Lynne Bowker

Jairo Buitrago Ciro

Machine Translation and Global Research

Towards Improved Machine Translation Literacy in the Scholarly Community

機器翻譯與全球研究

Maskinoversettelse

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

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

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.

Chapter 1

Scholarly Communication

> The triple disadvantage of having to read, do research and write in another language.
>
> – van Dijk (1994)

This chapter introduces the field of scholarly communication and considers how English came to be established as the principal language of global research dissemination in today's society. The traditional model of scholarly publishing is presented, and some of the challenges faced by researchers who are not native speakers of English but who wish to publish in English are explored. Translation, and more specifically machine translation, is analyzed as an option for researchers who have English as an additional language, and some advantages and disadvantages associated with using machine translation for research dissemination and information assimilation are presented. Finally, we address the emerging need for developing and teaching machine translation literacy skills.

What is scholarly communication?

Scholarly communication is the process by which academics, scholars, graduate students and other researchers share and publish their findings so that they are available to the wider research community, and beyond. As part of the scholarly communication system, knowledge is created, evaluated for quality, disseminated, and preserved for future reference.

Those working in the arena of scholarly communication today might be grappling with current hot topics such as copyright, intellectual property rights, metrics for measuring the impact of research, open access, or institutional repositories. However, the cornerstone of scholarly communication is essentially the exchange of knowledge. Researchers communicated before notions such as open access and institutional repositories came into existence, and we can expect that they will continue to do so even if current models evolve beyond recognition.

That being said, the world of scholarly communication is certainly in the midst of interesting times. On the one hand, in the years since World War II, the rate

Machine Translation and Global Research:
Towards Improved Machine Translation Literacy in the Scholarly Community, 7–36

doi:10.1108/978-1-78756-721-420191003

of scientific and technological research and discovery has expanded rapidly. New subject fields, specializations, and interdisciplinary areas of investigation abound, and with them come new terms and phrases to describe the new concepts, inventions, techniques, and practices. On the other hand, during the same period, the English language has increasingly asserted itself as the language of dominant use in many specialized fields. Both of these trends present opportunities and challenges for scholarly communication.

From a scientific boom to a scientific literature boom

The post-war flurry of scientific and technological activity bore witness to a corresponding explosion in the number of scientific journals that were published. According to Horguelin and Clas (1966, p. 16), this number more than doubled in the 20-year period after the war, rising from 24,000 to 60,000. In a more recent study, researchers looked at the rate at which science has grown in terms of number of publications and cited references since the mid-1600s (Bornmann & Mutz, 2015). Their bibliometric analysis revealed three essential growth phases in the development of science, and in each case, the growth rate had tripled in comparison with the previous phase. It rose from less than 1% up to the middle of the eighteenth century to 2–3% up to the period between the two world wars, and 8–9% between the end of World War II and 2010. Meanwhile, a researcher at the University of Ottawa estimated that more than 50 million scientific articles were published between 1665, when the first modern journal was established, and 2009, when he conducted his study (Jinha, 2010, p. 258). At this rate of production, scientists may soon be able to challenge fast food giant McDonald's by claiming "over 100 million served" … and counting!

As described by Delisle (2008, p. 43), a technological boom is inevitably accompanied by a terminological boom. Of course, this is not necessarily a bad thing; the birth of new fields, and thus new vocabularies, has long been seen as a defining aspect of scientific progress. However, Montgomery (2004, p. 1335) also draws attention to a number of linguistic challenges associated with such progress, noting, for instance, that increasing specialization has presented an ever-greater need and range of opportunities for the sharing of terminology:

> the power to examine, analyze, and manipulate phenomena at smaller and smaller scales has brought the province of the molecular, once reserved for chemists, into immediate relevance for botany, zoology, medicine, meteorology, many areas of geology, engineering, and so on. This has meant the adoption of terminologies appropriate to such scales of observation and analysis.

Meanwhile, the cover of a 2015 special issue of *Nature* dedicated to the subject of interdisciplinarity proclaims that "scientists must work together to save the world." Working together effectively means sharing a common vocabulary. According to Gray (2008), researchers who participate in team science need strong leaders who are comfortable in the various specialized languages in play

and who can help to translate, for instance, what a microbiologist is saying in terms that are meaningful to an epidemiologist.

The rise of English(es) as the international language of scientific communication

Meanwhile, during this post-war period of intense scientific activity, another linguistic trend has solidified: English has increasingly asserted itself as the international language of scientific and technical communication. Delisle (2008, p. 43), in a comparison of the English and French languages, observes that since the era of industrialization, English began to take the lead as the language of science and technology, and by the 1960s, French was recording a "lexical deficit" of thousands of words per year as compared to English. In other words, new terms were being coined in English at a much faster rate than they were being created in French. In another example, a bibliometric study of public health research published in Europe between 1995 and 2004 covered 210,433 publications found in the Science Citation Index and the Social Science Citation Index (with exclusions of overlap). Of the publications, 96.5% were published in English (Clarke et al., 2007).

Though English is just one of many languages spoken around the world, it is clear that it has become the leading language for scholarly publishing. Montgomery (2004) suggests that this is in part because World War II greatly advanced the military, economic, technological, and political sway of the United States, and thereby also increased its cultural impact. The uptake of English in technical circles has also been aided by the rise of big science in the United States. "English, in a sense, has ridden a great wave of cultural and intellectual affluence," says Montgomery (2004, p. 1333).

The Internet, too, was developed in the United States and was dominated by English in its early days. While it has started to become more linguistically diverse, the websites dealing with research products (e.g., pre-print archives, institutional repositories, and online journals) still tend to favor English to a high degree. Likewise, institutional sites, such as those for research organizations, universities, international programs, and so on, are also typically presented in English, or at least have an English version available.

In addition, there has been the globalization of science. Industrialization in developing areas of the world, such as Latin America, the Middle East, and parts of Asia and Africa, has motivated the spread of research in many fields. Nowadays, important scientific meetings and conferences are held all over the world, thereby prompting a need for a *lingua franca* or common language of communication. Similarly, if a researcher from Brazil wants to contact or collaborate with a researcher in India, there is a very good chance that they will use English to communicate.

According to Bennett (2013, p. 170), the Bologna Declaration, signed in 1999 by the education ministers of 29 European countries has been another contributing factor. Intended to boost the competitiveness of the European system of higher education and increase academic mobility, the agreement effectively

obliged universities throughout Europe and beyond to align their systems with that of the United Kingdom. As a result, academics throughout the Bologna zone are now obliged to publish in international journals to secure research funding and career advancement.

Another noticeable trend is the number of university programs that are being taught through the medium of English in countries where English is not an official language. For example, Wächter and Maiworm (2014, p. 16) investigated the number of English-taught programs on offer across non-English-speaking Europe, and according to their study, the number went up from 725 English-taught programs in 2001 to a staggering 8,089 programs in 2014 – an increase of over 1000%!

In other cases, a program may be taught in the language of the country in question, but obligatory courses in English may be part of the curriculum. One such example is provided by Ford, Faires, Hirsch, and Carranza (2017), who describe a newly established master's program in Library and Information Science offered at a university in Honduras where students are required to take a "Technical English Applied to Librarianship" course. According to Ford et al. (2017, p. 81), this course was included in the program for several reasons, which include the following:

- To facilitate reading the library and information science literature, since most of the information in this field is written in English.
- To improve English language proficiency so that – in a world in which most scientific information is published in English – librarians are equipped to help their patrons' access, understand, and use information.
- To enable librarians to communicate with their colleagues at similar institutions in other countries and to establish strategic links for sharing information, staying up-to-date in their knowledge, and participating in professional exchanges.

In addition, many universities in English-speaking countries are welcoming increasing numbers of international students. According to the Institute of International Education (2016), in the 2015/2016 academic year, the number of international students in the United States topped one million and had increased by more than 7% over the previous year. China, India, and Saudi Arabia were the three countries with the greatest number of students in the United States. Meanwhile, data released by Citizenship and Immigration Canada (2016) indicate that in the year 2016, there were 412,101 international students in Canada, which is up significantly from 179,149 in 2007. The top three countries from which these international students hailed in 2016 are China, India, and Korea. Indeed, not only are North American institutions open to welcoming international students, many of them are actively trying to attract them. For instance, in the University of Ottawa's (2014) most recent strategic plan, internationalization is presented as one of the four main goals for this institution. According to this plan, by the year 2020, the University of Ottawa seeks to double the number of international graduate students and increase the number of international undergraduate students by 50%, for a total of 3,650 international students, or 9% of the entire student body.

However, while there is much talk of how English is becoming the global language of research, another question that has been raised is whether researchers everywhere speak the same English. No single country or culture owns or has control of the features or direction of scientific English – it is truly the result of an evolving and cumulative international effort – and so it may actually be more correct to speak of English*es* or varieties of English. In fact, it has been well documented that when English is imported into a linguistic community, it is modified and adapted (e.g., Crystal, 2003). Accordingly, the variety of scientific English that is spoken by researchers in Korea is not precisely the same as that spoken by researchers in China, or India, and so on. Of course, there must be a strong common core. If varieties of scientific English diverge too far, there would be no reason to employ them. Still, there are enough differences to make communication challenging at times.

Another type of challenge may be faced by those who have only a limited command of English. While English as the common language of science has great potential to increase possibilities for study-abroad opportunities, scientific exchange, international collaboration, and career mobility, these avenues may not be equally open to scientists or students who have varying degrees of English proficiency, and they may be particularly difficult for those with limited English proficiency to access. Such researchers and learners may find it challenging to produce scientific articles in English, or even to search for and understand English-language publications. How are these members of the research community tackling the problems of publishing and engaging with the English-language literature in their field?

Scholarly publishing

As noted above, the vast majority of leading international journals are published exclusively in English, and those that do publish in other languages often require an English-language abstract to accompany the article. As a result, researchers around the world work toward submitting their research, written in English, to those leading journals in order to achieve international dissemination, reach, and impact for their work (Bennett, 2013). As pointed out by Moreno (2010, p. 59), publishing in English is not a luxury but a necessity in some countries. Using the example of Spain, Moreno explains that the systems used to determine whether candidates are qualified for contracts and tenure-track positions at universities take into account how many of their publications have appeared in journals that are indexed in prestigious databases such as the Web of Science. Meanwhile, in another example, Corcoran (2015, p. 1) points out that in Mexican universities, non-native English-speaking PhD students may be expected to publish an article in an indexed journal in their field in order to fulfill program requirements for graduation, while established faculty members are expected to do so in order to achieve career advancement. Likewise, Anazawa, Ishikawa, and Kiuchi (2012, p. 188) indicate that in Japan, nurses who wish to advance their career to become nurse researchers need to be prepared to contribute to the body of international literature in English. Similarly, in a survey of 1,717 Spanish scholars across a

range of domains, López-Navarro, Moreno, Quintanilla, and Rey-Rocha (2015) found a strong association between publication in English and the desire to be recognized and rewarded. There is even evidence to suggest that this trend toward needing English-language skills is unfolding not only in the international arena but also at the national level in some countries. Meneses Benavides (2011, p. 80) observes that in Colombia, authors who wish to publish in some of that country's national Spanish-language scientific journals must provide an abstract in English as well as in Spanish, in order to increase the visibility of the work.

Olohan (2016, p. 138) provides an excellent summary of the traditional model of scholarly journal publishing:

1. Authors submit previously unpublished research in the form of an article.
2. The article is reviewed by two or more academic peers.
3. On the basis of reports and recommendations written by the reviewers, the journal editor will accept or reject the article. Acceptance is usually conditional on the author making certain modifications to improve the quality of the article before it is published; however, in a case where there are too many deficiencies, the article may be rejected and resubmission may not be encouraged.

Given this model, journal editors and peer reviewers are frequently considered to be the gatekeepers for their discipline, admitting those research contributions that they deem to meet certain quality standards, and keeping others out. For speakers of languages other than English, this can present significant hurdles. Flowerdew (1999) surveyed 585 Cantonese-speaking academics about their experiences trying to publish in English, and one of the significant findings was that just over two-thirds of the respondents felt that they were at a significant disadvantage as compared to English speakers. Perhaps more worrying, however, was that close to one-third of the participants felt that prejudices by referees, editors, and publishers placed scholars who have English as an additional language at a disadvantage when writing for publication. In a similar vein, Mur Dueñas (2012) provides an account of the struggles that Spanish-speaking researchers in the field of business faced in attempting to publish their research in international journals, noting that out of 24 papers drafted and (re)submitted over a six-year period, only half were successfully published in the journal to which they were initially submitted. Moreover, in the reviewers' reports that requested a major revision, there were a high number of non-specific negative comments related to language or style.

Some academic researchers, such Benfield and Feak (2006, p. 1728) believe that the privilege of being a native English speaker comes with a responsibility to help researchers who are not native English speakers with "the burden of English." Likewise, for Drubin and Kellogg (2012, p. 1399), the negative attitude of editors and reviewers to imperfect English is unacceptable, and they call their fellow English-speaking scientists to task on this issue, emphasizing:

> the importance of recognizing and alleviating the difficulties faced by non-native speakers of English if we are to have a truly

> global community of scientists. For scientists whose first language is not English, writing manuscripts and grants, preparing oral presentations, and communicating directly with other scientists in English is much more challenging than it is for native speakers of English. Communicating subtle nuances, which can be done easily in one's native tongue, becomes difficult or impossible. A common complaint of non-native speakers of English is that manuscript reviewers often focus on criticizing their English, rather than looking beyond the language to evaluate the scientific results and logic of a manuscript. This makes it difficult for their manuscripts to get a fair review and, ultimately, to be accepted for publication.

Drubin and Kellogg (2012, p. 1399) go on to suggest that the communications advantage realized by native speakers of English obligates them to acknowledge and to help alleviate the extra challenges faced by their fellow scientists from non-English-speaking countries. They argue that it is inappropriate to reject or harshly criticize manuscripts from non-native speakers of English based on errors of grammar, syntax, or usage alone, and they recommend that native speakers of English should offer understanding, patience, and assistance when reviewing or editing manuscripts produced by non-native speakers of English. They emphasize that reviewers should look beyond linguistic errors and focus on evaluating the science. At the same time, however, Drubin and Kellogg (2012, p. 1399) recognize that non-native speakers of English must also do their part by endeavoring to produce manuscripts that are clear, simple, logical, and concise. This brings us back to the following question: how can a non-native speaker of English produce a manuscript of publishable quality?

English for research publication purposes

The most obvious solution – though certainly not the easiest or quickest one! – is for researchers with limited English proficiency to improve their English to a relatively high degree. It is quite telling that a new subfield of English for Academic Purposes has emerged in the wake of the trend toward the adoption of English as the international language of scientific publishing. As explained by Cargill and Burgess (2008, p. 75), English for Research Publication Purposes can be thought of as a branch of English for Academic Purposes that addresses the needs of professional researchers and graduate students who are required to publish in peer-reviewed international journals.

The fact that researchers around the world are willing, and in some cases even eager, to learn English for Research Publication Purposes can be seen in various ways. For instance, ElMalik and Nesi (2008) report that in Sudan, a major player in the field of tropical and preventative medicine research, a need has been identified for training to help researchers acquire the English language skills required for international publication. Corcoran (2015, pp. 6–7) provides a detailed description of a popular course for Mexican doctoral students on academic writing for

publication. In Spain, there are reports that researchers are pleading to have additional courses for training in English in all scientific areas (Moreno, Rey-Rocha, Burgess, López-Navarro, & Sachdev, 2012, p. 159). Meanwhile, a survey of the attitudes of Spanish researchers at one university toward the use of English in academia revealed a generally positive attitude in spite of the associated challenges (Fernández Polo & Cal Varela, 2009).

Nevertheless, there are some drawbacks to consider too. For instance, Burgess, Gea-Valor, Moreno, and Rey-Rocha (2014, p. 81) uncovered some resentment on the part of faculty members in Spain because "they see redirecting ever more limited resources in their faculties to training in English for research publication purposes as prejudicing the teaching of other skills essential to their discipline." Indeed, learning English may not be a feasible top priority in all situations. As reported by Anazawa, Ishikawa, Park, and Kiuchi (2012, p. 635), Japanese nurses are increasingly required to have knowledge of the research literature published internationally that is relevant to their clinical practices, education, and research activities. However, in a follow-up study, these researchers point out that even though nurses in Japan understand the need to review research published in English, their daily work does not require them to use English, and many consider that it is more important for them to be able to obtain nursing information quickly to provide quality care than it is to spend time mastering the English language (Anazawa, Ishikawa, Park, & Kiuchi, 2013, p. 59).

While developments in the field of English for Research Publication Purposes appear promising, this approach clearly does not offer a quick fix. Researchers must invest considerable time and effort – and sometimes money – to reach a high level of functionality in another language. Further discussion on the topic of English for Research Publication Purposes is beyond the scope of this chapter, but interested readers will find plenty of food for thought in the burgeoning literature in this field, such the special issue of the *Journal of English for Academic Purposes* edited by Cargill and Burgess (2008), or Flowerdew (2015), among others. Meanwhile, let's consider another option that is regularly presented to researchers with limited English proficiency: assistance from some type of *literacy broker*. The term "literacy brokers" was coined by Lillis and Curry (2010, p. 88) to refer to "all the different kinds of direct intervention by different people, other than named authors, in the production of texts." This could include translators, revisers, editors, or proofreaders of various types.

Translation of scientific research

Translation is almost as old as writing itself, and it has a long and colorful history. It is well documented, for instance, that translation has accompanied virtually every significant scientific and technological discovery (Byrne, 2012). Translation is the means of exporting these inventions and discoveries to other languages and cultures, and interested readers may like to consult Montgomery (2000) for an interesting collection of historical case studies of scientific translation.

However, Burgess et al. (2014, p. 75) note that for researchers who have English as an additional language, finding a translator with sufficient knowledge of

the discipline and the particular issues addressed in a research article is often a major challenge. As Olohan (2016, p. 136) so plainly puts it, reading, understanding, and translating scientific research articles can be a real challenge. They are written by specialists, for specialists, and their linguistic and textual choices may not be easily comprehensible to a non-specialist reader. The authors assume that their readers are familiar with the existing body of knowledge in their field and they write accordingly. Therefore, they may not explain complex concepts or ideas, leaving the uninitiated reader struggling to understand.

To facilitate publication by non-native speakers of English, some publishers may offer translation and editing services to potential authors who are seeking to publish in English-language journals. One such example is Elsevier, one of the world's largest scientific publishers. Elsevier offers translation into English from 15 different languages: Arabic, Chinese, Czech, Dutch, French, German, Italian, Japanese, Korean, Polish, Portuguese, Romanian, Russian, Spanish, and Turkish.

On the website, Elsevier describes the translation process as follows:

> When you submit a document it will first be translated and then reviewed by published experts in your discipline. These translators have extensive experience with the research and writing conventions of your field, in both languages. In addition, your paper is reviewed and polished after translation to ensure the highest quality. Within 12 business days, you will receive an email notifying you to download your finished document.[1]

Elsevier goes to great pains to emphasize that all the specialist translators have PhDs or are PhD candidates at top universities in the United Kingdom or the United States, and that they have excellent communication skills. The translators undergo rigorous training and frequent performance reviews. In addition, each manuscript is reviewed by a managing editor with a track record of scientific publication. Translators are also bound by a confidentiality agreement, so authors may be certain that their unpublished research is secure.

Of course, quality comes at a price, and would-be authors can expect to pay accordingly for this type of specialized translation service. According to Burgess et al. (2014, p. 76), the costs of translation and editing services are all too often prohibitive for scholars with non-native proficiency in English, even for those who come from countries perceived as being relatively well off, such as Sweden (Olsson & Sheridan, 2012, p. 45). The translation industry tends to price services according to a price per word multiplied by the word count. The price per word (or per character, in the case of languages such as Chinese, Japanese, or Korean) will vary according to language combination, with some language pairs commanding a higher price than others. In large part, it depends on the number

[1]http://webshop.elsevier.com/languageservices/translationservices/pages/howdoesitwork.html

of suitably qualified and experienced translators available for the language pair in question as well the volume of work for those pairs. Similarly, translation prices will vary depending on the level of specialization of the text, meaning that academic research papers are likely to be at the high end of the price scale. The deadline can also affect the price, and a short turnaround time or rush job usually comes with a surcharge.

Elsevier has a transparent pricing policy. Prices are based on the length of the document and do not include applicable taxes. Customers can go to the website, enter the language combination and text length, and click a button to calculate the price for the job. As an example, in mid-2018, translating a 10,000-word text from Spanish into American English had a price tag of US$1,131. Meanwhile, translating a 10,000-character text from traditional Chinese into British English would cost US$1,800.

Taylor & Francis is another leading publisher of scholarly journals, featuring publications in the most areas of humanities, social sciences, behavioral sciences, and science, technology, and medicine sectors. On their website,[2] they note the following:

> If you already have a paper in Portuguese, Spanish, Chinese, or Japanese, Taylor & Francis' Translation Service will give you an accurate and high-quality translation in English ready for submission. Each manuscript is assigned by area of study to ensure accurately translated field-specific terms. When your paper is returned to you, it will be indistinguishable from papers written by researchers who are native English speakers.

The price quoted by Taylor & Francis for having 10,000-word text translated from Spanish to English in 12 days is US$1,185, while the price for a Chinese text is US$1,757.

This trend continues with Wiley, a publisher that

> offers translation services from Portuguese, Spanish, and Chinese into English, as well as reverse translation. Our experienced translators are all academics with advanced degrees who can provide an accurate, high-quality translation in your field.[3]

At Wiley, a Spanish-to-English translation of a 10,000-word text will be delivered in 12 days for the price of US$1,422, while a Chinese-to-English translation of similar length will cost US$1,584.

While some readers may be shocked by these prices, the rates charged by these publishers are competitive with the general translation market. For example, American Journal Experts[4] is an independent firm that specializes in providing

---

[2]http://www.tandfeditingservices.com/en/services/translation/
[3]http://wileyeditingservices.com/en/translation-service/
[4]http://www.aje.com/us/services/translation/

Table 1: Comparison of the cost of professional translation services in 2018.

	Spanish-to-English (10,000 words, 12 days*, 16 days)**	**Chinese-to-English (10,000 characters, 12 days*, 16 days**)**
Elsevier*	US$1,131	US$1,800
Taylor & Francis*	US$1,185	US$1,757
Wiley*	US$1,422	US$1,584
American Journal Experts**	US$1,185	US$1,580

author services, including formatting, English-language editing, and translation into English from Chinese, Japanese, Portuguese, or Spanish. In June 2016, it was announced that American Journal Experts would be collaborating with Cambridge University Press to provide editing (but not translation) services.[5] The price quoted by American Journal Experts for a 10,000-word Spanish-to-English translation is US$1,185, while a translation from Chinese is priced at US$1,580. The turnaround time for American Journal Experts is 16 days, as compared with the 12-day delivery time promised by Elsevier, Taylor & Francis, and Wiley. For ease of reference, the price comparisons are summarized in Table 1, highlighting that all these companies charge similar rates.

There is absolutely no doubt that professional translators add great value to the translation process. As we have established, scientific translation is an extremely demanding task, and translation specialists are highly educated and rigorously trained. There is absolutely no question that they deserve to be paid fairly for their work. However, it does not change the fact that translation costs may be prohibitive for some researchers. Research institutions based in English-speaking countries are often comparatively wealthy. Yet researchers and graduate students in developing countries, where English is not typically the native language and where the need for translation is therefore greatest, are the ones who can least afford to pay for professional translation services. Are there any other options available to non-Anglophone scholars?

All of the above-mentioned publishers, as well as American Journal Experts, also offer English-language editing services at approximately half the cost of their translation services. However, in order to benefit from editing services, the manuscript first needs to be prepared in English of a reasonable quality. For researchers with limited English proficiency, this expectation may still be out of reach. With this in mind, some scholars may be driven to consider even less expensive options.

[5] http://www.cambridge.org/academic/author-services/

Convenience editing and translation in a gig economy

In some cases, researchers may approach an English-speaking colleague in their field or a language teacher at their institution with a request for editorial assistance – a process sometimes dubbed as "convenience editing." As we have established above, scientific writing requires specialized subject knowledge, and Willey and Tanimoto (2015), for example, recount that English teachers who are asked to edit academic articles in specialized areas that are unfamiliar to them experience considerable stress and uncertainty. The revisions made tend to be tentative and require lots of back-and-forth conversations with the authors, which is a time-consuming process for both groups. What is more, as reported by Luo and Hyland (2016), the work done by the English teachers who act as literacy brokers or text mediators in these situations is not typically remunerated or even appropriately recognized by the institution, thus making it an unattractive undertaking for these language teachers. Meanwhile, a genuinely collaborative project with an English-speaking colleague in the same field could be a good option for obtaining support with the preparation of an English-language article. However, as reported by Tietze (2008, p. 382), some non-Anglophones find themselves in such a desperate situation that they may go as far as offering first or second authorships for articles simply to obtain help with linguistic editing. This strategy may count against them when it comes time to apply for tenure or promotion.

In recent years, many industries have seen a rise in the number of people who offer their services on a freelance basis, and often at discounted rates. Ride-share organizations, such as Uber or Lyft, are examples of this trend. Referred to as a "gig economy," it comprises a labor market characterized by the prevalence of short-term contracts or freelance work rather than permanent jobs. Members of this precarious workforce may be professionals (e.g., sessional lecturers at a university); however, there are also many examples where non-professionals offer their services and undercut professionals. Fueled in part by the fact that it is easy to do translation and editing remotely and electronically, these are fields where it is easy to find non-professionals willing to offer their services for cheaper rates.

Websites such as Gigbucks[6] and Fiverr[7] offer free and easy-to-use platforms for gig workers and those seeking their services. How does it work? At Gigbucks, for example, you can advertise your services in the form of a "gig" for a price as low as US$5 and up to US$50. Categories of work include graphic design, photography, programming, and translation, among others. Purchasers can read your offer and order your services. Once you accept the offer and deliver the service, the purchaser will pay (e.g., through PayPal). Buyers can also rate and review the quality of the service. Browsing in the "Translation Services" category of the Gigbucks site, we see hundreds of advertised gigs such as the following:

- I will translate text from English to Spanish and vice versa, 500 Wds. for $5. Expected delivery: 2 days.

[6]https://gigbucks.com/#/explore/Translation-Services
[7]https://www.fiverr.com/categories/writing-translation?source=category_tree

- I will translate 1,000 words Vietnamese to English for $5. Expected delivery: 2 days.
- I will translate up to 2,000 words from Brazilian Portuguese to English for $5. Expected delivery: 3 days.

If these offers seem too good to be true, they probably are… In 2017, two professional translators – Natalie Soper[8] and Hannah Keet[9] – conducted a "mystery shopper" exercise by purchasing translation services from the gig site Fiverr. The full details of their experiment are reported on their respective blogs. In brief, they selected a 500-word French article on tourism (which is their area of specialization) and then contracted four translators who offered their services for French-to-English translation of up to 500 words for US$5. Soper and Keet then analyzed the translated texts that they received and found them all to be suspiciously similar to the output of Google Translate. As recounted by Keet:

> … the Fiverr translators depended on Google Translate, adjusting it slightly where they thought appropriate. None of them researched any part of the text or added glosses for aspects that English language readers would not understand or be aware of, which a professional translator will do as a matter of course. Although these translations only cost $5 each, we got neither speed, quality nor good customer service – we basically paid for Google Translate with a few tweaks, some of which weren't even appropriate.

It would be very easy to dismiss the idea that a scholar would turn to a gig site for affordable translation services. We might be tempted to pose the question as to who could possibly be so naive. Part of the answer may lie in the fact that people in high pressure situations – such as "publish or perish" – sometimes make desperate decisions. Predatory publishing is a good example. Predatory journals tend to provide little to no peer review or editorial oversight and seek to exploit the open access author-pays model for their own profit. With the explosion of online publishing and increasing use of the article processing charge business model, predatory publishers are becoming more common and sophisticated, and according to Eriksson and Helgesson (2017), the practice of publishing papers for profit, without any genuine concern for content, but with the pretence of applying authentic academic procedures of critical scrutiny, brings about a worrying erosion of trust in scientific publishing. While experienced researchers may be able to spot predatory practices more easily, Gillis (2017) recounts a number of examples of how predatory publishers have succeeded in duping less established researchers into publishing in scam journals.

In a similar fashion, those who are inexperienced with translation or who have limited proficiency in another language could easily fall into a comparable trap

[8]https://bellingua.co.uk/2017/03/01/using-fiverr-for-translations-part1/
[9]https://hannahkeet.co.uk/2017/03/01/fiverr-do-you-get-what-you-pay-for-part-2/

with gig translation services. If that argument does not seem convincing, consider the fact that the experiment conducted by Soper and Keet was inspired by the fact that a friend approached Soper for advice on where to obtain professional translation services. When the recommended services were deemed to be too expensive, the friend confided that her office was instead planning to seek services from the Fiverr gig site, thus prompting Soper and Keet to undertake their investigation to debunk this strategy.

It would therefore seem clear that gig translations do not offer a good return on investment for scholars wishing to publish in English, since people who use these services risk paying for machine translation output with few, if any, enhancements. With this in mind, we are prompted to ask whether it would be a better strategy for researchers to consider using machine translation systems directly and investing time in learning how to use them more effectively?

Publishing strategies and their pros and cons: A summary

In the preceding sections, we have discussed a number of possible strategies available to scholars who have English as an additional language and who wish to publish the results of their research. Table 2 summarizes these strategies and considers some of the main strengths and potential challenges associated with each.

While scholars can certainly publish in languages other than English, they are often reluctant to do so, even if their employment conditions (e.g., tenure requirements) permit it. A study by Desrochers and Larivière (2016) found that English-language publications produced by German, French, and Quebec scholars received on average three times as many citations as did articles that were published in German and French. According to Mur Dueñas (2012) many scholars consider that even if their articles are published in low-impact national journals, they will reach a wider audience if they are written in English than if they are written in another language. Taking the example of the field of business and economics in Spain, Mur Dueñas (2012, p. 145) observes that a number of Spain's national journals in this field are encouraging publications in English rather than in Spanish.

As summarized in Table 2, a variety of options are available to scholars who are not native English speakers but who wish to publish in English; however, depending on their specific circumstances, these options may be more or less realistic. For instance, a scholar who has a reasonable proficiency in English and who is based at a well-off institution may be able to provide a solid draft and pay for an English language editor. A scholar whose institution offers courses in English for Research Publication Purposes may be able to take advantage of such courses. A scholar who has a network of English-speaking collaborators may be able to ask them for editorial support. But what about scholars whose level of English is very low, who are based at an institution without much financial or linguistic support, and who are not part of an international research group? What options are left for these scholars? The most appealing avenue may be to give machine translation a try.

Table 2: Summary of the main publication strategies available to researchers who have English as an additional language.

Publishing strategy for scholars who have English as an additional language	Strengths	Challenges
Publish in languages other than English (e.g., national journals)	• Research is published and available	• May not be as widely read by the international community • May not offer strong support for career recognition and advancement • May still require abstract/keywords in English
Publish in less prestigious English-language journals	• Research is published and available	• May not be as widely read • May not offer strong support for career recognition and advancement • Risk that journals with low(er) standards may be predatory journals
Take courses in English for research and publication purposes (ERPP)	• Results in good command of ERPP	• Requires a long-term investment of time, effort, and sometimes money; not a quick-fix • Time spent on learning English is time not spent on research or other tasks, so productivity may drop
Hire a professional translator	• Results in high-quality text	• Professional translation services are very expensive
Hire a professional English-language editor	• Results in high-quality text	• Though usually cheaper than translation, professional editing services are still expensive • Requires a reasonably good draft as a starting point
Collaborate with English-speaking colleagues	• Affordable • May result in higher-quality text	• Tenure and promotion decisions may require evidence of independent scholarship • May lead to opportunistic relationships rather than true collaborations • English teachers may not be comfortable with specialized content of research
Use amateur "gig" translation or editing services	• Affordable	• Results are typically of a low quality that may not be publishable • May simply be (lightly edited) machine translation output
Use machine translation	• Affordable and fast	• May result in low-quality output *if users are not aware of how to optimize the tools*

Indeed, based on a survey conducted among a group of 46 Spanish-speaking health sciences researchers at a university in Colombia, it appears that many (87.5%) are already experimenting with machine translation at various points in the scholarly communication process – whether for searching, reading, or drafting publications (Buitrago Ciro, 2018). However, when asked, most respondents indicated that they may not be using the technology optimally, and none claimed to be aware of pre- or post-editing techniques. In contrast, the overwhelming majority (95%) of these machine translation users indicated that they would like to learn more about how to use this tool effectively for scholarly communication purposes. The following sections will explore in more detail the potential and challenges of using machine translation in a scholarly communication context.

Machine translation and scientific research: Horses for courses

Machine translation systems were once available only to large or wealthy organizations who dealt with high volumes of translation. Examples of early implementations of machine translation systems include the use of the SPANAM/ENGSPAN system by the Pan-American Health Organization in the 1980s (Vasconcellos & León, 1985), or the initial MÉTÉO machine translation system that was used to translate weather forecasts from English to French for Canada's ministry of the environment in the late 1970s (Chandioux & Guérard, 1981). However, since the launch of free online machine translation systems such as Google Translate and Microsoft's Bing Translator in 2007, machine translation has become easily available to anyone with an Internet connection. In Chapter 2, we will delve into more details about how machine translation systems work and how their output can be optimized. For the moment, however, let's consider some potential applications of this technology in the context of scholarly publishing. In particular, we will consider the use of machine translation for both dissemination purposes and helping users to discover and assimilate information.

Parlez-vous le français-robot? The limits of machine translation for knowledge dissemination

There is no shortage of literature – scientific and popular – pointing out the deficiencies of machine translation. A quick search of ProQuest's Canadian Major Dailies[10] newspaper database using the search term "machine translation" retrieves articles with headlines such as "Computers fall short as translators," "Humans can out translate machines," and "Translation is hard for software," not to mention multiple references to the theme "Lost in translation." Though the technology has come a long way since its initial inception during World

[10]http://www.proquest.com/products-services/Canadian-Major-Dailies.html

War II, in most contexts, the quality of translations produced by machines still typically falls short of that produced by professional translators. For instance, a recent attempt by the Government of Canada's Translation Bureau in 2016 to introduce the Portage machine translation system for use by public servants generated an enormous backlash by citizens concerned that it would have serious negative effects on the quality of French language communications within Canada's public service. "*Parlez-vous le français-robot?*" asked Jean Delisle and Charles LeBlanc (2016), two university professors of translation who were instrumental in leading the charge against the implementation of Portage, a statistical machine translation system developed by researchers at Canada's National Research Council (2015). In the end, public pressure was so strong that the project was delayed until the Office of the Commissioner of Official Languages could conduct an investigation. When things finally moved forward, the system was re-branded as a language comprehension tool instead of a machine translation tool, and there were a number of constraints on how it could be used. It is still too early to determine if there will be any lasting effects on linguistic quality as a result of using Portage in the public service, but planned periodic evaluations of the system and its implementation will no doubt prove to be enlightening in this regard.

Perhaps we should not be surprised that machine translation is challenging for computers. As expressed by Arnold (2003, p. 119), "Part of the reason why translation is difficult for computers is that translation is just difficult: difficult even for humans." However, part of the question of whether machine translation is deemed to be successful or not depends on users' needs and expectations. For instance, whereas machine translation was soundly rejected for widespread use within the Canadian government, it has long been incorporated into the translation processes used within the institutions of the European Union, such as the European Commission and the European Parliament (Bonet, 2013). Various systems, approaches, and processes have been employed – some with more success than others – but the use of machine translation continues to thrive in the European Union. Of course, whereas Canada has only two official languages, the European Union has 24 (and counting), making it all the more appealing to seek technological support to cope with the enormous volume of text and myriad language combinations that are present. Moreover, within the European Union context, texts intended for dissemination must be post-edited, which means that a professional translator takes the raw machine translation output and corrects any errors. Meanwhile, raw machine translation may be used for "gisting," which means that a user can employ machine translation for personal use in order to get the gist or comprehend the general idea of the meaning of a text that has been written in another language.

In spite of the inherent risks, it is possible that some researchers who have English as an additional language may experiment with machine translation as a means of converting their texts into English, though this is not typically advised by publishers. On the website of the American Journal Experts, the previously mentioned independent company that provides professional translation services for authors, there is a short discussion about machine translation

entitled "Translating Academic Writing: Human vs Machine."[11] The page notes the following:

> With the increasing number of research articles by non-native English speakers and a lack of spare time in researchers' schedules, automated translation may seem like an appealing option. Here are some issues you should be aware of. […] To avoid obscuring or altering your meaning when translating your manuscript, consider seeking only human translation help.

It then goes on to provide an example of a French language scientific abstract that has been translated by both Google Translate and a professional translator, identifying a number of flaws in the machine-translated version, including sentence fragments, illogical ordering of phrases, and non-domain-specific terminology, among others.

Meanwhile, on the Elsevier website[12], the use of machine translation is similarly dissuaded:

> If you've ever used Google Translate, you'll know how easy it is to structure sentences badly in another language. Direct translation can result in non-native sounding sentences; while this is forgivable when spoken, it's a no-no in an academic article. At best, it could make your work look sloppy, but at worst, it could change the meaning of what you're trying to say. This could mean you run the risk of your submission being rejected because it's unclear.

While American Journal Experts and Elsevier might be considered to have a vested interest in encouraging authors to use professional translation services, similar recommendations have been made by others, including members of the library and information science community. For instance, as a new addition to the sixth edition of the *Introduction to Reference Sources in the Health Sciences*, there is a section entitled "A Review of Online Products for Language Translation" that has been prepared by Beatriz Varman (2014), an experienced librarian at the Texas Medical Center Library in Houston. In it, Varman acknowledges that translation is a challenging task, and indicates that while machine translation is faster and cheaper than professional translation, the accuracy of machine translation is lower. In the opinion of Varman (2014, p. 172), machine translation may be adequate for "simpler translation needs such as words, phrases, and non-technical text," but she goes on to caution that "to have a reliable translated document in a specific discipline, human translation is currently the right choice."

---

[11]http://www.aje.com/en/arc/translating-academic-writing-human-vs-machine/

[12]http://blog.webshop.elsevier.com/tips-for-writing/why-its-best-to-ask-a-professional-when-it-comes-to-translation/

From post-editing to self-post-editing: A promising way forward?

If current thinking suggests that unedited machine translation is not of publishable quality, does post-edited machine translation represent a viable option? The task of post-editing and its role in professional translation practice has started to garner considerable attention in recent years (e.g., Garcia, 2011; Koponen, 2016). As mentioned above, post-editing typically refers to a situation where a language professional takes the raw machine translation output and corrects any errors in order to bring it up to an acceptable quality. Indeed, post-editing can be carried out to various levels depending on user needs. Rapid or minimal post-editing focuses on correcting only errors of meaning, whereas full or maximal post-editing also addresses any stylistic problems (Allen, 2003, p. 301). The intent of maximal post-editing is to produce a text that is of comparable quality to a professional translation. While there is some evidence to suggest that post-editing a text may be faster and cheaper than translating it from scratch, most experiments in this regard have been carried out with relatively short general texts, rather than with scholarly publications (e.g., Bowker & Buitrago Ciro, 2015). In others, productivity gains are only seen when certain conditions are met: the machine translation system is domain-specific, texts are first pre-edited to remove ambiguities, and post-editing is carried out by experienced professional post-editors (Garcia, 2011, p. 228). In fact, depending on the quality of the raw machine translation, post-editing may actually prove to be *more* challenging or time-consuming than translation. This was found to be the case for certain language pairs in the European Union's translation unit (e.g., Leal Fontes, 2013, p. 11).

More interesting to our readers, perhaps, are investigations where post-editing is carried out by people who are not language professionals. Schwartz (2014), for instance, found that a monolingual post-editor who is not familiar with the source language, but who is a domain expert, can be a highly effective post-editor. This means that scholars with limited English proficiency could potentially seek post-editing assistance from a colleague in their discipline who is a proficient English speaker. However, this type of literacy broker arrangement could still fall into the category of convenience editing that was described previously and could create opportunistic relationships unless the project in question is a truly collaborative one.

The most interesting question, however, is whether a scholar who is not a native speaker of English can use a combination of machine translation and *self*-post-editing to produce a viable text for publication in English. Indeed, this question has been the subject of recent investigations by a four-person research team comprising Marie-Josée Goulet, Sharon O'Brien, Carla Parra Escartín, and Michel Simard, who have published their findings in a selection of papers (Goulet, Simard, Parra Escartín, & O'Brien, 2017; O'Brien, Simard, & Goulet, 2018; Parra Escartín, O'Brien, Simard, & Goulet, 2017). Following a survey of the literature that compares the experience of writing in a dominant language and a foreign language, O'Brien et al. (2018) conclude that not only is academic writing in a foreign language perceived as being a burden but also that it may be easier

to think in one's dominant language while writing an academic paper and that writing in the dominant language may be more productive. Accordingly, O'Brien et al. (2018) set out to explore the potential of machine translation and self-post-editing as a means to support the academic writing process for authors who have English as an additional language.

In O'Brien et al.'s (2018) experiment, 10 non-Anglophone scholars with a variety of native languages and different areas of subject expertise were asked to write a 500-word academic abstract of some of their recent research. Half of the abstract was to be prepared in their native language, while the other half was to be prepared in English. The non-English part of the abstract was then translated into English using Google Translate. Participants were then instructed to revise the entire text to create a well-formed academic English abstract.

The results of this small-scale experiment were mixed. Some of the participants considered themselves to have a good level of competence in English and already had experience writing academic papers in English, while others were less confident in their ability to write in English and had less experience doing so. Accordingly, some of the participants found drafting the paper directly in English to be easier than post-editing the English language machine translation output, while others found the reverse to be true. Overall, seven of the 10 participants felt that the level of quality achieved when post-editing was equal to or better than the level of quality achieved when writing in English as a foreign language. Half of the participants indicated that they would consider using a combination of machine translation and self-post-editing regularly to produce future publications in English.

Meanwhile, in addition to gathering the self-reported perceptions of the participants, O'Brien et al. (2018) also engaged a professional reviser to revise the abstracts. This reviser was unaware that half of each text was machine-translated and half was written in English as a foreign language. The number of edits made in each half of the texts was comparable. Overall then, O'Brien et al. (2018, p. 250) conclude:

> The professional reviser's assessment suggests that there is no evidence that one writing method systematically produces better quality than the other. Furthermore, this assessment reflected that of the participants themselves. We take from this that the use of MT [machine translation] and self-PE [post-editing] as a writing aid did not negatively affect the perceived quality of the written product for this particular set of participants and texts.

In a follow-up to this study, reported in Parra Escartín et al. (2017), the researchers focused on one specific source language (Spanish) and one specific subject field (Medicine), though this small experiment included only five participants. The analysis showed that overall, without training in post-editing, the medical experts were able to implement a number of essential corrections and also some preferential corrections. However, a professional proofreader also determined that other essential corrections had been overlooked. Parra Escartín et al. (2017) conclude

that while researchers can successfully post-edit their own scientific papers to some degree, they cannot necessarily produce final drafts that will be suitable for publication.

Of course, both O'Brien et al. (2018) and Parra Escartín et al. (2017) are quick to acknowledge that these experiments were carried out on a very small scale and therefore may not be generalizable. Both groups also identified a range of issues for further investigation, such as domain-specific machine translation, controlled authoring, and language combinations, among others. On the whole though, the idea that some researchers could use machine translation and self-post-editing at least as a partial solution for publishing in English seems worthy of consideration and additional exploration. Indeed, perhaps machine translation and self-post-editing could be part of a staged process where the result of such an effort could produce a draft that is viable for editing by a third party for a more reasonable time- or cost-investment than a raw machine translation or a text drafted directly in English by a non-Anglophone scholar. We will return to the idea of scholars who have English as an additional language learning how to post-edit their own work in Chapter 3. In the meantime, preparing a publication in English is only one way in which scholars can consider using machine translation. As noted by Buitrago Ciro (2018), researchers who are not native speakers of English also turn to machine translation to help them engage with the existing scholarly literature in their field.

A helping hand when searching for information?

In the early stages of a research project, researchers typically conduct a literature survey to determine the state of the field and to identify work related to their own research or gaps in our collective knowledge of a subject that need to be filled. Given that the overwhelming majority of research articles in scholarly journals are published in English, non-Anglophone researchers may struggle to identify English language search terms that will help them to locate information that is relevant to their project, particularly if they are students or early career researchers who are still learning the ropes in their chosen field. Of course, this may depend on how or where they study. As noted above, there has been an increase in the number of university programs that are being taught through the medium of English in countries where English is not an official language. Indeed, scholars who are first introduced to their field through English may actually be more familiar with terms in that language than in their native language (e.g., O'Brien et al., 2018, p. 248). However, for those who complete their studies in their native language first, the terms may come less readily in English, at least to begin with.

Knowledge organization systems such as library catalogs and databases play a significant role in aiding information discovery by modeling the underlying semantic structure of a domain, providing a semantic road map of individual fields and the relationships among and across fields, and relating concepts to terms. Indexing languages are related to knowledge organization systems and constitute formalized controlled languages that are designed and used to describe

the subject content of documents for information retrieval purposes. In addition, many scientific databases include a less controlled means of describing the subject content of documents: author-supplied keywords.

Keywords are terms or phrases chosen by the author to capture the most important aspects of an article. They are typically included just below the abstract in a publication. Keywords help databases to create accurate search results. In other words, most databases use keywords to decide whether to display an article to potentially interested readers. Keywords make an article searchable, and so it is important to include relevant and well-chosen keywords that will help other researchers to find the article.

If scholars have learned about their research domain through their own language, then they probably start by looking for material in that language. However, since most current research is published in English, how can they take the next step of looking for comparable or related material in English? Can machine translation help with this search process? Kit and Wong (2008, p. 320) suggest that machine translation may be "good enough to serve most of the translation demands for the purposes of information access," including database access, though they did not expressly test this application of machine translation in their study. Therefore, we conducted a very small experiment of our own to establish whether this seems to be a promising avenue. It has been published in Bowker (2018) and the main thrust is summarized below.

As mentioned previously, some non-English-language journals do provide abstracts and keywords in English. However, others may not, such as national journals or journals run by individual university departments, which are often produced by and for their own graduate students who are therefore relatively new to the domain. To test whether machine translation can help scholars at the stage of database searching, we identified two journals in the field of Information Science in which the articles, as well as the abstracts and keywords, are provided only in Spanish: *e-Ciencias de la Información*,[13] published biannually by the School of Library Studies and Information Sciences at the University of Costa Rica, and *Métodos de Información*,[14] published biannually by the association of information professionals of Valencia in Spain.

For each journal, we randomly selected one article from each issue published in the five year period from 2013 to 2017, for a total of 20 articles. From each article, we extracted the list of the author-supplied keywords, which were copied into a spreadsheet and sorted alphabetically. After eliminating duplicates, we were left with a list of 71 Spanish keywords which we then translated into English using Google Translate. Next, we used the translated keywords to conduct a subject search in the Library, Information Science & Technology Abstracts (LISTA) database, which has one of the broadest ranges of coverage for the domain of library and information science (Vinson & Welsh, 2014).

[13]https://revistas.ucr.ac.cr/index.php/eciencias
[14]http://www.metodosdeinformacion.es/mei/index.php/mei

Of the 71 translated keywords, 37 (52%) returned relevant search results (i.e., articles on a similar topic to the corresponding Spanish article from which the original keywords were taken), while 34 (48%) did not. These 37 productive keywords appear to have been well translated by the machine translation system. Among the 34 translated keywords that did not return any results, 23 (68%) appear to be appropriately translated but simply not in alignment with the descriptors used in the LISTA database. For the remaining 11 (32%) keywords that did not return results, it appears that translation-related problems stemming from orthographic variation (e.g., "e-book" instead of "electronic book"), synonymy (e.g., "information competences" instead of "information skills"), or differing syntactic preferences and semantic field coverage have interfered with the information retrieval process.

While keeping in mind that this investigation was conducted on a very small scale – using just two journals, 71 keywords, one machine translation system, one language pair, and one bibliographic database – the results nonetheless appear promising. Globally, only 11 out of 71 (15%) of the author-supplied keywords were translated in a way that led to no results being retrieved from the LISTA database. A much higher proportion of the author-supplied keywords did not generate results because they did not align with the descriptors used in the LISTA database, rather than because the keywords were poorly translated.

If machine translation seems to be a reasonable tool for assisting with searching and information discovery, could it also be helpful for the next step, which is usually for researchers to read and try to understand the content of the retrieved articles as part of a literature survey or an effort to situate their work within the broader field?

A "good-enough" solution? The potential of machine translation for information assimilation

As we have already seen, machine translation systems have proved useful for gisting in some contexts, and Kit and Wong (2008, p. 306) enquire whether machine translation might be a "good-enough" solution when conducting a literature search. Indeed, over a quarter century ago, Hutchins and Somers (1992, p. 157) had already identified the following practical way to use raw machine translation output, even if it was of a relatively low quality:

> Experts in scientific fields need access to current documentation in languages they cannot read. The output from an MT [machine translation] system is unlikely to be very good, but for technical readers who know enough about the field, who know what is going on generally in this science, and who can maybe even guess roughly what the article is about, it may well provide sufficient material to get at least some idea of the content of the text. In particular, they should have enough information to say whether they do or do not want this or that paragraph translated "properly." It is an economically sound use of low quality MT output, and indeed, for many

> people with financial and time constraints it is better to have a crude translation than no translation at all. It may not be what the designers had in mind, but it is clearly a valued and practical use.

Does this observation still hold true today? In a 2013 trend report released by the International Federation of Library Associations and Institutions (IFLA, 2013, p. 7), machine translation is identified as a technology that will transform the global information environment, but IFLA also raises questions about the use of machine translation in a research setting:

> Advances in automated machine translation are changing the way we communicate with one another, and breaking down language barriers. In regions where local educational content is limited, students will have access to translated materials from overseas. Researchers and users will be able to read in their own language any book, article, online blog ever written. […] Automated machine translation will change the way we communicate, but will it increase our understanding?

As mentioned above, though Varman (2014) is skeptical about the value of machine translation for some tasks, she does acknowledge that it could be useful for gisting purposes. Accordingly, she devotes a section of her chapter in the *Introduction to Reference Sources in the Health Sciences* (6th edition) to providing a brief introduction to some of the free online translation resources available, including Google Translate and Microsoft's Bing Translator, along with some lesser known tools. The descriptions are short, providing essential details about how to access these tools, as well as information about the languages available. There is also a reminder about their limitations.

Similarly, in a 2009 issue of the *Canadian Journal of Dental Hygiene* (Canadian Dental Hygienists Association (CDHA), 2009, p. 151), there is a short notice about machine translation that has been prepared by the editors. In it, they recognize that some of their readers may sometimes be in need of translation, and they provide the following recommendation as to how machine translation systems should be used:

> Machine translation (MT) enables users to have a general understanding of a piece of foreign text. It is a useful online tool and should only be used for a general idea as to the meaning of the original text. MT is not perfect. Its results do not compete with human translation. But not all of us have a gift for translation and sometimes a translation, even with an error, can come in handy.

Meanwhile, ProQuest, a well-known provider of databases such as the Biological Science Database, the Materials Science & Engineering Database, and MEDLINE, among others, has partnered with a machine translation developer to offer integrated machine translation for database users. The Language Engineering

Company has produced a number of commercial machine translation products, one of which can be accessed through ProQuest Dialog, which according to ProQuest's website offers "a user-friendly, flexible interface with specialized workflow tools enabling the novice searcher and expert information professional to quickly find, organize, and share the right information."[15]

The website goes on to note that when using a ProQuest product, users may sometimes retrieve records in a language other than their own or their clients' and they may need to get a quick translation. ProQuest Dialog makes it possible to do this on-the-fly. The real-time machine translation feature is built into the interface and is provided as a convenience at no additional charge. However ProQuest does add the caveat that this machine translation option is in no way meant to replace human translation.

Essentially, when users retrieve a record of potential interest, they can read the associated abstract in the language in which it is retrieved, or they can click on the "translate" button and have the abstract (or in some cases, the full text) automatically translated by the Language Engineering Company's machine translation engine into one of 14 languages, including Arabic, Chinese (simplified and traditional), English, French, German, Italian, Japanese, Korean, Polish, Portuguese, Russian, Spanish, and Turkish. This approach offers an easy and cost-effective method to filter content for relevance.

Likewise, EBSCO, another major research database vendor, offers a similar service. As explained on the EBSCO website, if English is the language in which an article has been written, the full text is presented in English by default. However, if the translate feature is enabled, users may click a button to translate a full-text article from English into one of the 34 languages. EBSCO provides the following qualifying statement about machine translation:

> Machine translation is considered a "gisting" application, producing translations that enhance the end-user's understanding of the original document. It does not produce the same level of translation that a human translator could provide.[16]

Indeed, it seems that researchers do have an interest in using machine translation in this way. For example, Kit and Wong (2008) carried out a comparative evaluation of six free online machine translation systems to determine how useful they were for translating legal texts. In undertaking this study, they specified that their intent was

> to provide legal translators, law librarians, and law library users with the most reliable information of this kind so far about a number of popular online MT [machine translation] systems suitable for legal translation. (Kit & Wong, 2008, p. 320)

---

[15]http://www.proquest.com/products-services/ProQuest-Dialog.html

[16]https://help.ebsco.com/interfaces/EBSCO_Guides/EBSCO_Interfaces_User_Guide/translate_language_of_interfaces_and_full_text

At the end of their investigation, Kit and Wong (2008, p. 320) conclude that while machine translation was not suitable for producing texts of publishable quality, it served quite well when the user simply wanted to obtain the rough idea or the subject of a text, or to locate information (i.e., translation within multilingual systems of information retrieval, information extraction, database access, etc.).

Another group that has actively investigated whether machine translation can be helpful to users who have English as an additional language is based at the University of Tokyo's Graduate School of Medicine. Anazawa, Ishikawa, and Kiuchi (2012, p. 189) identify the following situation for nurses working in Japan:

> The trend toward evidence-based nursing increased the need for Japanese clinical nurses to read international literature written in English. Nursing research is commonly conducted in clinical settings in Japan. Nurses need to increase their knowledge of the research by finding previous related literature. This involves obtaining and examining information from various sources, and nurses must frequently conduct extensive searches in English language databases to identify research papers with strong findings.

However, given that few nursing workplaces in Japan offer continuing education programs in the English language, Anazawa, Ishikawa, and Kiuchi (2012, p. 189) posit that machine translation may be a useful technology. In a survey of 250 Japanese nurses, Anazawa, Ishikawa, and Kiuchi (2013a, p. 24) discovered that over 70% had indeed tried using online machine translation to help them understand English-language publications in their field. Overall, half of the respondents found it to be "usable to some degree," while an additional 6.6% found it to be "very usable." The main problem identified was incorrect domain-specific terminology, followed by problems with sentence structure.

An emerging need: Machine translation literacy

Another finding from the series of studies carried out by Anazawa and colleagues was that the users who had extremely limited proficiency in English found the machine translation output to be less useful than those who had moderate proficiency. This leads Anazawa, Ishikawa, Park, et al. (2013, p. 64) to suggest that, for optimal use of online machine translation, a combination of educational approaches is needed: "Enhancing and motivating English language learning is required, but training nurses how to use automated translation technology offered via the Internet is also an element to consider." This is emphasized again in Anazawa, Ishikawa, and Kiuchi (2013b, p. 386), where the authors note that nurses need to be trained if they are to make better use of machine translation tools in their work:

> Our results highlight that MT [machine translation] systems are currently ancillary tools, considering their low level of performance, and changes must be made in educating nurses to achieve optimal use of this technology.

Likewise, Varman (2014, p. 181) recommends that "when using machine translation, it is worthwhile to follow the tips suggested on individual websites to achieve the best results," while the previously mentioned short editorial in the *Canadian Journal of Dental Hygiene* also offered a brief list of tips for getting the most out of machine translation (CDHA, 2009, p. 151). Meanwhile, Kit and Wong (2008, p. 306) insist that it is not helpful to simply engage in "blaming MT [machine translation] for its uselessness." Instead, they allude to the need for training in the appropriate use of machine translation, noting the currently available machine translation technologies *if properly utilized* may be good enough to serve most of a researcher's needs for the purposes of information access and assimilation. Kit and Wong (2008, p. 321) emphasize this point once again in their closing remarks:

> As a last word, we would say that whether an existing MT [machine translation] system is useful or not depends not only on how well it can translate but also largely on how it is utilized. When there is real demand for translation and the suitability, strengths, and weaknesses of available MT systems are well understood, why not incorporate online MT services into one's working environment…?

Parra Escartín et al. (2017, p. 261), also point out that their study on self-post-editing "aimed at identifying the type of edits that medical practitioners make when they engage in the self-post-editing process without any prior training in MT [machine translation] or post-editing." They suggest that training and practice could make medical researchers better post-editors.

Overall then, we can infer that some kind of digital literacy education that relates to the use of machine translation tools would be beneficial. It is not unusual for new "literacies" to emerge as our societal needs evolve (e.g., computer literacy, information literacy, and media literacy). Free online machine translation has been easily accessible to researchers, among others, for about a decade, and it shows no signs of disappearing. It is true that online machine translation systems are relatively easy to use: just select a tool, type or copy-and-paste a text, choose a language pair, and click a button. However, this does not mean that users are equipped to apply these tools successfully to their learning activities. Training in the critical and effective use of machine translation can help.

Martin (2006) describes digital literacy as the awareness, attitude, and ability of individuals to appropriately use digital tools to identify, access, manage, integrate, evaluate, analyze and synthesize digital resources, construct new knowledge, and communicate with others, as well as to reflect upon this process. This definition emphasizes that critical thinking, rather than technical competence, is the core skill of digital literacy. But where does machine translation literacy fit in?

Though researchers such as Anazawa et al. (2013b) and Kit and Wong (2008), among others, have flagged the need for developing machine translation literacy, a search of the literature suggests that very little work has been done in this area to date. Questions remain, therefore, as to how machine translation literacy skills can be developed, and who can design and deliver relevant training.

As mentioned in the introductory chapter, librarians working at universities or other research institutions are a group that is well placed to step into the key role of developing and delivering a coherent program of machine translation literacy training to researchers and students. One of the ways in which librarians support researchers is by promoting and providing training in the development of different types of literacies, such as information literacy, media literacy, and digital literacy. Therefore, it seems reasonable to suggest that instruction in machine translation literacy can be usefully delivered by librarians also, but first they themselves need to be equipped to do so.

Ishimura and Bartlett (2014) examine the readiness of academic librarians for meeting the needs of international students, a group for whom limited English proficiency remains one of the principal barriers to success. They summarize a host of ways in which librarians have tried to prepare themselves to better serve international students, including taking courses in cultural sensitivity training, learning to avoid technical library jargon, and developing appropriate listening techniques (e.g., for deciphering foreign accents). There is little doubt that, collectively, academic librarians are committed to supporting international students. Yet Ishimura and Bartlett (2014, p. 313) question whether they are truly prepared to address all the needs of this group, noting that:

> Librarians are expected to teach information skills and adapt to the changing student body. However, it does not necessarily follow that librarians are equipped to teach effectively in these circumstances.

In their survey of 254 library professionals in North America, Ishimura and Bartlett (2014, p. 315) found that 92% of participants acknowledged that international students have needs that differ from those of domestic students, identifying communication skills, library systems, and research skills as areas that pose particular challenges for international students. Moreover, many respondents recognized a need to strengthen their own skills in ways that would meet international students' unique needs, and they showed strong interest in having opportunities for further training in this regard.

Machine translation literacy training was not specifically identified in Ishimura and Bartlett's study as a skill that librarians could develop to provide better and more targeted instructional services to international students, or indeed to international faculty; however, we agree with Baron and Strout-Dapaz (2001) who note that customized training and outreach can help to empower international students (and by extension, any researcher working in a non-native language) and reduce their research-related anxieties. Given that researchers who use English as an additional language do struggle with translation-related issues, it is likely that some form of machine translation literacy training could

be useful. Moreover, as we learned earlier, machine translation literacy instruction could also be useful to researchers who are native speakers of English as a means of equipping them to help alleviate the burden placed on their non-Anglophone peers.

The introduction of machine translation literacy can be viewed as a societal adaptation rather than solely as a response to a new technology. Over the past decade, machine translation literacy has begun to change from a technical or specialist literacy into an everyday literacy that is starting to have implications for the way that our society communicates and goes about activities relating to research, discovery, and innovation. Therefore, teaching machine translation literacy has to do with preparing researchers and students for future participation in an evolving society where English is emerging as the international language of scholarly communication, and digital technologies are becoming more deeply embedded in our structures and processes.

Key points from this chapter

- The exchange of knowledge between researchers is at the core of scholarly communication.
- The rate of scientific publication has increased tremendously since the end of World War II.
- English has become established as the international language for scientific communication and scholarly publishing despite the fact that fewer than 6% of the people in the world are native English speakers.
- Researchers who use English as an additional language may struggle to engage fully and effectively with the process of scholarly communication.
- While learning English may be the ultimate goal for non-Anglophone researchers, translation may be an interim solution.
- Professional translation offers a high quality, but it is expensive and time-consuming, while machine translation is faster and cheaper, but the quality is lower.
- Non-professional translators working in the "gig economy" are willing to offer translation services for a low fee, but the quality is often poor and may in fact consist of (lightly edited) machine translation output.
- Current machine translation systems do not offer a viable solution for generating publishable-quality research articles without post-editing, but recent studies suggest that for some non-Anglophone scholars, a combination of machine translation and self-post-editing could prove to be a useful aid for producing a text in English.
- Machine translation systems can also be used to help researchers who have English as an additional language to search for and understand the content of articles and so to guide and filter literature searches.
- There is an emerging need for training in machine translation literacy to enable users to get the most out of their interactions with machine translation tools.
- Academic librarians and other information professionals are well placed to develop and deliver training in machine translation literacy.

To find out more about …

English as the international language of scholarly communication

- Jenkins, J. (2014). *English as a Lingua Franca in the international university: The politics of academic English language policy*. Abingdon: Routledge.
- Jenkins, J., Baker, W., & Dewey, M. (Eds.). (2018). *The Routledge Handbook of English as a Lingua Franca*. London: Routledge.
- Montgomery, S. (2013). *Does science need a global language? English and the future of research*. Chicago: University of Chicago Press.
- Plo Alastrué, R., & Pérez-Llantada, C. (Eds.). (2015). *English as a Scientific and Research Language: Debates and Discourses: English in Europe*. Berlin: De Gruyter Mouton.

English for Research and Publication Purposes (ERPP)

- Flowerdew, J. (2015). Some thoughts on English for Research Publication Purposes (ERPP) and related issues. *Language Teaching*, *48*(2), 250–262.

Machine translation

- Ping, K. (2009). Machine Translation. In M. Baker and G. Saldanha (Eds.), *Routledge Encyclopedia of Translation Studies* (2nd edition) (pp. 162–169). London and New York: Routledge.

Scientific and technical translation

- Olohan, M. (2016). *Scientific and Technical Translation*. London and New York: Routledge.

Scholarly communication

- De Silva, P. U. K., & Vance, C. K. (2017). *Scientific Scholarly Communication: The Changing Landscape*. Cham: Springer.

Chapter 2

Machine Translation

> Machine Translation is one of the most challenging of research activities, involving the application of complex theoretical knowledge to the building of systems whose successes and failures can be judged by laymen in the simplest of terms.
>
> – Hutchins and Somers (1992)

This chapter introduces some of the main approaches to machine translation and also presents a range of different challenges that machine translation systems face. This content is not aimed at machine translation researchers, and no prior knowledge of linguistics, statistics, or machine translation is required. The goal is not to turn you into a computational linguist, but rather to help you to appreciate some of the difficulties involved in getting computers to translate. We also analyze some advantages and limitations of using sub-languages and controlled languages as a means of improving machine translation output. By learning more about how machine translation systems work, and understanding some of the complexities that are inherent in language, you can learn how to create texts that are less ambiguous and easier for computers to translate.

A brief history of machine translation

Machine translation is an area of research and development where computational linguists try to find ways of using computer software to translate text from one natural language (e.g., Spanish) to another natural language (e.g., English). Since natural languages are highly complex, machine translation is an extremely difficult task. Many words have multiple meanings, sentences may have various possible interpretations, and certain grammatical structures in one language might not exist in or map clearly to another language. Moreover, there are extra-linguistic factors involved in successful translation such as real-world knowledge.

Surprisingly, given the complexity of the task, machine translation was one of the earliest applications of computers. Inspired in part by the successful use of cryptographic techniques for codebreaking during World War II, Warren Weaver,

Machine Translation and Global Research:
Towards Improved Machine Translation Literacy in the Scholarly Community, 37–54

doi:10.1108/978-1-78756-721-420191004

a mathematician who worked for the Rockefeller Foundation, issued a memorandum in 1949 that brought the idea of machine translation to the attention of the public. This document, which came to be known as Weaver's Memorandum, was the main launching point for research in machine translation. Over the next decade, many research groups were established, numerous conferences and demonstrations were held, and funding was plentiful. All around the world, hard-working and enthusiastic researchers were optimistic that computers would soon be churning out fully automatic high-quality machine translations.

However, by the early 1960s, it was clear that cryptographic techniques were too limited and that translation was far more complex than simple word-for-word substitution. The enthusiasm and optimism that had accompanied machine translation research in the early days began to wane, and a major blow was delivered in 1966 in the form of a report issued by the Automatic Language Processing Advisory Committee or ALPAC. Commissioned by government sponsors of machine translation in the United States, this committee examined the progress and prospects of machine translation research and concluded that it did not make sense to continue investing in machine translation. The ALPAC report noted that early systems had not delivered on their promises and that there was an ample supply of human translators to meet translation demands. Although it was later criticized as being biased and short-sighted, the report was nonetheless influential. As a result, financial support for machine translation research virtually dried up, and the public perception of machine translation was damaged for many years to come.

While activity in this sector was significantly reduced, some research teams continued to work on the challenge of machine translation. The 1970s saw some successes, particularly with systems developed to work in very specialized areas with restricted vocabularies and limited syntax. A well-known example is the MÉTÉO machine translation system that was developed in Canada to translate weather forecasts from English into French for the ministry of environment (Chandioux & Guéraud, 1981). Meanwhile, the Commission of the European Communities, a forerunner of today's European Commission, successfully began using the Systran machine translation system (Pigott, 1986).

These few successes, coupled with a genuine need for more translation, led to a reawakening of interest in machine translation research in the 1980s. As we learned in Chapter 1, the years following World War II bore witness to a rapid growth in the number of activities relating to scientific and technological research and discovery, along with an accompanying increase in the volume of scientific publications. As the volume of text grew, so too did the demand for translation, which created a needs-driven market and a renewed level of interest and investment in machine translation research.

Along with continuing to investigate machine translation, researchers in the 1990s also turned their attention to developing computer-aided translation tools, such as translation memory systems. Operating on personal computers or over the Internet, computer-aided translation tools did not aim to carry out the task of translation per se but rather to support professional translators in this task by automating activities such as dictionary look up, or allowing translators to

compare their current translation job against texts they had translated previously to see if any parts of this prior work could be reused (Bowker, 2002). Gradually, machine translation and translation memory systems have become integrated, allowing professional translators to make use of both types of technology.

Prior to the year 2000, the main approach that had been used to develop machine translation systems was known as rule-based machine translation (RBMT). In a nutshell, researchers tried to program computers to process language in a way that resembles how human beings process language by incorporating grammar rules and large dictionaries. However, around the turn of the new millennium, there was a significant paradigm shift whereby researchers decided to approach machine translation in a different way, one that allowed computers to play to their strengths. As a result, new corpus-based approaches such as example-based machine translation (EBMT) and statistical machine translation (SMT) began to gain prominence (Carl & Way, 2003; Koehn, 2010). Most recently, neural machine translation (NMT) has emerged as another promising and active area of research in this field (Forcada, 2017). Meanwhile, over the years, some of these different approaches have been combined in hybrid models, which attempt to utilize the strengths of each (Costa-jussà & Fonollosa, 2015).

A more detailed description of the rich history of machine translation is beyond the scope of this chapter. However, former academic librarian W. John Hutchins – the self-appointed and highly respected historian of machine translation – has written extensively on this subject (e.g., Hutchins, 1986, 2000, 2010). Meanwhile, in the sections that follow, we will present a brief description of some of the main approaches to machine translation, looking in particular at a number of challenges that language and translation pose for computers. Our goal is not to provide highly technical details, and our intended audience is not composed of computational linguists or machine translation researchers. Instead, we simply aim to provide our target readers – including information professionals, people involved in the scholarly communication industry, and researchers who speak English as a native or as an additional language – with some basic insights into how machine translation systems work and where some of their main stumbling blocks lie. By gaining a deeper understanding of these issues, researchers, students, publishers, journal editors, peer reviewers, librarians, and the like, can improve their machine translation literacy. In turn, they will be better prepared to both share this knowledge with others, such as through machine translation literacy training, and interact more effectively with machine translation systems, such as by adapting their writing style to a translation-friendly one, or learning how to spot and fix typical machine translation errors.

Just follow the rules! Rule-based machine translation

The earliest machine translation systems, known as direct transfer systems, approached translation somewhat naively, essentially employing a look-up and replacement strategy that resulted in very literal word-for-word translations.

In this way, the Spanish phrase *el coche rojo* would be rendered into English as *the car red*, instead of as the desired construction *the red car*. From a linguistic point of view, what is missing in a direct transfer system is any kind of analysis of the internal structure of original or source text.

The failure of these first-generation direct transfer systems led to the development of more sophisticated linguistic models for translation. In particular, there was increasing support for the analysis of the source language texts into some kind of intermediate representation – a sort of representation of its meaning – which could then be used to generate a text in the target language. These second-generation machine translation systems, which were known as indirect transfer systems, approached translation in three main steps: analysis, transfer, and synthesis.

The task of the analysis module is to take a source language text and produce an abstract representation of that text. Part of the analysis includes parsing the sentences in the text to identify the correct part of speech (e.g., noun, verb, adjective, etc.) for each word as well as the relations between them. The general idea is that it will be easier for a computer to translate from the abstract representation rather than from an unstructured string of words in the source language. The challenge for a machine translation system is to infer the content from the source text because these texts are not always well written and because language contains many lexical and structural ambiguities. We will return to a more detailed discussion of language ambiguities later in this chapter, but here are a couple of examples of linguistic ambiguity to give you a general idea of the problem.

The post was delivered.

- *Lexical ambiguity*: Post = Letters? Post = Fence post?

Lucia told Mario that she had a baby over the telephone.

- *Structural ambiguity*: What is happening via the telephone = Talking? Delivering a baby?

Consider the first sentence, which contains an example of lexical ambiguity. Is *post* referring to the mail or a fence post? What about the second sentence where structural ambiguity is present? Were Lucia and Mario speaking over the phone, or did Lucia deliver the baby over the phone? It is important to keep in mind that a word or a sentence can be ambiguous for a computer even if it does not seem to be ambiguous for a human. People can draw on real-world knowledge to interpret meaning but RBMT systems cannot. They can only follow the linguistic rules.

Once an abstract representation has been produced for the source language text, the next task is for the transfer module to take that representation and produce something that the synthesis module can use to generate the target language text. The transfer stage brings to light another major challenge associated with translation, namely that different languages do not "package" information in the same way. That is why literal or word-for-word translations are not always possible. Again, some simple examples can help to make this point clear.

English: Nicolas walked across the street.
French: *Nicolas a traversé la rue à pied.* (Nicolas crossed the street on foot.)

English: I like pineapples.
Spanish: *Me gustan las piñas.* (Pineapples are pleasing to me.)

English: Esmée likes swimming.
German: *Esmée schwimmt gerne.* (Esmée swims gladly.)

As the above examples show, languages might use different parts of speech to express different elements of a sentence. In the first example, the meaning that is contained in the preposition *across* has been expressed using the verb *traverser* in French. In the second sentence, the subject pronoun *I* has been transformed into the indirect object *me* in Spanish. Finally, in the last example, the meaning associated with the verb *like* has been conveyed using the adverb *gerne* in German. Rule-based machine translation systems require complex rules to help them map the structures of one language onto the structures of another language during the transfer phase.

In the final phase, the synthesis module takes the target language representation produced by the transfer module and uses this to try to generate a well-formed sentence in the target language. A very simple example is that the system would apply grammatical rules to put the elements of the target language in the correct order. For instance, returning to our earlier Spanish-language example of *el coche rojo*, we would expect the machine translation system to reorder the words when producing the equivalent sentence in English so that the end result is *the red car* (and not *the car red*). This is because, in English, adjectives typically come before the noun instead of coming after the noun as they usually do in Spanish. However, not all instances are this straightforward. There are many cases where the content of the source language sentence could be expressed in multiple ways, all of which respect the rules of grammar. The challenge for the machine translation system is to choose the best option from among several possible or reasonable options.

The white dog chased the red car.
The red car was chased by the white dog.
The dog, which was white, chased the car, which was red.
There was a white dog which chased a red car.
There was a red car. It was chased by a white dog.

This rule-based approach, with analysis, transfer, and synthesis components, was the principal approach used by machine translation researchers and developers for about 50 years, from the period following World War II until around the turn of the new millennium. However, one of the biggest drawbacks of this approach is that it requires individual analysis and synthesis modules for each language, and a transfer module for each pair of languages. So for an organization such as the European Union, which has 24 working languages, it would be necessary to develop 24 different analysis and synthesis modules, along with 552 transfer modules to allow translation between all the possible language

combinations. Keeping in mind that a single transfer module can contain upwards of 10,000 rules, a staggering amount of effort would be required to build all these modules! This problem was one of the impetuses that spurred researchers to look for new ways of tackling machine translation that would not be so dependent on linguistic knowledge. As a result, the late 1990s and early 2000s saw a fundamental paradigm shift in the underlying approaches to machine translation. Instead of trying to program computers to process language the way that humans do (i.e., by following linguistic rules), researchers decided to see if they could bypass linguistics and instead get the computers to focus on tasks that they are very good at, such as pattern matching and number crunching. "With friends like statistics, who needs linguistics?" declared Fink, Kummert, and Sagerer (1995), ushering in a new era of machine translation research.

With friends like statistics, who needs linguistics? Corpus-based approaches to machine translation

The fundamental idea behind corpus-based approaches – sometimes referred to as data-driven approaches – is that, instead of being based on linguistic rules, translation is based on a very large database of examples of texts that have been translated by professional human translators. These pairs of texts are aligned, meaning that each sentence in the source language text is linked to its equivalent sentence in the target language text. The machine translation system can consult this resource, known as a parallel corpus, to determine how a particular word, phrase, or sentence has been translated in the past, and then use this information to propose a translation for the new text that is to be translated.

Two main types of corpus-based machine translation have emerged. The first is known as example-based machine translation, where the computer, when faced with a new sentence to translate, simply consults the parallel corpus to find examples of how that sentence (or at least, some of its parts) has been translated before. The idea is appealing because, in many cases, we are creatures of habit. Although language offers many possible ways of expressing an idea, we often use set expressions or particular turns of phrase over and over again. For example, English speakers will tend to ask "What time is it?" rather than "What is the hour?" or "How late is it?" even though these latter two options are perfectly acceptable from a linguistic point of view. On the other hand, sometimes we really are creative in our use of language. Just think of advertising campaigns or other types of text designed to catch our attention or distinguish themselves by their originality.

For example-based machine translation to work well, the parallel corpus that is provided for the machine translation system to consult must contain the right types of texts because a turn of phrase that is appropriate in one situation might not be the best choice in another. For instance, one well-known parallel corpus that has been used to feed example-based machine translation systems that translate between English and French is the Canadian *Hansard*. The Canadian *Hansard* contains the transcripts of parliamentary debates in Canada. Given the bilingual nature of the Canadian federal government, two equivalent versions of the Canadian *Hansard* are maintained, one in English and the other in French. In addition to being translated

and aligned, the large size of the *Hansard* and the fact that new material is always being added makes it an attractive parallel corpus for machine translation applications. However, a limitation of the *Hansard* for general translation purposes is that its contents reflect the language that is used in a very specific setting: Canada's House of Commons. Therefore, a machine translation system attempting to translate a sentence containing the word *hear* might propose the rather odd equivalent *bravo*. This is because politicians often express their support for a particular issue with a chorus of "Hear, hear!" As a result, *bravo* appears as the most common French-language translation for the English word *hear* in the Canadian *Hansard* corpus.

The Canadian *Hansard* contains a very specific text type, but could the problem be solved by giving an example-based machine translation system a corpus that contains a much wider and more representative range of texts? In fact, having many possible choices leads us back to the same overall problem type faced by rule-based machine translation systems: how does the computer know which option to pick? In principle, statistics could be helpful in this situation. In a large and more representative corpus, a computer might observe that as a French-language equivalent for *hear*, the verb *entendre* occurs much more frequently than the exclamation *bravo*, and on this basis choose the correct option. Thus, *statistical machine translation*, and in particular, phrase-based statistical machine translation, emerged as another type of corpus-based approach. A key distinction between example-based machine translation and statistical machine translation is whether the translation is selected from within the parallel corpus by comparing examples or by statistical means.

Statistical machine translation systems are trained using parallel corpora and make substantial use of probability calculations. In simple terms, a statistical machine translation system contains three components: a translation model, a language model, and a decoder. The basic steps undertaken by a phrase-based statistical machine translation are as follows. First, the source text is segmented into phrases, which for a statistical machine translation system can be any sequence of words, even if the combination is not linguistically motivated. For example, consider the phrases that a statistical machine translation system would identify in the sentence below. While some of the phrases coincide nicely with units that people would consider to be a coherent linguistic unit, such as *white dog*, others do not, such as *dog is*.

***Sentence*: The white dog is the smallest one.**

- *Phrases identified by a statistical machine translation system*:
  the white
  white dog
  dog is
  is the
  the smallest
  smallest one

In the next step, each of these phrases is translated into the target language. Finally, the phrases are reordered. At each step, statistical machine translation systems rely heavily on probability calculations. They use algorithms that give preference to sequences of words that are probable translations of source words,

with a probable reordering scheme, and they generate sequences of words in the target language with high probability.

A more detailed yet very accessible explanation of how statistical machine translation works can be found in Kenny and Doherty (2014, p. 278), who summarize the overall idea as follows:

> It [statistical machine translation] is based on an intuitively simple strategy: rather than trying to encode a priori in the form of dictionaries, grammars and knowledge bases, all the linguistic and world knowledge required to translate a text from one language into another (the approach taken in rule-based and knowledge-based MT), simply *learn* how to translate from already existing human translations. In practice, such learning involves the induction of statistical models of translation from parallel corpora, that is, source texts and their human translations. In the terminology of SMT [statistical machine translation], we say that translation models are trained on these parallel corpora. SMT systems also rely on so-called language models, or monolingual models of the target language; so rather than just ask whether "the house" is a likely translation of *la maison* (the answer to which question should come from the translation model), the SMT system also needs to ask whether "the house" is a likely sequence in English in the first place. Language models can be trained on the target language side of a parallel corpus, or on larger monolingual corpora of target language text.

The idea is very appealing in principle, but statistical approaches come with their own set of challenges. First, as was the case with the example-based machine translation systems, statistical machine translation systems tend to be more useful for translating texts that strongly resemble the corpus texts that were used for training. In other words, the translation and language models that a statistical machine translation system creates reflect the data on which they were trained so that a system trained on medical texts, for example, will be more useful for translating medical texts than it will be for texts from other domains. Moreover, for them to truly work, statistical machine translation systems need enormous training corpora that contain millions of examples because, even in a multimillion word corpus, there may be words that appear only once or twice, or not at all. This means that there can be a genuine challenge in getting reliable statistics. Finding corpora that are bilingual, aligned, balanced, and representative with regard to text type *and* contain millions of words is a very tall order. And that is the case even for relatively widely spoken languages such as English and Spanish. Imagine trying to find such a corpus to translate between less widely used languages, such as Hungarian and Danish, for instance.

What's next? The rise of neural machine translation

Very recently, there has been a surge of interest in another approach to machine translation, known as neural machine translation. A neural network is a sort of

information processing system that is inspired by the way that biological nervous systems, such as the brain, process information. It comprises a large number of highly interconnected processing elements that work in unison to solve specific problems. Like people, neural networks learn by example. A neural network is configured for a specific application, such as pattern recognition, through a learning process. Neural networks are typically organized in layers. Layers are made up of a number of interconnected nodes which contain an activation function. Patterns are presented to the network via the input layer which communicates to one or more hidden layers where the actual processing is done via a system of weighted connections. The hidden layers then link to an output layer where the answer is shown. Neural networks contain some form of learning rule which modifies the weights of the connections according to the input patterns that it is presented with. In this way, neural networks learn by example.

An article in *The Atlantic* reported on a humorous story of attempts to use neural networks to suggest new recipes or to name new colors of paint (Meyer, 2017). After feeding a neural network thousands of cookbooks, a researcher asked it to suggest new recipes, one of which turned out to be a recipe for chocolate cake that contained a cup of horseradish! In another experiment, the researcher fed the neural network the names of almost 8,000 colors of paint and asked it to propose new possibilities. The neural network complied, generating gems such as Rose Hork, Burf Pink, Stanky Bean, and Stummy Beige.

Entertainment potential aside, given their ability to derive meaning from complicated or imprecise data, neural networks can be used to identify patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A trained neural network can be thought of as an "expert" in the category of information that it has been given to analyze. This expert can then be used to provide projections given new situations of interest. A main difference between neural machine translation and statistical machine translation is that when researchers present training material to the deep learning algorithms in a neural network, they do not necessarily tell them what to look for. Instead, the neural machine translation system finds patterns itself, such as contextual clues around the source sentence. Forcada (2017, p. 296) likens the process to predictive text completion: for each word from the source text, the most likely output word is predicted while the target text is being constructed (assuming that the neural machine translation system has been adequately trained beforehand). The specifics of the process, however, remain mysterious in many ways.

In terms of success, early results suggest that neural machine translation works best in restricted fields for which it has a lot of training data. Meanwhile, it struggles to deal with rare words and very long sentences (Koehn & Knowles, 2017). For language pairs that do not have an extremely large volume of training data available, neural machine translation presents the possibility of using indirect training data from a pivot language. For instance, if there is not much data available for the Hungarian–Danish language pair, a neural machine translation system could learn how to translate in steps by going from Hungarian to English and from English to Danish.

While neural machine translation certainly represents an exciting development in the field, Castilho et al. (2017) determine that it is not yet at the stage

where it can outperform statistical machine translation, and they warn that it is important not to oversell this technology. Indeed, each of the main approaches to machine translation discussed in this chapter presents different strengths and weaknesses. While it may be convenient for pedagogical purposes to describe them individually, in reality many contemporary machine translation systems are actually hybrids that incorporate various techniques. Moreover, regardless of the approach employed, machine translation systems continue to grapple with the fact that language is inherently ambiguous. Now that we have a basic understanding of some of the ways in which computers have tried to tackle the difficult job of translation, let's take a closer look at some of the complexities of language with a view to considering how we, as text creators and users of machine translation, can be better prepared to interact more effectively with these systems.

It's all so confusing! The ambiguity of language and the challenges it poses for machine translation

The overarching problem faced by machine translation system developers is the fact that natural language is inherently ambiguous. Words can have multiple meanings, and grammatical structures can be interpreted in different ways. Below we present some of the most common types of ambiguity that pose challenges for machine translation, but this is by no means a comprehensive list.

Homonymy and polysemy

Sometimes a word form in a given language may have multiple meanings. In some cases, the meanings are quite unrelated. For instance, in English the noun *bank* may refer to a financial institution or to the side of a river. In French, these different meanings are represented by two completely different words: *banque* (financial institution) and *rive* (river bank). It can happen in the other direction too. The French noun *avocat* can be translated as either *avocado* or *lawyer*, depending on the desired meaning. Homonyms are essentially linguistic accidents. There is no logical reason why the same lexical item should be used to refer to both a financial institution and the side of a river. Therefore, choosing the wrong equivalent is bound to result in a nonsensical text.

> The salad came with chopped ***lawyers*** and tomatoes on the side.
> The criminal was represented in court by the city's most prominent ***avocado***.

Polysemy also refers to a situation where a single word has multiple meanings, but this time, the meanings tend to be related in some way. For instance, the noun *wood* can refer to a piece of a tree or to a large group of trees in one area.

Word category ambiguity

Word category ambiguity, also known as homography, describes a type of lexical ambiguity whereby a given word can belong to more than one grammatical

category. In other words, it can have more than one part of speech. Examples abound! In English, *test*, *flower*, *dance*, *drive*, *play*, *dog*, *drink*, *comb*, *laugh*, *cup*, and *eye* are just a few of the many examples of words that can be either a noun or a verb. What about words such as *minute* or *polish*? Consider the following examples:

> Pause a ***minute*** before reporting on the events in ***minute*** detail.
> ***Polish*** your shoes with the black shoe ***polish*** before going to the ***Polish*** food festival.

In the first example, a person has no trouble distinguishing between the meaning of *minute* being used as a noun and *minute* being used as an adjective. Similarly, the verb *polish*, the noun *polish*, and the adjective *Polish* are easily differentiated and understood. In some cases, we even use different pronunciations and capitalization as clues to help us distinguish between these forms. However, such differences may not be immediately clear to a machine translation system. It has no knowledge of pronunciation, and it may not be able to determine easily whether a word at the beginning of a sentence is a proper noun, or whether it has a capital letter only by virtue of its position at the start of the sentence.

When it comes to translation, a homograph may be translated into the target language using different words for different grammatical categories. For instance, when translating from English to French, the verb *polish* is rendered by a form of the verb *polir*, the noun *polish* becomes the noun *cirage*, and the adjective *Polish* is translated by a form of the adjective *polonais*. A well-known humorous example where a machine translation system incorrectly identified a word's part of speech is the example *Bill Gates*. When translated into French, the proper noun *Bill* became *Facturez*, which is an imperative verb form. As a result, the translated meaning became "Send the bill to Gates."

Structural ambiguity

Lexical ambiguities involve analyzing and transferring the meanings of individual words. However, in structural or syntactic ambiguity, the same sequence of words can be interpreted as having different syntactic structures. In many cases, these sentence-level ambiguities arise because individual words can serve different functions in a sentence, leading to different possible interpretations.

One common type of structural ambiguity is known as prepositional phrase attachment ambiguity. This occurs when it is not clear to what other part of the sentence a particular prepositional phrase should be attached. Take the following example:

> ***Original sentence*: Peter read about the train crash in the newspaper.**
> - *Interpretation 1*: A train crashed. Peter read about it in the newspaper.
> - *Interpretation 2*: A train crashed in the newspaper. Peter read about it.

For human readers, it is quite obvious that the prepositional phrase *in the newspaper* is meant to attach to the verb *read*. However, a machine translation system might attempt to attach the prepositional phrase to the noun phrase *train crash*, thus producing a different (if not logical) interpretation.

Newspaper headlines are an excellent source of examples of both lexical and structural ambiguity. Because they are written in a telescopic style, headlines often omit logical connectors and other disambiguators.

- Juvenile court to try shooting defendant
- Two sisters reunited after 18 years at checkout counter
- Stolen painting found by tree
- Miners refuse to work after death
- Kids make nutritious snacks
- Local high school dropouts cut in half
- Drunk gets nine months in violin case

In the case of newspapers, this strategy is often used deliberately to grab attention or generate a humorous effect. However, in other types of texts, if a machine translation system incorrectly interprets the meaning of the sentence, the translated version will likely be nonsensical and result in frustration, rather than generate a laugh.

Another type of structural ambiguity is called noun stacking, or modifier stacking, and English is a language where this is particularly prevalent. Consider constructions such as *liquid oxygen tank* or *silver oxygen tank*. Do you notice any potential ambiguity here? For a human, there really isn't any. Our real-world knowledge tells us that oxygen can come in liquid form, but a tank cannot. Meanwhile, a tank can be silver in color, but oxygen cannot. No problem! But what about for a machine translation system working into French? Without real-world knowledge, the machine might consider both of these options to be equally viable translations for *liquid oxygen tank*:

> *réservoir d'oxygène liquide* (literally, a tank of liquid oxygen); or
> *réservoir liquide d'oxygène* (literally, a liquid tank of oxygen).

And if problems can occur when trying to work out the relations in stacks of just two or three modifiers, think of the potential mangling that could occur when nine or ten begin to pile up! An example from the Canadian newspaper *Prince George Citizen* has earned a spot in the noun-stacking hall of fame after being cited in *The New Yorker* (June 27, 2011) under the heading "The bureaucratic mind at work":

> The board also gave third reading to a Foothills Boulevard Landfill gas emission reduction credits transfer contract authorization bylaw.

Anaphora, idioms, and so on!

The linguistic hurdles outlined above are just the tip of the iceberg for machine translation systems. Anaphora resolution is another problem, this time having to

do with correctly identifying the relationship between a pronoun and its antecedent noun. This is particularly critical when translating between languages that do not have linguistic genders (e.g., English) into languages that do (e.g., French, Spanish, and German). Take the example, *the glass fell off the table and it broke*. In this sentence, a person can quickly determine that the pronoun *it* is referring back to *glass* and not to *table*. However, without real-world knowledge, a machine translation system is likely to identify the nearest preceding noun as the antecedent, which in this case is *table*. This becomes a problem when translating into French because there is no non-gendered form for the pronoun *it*. There is only a masculine form (*il*) and a feminine form (*elle*). If the pronoun refers back to a masculine noun, such as *glass* (*le verre*), then the masculine pronoun *il* should be used. However, if the pronoun refers back to a feminine noun such as *table* (*la table*), then the feminine pronoun *elle* is the correct choice.

Meanwhile, machine translation systems also have to contend with idiomatic expressions, which frequently cannot be translated literally but must be changed for a different expression entirely. So while, English speakers might have *frogs* in their throats, French speakers have *chats* (cats) in theirs. Similarly, while an English speaker might be told that "*There's no use crying over spilt milk*," a French speaker will simply have to accept that "*Les carottes sont cuites!*" (The carrots are cooked!). What's a poor machine translation system to do?

Change the input, change the output: Machine translation and controlled languages

The preceding examples and discussion make it clear that one of the biggest disadvantages of machine translation systems is the fact that computers do not have real-world knowledge. They quite literally cannot tell the difference between a *lawyer* and an *avocado*. If we give a computer a text that says "*The teen saw the bird with the binoculars*," the computer has no way of knowing that it is far more likely to be the teen, and not the bird, who is in possession of binoculars. What's more, although machine translation researchers are working hard and will undoubtedly continue to make breakthroughs that improve this technology, it is highly doubtful, in the short term at least, that computers will acquire this type of knowledge.

However, if humans and computers work more closely as a "team," the output of machine translation systems can be improved. Computers may not possess real-world knowledge but human users do. It takes almost no effort on our part to resolve the vast majority of ambiguities that we encounter every day when using natural language. Most often, we undertake this sort of resolution when we are reading a text, rather than when we are writing it. But what if we changed our approach? What if we consciously chose to construct our texts in such a way that many of the ambiguities were clarified *before* the text is submitted to a machine translation system?

For instance, what if we chose to write:

> *A tank for liquid oxygen*, instead of a *liquid oxygen tank*;
> *The glass fell off the table and the glass broke*, instead of *the glass fell off the table and it broke;*

The teen used the binoculars to see the bird, instead of *the teen saw the bird with the binoculars;*
You can't change what's happened, instead of *there's no use crying over spilt milk?*

The idea of controlling the input to a machine translation system is not new. In fact, machine translation researchers realized quite early on that that their systems had better success when they were designed to work in highly specialized domains where the vocabulary and grammar were restricted (Lehrberger, 1982). Recall the above-mentioned example of the MÉTÉO system. Widely touted as one of the most successful machine translation systems ever to be developed, its sole job was to translate weather forecasts from English into French. As it turns out, there is a limited amount of vocabulary used to describe weather. Moreover, weather bulletins also use a very restricted syntax. Instead of using long and complex structures, weather forecasts use a more abbreviated style, as illustrated in Table 3.

Owing to the limited vocabulary and restricted syntax, weather forecasts present a sort of natural sub-language, which is easier for a computer to tackle than unrestricted language. The success associated with such purpose-built machine translation systems is evident. As noted above, rule-based, example-based, statistical, and neural approaches to machine translation all achieved better results when working in particular domains or on specific types of texts. When there are fewer choices to make, computers have a better chance of getting things right. However, in spite of the proven success associated with the adoption of a sub-language approach to machine translation, it is unlikely that dedicated systems can be designed and constructed for every possible subject field and subfield, every text type, and every language pair. Many people work with different types of texts, deal with different subject matter, and have needs that extend beyond highly restricted sub-languages.

Therefore, rather than focusing on building systems that are only meant to process a restricted sub-language, such as the language of weather forecasts, many machine translation researchers have aimed to build "try anything" systems that

Table 3: Typical and atypical sentence constructions for weather forecasts.

Grammatically correct but atypical format for weather forecasts	**Typical format for weather forecasts**
This evening it will be cloudy and there is also a possibility that there might be some light rain	Cloudy with a chance of showers this evening
It will be very windy, and the wind will be blowing in gusts from a southerly direction	Strong winds gusting from the south
The temperature will descend to a low point of 20 degrees during the night	Low of 20 degrees overnight

are capable of taking unrestricted natural language as their input. However, there have also been efforts to improve the output of this type of more general system by employing controlled language when creating the input text.

As explained by Kittredge (2005, p. 441), whereas a sub-language is a naturally occurring restricted subset of language (e.g., weather forecasts), a controlled language has been consciously engineered to meet a special purpose, such as writing technical documentation for non-native speakers. A controlled language has two essential characteristics. First, the grammar of the controlled language is more restrictive than that of the general language. Second, the vocabulary of the controlled language typically contains only a fraction of the words that are permissible in the general language. This means that authors who write in a controlled language have fewer choices available when writing a text. In this way, a controlled language has something in common with the controlled vocabularies that are developed and applied by information professionals who do subject cataloging or controlled indexing.

In library and information science, a controlled vocabulary is a carefully selected list of words and phrases that are used to tag documents, such as research articles, so that they may be more easily retrieved by a search. As explained by Taylor and Joudrey (2009, p. 333):

> People writing about the same concepts often do not use the same words to express them, and people searching for the same concept do not think of the same words to search for it.

Because there are different ways of describing the same concept, the purpose of a controlled vocabulary is to take the guess work out of searching. For instance, if you are interested in finding information about World War I, what term would you use to search? World War I? The First World War? The Great War? The war to end all wars? Controlled vocabularies introduce consistency and solve problems such as polysemy, homography, and synonymy by ensuring that there is a one-to-one relationship between the concepts and the authorized terms. In essence, controlled vocabularies reduce the ambiguity that is inherent in natural language. By drawing together all items on the same topic under a single word or phrase, a controlled vocabulary can make searching for information much easier. Some well-known controlled vocabularies include the Library of Congress Subject Headings and the Medical Subject Headings. Similarly, periodical indexes, both print and digital, also allow users to search for articles using a controlled vocabulary. For more information on subject analysis and systems for vocabulary control, see Taylor and Joudrey (2009) and Rowley and Hartley (2008).

Returning to the broader notion of controlled languages for document authoring, one of the best-known examples is AECMA[1] Simplified English, a controlled language used for writing technical manuals in the aerospace industry. AECMA Simplified English has approximately 950 basic "approved" words,

[1]AECMA is the French acronym for the European Association of Manufacturers of Aerospace Equipment.

which have well-defined meanings and selected parts of speech. In addition, there are about 55 rules governing word usage and sentence construction (e.g., "You must break up noun clusters of four or more words by rewriting, hyphenating, or a combination of the two"). Controlled languages have proved useful not only for aerospace and automotive product documentation but also for telecommunications and software manuals, to give just a few examples.

Though not all controlled languages have been developed with a view to machine translation, some companies have certainly had success with this approach. A well-known controlled language that has also been applied with the intent of facilitating machine translation is Caterpillar Technical English (CTE) (Lockwood, 2000). Caterpillar is one of the world's leading manufacturers of construction and mining equipment, and their product support network operates in 35 languages. With expanding volumes of translation at Caterpillar, the company wanted to integrate machine translation into their document production process. However, a review of available machine translation systems revealed that none could produce adequate quality. Rather than investing in post-editing in 34 different languages, Caterpillar decided to do as much English-language "pre-editing" as possible at the authoring stage of the document cycle in order to improve the machine translation output and reduce the time and cost associated with post-editing texts in all the other languages. Lockwood (2000, p. 199) reports that this new approach to authoring and translation involved a substantial reengineering of Caterpillar's internal processes; however, the resulting CTE delivered numerous benefits for both human and machine translation at Caterpillar. Consistent source-text authoring not only improved the output of machine translation but it also improved the consistency and cost-effectiveness of human translation. Moreover, controlled authoring in CTE improves the quality and consistency of documentation for English-language users too.

Kittredge (2005, p. 443) provides an excellent summary of some of the main advantages and limitations of controlled languages. On the plus side, the clear and unambiguous nature of controlled language documents leads to fewer errors and hence to improved safety during the use and maintenance of products. Moreover, users of controlled language documents have fewer questions, which lowers product support costs. The relative simplicity and clarity of controlled language documents may even reduce the need for translation. For instance, many international workers might not fully understand the manuals written directly by North American engineers; however, they may achieve a high level of understanding when reading the texts produced by technical writers who have been trained in applying some form of controlled English. Meanwhile, in cases where translation is required, controlled language documents lend themselves more readily to human or machine translation, thanks to the reduction of ambiguity and complex syntax, and to the consistent use of authorized vocabulary.

On the other hand, controlled languages require a significant amount of time and effort to design and use correctly. Setting up a new controlled language, or adapting an existing standard for a new document producer, requires the intensive collaboration of domain experts (who can clarify the intended meanings), technical writers, and users. Unless suitable care is taken, there is a potential risk that efforts

to simplify a document will obscure important nuances of meaning, or otherwise distort the author's intended meaning. Several design iterations may be needed to reach consensus among all parties involved in the document's life cycle. Even when consensus is reached, it may take time to make all the required adjustments in the organization's business process. In addition, writing in a controlled language is a highly specialized skill that must be acquired by the technical writers. Owing to this learning curve, the cost per page of writing and editing controlled language documentation may initially be considerably higher than for traditional documentation. Clearly, such investment is justified only when the user community is large and there are economic or other benefits associated with setting up and enforcing the standard. Whereas the aerospace and automotive industries appear to have reaped the benefits of controlled language, other smaller or less global industries may not.

It is clearly not realistic to think of designing and implementing highly controlled languages for scientific research articles. On the one hand, there would need to be different controlled languages developed for each domain, or even subdomain, which would be incredibly time-consuming and expensive. In addition, controlled languages may prove to be too restrictive, not allowing for the description of newly discovered concepts. Moreover, even if it were possible to develop these languages, it would be extremely difficult to train every single researcher to use this controlled language when preparing their research articles, or to train every editor to enforce it. However, short of implementing a controlled language, it is still possible to encourage researchers to write in a more translation-friendly way for the benefit of themselves and their readers. This less restrictive yet still translation-friendly way of writing will be explored in more detail in the following chapter.

Key points from this chapter

- Machine translation research began in earnest following World War II, but it proved to be more challenging than researchers expected.
- Various approaches to machine translation have been tried, including rule-based methods (where researchers try to program computers to process language using grammar rules), corpus-based approaches that use pattern-matching and number-crunching techniques, and neural networks that learn by example.
- Each approach to machine translation has strengths and weaknesses, but an overarching challenge is that language is full of ambiguities at both the word level and the level of larger linguistic units (e.g., phrases or sentences).
- Some constructions that might not appear to be ambiguous to people are nonetheless ambiguous to machine translation systems because these systems cannot rely on real-world knowledge to help with the interpretation of meaning.
- One way to improve the output of machine translation systems is to control the input that they receive, and regardless of the underlying approach used, machine translation systems are more successful when they work in a restricted domain or with a restricted text type.
- There are some naturally occurring sub-languages, such as weather forecasts, but these are relatively rare. Most users want machine translation systems to work with a broader range of domains or text types.

- Controlled languages are consciously engineered to have a restrictive grammar and a small number of authorized words, but they are challenging and expensive to design and apply.

To find out more about …

Challenges for machine translation systems

- Arnold, D. (2003). Why translation is difficult for computers. In H. Somers (Ed.), *Computers and Translation: A Translator's Guide* (pp. 119–142). Amsterdam and Philadelphia: John Benjamins.
- Forcada, M. L. (2010). Machine translation today. In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of Translation Studies* (Vol. 1, pp. 215–223). Amsterdam and Philadelphia: John Benjamins.

Controlled language

- Crabbe, S. (2017). *Controlling Language in Industry: Controlled Languages for Technical Documents*. Cham, Switzerland: Palgrave-Macmillan.

Example-Based Machine Translation (EBMT)

- Carl, M. & Way, A. (Eds.). (2003). *Recent Advances in Example-Based Machine Translation*. Dordrecht: Kluwer Academic Publishers.

History of Machine Translation

- Hutchins, W. J. (2010). Machine Translation: A Concise History. *Journal of Translation Studies, 13*(1-2), 29–70.

Neural Machine Translation (NMT)

- Forcada, M. L. (2017). Making sense of neural machine translation. *Translation Spaces, 6*(2), 291–309.

Rule-Based Machine Translation (RBMT)

- Quah, C. K. (2006). *Translation and Technology*. Basingstoke: Palgrave-Macmillan.

Statistical Machine Translation (SMT)

- Koehn, P. (2010). *Statistical Machine Translation*. Cambridge: Cambridge University Press.

Sublanguage

- Somers, H. (2003). Sublanguage. In H. Somers (Ed.), *Computers and Translation: A Translator's Guide* (pp. 283–296). Amsterdam and Philadelphia: John Benjamins.

Chapter 3

Expanding the Reach of Knowledge Through Translation-Friendly Writing

> Machine translation will change the way we communicate, but will it increase our understanding?
>
> – International Federation of Library Associations and Institutions (2013)

This chapter explores strategies and techniques that all authors can use to produce texts that are more reader-friendly and, by extension, more translation-friendly. This chapter is aimed at *both* native and non-native speakers of English in the scholarly community. As we learned in Chapter 1, native English speakers have a major advantage when it comes to researching and writing articles for publication, but these scholars can take steps to alleviate some of the extra burden faced by their colleagues who do not speak English as a native language. If you are an English-speaking scholar, one way that you can do this is by conducting peer reviews in a more empathetic manner, and another is by preparing your own texts in such a way that they can be more easily understood and/or translated. Meanwhile, if you speak English as an additional language but would like to publish in English with the help of machine translation, you can also take steps to improve the output of machine translation systems by making your input texts more translation-friendly.

Collective action for the common good

The idea of writing in a clear and easily understandable language is not new. There is a vast body of literature addressing concepts such as International English, Standard English, Common English, Global English, Globish, Basic English, Plain English, and more. Some of these have developed along the lines of a sort of controlled language, using a subset of the full grammar and vocabulary available in English. Other attempts at facilitating communication have taken a domain-oriented approach, such as business English or simplified

Machine Translation and Global Research:
Towards Improved Machine Translation Literacy in the Scholarly Community, 55–78

doi:10.1108/978-1-78756-721-420191005

technical English. All these efforts have both their supporters and detractors, and we have no doubt that our modest suggestion for adopting a more translation-friendly way of writing for scholarly communication will also draw both compliments and criticism. As the saying on a coffee cup in our departmental lounge so aptly reminds us, "Pobody's nerfect!," but we need to begin somewhere, and this book represents a starting point for a discussion on machine translation literacy in the context of scholarly communication.

Before diving into the details of what translation-friendly writing might look like, we want to clarify that we are not debating whether there *should* be a *lingua franca* for scholarly communication – whether this be English or any other language. That is a different discussion entirely, and one that is largely beyond the scope of this book (though we touch on it very briefly in Chapters 1 and 4). In the context of this chapter, we are rooting ourselves in the existing situation in scholarly communication, where English has, for better or worse, currently taken hold as the dominant language for publication. Given this state of affairs, we are asking whether there are modest and practical ways, such as writing in a more translation-friendly fashion, to make this situation more workable for the benefit of all members of the scholarly community. Improving machine translation literacy will not solve all the problems, but we hope that it might nonetheless help to move the conversation forward incrementally.

As participants in and supporters of the scholarly communication process, it is time for us to take stock of the current reality, which includes the following notions that have been discussed up to this point in the book. Being aware of and understanding this general situation is an important precursor toward developing machine translation literacy:

- All natural languages are ambiguous to some degree.
- Readers who are highly proficient in a given language can disambiguate texts written in that language relatively easily, while those who are less proficient have more difficulty with the disambiguation process.
- Most research articles are published in English.
- Many researchers are not native speakers of English.
- Researchers who have English as an additional language are increasingly turning to machine translation systems to locate literature relevant to their research, to translate research material into their own language for assimilation, and to prepare articles in English for dissemination.
- Machine translation systems have difficulties processing texts that contain linguistic ambiguities (including ambiguities that do not typically pose problems for people).
- Reducing the linguistic ambiguities in a text makes the text easier to read (for any reader, including researchers who have English as an additional language) and translate (for a machine translation system or a professional translator).

Of course, it would be ideal if everyone approached every writing task with the aim of communicating in the clearest and the most effective manner possible. Indeed, when writing up the results of scientific research, most researchers

undoubtedly do strive to communicate effectively, and they probably do quite a reasonable job when the readers in question are highly proficient in the language and subject of the article. But without even realizing it, many authors may be making things more difficult than necessary for their colleagues or students who are speakers of other languages. Once researchers begin to consciously consider who is consulting their texts, and how these texts are being processed (e.g., through an additional language, or with the help of machine translation), they can take steps to ensure that their texts will better meet the needs of their readers. And in so doing, they will be making sure that these texts better serve their own needs too. After all, a critical part of any research project is the dissemination and sharing of knowledge. If the results of the research cannot be easily found or understood, then this research is less successful than it could be if others were able to find and consult it effectively.

While the authors of research articles certainly play a key role in making sure that their research is being communicated in an effective manner, they are not the only people to have input into the scholarly communication process. As noted in Chapter 1, peer reviewers, journal editors, and publishers have opportunities to provide feedback or to promote techniques for communicating more effectively. Meanwhile, academic librarians and others working in the information professions also have the opportunity to play an important role by promoting machine translation literacy among researchers across the spectrum of academia and beyond. What could easily be perceived as a "problem" that non-Anglophone researchers must solve by improving their English could perhaps be more helpfully recast as a challenge that all members of the scholarly scientific community can work to address. Recall the arguments made by Drubin and Kellogg (2012) and Benfield and Feak (2006), who suggest that the communications advantage held by native English speakers obligates them to acknowledge and to help alleviate the extra challenges faced by their fellow researchers who are using English as a second (or perhaps third or fourth) language, such as by being more generous as editors and peer reviewers. While these authors did not touch on matters related to machine translation, we feel that this is a natural extension of their sentiments. By educating researchers and students about the challenges and limits of machine translation, and about ways to address these by reducing linguistic ambiguities in academic writing, librarians and other members of the scholarly communications community can play a pivotal role in improving the situation across the board.

Indeed, there is ample evidence to suggest that this community as a whole is already taking steps to promote and support good writing, even though these efforts may not specifically aim to facilitate machine translation. As we have seen, some universities now offer courses on writing English for Research Publication Purposes. In addition, there are plenty of books that provide advice on how to write well, such as *Writing Scientific Research Articles* (2nd edition) by Cargill and O'Connor (2013). University libraries, such as the Duke University Medical Center Library and Archives[1] and the University of Southern California

[1]http://guides.mclibrary.duke.edu/scientificwriting

Libraries,[2] offer LibGuides (i.e., web pages that pull together all types of information about a particular subject) on scientific writing and writing research articles. Librarians at the Mount Sinai Hospital in Toronto have initiated a project to teach plain language principles to medical researchers so that patients can understand their work more easily, noting that "As librarians' roles evolve and become more integrated in the circle of care, they become a natural fit for teaching information literacy skills [to medical researchers]" (Horn, 2018, p. 105). Meanwhile, the websites of scientific publishers, such as Elsevier and Taylor & Francis, offer plenty of information such as "guides to getting published," "writing a journal article," "research writing skills," and the like. Some publishers, such as Wiley,[3] even offer tips on writing for search engine optimization to ensure that other researchers can find your articles more easily.

One group that is working actively to apply plain language techniques to scientific communication is Cochrane.[4] This non-profit, non-governmental organization draws on a large pool of volunteer experts to conduct systematic reviews of healthcare interventions and diagnostic tests. These reviews are written up for medical experts; however, each review is also accompanied by a Plain Language Summary (PLS) that a general audience without a medical or research background could understand and be able to use (Glenton, 2017). Though the medical experts receive some guidelines[5] to follow when preparing the plain language summaries, Jelicic Kadic et al. (2016) examined 1,738 summaries and determined that

> Cochrane plain language summaries are highly heterogeneous with a low adherence to the PLEACS [Plain Language Expectations for Authors of Cochrane Summaries] standards. Therefore, there is much room for improving the content and consistency of the PLS [Plain Language Summary]. A standardization of PLSs is necessary to ensure delivery of proper and consistent information for consumers.

As a result, Cochrane launched a pilot program in 2016 in which they set out to develop a template that could be distributed to PLS writers. The motivation for developing the template was multifold: "To ease production, comprehension, and translation to non-English languages through standardization of content, structure and language of the PLS, and develop writing guidance" (Wood & Tovey, 2016). With regard to translation, Cochrane does not appear to use machine translation at the present time. Rather, according to Jelicic Kadic et al. (2016):

> Since all Cochrane reviews are prepared and published in English, Cochrane has recognized the need to promote evidence-informed

[2]http://libguides.usc.edu/writingguide
[3]https://authorservices.wiley.com/author-resources/Journal-Authors/Prepare/writing-for-seo.html
[4]https://community.cochrane.org/
[5]Plain Language Expectations for Authors of Cochrane Summaries (PLEACS).

> health care by publishing its high-quality content in languages other than English. At the moment, there are 13 Cochrane translation teams around the world, managing translations into Chinese (simplified and traditional), Croatian, French, German, Japanese, Korean, Malay, Polish, Portuguese, Russian, Spanish and Tamil. Most of them translate only PLSs. Therefore, a PLS is now an important knowledge translation tool, not only in English, but in numerous languages worldwide. Therefore, it is now more important than ever to ensure high quality and homogeneity of the PLSs, and their adherence to standards.

In examining the template[6] currently used to guide the production of Cochrane Plain Language Summaries, we can see that it focuses primarily on the overall structure of the summary, providing suggested subheadings and indicating the type of content that should be included in each section. There is less emphasis on linguistic issues, though some general recommendations are made, such as: "Avoid acronyms and jargon." In addition, some set phrases are provided for authors to use in each section if appropriate. For example, in the section entitled "What is the aim of this review?," authors receive the following advice:

> People do not always understand that the results of a plain language summary come from a systematic review rather than a single study. Some also wrongly assume that the review authors have carried out the studies themselves. We therefore suggest that you use an introductory sentence such as: "The aim of this Cochrane Review was to find out if [….]. Cochrane researchers collected and analysed all relevant studies to answer this question and found [X#] studies." (Cochrane, 2018, p. 3)

The efforts made by Cochrane with regard to the production of plain language summaries and their translations are inspiring. However, their template applies only to the production of summaries of systematic reviews for their own Cochrane Library and would be difficult to apply more broadly to the production of scientific literature in general. In addition, Cochrane's current template focuses more on helping writers determine appropriate content for each section of the summary, with considerably less attention being paid to linguistic issues.

We saw in Chapter 2 that controlled languages can dramatically improve the quality of translated text, though designing and implementing a controlled language is very demanding. However, a number of studies have demonstrated that simply trying to remove ambiguities and constructions that could pose difficulties for a machine translation system can still lead to improved output (e.g., Aikawa, Schwartz, King, Corston-Oliver, & Lozano, 2007; Bowker & Buitrago Ciro, 2018; Gerlach, Porro, Bouillon, & Lehmann, 2013; Miyata & Fujita, 2017;

[6]https://www.cochrane.no/plain-language-summary-format

Sereten, Bouillon, & Gerlach, 2014). This less rigorous approach is sometimes referred to as pre-editing because it is applied to a source language text before the text is sent to a machine translation system. Investigations into pre-editing have shown that this will produce not only a more readable input text but also a more accurate and readable machine-translated version. In contrast, post-editing refers to efforts to improve the target language text after it has been translated by a machine translation system. Often, both activities are required in order to arrive at a translated text that is of publishable quality. However, pre-editing on its own can still improve the usability of a text for non-native speakers as well as improve the translatability of a text.

Therefore, general guidance on how we can write in a way that facilitates translation will be valuable. While pre-editing is the technical term used within the machine translation community, for a more general audience, we prefer to use the term translation-friendly writing. In the following sections, we will offer some suggestions for how members of the scholarly community can begin to prepare their research articles in a more translation-friendly way.

Academic writing style: Set in stone or open to change?

For most academic disciplines, the research article is the most prestigious genre. It is beyond the scope of this chapter to provide a detailed overview of the historical developments of the genre of scientific articles in English; however, a detailed analysis of this genre can be found in Swales (2004). In brief, and as summarized by Olohan (2016, p. 149), the evolution of the scientific article over the past few centuries has been interesting because changes in the form and content of articles reflect changes in how scientific research is carried out. For instance, while science was once considered an amateur, individual occupation for curious and wealthy gentlemen, it has evolved into a global industry where research is pursued by teams of highly trained specialists heavily backed by government and corporate funding. In line with these developments, research articles in the English language in particular have become more abstract, more densely packed with information, more focused on their argumentation, and more centered on the research than on the researcher.

While academic writing may indeed be characterized by certain features, it is not a static text type. Language is a living thing and if researchers in the scholarly community make a collective decision to alter the way that they write for the benefit of their peers, and by extension for society at large, then they have it within their power to do so.

Hyland and Jiang (2017) conducted a corpus-based investigation of academic writing style across several disciplines over a 50-year period. While they did not find that the style has changed dramatically across the board, they did observe that academic writing in engineering and biology appears to show a noticeably increased use of informal features such as first person pronouns and sentences beginning with conjunctions (e.g., “and,” “but”). In a similar vein, a recent blog post on the site of the Centre for Corpus Approaches to Social Science at the University of Lancaster in the United Kingdom suggests that academic writing

Table 4: Examples from the BNC 1994 and the BNC 2014 that illustrate how academic writing is becoming more colloquial.

Extract from BNC 1994	Extract from BNC 2014
Experimentally **one cannot** set up just this configuration because of the difficulty in imposing constant concentration boundary conditions (Section 14.3). In general, the most readily practicable experiments are ones in which **an initial density distribution is set up** and there is then some evolution of the configuration during the course of the experiment.	No doubt **people might object** in further ways, but in the end nearly all these replies boil down to the first one **I** discussed above. **I'd** like to return to it and ponder a somewhat more aggressive version, one that might reveal the stakes of this discussion even more clearly. Very well, someone might say. Not reproducing may make sense for most people, but my partner and I are well-educated, well-off, and capable of protecting our children from whatever happens down the road. Why **shouldn't** we have children if we want to**?**

Source: Hawtin (2018).

has shown signs of changing in the 20-year period between 1994 and 2014 (Hawtin, 2018). More specifically, a comparison of the contents of the original British National Corpus (BNC) that was compiled using data from 1994 and the content from the BNC 2014 reveals that academic writing has become more colloquial. Hawtin (2018) provides examples taken from the BNC 1994 and the BNC 2014 respectively, as illustrated in Table 4.

According to Hawtin (2018), the example on the left is typical of academic writing in the 1990s. First, it illustrates avoidance of using first and second person pronouns and contractions, both of which have increased in use in the BNC 2014. In addition, it includes a passive construction, the use of which has decreased in the BNC 2014. Meanwhile, in the example on the right taken from the BNC 2014, we see active sentence constructions, first person pronouns, verb contractions, negative contractions, and a question. Hawtin (2018) suggests that a more colloquial academic writing style is emerging and she posits that this is the result of researchers seeking to make their work more accessible to a wider audience:

> The colloquialisation of language can make messages more easily understood by the general public because, whilst not everybody is familiar with the specifics of academic language, everyone is familiar with spoken language.

Though Hawtin (2018) specifically mentions the use of colloquialization as a means to reach the general public, we believe that other changes in academic writing could result in this work becoming more easily accessible to researchers who have English as an additional language. Important next steps include

raising awareness in the scholarly community about the benefits of writing in a translation friendly way, as well as providing some tips about how this could be done.

What is the purpose of translation-friendly writing?

Translation-friendly writing is essentially writing that uses plain and simple language. The basis for an accurate, clear, consistent, and easy-to-read translation is an accurate, clear, consistent, and easy-to-read source text. We have seen that translation is challenging under the best of circumstances, so it stands to reason that it will be that much more difficult if the original source material is confusing or poorly written.

Plain language is writing – in any language – that is designed to ensure that readers can understand a text as quickly, easily, and completely as possible. The Plain Language Association International (PLAIN)[7] is one group that advocates the use of plain language in a range of document types, including government communications. They have devised a series of recommendations for how to write using plain language, which can be downloaded from their website. Moreover, this group does not work only in English; they have members working in a range of other languages too, including Dutch, Finnish, French, Hungarian, Italian, Malay, Norwegian, Portuguese, Spanish, and Swedish. PLAIN is just one such organization that is concerned with facilitating clear communication. Many writing and editing services have their own sets of guidelines for writing clearly, and many translation service providers give advice about how to write in a way that will facilitate translation by both professional translators and machine translation systems. The information provided below is therefore not unique or original. Rather, we have identified a selection of the strategies that are most commonly recommended to act as a starting point for scholars who are looking for guidelines on how to write a translation-friendly text. Additional sources of information are provided at the end of the chapter.

Our choice of the word guidelines is quite deliberate. We are not proposing to impose a highly regulated style of writing on scholars around the world. For one thing, it would be impossible to enforce, and for another, it may not be desirable in every single situation. Rather, what we would like to encourage is a mindset where scholars who are writing articles for publication think more carefully about:

1. *who* they are writing for;
2. *how* these readers might be accessing the text; and
3. *what* they want these readers to take away from the text.

Who: If you are writing for your peers, meaning other scholars in your field, then writing in plain language will help to ensure that all colleagues – even those who speak a different native language – can access the ideas in your text more easily.

[7]http://plainlanguagenetwork.org/

As an added benefit, a clear text might be easier for graduate students or early-career scholars in the field to understand as well.

How: If you are writing for a global audience of your peers, then be aware that many of your readers will be accessing your text either through a foreign language directly (meaning that they are reading your text in their second or third language), or via a translation (either human or machine). In either case, the content will be easier to access if the original text has been prepared in plain language.

What: If your main reason for writing the article is so that your peers can understand, contextualize and build on your findings, then presenting the material clearly and unambiguously is paramount.

It might sound as though writing in this fashion will result in a text that is bland or boring. Perhaps it will. But if your purpose is to communicate the findings clearly, rather than to entertain readers or market a product, then a bland style might not be detrimental. Just because the word choice isn't sizzling, doesn't mean that the ideas are not! If you want readers to really appreciate the content, then deliver it to them in an accessible style.

Of course, scholars write many different types of texts: textbooks, grant applications, annual reports, popularized articles, patent applications, blog posts, project websites, etc. We are not suggesting that all scholarly writing has the same audience or purpose, or that it should all be written in the same style. Rather, we simply want to encourage authors to think carefully about whom they are trying to reach when writing an academic article and how this audience may be accessing and interacting with their text. To serve its main purpose, does your academic article really need that witty title, that clever wordplay, that culture-specific metaphor, that telescoping acronym, or that unwieldy syntactic construction? If you think that it absolutely does, then by all means, leave it in! But if you feel that they are not strictly required for the type of text that you are writing, then consider leaving them out.

Ten guidelines for translation-friendly writing

The 10 guidelines given below are presented with examples in English. English speakers can apply these guidelines to their academic writing to make it easier for scholars who are not Anglophones to either read the English text directly or to obtain higher-quality machine translation output for the purpose of assimilating the information. Meanwhile, non-native speakers of English who would like to use machine translation to generate an English-language draft of an article for publication can use the same general approach – writing in plain language – when preparing a source text in their own language. A text that is prepared in plain language *in any language* will be more translation-friendly than one that is not.

Many of the guidelines presented below will likely seem obvious. Perhaps even too obvious to bother pointing out? However, writing style may not always be top of mind when you are preparing your findings for publication. A large part of our goal in promoting machine translation literacy is to make you aware of the effects of your language choices, many of which may be unconscious.

By thinking more specifically about your audience, you can make stylistic decisions that will improve the accessibility of your ideas. In suggesting that you adopt a more reader-friendly or translation-friendly style, we are not suggesting that you should "dumb down" the content. Rather, we are encouraging that wherever you can, you should seek to make that content clear and comprehensible to the widest possible range of researchers. If your aim in publishing research is to enable others to learn from it and to build on it, then it makes sense to do your best to make this work accessible.

1. Use short sentences

Shorter sentences are much easier to understand and translate than long complex strings of information joined by commas and semicolons. To avoid confusing readers or tripping up a machine translation system, use short clear sentences that express one idea at a time. It can also be helpful to follow the standard word order, which in English is subject–verb–object. A common recommendation is to aim for sentences that have between 20 and 25 words, though, of course, you should not sacrifice clarity for brevity. A good trick for keeping sentence length reasonable is to read the text aloud. If you have to take multiple breaths or read a sentence more than once to work out how its various components fit together, chances are that complications will arise during the translation process.

Instead of: Erdheim-Chester disease (ECD) is an extremely rare hematopoietic neoplasm that represents clonal proliferation of myeloid progenitor cells and may involve bone and one or more organ systems, primarily affecting adults in their fifth to seventh decades of life, with a slight male predominance. (43 words)

Try: Erdheim-Chester disease is an extremely rare hematopoietic neoplasm that represents clonal proliferation of myeloid progenitor cells. This disease may involve bone and one or more organ systems. This disease primarily affects adults in their fifth to seventh decades of life, with a slight male predominance. (16 words + 11 words + 18 words = 45 words)

2. Use the active voice rather than the passive voice

Using the active voice means clearly identifying the agent who is performing the action. As a result, the sentence is more direct and easier to understand and translate. In an active sentence, the person or thing responsible for the action comes first. In a passive sentence, the person or thing acted on comes first, and the actor is added at the end or sometimes omitted entirely. For many years, use of the passive voice has been considered a typical characteristic of academic and scientific writing. However, as we learned earlier in this chapter, evidence from the BNC collected in 2014 suggests that researchers have begun to use the passive voice less frequently in academic writing (Hawtin, 2018). From a translation point of view,

the active voice is more translation-friendly and will likely generate better machine translation output.

Instead of: *Data were summarized* on the basis of methodological themes of interest.

Try: *We summarized the data* on the basis of methodological themes of interest.

3. Avoid long noun strings or modifier stacks

When connecting elements are omitted from noun strings, readers – or machine translation systems – must infer the relationship between the words. Earlier in the book we saw the example of *liquid oxygen tank*, which raises the question about the relationship of *liquid* to the other elements of the phrase. Is the *oxygen* liquid, or is the *tank* liquid? If you have to read a sentence several times to work out its meaning, chances are that there will be further complications when it is translated into a different language. When this happens, the original meaning may be misinterpreted, or the translation may appear too literal. Where possible, try to break up long strings of nouns or modifiers by clarifying the relationships between them.

Instead of: Our findings provide a preliminary foundation for carrying out *copy number variation-based association studies* with economically important phenotypes of fine wool sheep in the future.

Try: Our findings provide a preliminary foundation for carrying out *association studies based on copy number variation* with economically important phenotypes of fine wool sheep in the future.

4. Use relative pronouns such as "that" and "which"

Understanding the relationship between different elements in a sentence is key to understanding its overall meaning. Sometimes, elements of a sentence are considered optional. In English, for example, the relative pronouns "that" and "which" may sometimes be omitted. However, when writing in a translation-friendly way, it is best to include such optional elements because they can help to clarify the relationships between different sentence elements and thus improve understanding.

Instead of: Our findings indicate hypoalphalipoproteinemia in male patients carrying the $BRAF^{V600E}$ mutation favors the formation of lipid-laden histiocytes.

Try: Our findings indicate *that* hypoalphalipoproteinemia in male patients carrying the $BRAF^{V600E}$ mutation favors the formation of lipid-laden histiocytes.

Instead of: MCPyV as well as Epstein-Barr virus, normally connected with humans under the form of subclinical infection, are thought to be involved at various degrees in several neoplastic and inflammatory diseases.

Try: MCPyV as well as Epstein-Barr virus, *which are* normally connected with humans under the form of subclinical infection, are thought to be involved at various degrees in several neoplastic and inflammatory diseases.

5. Avoid wordiness

The previous guideline suggests that optional pronouns should be included because they play an important role in disambiguating the relationships between different elements in a sentence. However, as a general rule, simple and precise sentences that do not contain any unnecessary words or information will be easier to translate. Fewer words mean fewer opportunities to make a mistake! All languages have impressive vocabularies, and some types of writing benefit from flowery descriptions that make use of a wide range of the words and structures available in a language. However, if the main goal of a scientific article is to share research findings clearly and effectively, rather than to entertain or impress readers with our expansive vocabulary, then eliminating unnecessary words and empty phrases is a more translation-friendly strategy. Focus on what is truly important and don't include unnecessary language.

Instead of: A spectrum analyzer is a versatile measurement instrument allowing a user to observe a frequency spectrum, whether it be broad or narrow.

Try: A spectrum analyzer displays signals over a given frequency range.

6. Use nouns instead of personal pronouns

A pronoun is a word that can be substituted for a noun. A pronoun usually refers back to a noun used earlier in the text. Often, the use of a pronoun can improve the readability of a text by making it sound less clumsy and repetitive. When seeking to write in a translation-friendly way, the key is to find the right balance. A text with no pronouns is unlikely to read well. However, since clear and precise communication is the overriding goal, there may be times when repeating the initial noun or using another noun instead of substituting a personal pronoun will help to ensure that the message is communicated correctly during translation. Recall our earlier discussion about the challenge of translating from a non-gendered language, such as English, into a language that uses grammatical genders, such as French, Spanish, or German. The English pronoun *it*, for example, may be rendered into French as either the masculine form *il*, or the feminine form *elle*, depending on the gender of the noun to which it is referring. Using the wrong gender for a pronoun could alter the meaning of the sentence. Therefore, it is worthwhile keeping this challenge in mind and using pronouns judiciously, rather than as a matter of course. The best way to ensure that the message will be transferred correctly is to use a noun, rather than a personal pronoun.

Instead of: The purpose of this research is to identify verbal dyspraxia. This subtype of autism is a motor-speech problem. *It* disables oral-motor movements needed for speaking.

Try: The purpose of this research is to identify verbal dyspraxia. This subtype of autism is a motor-speech problem. *Verbal dyspraxia* disables oral-motor movements needed for speaking.

Instead of: Pulmonary Langerhans cell histiocytosis is an interstitial primary pulmonary disease, characterized by Langerhans cell proliferation. *It* is easily misdiagnosed in children.

Try: Pulmonary Langerhans cell histiocytosis is an interstitial primary pulmonary disease, characterized by Langerhans cell proliferation. *This disease* is easily misdiagnosed in children.

7. Use terminology consistently

In a similar vein, writing in a translation-friendly way means using specialized terminology consistently, rather than substituting synonyms. Once again, because the clarity of the content is paramount in a research article, it is better to have a text that is clear if somewhat repetitive, rather than one that risks skewing the message by introducing unnecessary synonyms. Ideally, a single concept should be represented by a single term. Therefore, when referring to the same concept, always use the same term.

Instead of: Neurodegenerative disorders, including *amyotrophic lateral sclerosis, or motor neuron disease*, have been associated to alterations in chromatin structure resulting in long lasting changes in gene expression. *Amyotrophic lateral sclerosis* is predominantly a sporadic disease and environmental triggers may be involved in its onset. In this paper, we demonstrate that […] are selectively altered in cellular and animal models of *motor neuron disease*. These results reinforce the idea that epigenetic therapy may represent a potential and attractive approach for *amyotrophic lateral sclerosis* treatment.

Try: Neurodegenerative disorders, including *amyotrophic lateral sclerosis*, *or motor neuron disease*, have been associated to alterations in chromatin structure resulting in long lasting changes in gene expression. *Amyotrophic lateral sclerosis* is predominantly a sporadic disease and environmental triggers may be involved in its onset. In this paper, we demonstrate that […] are selectively altered in cellular and animal models of *amyotrophic lateral sclerosis*. These results reinforce the idea that epigenetic therapy may represent a potential and attractive approach for *amyotrophic lateral sclerosis* treatment.

As the above example shows, sometimes there is a legitimate synonym for a term and you might wish to make readers aware of this alternative way of describing the concept. In such a case, identify the synonym the first time that the concept is introduced, but avoid alternating between synonyms throughout the document.

8. Choose unambiguous words

Technical or specialized terminology rarely causes difficulty for researchers who are familiar with the specialized language of their field and know how to use it appropriately. Indeed, within a specialized field, clear communication is aided by the correct use of specialized terminology, so researchers should use the correct technical terms. However, a text does not consist solely of specialized terms. There are lots of general language words in research articles too. As we saw in an earlier chapter, words can sometimes have more than one meaning. In many cases, the translation of the word will be different depending on which meaning is selected. For instance, the English noun *ball* can refer to an object that is thrown or kicked in a sport, or to a formal dance party. *Ball* can therefore be translated into French as either *ballon* or *bal*, depending on the intended concept. To facilitate translation, choose words whose meanings are unambiguous whenever possible.

Instead of: We observed that the beams had a tendency to *bow*.
Try: We observed that the beams had a tendency to *bend*.

In this example, the English word *bow* has numerous possible meanings: a weapon used by an archer, a device used to play a stringed instrument, a hair decoration, the front of a ship, a polite gesture of acknowledgment, and so on. Assuming that a machine translation system could recognize that *bow* is being used as a verb in the sentence, rather than as a noun, there are still multiple interpretations, such as an actor bowing to the audience, or a sports team bowing to defeat, or a violinist bowing his instrument. In contrast, the word *bend* is likely to cause less confusion in translation because its meanings are more closely related (e.g., a *bend* in the road, *bending* a piece of metal or wood). While there can be different translations (e.g., *plier*, *courber*, or *tordre* in French), these tend to represent different nuances of one overall meaning, rather than referring to completely different concepts. Therefore, it is likely that the intended message will be understood.

9. Avoid abbreviated forms

In writing, long words or phrases may be abbreviated to save time, effort, and space. Though it is tempting to use abbreviated terms in scientific writing, there is a risk that communication will be garbled rather than clarified if, for example, an abbreviation is unfamiliar to the reader or to the machine translation system. The English-language abbreviation MS, for instance, could be a short form for

multiple sclerosis, *master of science*, *manuscript*, or even a polite term of address for a woman (Ms, rather than Miss or Mrs). When translating into French, a machine translation system might leave the abbreviated form untranslated, or it could translate as *SEP* (*sclérose en plaque*) or as *Mme*. Time saved by using an abbreviated form in the source text may generate confusion for readers of the translated text. Unless the complete term is terribly unwieldy and must be repeated often, it is more translation-friendly to use the full form.

Instead of: Electromagnetic interference is a disturbance that affects an electrical circuit due to either electromagnetic conduction or electromagnetic radiation emitted from an external source. *EMI* problems have been increasing with the proliferation of mobile electronic systems, wireless communication systems, and computer networks.

Try: Electromagnetic interference is a disturbance that affects an electrical circuit due to either electromagnetic conduction or electromagnetic radiation emitted from an external source. *Electromagnetic interference* problems have been increasing with the proliferation of mobile electronic systems, wireless communication systems, and computer networks.

10. Avoid idiomatic expressions, humor, and cultural references

People do not simply speak different languages. Rather, they come from other cultures, draw on other references, and have other ways of seeing the world that are completely valid yet different from our own. Idiomatic expressions from one language cannot usually be translated literally and are typically replaced with an entirely different expression in another language. For instance, the French expression "*se faire rouler dans la farine*" translates literally as "to be rolled in the flour." However, a more typical rendering of this expression into English might be something along the lines of "to have the wool pulled over your eyes." Humor is an essential component of some text types, but it is not easy to translate. Does it play an essential role in a research article, or can it be left out? Meanwhile, analogies and metaphor can be powerful tools for explaining or understanding complex ideas, but if they draw on culture-bound references, then these may end up being confusing rather than helpful. Consider the following opening paragraph from a scientific article entitled "Charge and Spin Density Waves" (Brown & Grüner, 1994a/b). Beside it, you can find the French translation – or rather, adaptation – produced by an unnamed professional translator.

On a hot July afternoon the Mall in Washington, DC, is overrun with sightseers. They move earnestly in zigzag patterns carrying their coolers, bouncing from museum to monument to cafeteria. Most of the streets bordering the lawns are flat, and as

Au Champs-de-Mars, les touristes et les promeneurs se déplacent sans cesse, au fil des allées, entre la tour Eiffel, l'École militaire, le pont d'Iéna et les rues adjacentes. Le site est plat, et des nombres égaux de promeneurs se dirigent dans toutes les directions.

many tourists stroll in one direction as in the other. Suddenly a drum roll is heard: a marching band is assembling. On the roads, displacing the confused crowd, are gathering serried ranks of uniformed high school students. Soon the band is mustered in neat rows, hardly disturbed even by a child trying to hide between the trumpeters' legs from a pursuing parent. As the tourists watch, the band starts to play and then marches forward with a clash of cymbals. The wanderers on the Mall imitate rather closely the behavior of electrons in common metals.	Soudain un roulement de tambour retentit : la fanfare des Beaux-Arts commence à jouer. Les musiciens se mettent en rangs, repoussant la foule désordonnée, et la fanfare bien en rangs, à peine perturbée par un enfant qui rejoint ses parents en se faufilant entre les trompettistes, s'avance au signal d'un coup de cymbales. Les visiteurs du Champ-de-Mars sont comme les électrons des métaux usuels.

Let's start by considering the English text. To explain the behavior of electrons in common metals, the authors use an analogy to a scene from everyday life. But who exactly can relate to this scene? It refers to a particular location in the capital city of the United States, as well as to concepts that are common in American culture: coolers, a marching band, high school students … But how easy is it for English speakers from other countries and cultures, such as Ireland or South Africa, to relate to this analogy, never mind readers from non-English-speaking countries such as France? In looking at the French version, we can see that the translator made a tremendous effort to adapt the text and make it much more relatable for a French audience by replacing landmarks in Washington, DC, with landmarks in Paris and dropping references with no typical cultural equivalent, such as *coolers*. A machine translation system would not be able to make such adaptations, and French readers would be left with a French language description of the scene in Washington, DC. But even if the machine translation system were able to substitute the Champs-de-Mars for the Mall, would the text be readily accessible to French-speaking readers outside France such as those residing in Quebec or in Cameroon, for example? In fact, the original English text with its culture-bound analogy is neither reader-friendly nor translation-friendly if its intention is to address a global audience.

We can't claim that the above list of guidelines is complete or that every guideline will be relevant to every line of every research article that you write. Moreover, these particular recommendations have been suggested with English in mind, and there may be other plain language recommendations that are more pertinent when writing in other languages. However, the main thing that we want to emphasize here is the importance of taking the time to see your text from the point of view of your readers, some of whom may be accessing it through another language. Even a small effort on your part to make your text more translation-friendly could result in a big payoff for your readers in terms of its comprehensibility. Are you ready to give it a try?

Taking the leap?

Even if you agree that the above suggestions for writing in a translation-friendly way make sense, we are not suggesting that changing the way that you write will be easy. It would be much simpler to continue writing in your usual way even if the result is not optimal. Change is hard! If you feel that modifying your overall academic writing style in one fell swoop is too daunting, a more workable approach could be to start small and continue on from there. One type of text that is regularly produced by all kinds of researchers and which could be a good starting point for practicing how to write in a more translation-friendly way is the academic abstract. If you like the idea of presenting your own work in a more translation-friendly way, then a feasible short-term goal could be to start by modifying the way that you write the abstracts associated with your articles. Given the importance of academic abstracts in the scholarly communication process, even making this modest effort has the potential to greatly improve the experience of researchers who have English as an additional language.

Abstracts

An abstract is a summary of a research article, usually between 100 and 500 words in length. While it is associated with a research article, an abstract must also function as a standalone text. It typically appears immediately preceding the main text of the article, though sometimes it is included at the end. This summary should allow readers to identify the basic content of an article quickly and accurately to determine its relevance to their own research interests. This will help them to decide whether they need to read the entire document (and perhaps therefore whether they need to have it translated). In addition, abstracts are usually indexed. The result of the indexing process is a series of subject-specific bibliographic databases which record the publications of the subject field and make them available for future searching. In this way, abstracts play an important role in helping researchers to locate articles as part of a literature search.

Scholarly journals almost always require an author to submit an abstract along with an article, and journals may provide guidelines for the abstract's form or content. There are different formats for abstracts, but the two that are most commonly found in academic journals are structured abstracts and informative abstracts. Structured abstracts are divided into a number of sections such as *Purpose*, *Approach*, *Findings*, and *Practical Implications*, or *Objective*, *Methods*, *Results*, and *Conclusions*. Meanwhile, informative abstracts contain similar information but are presented in a single paragraph format. Publishers may provide guidelines for authors to follow when preparing abstracts; however, in many cases, these guidelines tend to focus on content rather than form.

Though we have established that it would not be feasible to implement a controlled language for writing abstracts, we believe that there are advantages to be gained by writing abstracts in a way that will reduce ambiguity and thus increase

translatability. By creating reader- and translation-friendly abstracts, authors can expect to reap the following benefits:

- Abstracts will be easier for indexers to index accurately for inclusion in research databases, thus making the articles more findable.
- Abstracts will be easier for all readers (both native and non-native speakers) to understand, thus ensuring that knowledge is communicated more effectively.
- Abstracts will be easier to translate (for both machine translation systems and human translators), thus expanding the reach of the research while ensuring that the results are being communicated more effectively to speakers of other languages.

We should also point out that, where abstracts are concerned, the shoe might occasionally be on the other foot when it comes to providing text in a non-native language. While many journals seek articles written in English, some may nonetheless require abstracts to be provided in other languages. One example of a journal that adopts this practice is the *Canadian Journal of Information and Library Science*, which publishes articles in either English or French but requires the author to provide a title, abstract and keywords in both languages.[8] Though there are a few scientific journals that publish in this type of bilingual format, a more common practice is for English-language journals to publish abstracts of articles in a second or third language. For instance, the *British Journal of Surgery* provides Spanish language abstracts for all its articles.[9] Meanwhile, as we have already seen, an increasing number of national journals require English-language abstracts to accompany articles written in the national language. For instance, the National Autonomous University of Mexico publishes the journal *Investigación Bibliotecológica: Archivonomía, Bibliotecología e Información*, in which the vast majority of articles are written in Spanish. However, all Spanish-language articles must be accompanied by an English-language abstract.[10]

Transforming an abstract into a translation-friendly text: An extended example

The preceding sections suggested 10 general guidelines for writing in a translation-friendly way. In this section, we put these examples into practice using one of our own research abstracts. This abstract was written before we began thinking about the importance of translation-friendly academic writing and machine translation literacy. Hence, there is room for improvement! We present the original abstract on the left, while the version on the right has been revised to make it more reader- and translation-friendly.

In Table 5, we discuss some of the modifications that were applied to the text.

[8]https://utorontopress.com/ca/canadian-journal-of-information-and-library-science
[9]https://www.bjs.co.uk/spanish-abstracts/
[10]https://www.elsevier.com/journals/investigacion-bibliotecologica-archivonomia-bibliotecologia-e-informacion/0187-358x/guide-for-authors

Original abstract not prepared in a translation-friendly way	Revised abstract prepared following translation-friendly writing guidelines
Purpose – This study was designed to investigate the potential benefits and limitations associated with aligning accreditation and academic program reviews in post-secondary institutions.	*Purpose* – This study investigates the potential benefits and limitations of aligning accreditation and academic program reviews in post-secondary institutions.
Design/Methodology/Approach – Two graduate programs subject to both external professional accreditation and institutional cyclical review underwent a harmonized review. The process was developed as a collaborative effort between the academic units, the professional associations, and the university's graduate-level quality assurance (QA) office. For each program, a single self-study was developed, a single review panel was constituted, and a single site visit was organized.	*Design/Methodology/Approach* – Two graduate programs that are subject to both external professional accreditation and institutional cyclical reviews underwent an aligned review. The academic units, the professional associations, and the university's quality assurance office collaborated to develop the process. For each program, we developed a single self-study, appointed a single review panel, and organized a single site visit.
Findings – Initial feedback from the academic units suggests that the harmonization of accreditation and program reviews is perceived as reducing the burden on programs with regard to the time and effort invested by faculty and staff as well as in terms of financial expenses. Fourteen recommendations emerged for ways in which a harmonized review process can be set up for success.	*Findings* – Initial feedback from the academic units suggests that aligned reviews reduce the time and effort required from faculty, staff, and other stakeholders. Aligned reviews also reduce financial expenses. Based on this experience, we developed fourteen recommendations for successfully conducting an aligned review.
Practical Implications – The results suggest that harmonized reviews are not only resource-efficient but also allow reviewers to provide more holistic feedback which faculty may be more willing to take on board for program enhancement.	*Practical Implications* – The results suggest that aligned reviews are resource-efficient. Aligned reviews also enable reviewers to provide holistic feedback. Faculty may be more willing to implement this feedback to enhance programs.
Originality/Value – The present study contributes to the existing body of knowledge about aligned reviews by summarizing potential benefits and limitations. It also proposes criteria for the success of harmonized reviews and suggests best practices.	*Originality/Value* – The present study contributes to the existing body of knowledge about aligned reviews by summarizing potential benefits and limitations. The study also proposes recommendations and potential best practices for conducting aligned reviews.
(214 words)	(182 words)

Table 5: Description of modifications applied to the text to make it translation-friendly.

Guidelines	**Examples from the abstract**
1. Use short sentences	We split some long sentences into several shorter ones, such as: • The results suggest that harmonized reviews are not only resource-efficient but also allow reviewers to provide more holistic feedback which faculty may be more willing to take on board for program enhancement. (32 words) • The results suggest that aligned reviews are resource-efficient. Aligned reviews also enable reviewers to provide holistic feedback. Faculty may be more willing to implement this feedback to enhance programs. (8 + 9 + 12 words = 29 words)
2. Use the active voice rather than the passive voice	We replaced several instances of the passive voice with the active voice, including the following: • This study was designed to investigate → This study investigates • A single self-study was developed, a single review panel was constituted, and a single site visit was organized → We developed a single self-study, appointed a single review panel, and organized a single site visit
3. Avoid long noun strings or modifier stacks	We determined that reference to *graduate-level* was not essential and were therefore able to reduce the length of the noun string *graduate-level quality assurance office* to simply *quality assurance office*
4. Use relative pronouns such as "that" and "which"	In an effort to clarify the sentence, we inserted an optional *that*: "two graduate programs *that are* subject to …"
5. Avoid wordiness	In several places, the language used in the original abstract was tightened up to reduce wordiness: • associated with → of • as a collaborative effort → collaboratively • the harmonization of accreditation and program reviews → an aligned review
6. Use nouns instead of personal pronouns	We replaced the pronoun *it* with *This study*
7. Use terminology consistently	The original abstract used both the terms *harmonize* and *align* to describe the process of combining the two types of reviews. This terminology was standardized so that only forms of *align* were used

Table 5: (*Continued*)

Guidelines	Examples from the abstract
8. Choose unambiguous words	We replaced the word *constituted*, which can mean either "made up of" or "established" with the less ambiguous word *appointed*
9. Avoid abbreviated forms	The original abstract contained the abbreviated form QA, and this was replaced by the full form *quality assurance*
10. Avoid slang, idiomatic expressions, humor, and cultural references	We replaced the colloquial expression *take on board* with the more standard term *implement*

Though we did not organize a formal evaluation to compare the readability and translatability of the two versions of the abstract presented above, we did some informal consultations. We took both the original and the translation-friendly versions of the abstract and ran them through two different machine translation systems (Google Translate and DeepL Translator) to get translations in French and Spanish. Next, we showed all the texts – both source and target – to a group of French- and Spanish-speaking colleagues. All agreed that the translation-friendly original, as well as its translations into French and Spanish, were easier to understand than the corresponding non-translation-friendly versions. Though our little exercise was not conducted in a very scientific way, the results are in line with the more rigorous evaluations carried out by ourselves (e.g., Bowker & Buitrago Ciro, 2018) and by researchers such as Aikawa et al. (2007), Gerlach et al. (2013), Miyata and Fujita (2017), and Sereten et al. (2014), which demonstrated that pre-editing improves the quality of machine translation output.

Up to this point in the chapter, the focus has been firmly placed on the idea of writing or editing a text in a translation-friendly way as a means of improving the machine translation output. We have suggested that applying translation-friendly writing techniques to abstracts' preparation could be a good place to begin given the important role that abstracts play in scholarly communication. Once researchers become accustomed to preparing their abstracts with (machine) translation in mind, these translation-friendly writing habits might eventually spill over into the research articles themselves, making these articles more amenable to translation also. Even though this book focuses mainly on English and has suggested techniques that are geared toward writing in a translation-friendly way in English, we have pointed out that the general idea of translation-friendly writing can be applied to all languages. Some resources about how to write for translation in other languages can be found at the end of this chapter. However, even if authors do their best to prepare a translation-friendly text, the output of machine translation will still need to be post-edited if the ultimate goal is publication, rather

than simply "gisting" or information assimilation. The following section contains some information on post-editing machine translation output.

Some notes on post-editing

As we have learned, post-editing refers to the process of "fixing up" a text that has been translated by a machine translation system in order to correct any errors and make the text sound more natural. In many contexts, post-editing is carried out by professional translators. However, as mentioned previously, such services, though cheaper than translation, may still be beyond the means of many non-Anglophone scholars who are trying to publish their work in English. Recently, however, a group of researchers, including Marie-Josée Goulet, Sharon O'Brien, Carla Parra Escartín, and Michel Simard, have begun to investigate the feasibility of having scholars post-edit their own work in a process they refer to as *self-post-editing* (Goulet et al., 2017; O'Brien, Simard, & Goulet, 2018; Parra Escartín, O'Brien, Simard, & Goulet, 2017). Though self-post-editing poses a number of challenges and may not be possible for everyone, it appears to be a promising avenue for further exploration.

Nevertheless, though we were able to offer some practical general suggestions for how to write in a translation-friendly way, it is more difficult to provide this type of guidance for self-post-editing. This is because the types of problems that will be present in machine translation output will be tied more directly to the specific language pair and the particular machine translation system that was used.

With regard to the language pair, it is well known that a given source language is likely to interfere with target language production. Language teachers are very familiar with the fact that speakers of Chinese who are learning English as an additional language will make different types of errors than speakers of Spanish who are learning English as an additional language. In fact, it is partly owing to this type of source language transference that different varieties of English have emerged internationally, such as Indian English, which describes the variety of English that is characteristic of people from India. In some areas of India, speakers blend elements from English and Hindi together, creating a sort of hybrid language that has become known as Hinglish. Meanwhile, in Europe, there is an ongoing debate about whether the variety of English spoken by non-Anglophones on this continent constitutes a new variety of English, sometimes referred to as Euro-English (Forche, 2012).

In the case of machine translation, the underlying structure of the source language could influence the way that the text is interpreted or translated, and different source languages will present different challenges to a machine translation system. For instance, in French and Spanish, prepositional phrases are commonly used in places where English would more naturally use a pre-modification structure. Therefore, when translating the Spanish phrase *sociedad de la información* into English, a machine translation system might render it as *society of information*, rather than as the expected phrase *information society*. However, the types of problems that could arise when translating from German into English are likely to be different. German is a synthetic language, while English is more

analytic. Synthetic languages compose (synthesize) multiple concepts into each word, while analytic languages break up (analyze) concepts into separate words. When translating a word from German to English, a machine translation system might have difficulty knowing where to "break" the word, and breaking it at wrong place could produce a completely different translation. For instance, the German word *Wachtraum* could be split into *Wach* + *traum* or *Wacht* + *raum*. The first would be best translated into English as *daydream*, while the second would mean something more along the lines of a *guard room*. The point here is that the language combinations in play will likely generate different types of errors.

Similarly, the subject matter of the text and the machine translation system that is used will also have an influence on the output. As we saw in Chapter 2, a system that has been trained on a corpus of government documents might struggle to translate a scientific text. Likewise, a rule-based machine translation system will generate different kinds of errors than a statistical machine translation system.

Since the conditions – such as language combination, text type, subject field, and machine translation system – are likely to be different for each researcher, it is not feasible for us to offer generic guidelines for self-post-editing in this book. Instead, each researcher will need to practice using machine translation for their texts and learn to spot and correct the types of errors that are produced in their texts. It may, however, be useful to connect with both English language teachers and English-speaking colleagues in your subject field. If you can get assistance identifying the typical types of errors produced by a given machine translation system for a few of your texts, you may find it easier to spot and correct such errors on your own moving forward. In Chapter 5, we will consider a framework for machine translation literacy instruction in which language teachers, subject experts, and academic librarians can work together to raise awareness about issues relating to the use of machine translation in the context of scholarly communication. First, however, we will take a look at the bigger picture by contemplating some of the broader implications of machine translation use in Chapter 4.

Key points from this chapter

- Academic writing style is not set in stone and it appears to be changing to become somewhat less formal.
- Abstracts play an important role in scholarly communication: they summarize the main elements of a research article, and these summaries are indexed in academic databases which scholars consult to find related research.
- Abstracts can therefore be a good place to begin practicing translation-friendly writing.
- Controlled language improves the quality of machine translation output but is costly and difficult to implement.
- Less rigorous pre-editing or writing in a translation-friendly way to remove constructions that pose difficulties for machine translation systems can still lead to improved output.

- Translation-friendly writing might sound like it could be bland or boring, but a key goal for academic writing is to ensure that readers can access the content easily.
- Translation-friendly writing encourages authors to keep their audience in mind and to consider how these readers might be accessing texts (e.g., through a second language or via translation).
- Guidelines for translation-friendly writing include: use short sentences; use the active voice; avoid long noun strings; include relative pronouns; avoid wordiness; use nouns instead of pronouns; use terminology consistently; choose unambiguous words; avoid abbreviations; and avoid idioms, humor, and cultural references.
- It is difficult to present generic tips for post-editing because machine translation errors are dependent on the language combination, text type, subject field, and machine translation system used.

To find out more about …

Abstracting

- Cremmins, E. T. (1996). *The Art of Abstracting (2nd edition)*. Arlington: Information Resources Press.

Self-post-editing

- O'Brien, S., Simard, M., & Goulet, M.-J. (2018). Machine Translation and Self-post-editing for Academic Writing Support: Quality Explorations. In J. Moorkens, S. Castilho, F. Gaspari, & S. Doherty (Eds.), *Translation Quality Assessment* (pp. 237–262). Cham: Springer.

Translation-friendly writing

- Directorate-General for Translation. (2016). *How to write clearly*. https://publications.europa.eu/en/publication-detail/-/publication/725b7eb0-d92e-11e5-8fea-01aa75ed71a1/language-en/format-PDF/source-75466029 Note: This document is available in the twenty-four languages of the European Union.
- Parrish, D. (ed). (2009). Writing for Translation: Getting started guide. *Multilingual Magazine*, October/November.
- Translation Centre for the Bodies of the European Union. (2015). *Writing for Translation*. https://publications.europa.eu/en/publication-detail/-/publication/7c6c9877-f714-499b-980c-6d57259d75ea/language-en/format-PDF/source-75466077 Note: this document is also available in French.

World Englishes

- Mukherjee, J., and M. Hundt (Eds.). (2011). *Exploring Second-Language Varieties of English and Learner Englishes*. Amsterdam: John Benjamins.
- Percillier, M. (2016). *World Englishes and second language acquisition: Insights from Southeast Asian Englishes*. Amsterdam: John Benjamins.

Chapter 4

Some Wider Implications of Using Machine Translation for Scholarly Communication

> A discussion of the ethics of machine translation is urgently needed.
> – Kenny (2011)

Up to this point in the book, we have focused mainly on very practical and concrete issues, such as the challenges that scholars who have English as an additional language are facing, how machine translation systems work, and ways to write in a more machine translation-friendly fashion. In this chapter, we will take a moment to consider briefly some of the wider ethical implications associated with the use of machine translation in the context of scholarly communication. The decision to use machine translation is ultimately a personal one that will depend on a range of factors, and our goal in this chapter is simply to make you aware of some of the wider implications of using this technology so that you can make informed choices.

Just because we *can* use machine translation, does that mean that we *should*?

In Chapter 3, we emphasized that this book does not explicitly seek to answer the question about whether or not there *should* be a *lingua franca* for scientific communication. Nevertheless, free online machine translation is now ubiquitous for anyone with an Internet connection: we can access it through search engines, via widgets embedded into web pages, and even in some word-processing packages. But does the mere fact that we can access this technology at the click of a button mean that we should do so without a second thought? Because machine translation makes translation faster, and more affordable than it used to be, should we rush to translate scholarly material without hesitation? There is a longstanding debate as to whether technology is neutral or non-neutral. Some argue that technology itself is neither good nor bad, and that what counts is the way in which we choose to use it. Others counter that we cannot merely use technology without also being influenced by it to some extent. Both sides of the technological determinism versus

Machine Translation and Global Research:
Towards Improved Machine Translation Literacy in the Scholarly Community, 79–86

doi:10.1108/978-1-78756-721-420191006

technological instrumentalism debate present impassioned arguments (e.g., Carr, 2010, pp. 46–48), and it is unlikely that this debate will be resolved any time soon. Nevertheless, we feel it would be remiss if we did not at least raise some of the broader issues touching on ethics and machine translation for readers to contemplate. There is value in raising awareness of issues because it allows people to reflect more consciously and to make informed decisions. This was the case, for instance, for academic journal editors who were interviewed by Flowerdew (2001) about their varied experiences dealing with manuscripts submitted by native and non-native speakers of English. At the project's end, Flowerdew (2001, p. 146) remarked:

> Indeed, a number of editors stated that the interview process itself had been valuable in raising the various issues, some of which they had not considered. To take just one example, those editors who had not considered the question of World Englishes may now be in a better position to appreciate this perspective on international publication.

In that general spirit, we present the following sections in an effort to raise the visibility of some of the broader implications of using machine translation in the context of scholarly communication. Indeed, we consider fostering awareness of such issues to be an important component of machine translation literacy instruction.

To translate, or not to translate?

In any research project, one of the first steps is to situate the project in the wider field of knowledge. This typically involves conducting some type of literature review to determine what is already known about the subject. Because research takes place around the globe, and researchers speak many different languages, machine translation can be viewed as a tool that can help to cut through language barriers and allow all researchers to stand on an equal footing by having access to work that has been published in languages other than their own. As we have already seen, researchers regularly use machine translation for this type of "gisting." Moreover, because they are domain experts who are already familiar with the concepts in their field, the resulting machine translation output is typically seen as being quite usable and helpful for comprehension purposes. Therefore, when viewed through the lens of information assimilation, machine translation is often seen in quite a positive light.

But what about the other main use of machine translation: dissemination? If machine translation for assimilation purposes makes it possible, in principle, for researchers to publish in their own language and still reach a wide audience, then machine translation for dissemination purposes could be seen to favor the opposite and to support the use of a common language for research publication. As we learned in Chapter 1, English has, for the time being at least, established itself as the overwhelmingly dominant language of scientific publication. The pressure to publish in this *lingua franca* is linked to a common belief that this choice will result in the research being more readily indexed, accessed, read, used, and cited. However, as Desrochers and Larivière (2016, p. 888) have so aptly put it, in countries where the national or predominant language is not English, "publish or perish" has

a twist: "publish in English or perish." While the option to publish in English was simply not easily available to many non-Anglophone scholars in the past – because professional translation and editing services were too costly, for example – recent advances in machine translation coupled with a better understanding of how to interact effectively with these systems through pre- and post-editing means that the option of publishing in English is becoming more attainable. However, just because it is possible to publish in English with the help of machine translation, does it mean that we should? Maybe yes, maybe no, maybe sometimes. The decision is ultimately a personal one for each scholar, and every publication opportunity brings with it the chance to reflect and make that decision anew. What are some of the bigger picture issues that scholars may wish to consider as part of their decision-making process?

We have already discussed in Chapter 1 the fact that researchers who publish in English tend to be cited more often, they have better career advancement prospects, and have more opportunities for building networks, but are there downsides to consider? Bennett (2013) is one author who draws attention to the bigger picture and the potential consequences of using English as the dominant language of scholarly communication. Though she does not address the use of machine translation in particular, she does comment on the fact that translation in general has the potential to destroy the epistemological infrastructure of a source text. As Bennett (2013, p. 171) points out, not only are certain concepts so local that there may not be a translation readily available for them in English, the original text may also be structured according to rhetorical norms that are quite unfamiliar to those operating in English. In fact, sometimes the academic discourse used in the source text is embedded in an epistemological paradigm that effectively cannot be mapped directly to the paradigm underpinning equivalent texts in English. In such a case, translating the text inevitably destroys the entire epistemological infrastructure and replaces it with another that is more in keeping with the Anglo-Saxon worldview. Bennett (2013, p. 171) describes this process as "epistemicidal" because it essentially implies the obliteration of an alternative way of constructing knowledge. To combat it, Bennett (2013, p. 189) suggests that translators need to do more as intercultural mediators, raising awareness of these issues among stakeholders on both sides of the linguistic divide. Clearly, playing the role of intercultural mediator is far beyond the capacity of any current machine translation system.

Meanwhile, Desrochers and Larivière (2016), for their part, suggest that language is more than just a vehicle for knowledge. They note that it also has symbolic power, prompting them to ask whether in countries where language and identity are intertwined, publication strategies can be more than simply a question of "impact factor" and can generate symbolic capital for researchers. They give the example of funding agencies that might prioritize local societal impact. The answer does not appear to be clear-cut, however.

This raises questions fundamental to science: are there still contexts where opting for a language other than English can play in a researcher's favor? What symbolic capital can be associated with publication in a national language? And how will this capital be measured in a researcher's evaluation or in an institution's ranking?

For the moment, Desrochers and Larivière (2016) acknowledge that in the absence of policies that grant clear value to national languages, researchers seem set to continue seeking opportunities to publish in English. However, Burgess, Gea-Valor, Moreno, and Rey-Rocha (2014) recount that there are some researchers

who are more strategic with regard to their publication decisions. They surveyed a group of historians in Spain and found that while these researchers do indeed recognize that external forces have converged to require them to publish in English if they want to attract funding and obtain promotions, they simultaneously

> appear to be fighting a quiet rear-guard action where they support publication in their own and other languages while making astute strategic use of resources that allow them to publish in English when they deem it necessary and desirable. (Burgess et al., 2014, p. 81)

Solovova, Vieira Santos, and Veríssimo (2018) also discuss the tension between the challenges and opportunities associated with the choice between publishing in English and another language. They retrieved data from the Web of Science on articles from three disciplinary areas (Linguistics, Library and Information Science, and Pharmacology) written in Portuguese and English by authors based in Portugal between 1998 and 2017. Results show that while the number of English papers is higher throughout, the rise in the number of Portuguese papers is steeper in more recent years. In the view of Solovova et al. (2018, p. 12), the fact that some of these authors seek to retain their Portuguese voices points to "spaces of resilience and contestation of some hegemonic practices," and they end by observing that "it remains to be seen through further research if the future will become conflated under a monolingual banner or will be steered towards plurilingual diversity."

Indeed, though much of this volume has focused on the potential of machine translation for helping scholars who are not native English speakers to publish their work in English, there is no reason why machine translation could not be used as a tool to promote a greater degree of multilingualism in academic publishing. Burgess et al. (2014, p. 82) close their article with the following suggestion:

> Finally, those L1 [Language 1] English speakers who already use a language other than English might make a greater effort to use it as a medium of publication so as to provide further support for multilingual publication and partially redress the imbalance in favour of English, the consequences of which their EAL [English as an additional language] colleagues confront.

Considered in this light, could machine translation be a tool that could help to turn the tide?

Looking beyond the machine

In more direct reference to translation and technologies, there has been relatively little written thus far on the subject of ethical considerations. Nevertheless, there is some evidence that this issue is now beginning to attract more attention (e.g., Dos Santos, 2016; Marshman, 2014). For the most part, this recent literature explores questions that relate to the use of translation tools by professional translators, including how this affects their professional identity, their bottom line, and the sustainability of the profession. However, a paper by Kenny (2011) has examined the subject of ethics and machine translation from a broader perspective, taking into

account other stakeholders, including commissioners and consumers of machine translated texts. (Note: In cases where machine translation is used for assimilation purposes only, the consumer and the commissioner are one and the same.)

According to Kenny (2011), there is an urgent need to engage in a discussion on the ethics of machine translation, and she traces this need to the advent of statistical machine translation, which she notes has begun to cause some malaise among translation scholars. As we learned in Chapter 2, the original rule-based approach to machine translation relied on an extensive set of grammatical and lexical rules written by trained computational linguists. However, statistical machine translation does not rely on grammars and lexicons. Instead, it seeks to "learn" probable translations of phrases from an already existing parallel corpus of source texts and their human translations. In addition to relying on human translations for its training data, statistical machine translation also relies on human translation for its legitimacy: the reason that developers train statistical machine translation systems with parallel corpora is because these corpora are assumed to contain good answers. What's more, notes Kenny (2011, p. 2), is that "they are assumed to contain good answers precisely because they contain translations performed by human beings." Though Kenny's paper preceded the introduction of neural machine translation, this newer technology also depends to a large extent on training data that consists of human translations. But what exactly is the ethical issue here?

The rub lies in the fact that the contribution made by human translators to the process of machine translation goes largely unnoticed. As described by Kenny (2011, p. 2), this technology:

> relies on the ingenuity of both human translators (who produce vital data) and statistically-minded computer scientists (who work out clever ways of using these and other data), and both sets of protagonists might expect to be acknowledged in discussions.

Kenny (2011, p. 6) suggests that ethical decisions are reasoned decisions that take others into account, and she asks whether the developers and users of statistical machine translation take into account the interests of translators such as by recognizing and acknowledging translators as a source of data? Kenny finds that, with few exceptions, the role of translators in creating vital data has been mostly downplayed or ignored. She also finds that there are more subtle ways of making translators invisible, such as through the use of metaphors. Kenny cites the example of "translation as a natural resource," where translation is understood as occurring in nature, and like other natural resources, is open to exploitation by those with appropriate technology and legal rights. Kenny (2011, p. 9) concludes by observing that:

> at a time when machine translation has never been as reliant on human translation as it is now, it is ironic that the role of translators in the creation of parallel data is often obscured.

Stupiello (2008), too, has expressed concern about the invisibility of translators, noting that "the illusion that the machine is able to translate may affect the way translators will be seen in the future, an impression that should be given careful consideration."

Free online machine translation can now be easily accessed by researchers around the world with the press of a button, and our goal in raising these issues is to remind scholars that the routine decisions that we make in our working lives – such as using free online machine translation systems – can affect others, such as human translators. Researchers who choose to use machine translation systems have some ethical obligations toward those whose intellectual production is re-consumed in the process, even if this is simply recognizing the fact that machine translation has not eliminated the need for human translators but is instead highly dependent upon it.

Privacy concerns

Much of the discussion surrounding machine translation revolves around the quality of the text. Putting quality aside for a moment, another issue to consider is that there may be unseen risks associated with using free online machine translation tools. According to Kamocki and O'Regan (2016, p. 4461), using such tools may entail some privacy risks of which users may be completely unaware, and of which machine translation service providers may be tempted to take advantage. For example, users may imagine that the data entered into a free online machine translation service simply disappears once the translation process is completed. However, machine translation service providers are typically interested in keeping this data and in possibly reusing it in the future (e.g., as training data). This is one reason why it is important for scholars who wish to employ machine translation to understand – at least in basic terms – how these systems work. As the capacity and availability of these free machine translation tools expand, so to does the amount of inadvertently disclosed sensitive data. Depending on the nature of their research, scholars may need to be concerned about the potential loss of intellectual property or proprietary data on free online machine translation sites. By using free online machine translation, they may be unintentionally broadcasting their intellectual property to the world before it is published officially.

Common Sense Advisory, an independent market research firm focusing on language services, has explored the risks associated with inputting sensitive information into online translation tools such as Google Translate and Microsoft's Bing Translator, among others. According to DePalma (2014), sensitive data can leak in two ways: in transit or at the site. First, the "wrong" people can see information in transit. This issue is not restricted to machine translation but is a symptom of increasing reliance on web-based services and the cloud. Users make machine translation requests over unencrypted connections or use open Wi-Fi hotspots that anyone could monitor. This exposes potentially sensitive information to whoever happens to be listening in.

Second, and less often considered, is what online machine translation providers do with the data that users input. These free online machine translation sites can use data in ways that researchers did not intend. While content ownership remains with the creator, free online machine translation providers may claim usage rights under their terms and conditions. For example, as reported by DePalma (2014), Google notes that it does not claim any ownership of the content that users submit or of the translations of that content that is returned by the machine translation system. However, if you follow the policy links, you learn the following:

> When you upload or otherwise submit content to our Services, you give Google (and those we work with) a worldwide license to use, host, store reproduce, modify, create derivative works (such as those resulting from translations, adaptations or other changes we make so that your content works better with our Services), communicate, publish, publicly perform, publicly display and distribute such content. (cited in DePalma 2014)

Moreover, the license to use the data continues even after a user ceases to use the service.

Both DePalma (2014) and Kamocki and O'Regan (2016) advocate increased user awareness: by being aware of potential dangers, researchers can take steps to limit them, such as by following safe network and Wi-Fi procedures (e.g., using secure HTTP and encrypted connections), or identifying an offline or secure machine translation option. For instance, some machine translation services can be purchased and it may be possible to negotiate how your data will be used in such instances; however, the fee may not be reasonable for individual scholars. There are also options such as using an open source system that can be installed locally, such as the Moses machine translation system,[1] though this will require a significant investment of time, along with a reasonable amount of technical know-how, to install, train, and run the system. It also requires a substantial volume of appropriate training data, which may not be something that a researcher has at hand.

A more realistic option for scholars may be to distinguish between those areas of their paper that contain potentially sensitive information (e.g., findings) and those that do not (e.g., literature review) and to use free online machine translation only to translate the non-sensitive portions. In this way, machine translation could offer at least a partial solution.

In conclusion, the decision to use a tool such as machine translation is ultimately a personal one, and it will depend on a host of factors. Our intention has not been to sway researchers in one direction or the other. Rather, in including this chapter, where we have introduced and briefly discussed some of the wider issues surrounding use of machine translation in the context of scholarly communication, we have aimed simply to raise awareness of these issues so that people who choose to use machine translation can be informed users.

Key points from this chapter

- Using machine translation in the context of scholarly communication may have unintended consequences.
- While there are many advantages to using a common *lingua franca* for global research dissemination, such as enabling the work of scholars to be indexed, accessed, read, used, and cited more readily, there are also drawbacks, such as the creation of an epistemological monoculture.
- Machine translation can be used to translate into English, thus enabling non-Anglophone researchers to participate more easily in the English-dominated

[1] http://www.statmt.org/moses/

publishing sphere; or, it can be used to translate out of English, thus attempting to partially redress the imbalance that currently favors English.
- Far from eliminating the need for human translators, current approaches to machine translation, such as example-based machine translation, statistical machine translation, and neural machine translation, rely heavily on data produced by human translators, though this work often goes unacknowledged.
- Material that is sent to a free online machine translation system may be kept and may be subject to usage rights by the provider.
- By considering the wider implications of the use of machine translation, scholars will be better able to make an informed decision about whether machine translation is the correct option for them for a given task.

To find out more about …

Ethics of machine translation

- Kenny, D. (2011). The ethics of machine translation. In S. Ferner (Ed.), *Reflections on Language and Technology – The Driving Forces in the Modern World of Translation and Interpreting. Proceedings of the 20th New Zealand Society of Translators and Interpreters National Conference*, 4-5 June 2011, Auckland, New Zealand: NZSTI. Available from: http://doras.dcu.ie/17606/

Machine translation and privacy concerns

- DePalma, D. (2014). Free Machine Translation can leak data. *TC World*, July 2014. http://www.tcworld.info/e-magazine/translation-and-localization/article/free-machine-translation-can-leak-data/

Publishing in a national language or publishing in English

- Bennett, K. (2015). Towards an epistemological monoculture: Mechanisms of epistemicide in European research publication. In R. Plo Alastrué & C. Pérez-Llantada (Eds.), *English as a Scientific and Research Language: Debates and Discourses: English in Europe*, pp. 9–36. Berlin: De Gruyter Mouton.
- Gotti, M. (2017). English as a Lingua Franca in the academic world: Trends and dilemmas. *Lingue e Linguaggi*, *24*, 47–72.
- Solovova, O., Vieira Santos, J., Veríssimo, J. (2018). Publish in English or Perish in Portuguese: Struggles and Constraints on the Semiperiphery. *Publications, 6*(2), 25.

Technological determinism and technological instrumentalism

- Carr, N. (2010). *The Shallows: What the Internet is Doing to our Brains*. New York: W. W. Norton & Company.

Chapter 5

Towards a Framework for Machine Translation Literacy

> The Internet is altering the nature of literacy, generating New Literacies that require additional skills and strategies.
>
> – Leu, Zawilinski, Forzani, and Timbrell (2015)

Up to this point in the book, we have considered how English has become the global language of scholarly communication as well as some of the challenges faced by researchers who are not native speakers of English but who wish to publish in this language. We have learned some of the history of machine translation and the main approaches used in this field. We have explored how techniques such as writing in a translation-friendly way can improve the output of machine translation systems, and we have briefly discussed post-editing texts that have been machine translated. We have also contemplated some of the broader implications of using machine translation for scholarly communication. Essentially, we now have all the building blocks to create a working definition and framework for machine translation literacy. Bringing these elements together into such a framework is the purpose of this final chapter.

Machine translation literacy: A working definition

While it may seem odd to offer a working definition of machine translation literacy in the final chapter of this book, rather than presenting it at the beginning, this has been a deliberate decision. When we started writing this book, we did not have a fully formed definition in mind. Rather, it was through researching the concepts, talking to others, and experimenting with these ideas that we began to develop a better understanding of what is involved in machine translation literacy in the context of scholarly communication.

In general terms, "literacy" means having competence or knowledge in a specified area. The areas of both scholarly communication and machine translation are currently in a state of flux; they are each evolving as the world around us – and

Machine Translation and Global Research:
Towards Improved Machine Translation Literacy in the Scholarly Community, 87–95

doi:10.1108/978-1-78756-721-420191007

our relationship with it – evolves too. With that in mind, we expect that any definition that we put forward will not remain definitive in the long term; however, we believe it is important to keep this conversation about scholarly communication and machine translation moving forward. To the best of our knowledge, this book represents a first attempt to propose and discuss the notion of machine translation literacy. We therefore hope that our working definition will be taken as just that – an initial effort to outline the key requirements for machine translation literacy in the broad context of scholarly communication. With that in mind, we propose the following working definition, and we invite feedback and suggestions for its refinement and improvement.

In the context of scholarly communication, machine translation literacy refers to a scholar's ability to:

- comprehend the basics of how machine translation systems process texts;
- understand how machine translation systems are or can be used (by oneself or by other scholars) to find, read, and/or produce scholarly publications;
- appreciate the wider implications associated with the use of machine translation;
- evaluate how (machine) translation-friendly a scholarly text is;
- create or modify a scholarly text so that it could be translated more easily by a machine translation system; and
- modify the output of a machine translation system to improve its accuracy and readability.

A preliminary framework for machine translation literacy instruction

What we are presenting here is intentionally called a framework because it is based on a cluster of interconnected core concepts with flexible options for implementation, rather than requiring a single, rigid program for delivery. Within the broad group referred to as scholars, individual members have different characteristics, aims, and needs. Native English speakers could be trying to increase the accessibility of their work through translation-friendly writing. Meanwhile, non-native speakers, who may have different native languages and varying degrees of proficiency in English, may want to use machine translation to help them find and access existing scholarly literature in their field (i.e., translating from English into their native language), or may be looking to use machine translation as an aid to help them produce a text for dissemination (i.e., translating from their native language into English).

In addition, "machine translation" is not a single entity either. We have learned that different machine translation systems may be based on different underlying approaches (e.g., rule-based, example-based, statistical, or neural) or on a combination of such approaches. Different tools are therefore likely to produce different translations. Moreover, even if we take a single system, it will not perform equally well for all language pairs or directions. A given system may have a large volume of training data available for English and French but a considerably smaller set of training data for English and Icelandic, for instance. So too have we

seen that systems may perform differently when dealing with texts from different subject fields. A system that has been trained on texts in the legal domain may not perform well when faced with a scientific text, for example.

Even though librarians (or others) who aim to offer machine translation literacy instruction may eventually wish to customize a program that is tuned to the specific needs of their patrons (e.g., focusing on a particular machine translation system or language combination), we feel that it might be helpful for us to outline in broad strokes what a machine translation literacy training program for scholars could look like.

In the sections below, we present four main modules that taken together could comprise a half-day interactive workshop that librarians could deliver to scholars who are seeking to become more informed about the (potential) applications of machine translation for scholarly communication. The first three modules are addressed to all scholars, including both native and non-native speakers of English, while the fourth is likely to be of interest mainly to those who wish to use machine translation as an aid for publishing in English. In the model presented below, we have chosen to weave the questions that touch on the wider implications of machine translation use into other discussions at places where we find them to be most pertinent. However, if desired, it would be possible to break such questions out into a separate module.

Module 1. Why think about machine translation in the context of scholarly communication?

Target audience: All scholars.

Goals: To set the scene regarding the current linguistic context of scholarly communication; to present the main ways in which machine translation is used in the context of scholarly communication; and to foster greater collegiality across the scholarly community with regard to scholarly publication.

Format: Presentation and discussion.

Allotted time: 30 minutes.

Content: Begin with a brief overview of the overall linguistic state of scholarly publishing, including a very brief discussion of the potential consequences of using a *lingua franca* for global scholarly communication. Introduce in a general way the main applications of machine translation in a scholarly communication setting. First, explain how a non-native English speaker can use machine translation to translate potential keywords from another language into English to use as search terms in a database. Next, discuss how machine translation can be used for assimilation purposes, such as the case of a non-native speaker of English using machine translation to read the English-language literature in their own language. Finally, present the possibility of a non-native speaker of English preparing a draft publication in their own language and using machine translation to obtain an English-language version for potential publication. End by raising awareness of the ways in which publishers, editors, peer reviewers, and authors – both native and non-native speakers of English – can work together to create a

more empathetic and collegial scholarly community, and how machine translation fits into this.

Module 2. Overview of machine translation systems

Target audience: All scholars.

Goals: To introduce some of the complexities of language and raise awareness about why translation is challenging for machines.

Format: Presentation and discussion.

Allotted time: 45 minutes.

Content: A brief history of machine translation will help to situate participants, and a general overview of the main approaches will give them an appreciation of both the complexity of language and the limitations of computer techniques for processing language. Briefly describe the main approaches to machine translation, moving from a rule-based approach where programmers tried to make computers process language the way humans do (i.e., using bilingual dictionaries and grammar rules), through example-based and statistical approaches where computers were allowed to do what they do well (i.e., pattern matching and number crunching), to the most recent neural approaches based on artificial intelligence techniques (i.e., neural networks and machine learning). Emphasize the role played by human translators in example-based, statistical, and neural approaches to machine translation. Do not get bogged down in too many technical details; providing examples to illustrate some of the complexities of language and translation between languages will be more effective than simply giving technical descriptions of the tools. As a result, scholars may be less frustrated by the imperfect results of machine translation and better able to spot problems in both the input and the output texts. Inform participants of the potential risks associated with sending sensitive data or intellectual property to free online machine translation systems.

Module 3. Translation-friendly writing and editing

Target audience: All scholars.

Goals: To make participants aware of the features of translation-friendly texts, to give them a chance to practice applying translation-friendly writing or editing techniques, and to foster greater awareness of the benefits of translation-friendly writing in the context of scholarly communication.

Format: Presentation and hands-on exercises. It will be beneficial either to hold this portion of the workshop in a computer lab, or to invite participants to bring their own device (e.g., laptop or tablet) with them.

Allotted time: 60 minutes.

Content: In advance, invite participants to bring one or more of their own research abstracts with them. For native English speakers, these abstracts can be in English, while for speakers of other languages they can be in their native language. Introduce and explain the main guidelines for writing in a translation-friendly way using

illustrative examples, keeping in mind that the general principles for translation-friendly writing can be applied in any language. Next, ask participants to work on their own abstracts by first identifying problematic areas in their texts (i.e., non-translation-friendly constructions), and then editing the text to make it more translation-friendly. Take both the original abstract and the more translation-friendly version and run these through a machine translation system. Ask a participant who is a speaker of the target language to assess the results, and discuss as a group. If there is time, do another round of revisions to make the text even more translation-friendly.

As a variation, even without running the texts through a machine translation system, participants can work in pairs or small groups to compare the readability and comprehensibility of the two source language versions (i.e., the original abstract and the translation-friendly version). As a general rule, a text that is considered to be translation-friendly will also be easier for readers to read and understand than one that is not translation-friendly.

Module 4. Self-post-editing machine translation output

Target audience: Scholars wishing to use machine translation for dissemination. (Recommended for participants with an advanced level of English.)

Goals: To examine machine translation output in English, to identify and correct problems in the text, and to explore and reflect on the potential of machine translation for producing scholarly texts for dissemination.

Format: Hands-on exercises. It will be beneficial either to hold this portion of the workshop in a computer lab, or to invite participants to bring their own device (e.g., laptop or tablet) with them.

Allotted time: 45 minutes.

Content: In advance, invite participants to bring a section of an academic paper that they have drafted in their own language. It should be a section of approximately 350–500 words that does not contain sensitive information. Remember that different machine translation systems will produce different translations. Ask participants to run their text through two or more different machine translation systems, and to compare the resulting English-language versions. Next, participants should select the version that they feel will be easiest to edit and work on fixing errors and improving the text.

This section of the workshop will consist mainly of individual work, but the instructor can circulate to answer questions or offer suggestions. The goal is to allow participants to experiment with machine translation and to evaluate its potential for helping them to produce an English-language version of their paper. Some participants may decide that editing a machine-translated version is less beneficial than drafting a text directly in English. However, others may find this to be a helpful way of working, and following the workshop they can continue to hone their skills independently or with the help of language teachers or editors. Remember that the overarching goal of the machine translation literacy workshop is to explore the potential of machine translation for facilitating scholarly communication, rather than to provide comprehensive training in self-post-editing.

At the end of the workshop, or a short while after, it will be helpful to conduct a survey to get feedback from the participants. This will allow you to customize, refine, modify, or expand the content of the workshop for future iterations.

Who will deliver the instruction?

As we have noted already, we feel that academic librarians are well placed to play a key role in developing, coordinating, and delivering instruction on machine translation literacy at academic and research institutions. The need for machine translation literacy is present in every discipline of scientific enquiry, and the library is an ideal venue for this type of cross-cutting instruction. Academic libraries already offer a range of workshops that attract researchers from a broad range of disciplines, such as workshops on research data management, managing bibliographic references, and navigating electronic resources, to name just a few. Moreover, librarians are already experienced in providing instruction in scholarly communication, information literacy, media literacy, and digital literacy, for example, and we see instruction in machine translation literacy as being a logical extension of this type of work.

With regard to the specific type of academic librarian who would be equipped to deliver machine translation literacy instruction, we recognize that different institutions have divided the work of academic librarianship differently. In some universities, it might be the scholarly communication librarian who could take on such a workshop, while elsewhere it might fall under the remit of a digital literacy librarian, or a research and instruction librarian. In our view, any academic librarian whose job involves providing direct support to researchers engaged in the scholarly communication process would be well positioned to play an active role in the development, coordination, and delivery of some type of program on machine translation literacy.

We firmly believe that academic libraries are dedicated to offering services that will improve the overall experience of both faculty and students in the academic sphere. As we learned in Chapter 1, even a brief perusal of the literature demonstrates that librarians are extremely adept at finding creative solutions to meet the ever-changing and expanding needs of their patrons, including the needs of international scholars. Hughes (2010), Knight, Hight, and Polfer (2010), and Lahlafi, Allen, and Bull (2013) are just a few librarians who have been attempting to rethink the academic library and its services for the international community in recent years. In a similar vein, Jackson and Sullivan (2011) describe over a dozen innovative library-based projects that aim to support the success of international students studying at academic institutions in the United States, including efforts to identify and address "deeper" issues of information literacy in international students. Given that libraries in general go to considerable lengths to understand and address the particular needs of their patrons, a librarian can customize the content of a machine translation literacy program accordingly. For instance, a university library that works with a high percentage of French-speaking scholars may choose to focus on a different machine translation system or a different set of examples than a university library whose research community has a high proportion of Korean-speaking scholars.

While we see academic librarians as being central figures in the effort to raise the level of machine translation literacy, they could certainly benefit from assembling a team that includes supporters from elsewhere on campus. For instance,

with regard to the more language-oriented components of a machine translation literacy session, such as the modules on translation-friendly writing and self-post-editing, academic librarians may wish to collaborate with faculty teaching English language or writing courses. Horn (2018) has demonstrated that librarians can successfully deliver plain language instruction with a view to improving information literacy; however, establishing links with English language teachers might provide opportunities for scholars to pursue additional or deeper training in translation-friendly writing and self-post-editing that goes beyond the basic machine translation literacy instruction.

Similarly, academic librarians might like to invite faculty members to participate in the design and delivery of machine translation literacy sessions. Faculty who are experienced in the use of machine translation in a scholarly communication context – whether for information assimilation or dissemination – could provide valuable advice, tips, or tricks to others to help them get the most out of this technology. The combination of subject expert and language teacher worked well for Benfield and Feak (2006), for example, who collaborated to offer workshops on academic writing to a range of medical researchers who were not native speakers of English.

Finally, academic librarians may also want to coordinate their efforts with the international office on campus, whether as part of a user needs analysis before offering a workshop, or to promote the workshop once it is ready to be delivered. Overall, the proposed framework will open the way for librarians, faculty, and other institutional partners to redesign and expand information, media, and digital literacy instruction sessions, to connect machine translation literacy with researcher success initiatives, and to create wider conversations about the relevance of machine translation literacy to scholarly communication on local campuses and beyond.

Finally, we should mention that though the focus of this book has been firmly placed on the use of machine translation by and for post-secondary researchers, we believe that school librarians or public librarians could adapt the concepts presented here to offer a modified machine translation literacy program to their patrons. School children, including those who may be newly arrived immigrants or those following language immersion or multilingual education programs, may also stand to benefit from some form of machine translation literacy instruction.

Key points from this chapter

- After following a training program on machine translation literacy, scholars will be able to:
  - outline the basics of the main approaches to machine translation;
  - describe how machine translation systems can be used for finding, understanding, or producing scholarly texts;
  - articulate broader issues surrounding the use of machine translation;
  - recognize and create translation-friendly texts; and
  - improve machine translation output.
- A framework for machine translation literacy has flexible implementation options, but one possible delivery format could be a half-day workshop along the following lines:

Module	Main contents and activities	Instructor	Suggested time (minutes)
Why think about machine translation in the context of scholarly communication?	• Current linguistic context in scholarly communication (English as *lingua franca*) • Main uses of machine translation in scholarly communication: finding research, reading research, publishing research • Ways to create a more collegial scholarly community, and the role of machine translation within it	Academic librarian	30
Overview of machine translation systems	• Main approaches to machine translation: rule-based, example-based, statistical, and neural machine translation, including the role of human translators in machine translation • Challenges for machine translation systems • Risks of using free online machine translation	Academic librarian	45
BREAK			15
Translation-friendly writing/ editing	• Main guidelines for translation-friendly writing • Identify and edit problem areas in an abstract • Compare original and edited versions for readability and comprehension; compare the machine-translated versions	Academic librarian in collaboration with language instructor	60
BREAK (or end of workshop for those not wanting to explore self-post-editing)			15
Self-post-editing machine translation output	• Translate an extract from a research paper using two or more machine translation systems • Compare versions to see which will be easier to post-edit • Identify and correct errors	Academic librarian in collaboration with language instructor or faculty expert	45

Concluding remarks

Writing this book has offered us a very satisfying opportunity to read many pages on research, communication, language, literacy, and technology, and to reflect on and compare our own – very different! – experiences as researchers and members of the scholarly communication community. Discussions with other researchers and research facilitators at home and abroad during this reading, reflecting, and writing process have been both enriching and challenging. As a result, producing this book has been both an eye-opening and a rewarding experience which has led us to new reflections on machine translation and its applications in the research process. We trust that the material put forward here will help and challenge others, and we especially hope that it will serve as a tool for further critical engagement with daily practice in the broad context of scholarly communication.

This page intentionally left blank

References

Aikawa, T., Schwartz, L., King, R., Corston-Oliver, M., & Lozano, C. (2007). Impact of controlled language on translation quality and post-editing in a statistical machine translation environment. In *Proceedings of the MT summit XI*, Copenhagen, Denmark (pp. 1–7). Retrieved from https://www.microsoft.com/en-us/research/wp-content/uploads/2007/10/MTS-2007-Aikawa.pdf

Allen, J. (2003). Post-editing. In H. Somers (Ed.), *Computers and translation: A translator's guide* (pp. 297–317). Philadelphia, PA: John Benjamins.

Anazawa, R., Ishikawa, H., & Kiuchi, T. (2012). The perceived need for Japanese nursing faculty members to learn English: Issues related to career development. *The Journal of Continuing Education in Nursing*, *43*(4), 188–191. doi:10.3928/00220124-20111101-01

Anazawa, R., Ishikawa, H., & Kiuchi, T. (2013a). Use of online machine translation for nursing literature: A questionnaire-based survey. *Open Nursing Journal*, *7*(1), 22–28. doi:10.2174/1874434601307010022

Anazawa, R., Ishikawa, H., & Kiuchi, T. (2013b). Evaluation of online machine translation by nursing users. *CIN: Computers, Informatics, Nursing*, *31*(8), 382–387. doi:10.1097/NXN.0b013e3182999dc2

Anazawa, R., Ishikawa, H., Park, M. J., & Kiuchi, T. (2012). Preliminary study of online machine translation use of nursing literature: Quality evaluation and perceived usability. *BMC Research Notes*, *5*(1), 635. doi:10.1186/1756-0500-5-635

Anazawa, R., Ishikawa, H., Park, M. J., & Kiuchi, T. (2013). Online machine translation use with nursing literature: Evaluation method and usability. *CIN: Computers, Informatics, Nursing*, *31*(2), 59–65. doi:10.1097/NXN.0b013e3182701056

Arnold, D. (2003). Why translation is difficult for computers. In H. Somers (Ed.), *Computers and translation: A translator's guide* (pp. 119–142). Philadelphia, PA: John Benjamins.

Baron, S., & Strout-Dapaz, A. (2001). Communicating with and empowering international students with a library skills set. *Reference Services Review*, *29*(4), 314–326. doi:10.1108/009 07 320 110408447

Benfield, J., & Feak, C. (2006). How authors can cope with the burden of English as an international language. *Chest*, *129*(6), 1728–1730. doi:10.1378/chest.129.6.1728

Bennett, K. (2013). English as a lingua franca in academia: Combating epistemicide through translator training. *The Interpreter and Translator Trainer*, *7*(2), 169–193. doi:10.1080/13556509.2013.10798850

Bennett, K. (2015). Towards an epistemological monoculture: Mechanisms of epistemicide in European research publication. In R. Plo Alastrué & C. Pérez-Llantada (Eds.), *English as a scientific and research language: Debates and discourses* (pp. 9–36). Berlin, Germany: De Gruyter Mouton.

Bonet, J. (2013, February). No rage against the machine. *Languages and Translation*, *6*, 4–5. Retrieved from https://www.scribd.com/document/341510534/Languages-and-Translations-Machine-Translations

Bornmann, L., & Mutz, R. (2015). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, *66*(11), 2215–2222. doi:10.1002/asi.23329

Bowker, L. (2002). *Computer-aided translation technology: A practical introduction*. Ottawa, Canada: University of Ottawa Press.

Bowker, L. (2018). Can machine translation assist scholars who have limited English proficiency with searching? *Catalogue and Index*, *192*, 48–51. Retrieved from https://archive.cilip.org.uk/cataloguing-indexing-group/newspublications/catalogue-index

Bowker, L., & Buitrago Ciro, J. (2015). Investigating the usefulness of machine translation for newcomers at the public library. *Translation and Interpreting Studies*, *10*(2), 165–186. doi:10.1075/tis.10.2.01bow

Bowker, L., & Buitrago Ciro, J. (2018). Localizing Websites using machine translation: Exploring the connection between user eXperience and Translatability. In S. Chan (Ed.), *The human factor in machine translation* (pp. 138–151). London: Routledge.

Brown, S., & Grüner, G. (1994a). Charge and spin density waves. *Scientific American, 270*, 50–56. doi:10.1038/scientificamerican0494-50

Brown, S., & Grüner, G. (1994b). Ondes de densité de charge et ondes de densité de spin. *Pour la Science*, *200*, 44–51.

Buitrago Ciro, J. (2018). Les bibliothécaires pourraient-ils appuyer les chercheurs? *Journal de l'Association des bibliothèques de la santé du Canada*. [*Journal of the Canadian Health Libraries Association.*], *39*(2), 114. doi:10.29173/jchla29379

Burgess, S., Gea-Valor, M. L., Moreno, A. I., & Rey-Rocha, J. (2014). Affordances and constraints on research publication: A comparative study of the language choices of Spanish historians and psychologists. *Journal of English for Academic Purposes*, *14*, 72–83. doi:10.1016/j.jeap.2014.01.001

Byrne, J. (2012). *Scientific and technical translation explained: A nuts and bolts guide for beginners*. Manchester: St. Jerome Publishing.

Canadian Dental Hygienists Association (CDHA). (2009). Machine translation. *Canadian Journal of Dental Hygiene*, *43*(4), 151. Retrieved from https://www.cdha.ca/pdfs/Profession/Journal/v43n4.pdf

Cargill, M., & Burgess, S. (2008). Introduction to the special issue: English for research publication purposes. *Journal of English for Academic Purposes, 7*(2), 75–76. doi:10.1016/j.jeap.2008.02.006

Cargill, M., & O'Connor, P. (2013). *Writing scientific research articles: Strategy and steps*. Chichester: Wiley-Blackwell.

Carl, M., & Way, A. (Eds.). (2003). *Recent advances in example-based machine translation*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Carr, N. (2010). *The shallows: What the internet is doing to our brains*. New York, NY: W. W. Norton & Company.

Castilho, S., Moorkens, J., Gaspari, F., Calixto, I., Tinsley, J., & Way, A. (2017). Is neural machine translation the new state of the art? *Prague Bulletin of Mathematical Linguistics*, *108*(1), 109–120. doi:10.1515/pralin-2017-0013

Citizenship and Immigration Canada. (2016). *International students by top 50 countries of citizenship*. Retrieved from https://open.canada.ca/data/en/dataset/6609320b-ac9e-4737-8e9c-304e6e843c17

Chandioux, J., & Guéraud, M.-F. (1981). MÉTÉO: Un système à l'épreuve du temps. *Meta*, *26*(1), 18–22. doi:10.7202/002213ar

Clarke, A., Gatineau, M., Grimaud, O., Royer-Devaux, S., Wyn-Roberts, N., Le Bis, I., & Lewison, G. (2007). A bibliometric overview of public health research in Europe. *European Journal of Public Health*, *17*(1), 43–49. doi:10.1093/eurpub/ckm063

Cochrane. (2018). How to write a plain language summary of a Cochrane intervention review. Retrieved from https://www.cochrane.no/plain-language-summary-format

Corcoran, J. (2015). *English as the international language of science: A case study of Mexican scientists' writing for publication*. Retrieved from https://tspace.library.utoronto.ca / handle/1807/70842

Costa-jussà, M. R., & Fonollosa, J. A. R. (Eds.). (2015). Special issue: Hybrid machine translation: Integration of linguistics and statistics. *Computer Speech and Language*, *32*(1), 1–160.

Crabbe, S. (2017). *Controlling language in industry controlled languages for technical documents*. Cham, Switzerland: Palgrave-Macmillan.

Cremmins, E. T. (1996). *The art of abstracting*. Arlington, TX: Information Resources Press.

Crystal, D. (2003). *English as a global language* (2nd ed.). Cambridge: Cambridge University Press.

Delisle, J. (2008). *La terminologie au Canada : Histoire d'une profession (Terminology in Canada: History of a profession)*. Montreal, Canada: Linguatech.

Delisle, J., & LeBlanc, C. (2016, February 7). Parlez-vous le français-robot? *Le Droit*, *103*(261), 14. Retrieved from https://www.ledroit.com/opinions/votre-opinion/parlez-vous-le-francais-robot-404dc1eca304d03e701eb6d6a671663c

DePalma, D. A. (2014). Free machine translation can leak data. *TC World*. Retrieved from http://www.tcworld.info/e-magazine/translation-and-localization/article/free-machine-translation-can-leak-data/

De Silva, P. U. K., & Vance, C. K. (2017). *Scientific scholarly communication: The changing landscape.* Cham, Switzerland: Springer.

Desrochers, N., & Larivière, V. (2016). Je veux bien, mais me citerez-vous? On publication language strategies in an anglicized research landscape. In I. Ràfols, J. Molas-Gallart, E. Castro-Martínez, & R. Woolley (Eds.), *Proceedings of the 21st international conference on science and technology indicators: Peripheries, frontiers and beyond* (pp. 887–892). Valencia, Spain: Universitat Politècnica de València. Retrieved from http://ocs.editorial.upv.es/index.php/STI2016/STI2016/paper/viewFile/4543/2327

Directorate General for Translation. (2016). How to write clearly. *EU Publications*. Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/725b7eb0-d92e-11e5-8fea-01aa75ed71a1/language-en

Dos Santos, L. M. (2016). Effect of technological developments on ethical position of translator. *English Language and Literature Studies, 6*(3), 42–46. doi:10.5539/ells.v6n3p42

Drubin, D. G., & Kellogg, D. R. (2012). English as the universal language of science: Opportunities and challenges. *Molecular Biology of the Cell*, *23*(8), 1399. doi:10.1091/mbc.E12-02-0108

ElMalik, A. T., & Nesi, H. (2008). Publishing research in a second language: The case of Sudanese contributors to international medical journals. *Journal of English for Academic Purposes*, *7*(2), 87–96. doi:10.1016/j.jeap.2008.02.007

Eriksson, S., & Helgesson, G. (2017). The false academy: Predatory publishing in science and bioethics. *Medicine, Healthcare and Philosophy*, *20*(2), 163–170. doi:10.1007/s11019-016-9740-3

Fernández Polo, F. J., & Cal-Varela, M. (2009). English for research purposes at the University of Santiago de Compostela: A survey. *Journal of English for Academic Purposes*, *8*(3), 152–164. doi:10.1016/j.jeap.2009.05.003

Fink, G. A., Kummert, F., & Sagerer, G. (1995). With friends like statistics who needs Linguistics – Statistische versus wissensbasierte sprachverarbeitung. *KI – Künstliche Intelligenz, (Statistical versus knowledge-based language processing. AI - Artificial Intelligence) 5*, 31–33.

Flowerdew, J. (1999). Writing for scholarly publication in English: The case of Hong Kong. *Journal of Second Language Writing*, *8*(2), 123–145. doi:10.1016/S1060-3743(99)80125-8

Flowerdew, J. (2001). Attitudes of Journal Editors to Nonnative Speaker Contributions. *TESOL Quarterly*, *35*(1), 121–150. doi:10.2307/3587862

Flowerdew, J. (2015). Some thoughts on English for research publication purposes (ERPP) and related issues. *Language Teaching*, *48*(2), 250–262. doi:10.1017/S0261444812000523

Forcada, M. L. (2010). Machine translation today. In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of translation studies* (Vol. 1, pp. 215–223). Philadelphia, PA: John Benjamins.

Forcada, M. L. (2017). Making sense of neural machine translation. *Translation Spaces*, *6*(2), 291–309. doi:10.1075/ts.6.2.06for

Forche, C. R. (2012). On the emergence of Euro-English as a potential European variety of English: Attitudes and interpretations. *Jezikoslovlje (Linguistics)*, *13*(2), 447–478. Retrieved from https://hrcak.srce.hr/91474

Ford, C., Faires, D., Hirsh, S., & Carranza, N. (2017). The significance of language study in library and information science: A comparison of two programs in the United States and Honduras. *Journal of Education for Library and Information Science*, *58*(2), 77–93. doi:10.12783/issn.2328-2967/58/2/3

Garcia, I. (2011). Translating by post-editing: Is it the way forward? *Machine Translation*, *25*(3), 217–237. doi:10.1007/s10590-011-9115-8

Gerlach, J., Porro, V., Bouillon, P., & Lehmann, S. (2013) Combining pre-editing and post-editing to improve SMT of user-generated content. In S. O'Brien, M. Simard, & L. Specia (Eds.), *Proceedings of MT summit XIV workshop on post-editing technology and practice*, Nice, France (pp. 45–53). Retrieved from https://archive-ouverte. unige.ch/unige:30952

Gillis, A. (2017). Beware! Academics are getting reeled in by scam journals. *University Affairs*, *58*(2), 18–27. Retrieved from https://www.universityaffairs.ca/features/feature-article/beware-academics-getting-reeled-scam-journals/

Glenton, C. (2017). *Assessing the feasibility and acceptability of approaches for improving the quality of plain language summaries in Cochrane Reviews*: *A pilot study*. Final report. Retrieved from https://community.cochrane.org/review-production/production-resources/plain-language-summaries/updates-plain-language-summaries

Gotti, M. (2017). English as a lingua franca in the academic world: Trends and dilemmas. *Lingue e Linguagge (Languages and Speech)*, *24*, 47–72. doi:10.1285/i22390359v24p47

Goulet, M.-J., Simard, M., Parra Escartín, C., & O'Brien, S. (2017). La traduction automatique comme outil d'aide à la rédaction scientifique en anglais langue seconde: Résultats d'une étude exploratoire sur la qualité linguistique. *ASp*, *72*, 5–28. Retrieved from https://journals.openedition.org/asp/5045

Gray, B. (2008). Enhancing transdisciplinary research through collaborative leadership. *American Journal of Preventive Medicine*, *35*(2 Suppl), S124–S132. doi:10.1016/j.amepre.2008.03.037

Hawtin, A. (2018). *Is academic writing becoming more colloquial*? Blog post on the site of the ESRC Centre for Corpus Approaches to Social Science, University of Lancaster, UK. Retrieved from http://cass.lancs.ac.uk/author/abigailhawtin/

Horguelin, P. A., & Clas, A. (1966). La traduction technique. *Meta*, *11*(1), 15–25. doi:10.7202/003113ar

Horn, D. (2018). Engaging learners: Interactive ideas for teaching health literacy. *Journal de l'Association des bibliothèques de la santé du Canada*. [*Journal of the Canadian Health Libraries Association.*]*, 39*(2), 105. doi:10.29173/jchla29379

Hughes, H. (2010). International students' experiences of university libraries and librarians. *Australian Academic & Research Libraries*, *41*(2), 77–89. doi:10.1080/00048623.2010.10721446

Hutchins, W. J. (1986). *Machine translation: Past, present, future*. Chichester: Ellis Horwood.

Hutchins, W. J. (Ed.). (2000). *Early years in machine translation: Memoirs and biographies of pioneers*. Philadelphia, PA: John Benjamins.

Hutchins, W. J. (2010). Machine translation: A concise history. *Journal of Translation Studies*, *13*(1 & 2), 29–70.

Hutchins, W. J., & Somers, H. L. (1992). *An introduction to machine translation*. London: Academic Press.

Hyland, K., & Jiang, F. (2017). Is academic writing becoming more informal? *English for Specific Purposes*, *45*, 40–51. doi:10.1016/j.esp.2016.09.001

Institute of International Education. (2016). *Fast facts*. Retrieved from https://www.iie.org/Research-and-Insights/Open-Doors/Fact-Sheets-and-Infographics/Fast-Facts

International Federation of Library Associations and Institutions (IFLA). (2013). *Riding the waves or caught in the tide? Navigating the evolving information environment*. Retrieved from http://trends.ifla.org/insights-document

Ishimura, Y., & Bartlett, J. C. (2014). Are librarians equipped to teach international students? A survey of current practices and recommendations for training. *Journal of Academic Librarianship*, *40*(3–4), 313–321. doi:10.1016/j.acalib.2014.04.009

Jackson, P. A., & Sullivan, P. (2011). *International students and academic libraries: Initiatives for success*. Chicago, IL: Association of College and Research Libraries.

Jelicic Kadic, A., Fidahic, M., Vujcic, M, Saric, F., Propadalo, I., Marelja, I.,… Puljak, L. (2016). Cochrane plain language summaries are highly heterogeneous with low adherence to the standards. *BMC Medical Research Methodology*, *16*, 61. doi:10.1186/s12874-016-0162-y

Jenkins, J. (2014). *English as a lingua franca in the international university: The politics of academic English language policy*. London: Routledge.

Jenkins, J., Baker, W., & Dewey, M. (Eds.). (2018). *The Routledge handbook of English as a lingua franca*. London: Routledge.

Jinha, A. E. (2010). Article 50 million: An estimate of the number of scholarly articles in existence. *Learned Publishing*, *23*(3), 258–263. doi:10.1087/20100308

Kamocki, P., & O'Regan, J. (2016). Privacy issues in online machine translation services – European perspective. In N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, & B. Maegaard (Eds.), *Proceedings of the tenth international conference on language resources and evaluation (LREC 2016)* (pp. 4458–4462). Paris, France: European Language Resources Association. Retrieved from https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/5043/file/Kamocki-ORegan_privacy_issues_in_online_machine_translation_2016.pdf

Kenny, D. (2011). The ethics of machine translation. In S. Ferner (Ed.), *Reflections on language and technology – The driving forces in the modern world of translation and interpreting. Proceedings of the 20th New Zealand society of translators and interpreters national conference*. Auckland, New Zealand: New Zealand Society of Translators and Interpreters. Retrieved from http://doras.dcu.ie/17606/

Kenny, D., & Doherty, S. (2014). Statistical machine translation in the translation curriculum: Overcoming obstacles and empowering translators. *The Interpreter and Translator Trainer*, *8*(2), 1–19. doi:10.1080/1750399X.2014.936112

Kit, C., & Wong, T. M. (2008). Comparative evaluation of online machine translation systems with legal texts. *Law Library Journal*, *100*(2), 299–321.

Kittredge, R. I. (2005). Sublanguages and controlled languages. In R. Mitkov (Ed.), *The oxford handbook of computational linguistics* (pp. 430–447). Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780199276349.013.0023

Knight, L., Hight, M., & Polfer, L. (2010). Rethinking the library for the international student community. *Reference Services Review*, *38*(4), 581–605. doi:10.1108/00907321011090746

Koehn, P. (2010). *Statistical machine translation*. Cambridge: Cambridge University Press.

Koehn, P., & Knowles, R. (2017). Six challenges for neural machine translation. In *Proceedings of the first workshop on neural machine translation* (pp. 28–39). Retrieved from https://sites.google.com/site/acl17nmt/accepted-papers

Koponen, M. (2016). Is machine translation post-editing worth the effort? A survey of research into post-editing and effort. *JoSTrans: Journal of Specialised Translation*, *25*, 131–148. Retrieved from http://www.jostrans.org/issue25/art_koponen.php

Lahlafi, A., Allen, S., & Bull, S. (2013). A positive library experience for international students? Results from a Business Librarians Association survey. *SCONUL Focus*, *57*, 28–33. Retrieved from http://shura.shu.ac.uk/7566/

Leal Fontes, H. (2013). Evaluating machine translation: Preliminary findings from the first DGT-wide translators' survey. *Languages and Translation*, *6*, 10–11. Retrieved from https://www.scribd.com/document/341510534/Languages-and-Translations-Machine-Translations

Lehrberger, J. (1982). Automatic translation and the concept of sublanguage. In R. Kittredge & J. Lehrberger (Eds.), *Studies of language in restricted semantic domains* (pp. 81–106). Berlin, Germany: Walter de Gruyter.

Leu, D. J., Zawilinski, L., Forzani, E., & Timbrell, N. (2014). Best practices in teaching the new literacies of online research and comprehension. In L. Gambrell & L. Mandel (Eds.), *Best practices in literacy instruction* (5th ed., pp. 343–364.). New York, NY: The Guilford Press.

Lillis, T. M., & Curry, M. J. (2010). *Academic writing in a global context: The politics and practices of publishing in English*. New York, NY: Routledge.

Lockwood, R. (2000). Machine translation and controlled authoring at Caterpillar. In R. Sprung (Ed.), *Translating into success: Cutting-edge strategies for going multilingual in a global age* (pp. 187–202). Philadelphia, PA: John Benjamins.

López-Navarro, I., Moreno, A., Quintanilla, M., & Rey-Rocha, J. (2015). Why do I publish research articles in English instead of my own language? Differences in Spanish researchers' motivations across scientific domains. *Scientometrics*, *103*(3), 939–976. doi:10.1007/s11192-015-1570-1

Luo, N., & Hyland, K. (2016). Chinese academics writing for publication: English teachers as text mediators. *Journal of Second Language Writing*, *33*, 43–55. doi:10.1016/j.jslw.2016.06.005

Marshman, E. (2014). Taking control: Language professionals and their perception of control when using language technologies. *Meta*, *59*(2), 380–405. doi:10.7202/1027481ar

Martin, A. (2006). Literacies for the digital age. In A. Martin & D. Madigan (Eds.), *Digital literacies for learning* (pp. 3–25). London: Facet.

Meneses Benavides, G. A. (2011). Common difficulties observed in abstracts written in English by people having Spanish as mother tongue. *Revista Escuela de Administración de Negocios (Journal of the School of Business Administration)*, *70*, 78–86. doi:10.21158/01208160.n70.2011.534

Meyer, R. (2017). When a robot names a new color of paint. *The Atlantic,* May 22. Retrieved from https://www.theatlantic.com/amp/article/527421/

Miyata, R., & Fujita, A. (2017). Dissecting human pre-editing toward better use of off-the-shelf machine translation systems. In *Proceedings of the 20th annual conference of the European association for machine translation (EAMT)* (pp. 47–52). Retrieved from https://ufal.mff.cuni.cz/eamt2017/program.php

Montgomery, S. (2000). *Science in Translation: Movements of Knowledge Through Cultures and Time*. Chicago, IL: University of Chicago Press.

Montgomery, S. (2004). Of towers, walls, and fields: Perspectives on language in science. *Science*, *303*(5662), 1333. doi:10.1126/science.1095204

Montgomery, S. (2013). *Does science need a global language?: English and the future of research.* Chicago, IL: University of Chicago Press.

Moreno, A. I. (2010). Researching into English for research publication purposes from an applied intercultural perspective. In M. F. Ruiz-Garrido, J. C. Palmer-Silveira, & I. Fortanet-Gómez (Eds.), *English for professional and academic purposes* (pp. 59–71). Amsterdam, the Netherlands: Rodopi.

Moreno, A. I., Rey-Rocha, J., Burgess, S., López-Navarro, I., & Sachdev, I. (2012). Spanish researchers' perceived difficulty writing research articles for English-medium journals: The impact of proficiency in English versus publication experience. *Ibérica*, *24*, 157–184.

Mukherjee, J., & Hundt, M. (2011). *Exploring second-language varieties of English and learner Englishes*: *Bridging a paradigm gap*. Amsterdam/Philadelphia, PA: John Benjamins.

Mur Dueñas, P. (2012). Getting research published internationally in English: An ethnographic account of a team of finance Spanish scholars' struggles. *Ibérica*, *24*, 138–156.

National Research Council (NRC). (2015). *Machine translation: Benefits and advantages of statistical machine translation and NRC's Portage*. Technical report, National Research Council of Canada. Retrieved from https://nparc.nrc-cnrc.gc.ca/ eng/view/object/?id=72d5a618-0cd3-4b68-b56f-41178020c9d0

Nature. (2015). Special issue: Interdisciplinarity – Scientists must work together to save the world. *Nature*, *525*(7569), 289–418. Retrieved from https://www.nature.com/nature/journal/v525/n7569/index.html

O'Brien S., Simard, M., & Goulet, M.-J. (2018). Machine translation and self-post-editing for academic writing support: Quality explorations. In J. Moorkens, S. Castilho, F. Gaspari, & S. Doherty (Eds.), *Translation quality assessment*: *From principles to practice* (pp. 237–262). Cham, Switzerland: Springer. doi:10.1007/978-3-319-91241-7_11

Olohan, M. (2016). *Scientific and technical translation.* London: Routledge.

Olsson, A., & Sheridan, V. (2012). A case study of Swedish scholars' experiences with and perceptions of the use of English in academic publishing. *Written Communication*, *29*(1), 33–54. doi:10.1177/0741088311428566

Parra Escartín, C., O'Brien, S., Simard, M., & Goulet, M.-J. (2017). Machine translation as an academic writing aid for medical practitioners. In *Proceedings of the MT summit XVI* (Vol. 1, pp 254–267). Retrieved from https:// drive.google.com/file/d/0B jdsNNRF9Xu WU dzd 1 ZqZWFXaTQ/view

Parrish, D. (Ed.). (2009). Writing for translation: Getting started guide. *Multilingual*, October/November. Retrieved from https://multilingual.com/downloads/screenSupp107.pdf

Percillier, M. (2016). *World Englishes and second language acquisition: Insights from south-east Asian Englishes*. Philadelphia, PA: John Benjamins.

Pigott, I. M. (1986). Current systran developments at the EC commission. *Terminologie et traduction (Terminology and Translation)*, *1*, 10–16. Retrieved from http://www.mt-archive.info/T&T-1986-Pigott.pdf

Ping, K. (2009). Machine translation. In M. Baker & G. Saldanha (Eds.), *Routledge encyclopedia of translation studies* (2nd ed., pp. 162–169). London: Routledge.

Plo Alastrué, R., & C. Pérez-Llantada (Eds.). (2015). *English as a scientific and research language: Debates and discourses: English in Europe* (pp. 9–36). Berlin, Germany: De Gruyter Mouton.

Quah, C. K. (2006). *Translation and technology*. Basingstoke: Palgrave Macmillan.

Rowley, J. E., & Hartley, R. (2008). *Organizing knowledge: An introduction to managing access to information* (4th ed.). Farnham: Ashgate.

Schwartz, L. (2014). Monolingual post-editing by a domain expert is highly effective for translation triage. In *Proceedings of the third workshop on post-editing technology and practice* (pp. 34–44). Retrieved from https://www.amtaweb.org/AMTA2014Proceedings/AMTA2014Proceedings_PEWorkshop_final.pdf

Sereten, V., Bouillon, P., & Gerlach, J. (2014). A large-scale evaluation of pre-editing strategies for improving user-generated content translation. In *Proceedings of the ninth international conference on language resources and evaluation (LREC 2014)* (pp. 1793–1799). Retrieved from http://www.lrec-conf.org/proceedings/lrec2014/summaries/676.html

Smith, R. (2018). Where's the arena? Speak into my phone. *The New York Times*, July 14, p. D1. Retrieved from https://www.nytimes.com/2018/07/13/sports/world-cup/google-translate-app.html

Solovova, O., Vieira, J. & Veríssimo. J. (2018). Publish in English or perish in Portuguese: Struggles and constraints on the semi-periphery. *Publications*, *6*(2), 25. doi:10.3390/publications6020025

Somers, H. (2003). Sublanguage. In H. Somers (Ed.), *Computers and translation: A translator's guide* (pp. 283–296). Philadelphia, PA: John Benjamins.

Stupiello, E. N. A. (2008). Ethical implications of translation technologies. *Translation Journal*, *12*(1). Retrieved from http://translationjournal.net/journal/43ethics.htm

Swales, J. M. (2004). *Research genres: Explorations and applications*. Cambridge: Cambridge University Press.

Taylor, A. G. & Joudrey, D. N. (2009). *The organization of information* (3rd ed.). Westport, CT: Libraries Unlimited.

Tietze, S. (2008). The work of management academics: An English language perspective. *English for Specific Purposes*, *27*(4), 371–386. doi:10.1016/j.esp.2008.05.002

Translation Centre. (2015). *Writing for translation*. Luxembourg: EU Publications. Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/7c6c9877-f714-499b-980c-6d57259d75ea/language-en/format-PDF/source-75466077 (Note: this document is also available in French).

University of Ottawa. (2014). *Destination 2020*. *Discover the future*: *The University of Ottawa's strategic plan*. Retrieved from https://www.uottawa.ca/about/sites/www.uottawa.ca.about/files/destination-2020-strategic-plan.pdf

Van Dijk, T. A. (1994). Academic nationalism. *Discourse and Society*, *5*(3), 275–276. doi:10.1177/0957926594005003001

Varman, B. (2014). Conferences, reviews, and translations. In J. T. Huber & S. Swogger (Eds.), *Introduction to reference sources in the health sciences* (6th ed., pp. 157–181). Chicago, IL: Neal-Schuman.

Vasconcellos, M., & León, M. (1985). SPANAM and ENGSPAN: Machine translation at the Pan American Health Organization. *Computational Linguistics*, *11*(2–3), 122–136. Retrieved from http://aclweb.org/anthology/J85-2003

Vinson, T. C., & Welsh, T. S. (2014). A comparison of three library and information science databases. *Journal of Electronic Resources Librarianship, 26*(2), 114–126. doi:10.1080/1941126X.2014.910407

Wächter, B., & Maiworm, F. (2014). *English-taught programmes in European higher education: The state of play in 2014*. Bonn, Germany: Lemmens.

Wood, J., & Tovey, D. (2016). *Plain language summaries*: *Pilot proposal*. Retrieved from https://community.cochrane.org/review-production/production-resources/plain-language-summaries/resources-plain-language-summaries

Index

Page numbers followed by "*t*" indicate tables.