPLE

Dear families: Here is the book.

This page intentionally left blank

Contents

This page intentionally left blank

List of Tables

This page intentionally left blank

About the Authors

Lynne Bowker holds a PhD in Language Engineering from the the University of Manchester Institute of Science and Technology (UMIST) in the United Kingdom. She is a Full Professor at the University of Ottawa in Canada, where she holds a cross-appointment between the School of Translation and Interpretation and the School of Information Studies. She is the author of *Computer-Aided Translation Technology* (2002, University of Ottawa Press) and co-author, with Jennifer Pearson, of *Working with Specialized Language: A Practical Guide to Using Corpora* (2002, Routledge).

Jairo Buitrago Ciro holds an ALA-accredited Master of Information Studies from the University of Ottawa in Canada. He has previously worked as an Academic Librarian at the Université du Québec en Outaouais in Canada and as the Director of the Library at the Universidad Simón Bolívar in Colombia. He is currently pursuing a Doctorate in Electronic Business at the University of Ottawa.

This page intentionally left blank

Introduction

Why this book?

Machine translation is hilarious! Just take a look at the Web series "Google Translate Sings" created by YouTube performer Malinda Kathleen Reese in which she takes songs by famous artists, runs the lyrics through Google Translate until they barely resemble the original text, and then performs the song with the new lyrics.[1] As Malinda promises, hilarity ensues.

But while we all appreciate a good multilingual chuckle, there are times when a language barrier is no laughing matter. For better or worse, English has emerged as the international language of scholarly communication – particularly in the domains of science and technology – despite the fact that only roughly 6% of the world's population speaks English as a native language. So what does this mean for the other 94%? It means that they need tools, techniques, and training to help them engage with and contribute to the scientific literature in their field for the betterment of our whole society.

Can machine translation help? Maybe. Sometimes. For some purposes. The answer is not clear-cut, but the question is certainly worthy of investigation. If we want the best and the brightest minds on the planet working together to solve problems such as climate change, cancer, and energy crises, then we need to make sure that they can effectively share their research findings with one another.

The goal of this book is to inform you about the ways that machine translation is used in the context of scholarly communication, and to teach you some ways in which this tool can be used more effectively. It's true that online machine translation systems are almost ridiculously easy to use …. Select a text, choose a language, click "Translate," and *voilà*! However, making critical and effective use of machine translation as part of the scholarly communication process is a little more complex. There is an emerging yet rapidly increasing need for machine translation literacy among members of the scientific research and scholarly communication communities. Yet in spite of this, there are very few resources to help these community members acquire and teach this type of literacy. This book is designed to fill that gap by introducing machine translation in an accessible way and providing guidance about how to use this technology effectively

[1]https://www.youtube.com/channel/UCP2-S6-M9ZvlY8t7cRn4O6A

Machine Translation and Global Research:
Towards Improved Machine Translation Literacy in the Scholarly Community, 1–6
Copyright © 2019 by Emerald Publishing Limited
All rights of reproduction in any form reserved
doi:10.1108/978-1-78756-721-420191002

for scholarly communication. It assumes no prior specialized knowledge of or experience with machine translation, and it does not require knowledge of another language. In most cases, English-language examples are used as a basis for discussion in the book, but examples from other languages are integrated and explained where practicable. Used as a learning resource, the book will enable members of the research and scholarly communication communities to develop the knowledge and skills they need to ensure that the results of scientific research are being disseminated in a form that can be accessed more effectively by scholars who are not native speakers of English. Specifically, you will be able to do the following:

- Understand the challenges faced by non-Anglophone researchers and students who must engage with the scholarly literature and disseminate their findings in English.
- Recognize some of the ambiguities in natural language.
- Appreciate the limitations of machine translation systems.
- Understand the purpose of abstracts for scholarly communication.
- Analyze abstracts and identify linguistic structures that will pose problems for machine translation systems.
- Apply strategies to minimize linguistic ambiguity.
- Generate your own scientific abstracts that will be suitable for machine translation.
- Appreciate some of the "bigger picture" issues associated with using machine translation, such as privacy concerns, or the effect that machine translation is having on the translation profession and on the nature of scholarly publications.
- Share your knowledge of machine translation literacy with others.

Who is this book for?

A number of different groups have a stake in the area of machine translation literacy and scholarly communication, so this book will be of interest to a wide range of readers. However, it might first be useful to specify who this book is *not* intended for. The book is not aimed at machine translation researchers or developers, though they might find it interesting to learn more about how users interact with these tools. It is also not intended as a "how to" manual for professional translators who use or who are considering the use of machine translation tools for many of their professional activities; however, as discussed below, translators may find this book useful in contexts where they have the opportunity to support researchers as part of the scholarly communication process.

So then who *is* this book for? This book should be a useful resource for any higher education program in which academic writing, scholarly publishing, or information or digital literacy is taught as core or optional course units. Likewise, it can be helpful to professional associations that aim to provide professional development or continuing education in these subject areas for their members.

However, it can also function well as a self-study resource for emerging or established researchers or members of the broader scholarly communication community. Let's consider how this book will benefit these various groups.

First, there are a number of different types of information professionals for whom this book is relevant. Among these, academic librarians are a key group. A significant role of academic librarians is to support the scholarly communication process by supporting the research of students, faculty, and other scientific investigators at a university or research institution. One of the ways in which they do this is by promoting and providing training in the development of different types of literacies, such as information literacy, media literacy, and digital literacy. Hopefully, we will soon see machine translation literacy added to this list. While researchers and scholars provide the focus and the purpose for scholarly communication, librarians can be forceful advocates for positive change. Outreach to other members of the scholarly community is one important activity that librarians can undertake to promote positive change that will advance the scholarly communication system. Moreover, because the need for machine translation literacy is present in every discipline of scientific enquiry, librarians are well placed to step into the key role of developing and delivering a coherent program of machine translation literacy training to students and researchers across an entire research or higher education institution.

Of course, for librarians to be equipped to deliver training in machine translation literacy, they first need to learn about it themselves. Many librarians avail themselves of continuing education and professional development opportunities offered by professional associations, such as the American Library Association, the Association of College and Research Libraries, the Canadian Association of Research Libraries, the Canadian Association of Professional Academic Librarians, the Chartered Institute of Library and Information Professionals in the United Kingdom, the Australian Library and Information Association, or the Library and Information Association of New Zealand Aotearoa. These professional associations, who facilitate the training of trainers, are therefore another key target audience for this book, whose contents can be used as a resource for workshops, webinars, or other continuing education opportunities for library association members.

Professional abstractors are another type of information professional who will find the contents of this book pertinent to their professional activities. Abstracting–indexing organizations employ abstractors to generate content for their research databases, and these abstract writers will benefit from having an enhanced understanding of how machine translation will affect their texts and how they can prepare texts that can be translated more easily and accurately by machine translation systems.

Along with practicing information professionals, professors of library and information science will also find this book relevant. Its contents might logically find a home in a variety of different library and information science courses, including but not limited to those that deal with information literacy,

digital literacy, scholarly communication, indexing and abstracting, information retrieval, research and evaluation, and international librarianship. Indeed, while this book focuses on the use of machine translation for scholarly communication, other groups, such as school children or newly arrived immigrants, may benefit from similar instruction in machine translation literacy. Therefore, some of the concepts presented in this book may also be relevant for courses on school librarianship or public librarianship. By introducing library and information science students to the concept of machine translation literacy at an early stage of their career, professors will ensure that these students are well prepared to both apply and teach machine translation literacy skills once they become practicing information professionals.

Next, this book will be of interest to the teams of people outside the information professions proper who also support authors in the scholarly communication process. This includes peer reviewers, journal editors, scholarly publishers, and sometimes even translators, editors, or other "literacy brokers," whose job is to support the effective dissemination of research results to the wider scientific community. At various stages of the scholarly communication process, there are opportunities to give feedback to authors, including suggestions about how to write more effectively with a view to improving machine translation output. In many cases, scientific researchers – and particularly those who are native speakers of English – are simply not aware of the fact that other researchers may be running their texts through machine translation systems, nor of what happens to their texts once they have been processed by a machine translation system. By alerting these authors and making constructive recommendations for improving the translatability of their texts, peer reviewers, editors, and others involved in the scholarly communication process can contribute to a movement to raise awareness about how researchers who have English as an additional language, rather than a native language, might use these texts, and how authors can make them more usable. By developing their own machine translation literacy (i.e., by gaining a deeper understanding of machine translation and understanding how to work effectively with this type of tool), these groups will be better able to support the scholarly communication process at a broad level and to ensure maximum participation from researchers around the world.

This brings us to the researchers themselves who are seeking to disseminate their findings as effectively as they can to the broadest possible range of interested readers. To begin with, this includes native English-speaking researchers who, for the moment at least, have the luxury and the privilege of being able to disseminate the results of their research in their own language. To make sure that these results are maximally accessible to their peers and students who have English as an additional language, native English-speaking researchers will benefit from learning more about how machine translation may affect their texts, as well as learning how to produce texts that will be more amenable to machine translation (or even simply to ensure that their untranslated texts are easier to read for those who are not native English speakers). The best way to advance our collective knowledge and solve issues of global importance is to make sure that everyone who is

capable of contributing is able to participate in the conversation. By increasing their own level of machine translation literacy and preparing translation-friendly texts, native English-speaking researchers can do their part to facilitate specialized communication among scholars around the world.

While we would dearly like for researchers who have English as an additional language to be able to benefit directly from this book as well, we fully recognize the irony that, by writing it in English, we are contributing to the reinforcement of the current dominance of English as the international language of scholarly communication. However, we hope that by bringing to light some of the difficulties faced by non-Anglophone researchers, and some possible ways to address them by changing the way that researchers prepare texts and interact with machine translation systems, we can help to improve their situation, even indirectly. As we will learn later in the book, machine translation has still not reached a stage where its raw output can be used to generate publication-quality material; however, the guidelines in this book may act as a useful starting point that non-Anglophone researchers can employ to produce translation-friendly drafts that can later be machine translated and (self) post-edited. Moreover, machine translation can be an extremely valuable tool for them in the earlier stages of the research process, such as when they are searching for relevant research material as part of a literature survey. Meanwhile, given that machine translation technology continues to advance and may soon be even more useful for helping researchers to translate their articles from Chinese or Arabic or any other language into English, non-Anglophone researchers will benefit from developing sound machine translation literacy skills in order to be able to take full advantage of the future advances in this field.

Some of this book's contents may also be usefully incorporated into courses on English for Research Publication Purposes or courses on scientific or academic writing and publishing. Though machine translation literacy may not be the main focus of such courses, it will surely be useful for these students to keep machine translation applications in mind when preparing their own texts for publication.

Finally, though the general public is not intended to be a primary audience, this book may nonetheless be of interest to members of the wider public who are curious about machine translation and its role in the dissemination of information. As more and more countries welcome newcomers who speak other languages, and as a growing percentage of the world's population gains access to the Internet and the wealth of information that is stored there – in an increasingly diverse number of languages – people from all walks of life are coming into contact with machine translation and may be interested in learning how to make it work more effectively for them. For instance, the 2018 FIFA World Cup held in Russia was dubbed by *The New York Times* as "the Google Translate World Cup" because of the multitude of ways that this tool was used:

> Across Russia for the last month, fans (and journalists) have used translation apps for everything: asking for directions, chatting with taxi drivers, getting slightly nerve-racking haircuts, checking

> into hotels, making friends, even flirting. The app's camera function – which can scan and translate text – has allowed visiting fans to decode menus, decipher signs and read the names of subway stations, even if the Cyrillic alphabet remains a mystery to them. (Smith, 2018, p. D1)

Finally, at the very least, readers may be pleased to have learned about "Google Translate Sings" and the hours of entertainment that it can provide!

What's in this book?

Chapter 1 sets the scene by providing an overview of the state of scholarly communication and the evolution of English as the international language of research dissemination. Here, you will learn about the traditional model of scientific publishing as well as some of the challenges faced by researchers who are not native speakers of English but who wish to publish in English. Translation, and more specifically machine translation, is explored as an option for researchers who have English as an additional language, and some advantages and disadvantages associated with using machine translation to search for information, to assimilate scholarly material, and to disseminate research are considered. Finally, you will understand why there is an emerging need for a new type of digital literacy – machine translation literacy – to be developed.

Chapter 2 provides you with a more detailed look at the world of machine translation, beginning with a brief history of the field. Next, different approaches to machine translation are introduced, along with examples that reveal common types of errors and demonstrate why translation is difficult for machines. By understanding more about how machine translation systems work, and the associated challenges, users can devise ways of interacting with these systems to improve their output. Here, you will find out more about one such option – controlled language – and learn about its advantages and limitations in the context of scholarly communication.

Chapter 3 introduces the notion of writing for translation, and, in particular, writing with *machine* translation in mind. If authors recognize how machine translation is likely to be used by researchers who have English as an additional language, they can write in such a way as to improve the translatability of their abstracts and make it easier for everyone to understand the machine-translated contents. Here you will pick up some strategies and tips for writing abstracts in a way that is machine-translation friendly.

Chapter 4 steps away from the details of how machine translation systems work and how users can interact with them more effectively to consider the bigger picture and to explore briefly some of the wider implications associated with the use of machine translation in the context of scholarly communication. The decision to use machine translation will depend on a range of factors, and topics discussed in this chapter can help you to make informed choices.

Chapter 5 introduces a working definition and a framework for machine translation literacy that could be used by information professionals or other groups to design and promote effective instruction in machine translation literacy.