Fachkommunikation

Jörg Porsiel (Ed.)

Machine Translation

What Language Professionals Need to Know

BDÜ Fachverlag

Jörg Porsiel (Ed.)

Machine Translation

What Language Professionals Need to Know

BDÜ Fachverlag

Die Deutsche Bibliothek – CIP Einheitsaufnahme

Jörg Porsiel (Ed.): Machine Translation
What Language Professionals Need to Know

ISBN: 978-3-938430-94-1

verlegt von der BDÜ Weiterbildungs- und Fachverlagsgesellschaft mbH, Berlin,
einem Unternehmen des Bundesverbandes der Dolmetscher und Übersetzer e. V. (BDÜ)

Layout and design: Thorsten Weddig, Essen
Cover image: Tatiana Shepeleva/Fotolia.com
Printed by: Schaltungsdienst Lange oHG, Berlin

Contents

Dr. Aljoscha Burchardt, Jörg Porsiel

Preface

Machine translation: What can—and can't—it do?

To get straight to the point: Different approaches to machine translation (whether rule- or phrase-based, hybrid or, most recently, neural) have different results. Depending on the language pair and/or field, each approach generates different translations, some more and some less useful. The key factors in this are the quality and type of source text. The different approaches cannot be compared directly, so blanket statements like "system X is better than system Y" should always be taken with a pinch of salt.

Machine translation (MT) has experienced highs and lows since its inception in the 1930s. Critics of the technology like to recall the scientists' famous response when asked about advances in quality: "translators will be a thing of the past in five years"—the same answer has been given for roughly 80 years. However, after a long spell in the shadows, MT has been on everyone's lips again in recent years. Among other things, this is because computer and processor performance are no longer limiting factors, technically or financially, and because sufficient bilingual data is now available to generate the systems, some even in the required quality.

In spite of this, popular opinion could not be more varied: Some suggest that MT simply does not work properly, and is therefore irrelevant for professional translation work. They consider research in this field as *l'art pour l'art*. Their position is backed up by the continued poor quality of some translation results, and the fact that many potential users lack the required technology and suitable staff. However, the fact that Google alone machine translates roughly 100 billion words every day would appear to contradict this.[1] Also,

1 https://blog.google/products/translate/ten-years-of-google-translate/

relatively recently, marketers have been making the grand promise that MT will soon offer the same quality as human translation.[2] A third, equally widespread opinion is that language and translation themselves are not worth investing in, or even worth focusing on. Of all of these points of view, it is probably most important to disabuse holders of the last opinion.

A brief history

The history of machine translation (MT) in the 20th century can be traced back to the 1930s, over 80 years ago. The development did not really pick up speed and attract broader academic attention until the end of the Second World War, when the British and Americans attempted to translate German radio messages virtually simultaneously. These efforts were intensified during the Cold War. For the Americans, this culminated in the Georgetown University-IBM experiment in 1954, whose results elicited euphoric predictions. At the time, both scientists and the general public assumed that the big breakthrough was just around the corner: an end to all communication or translation problems, as rule-based machine translation (RBMT) seemed to be the solution to all linguistic problems. Perhaps that is one reason why we are now often confronted with *science fiction* rather than *science facts*.

However, the ALPAC report poured cold water on this illusion with its conclusion "*There is no immediate or predictable prospect of useful machine translation.*"[3] As a consequence, research and funding for machine translation in particular decreased drastically for decades. New approaches were introduced in the 1980s, including statistical machine translation (SMT). However, as it too failed to make the final breakthrough, improving some areas while weakening others, a hybrid approach was the next logical step. The assumption was that combining RBMT and SMT would enhance their respective strengths and more or less cancel out the respective weaknesses. This hypothesis has not been fulfilled yet either. The newest approach is neural

2 https://research.googleblog.com/2016/09/a-neural-network-for-machine.html

3 John R. Pierce, John B. Carroll, et al.: Language and Machines—Computers in Translation and Linguistics. ALPAC report, National Academy of Sciences, National Research Council, Washington, DC, 1966: www.nap.edu/read/9547/chapter/1

machine translation (NMT), which seeks to combine MT with artificial intelligence (AI). Will this latest attempt succeed in solving all or at least most linguistic and translation problems? We will see …

What is the aim of this book?

The aim of this book is to concisely and comprehensibly present the current status quo of MT under a variety of aspects, beginning with the technology and potential of the different system types, through cost-benefit considerations and invoicing models for post editing, right up to business models. It will serve as a guide for decision makers and linguistic experts and help them make decisions on introducing or using machine translation in their own professional environments.

We will explain what can—and cannot—be achieved under specific constraints, and why. We will discuss the potential this technology offers and the personnel, financial, textual, linguistic and technical problems its implementation and use can entail. We focus on presenting a neutral discussion of the subject by language and MT experts based on the latest findings. An objective assessment of the advantages and disadvantages of MT as a whole is essential to establish a tailored and needs-based process chain for machine translation.

Finally, we will also explain potential risks of using MT on the Internet and the currently largely ignored legal aspects of copyright for source and target texts, as well as the threat to data and information security online.

This is intended to help readers make business decisions on the use of MT based on the latest facts and knowledge, or at least to identify the additional expertise they need as a basis for such decisions, and where they can find more information and sources.

Intended audience

This book is intended for decision makers in medium and large (multinational) companies, who either employ their own translators or contract language

service providers (LSPs). It is also aimed at LSPs that want to or must use MT due to increasing customer demand, and of course at other interested parties such as IT experts, and not least at professional translators who want to keep up with the cutting edge of technological innovations in their professional environment.

MT: a Swiss Army knife or panacea for translation?

Nevertheless, this book will be a disappointment for anyone who sees machine translation as the universal solution to all problems, and believes that all they have to do is install any MT software (preferably the cheapest), without additional supporting measures like terminology and data management, controlled language, incorporating pre- and post-editing by suitably qualified personnel, and then expects to be able to translate any text from any language to any other language, eliminating the need for translators entirely.

This vision is fuelled in particular by (marketing) buzzwords like "fit for purpose translation" or "good enough translation", which suggest that it is all simply a question of the right MT approach and the volumes of data used, ignoring the complexity of languages, translation and MT. However, what the marketers generally neglect to explain is who decides what is "fit" or "good enough" and for whom, and how decisions like this are made. Formal specifications for the required translation properties and corresponding error metrics such as MQM (Multidimensional Quality Metrics)[4] are not yet commonplace and presuppose in-depth knowledge of language, translation, workflow used et cetera. In addition to this, most end users of translations are still exclusively accustomed to "human quality" and therefore (at best) amused to (often) irritated when presented with gist translations, as they have not learned to deal with this quality yet.

4 Multidimensional Quality Metrics: www.qt21.eu/launchpad/content/multidimensional-quality-metrics

On the other hand, professional translators are at the opposite end of the scale: they are sceptical and suspicious of the development, to say the least. Many fear that they will lose their jobs or at least be downgraded to mere badly paid assembly line tasks and job work as post-editors, responsible for cleaning up the mess made by the machines.

However, what machine translation definitely can do, is process/translate (very) high volumes of text in (extremely) short times; and do so consistently, depending on the technology used.

What makes MT so difficult?

We can only briefly mention a few problem areas as examples here. First of all, language and communication are highly complex systems, which humans apparently manage effortlessly, but which are very difficult to express in algorithms, and even then can only be expressed partially. A key feature of language is that it is often vague and ambiguous in everyday use. A subject can be expressed in various ways, and words or sentences can often have different meanings depending on the context. Thanks to the knowledge we accumulate over time and our intelligence, humans can generally establish clarity or at least make plausibility assumptions, for example what the two pronouns "she" and "her" in the sentence "Whenever Maria feeds her cat, she bites her" refer to, or what the problem with this specific delivery is: "The other day police stopped a truck carrying a consignment of forged pistons." The machine processes each individual sentence without considering its context and in particular without "knowledge" of the context, forcing it to constantly make decisions based on insufficient information. This heightens translation problems that are often challenging, even for humans.

Specific terminology and additional stylistic requirements are often important, even though they are not necessarily known or formally stipulated. While translators generally know from experience what the customer needs (text type, quality level, target group etc.), the machine lacks this knowledge: for example the simple question of whether customers are to be addressed formally or informally (e.g. "*Sie*" or "*Du*" in German). The MT workflow itself compounds the challenges of language and specifications. Depending on the technology, it involves steps like data preparation, training and tuning of the MT engines, terminology management, controlled language, pre- and post-

editing, continuous qualified feedback to the system administrator at every level, as well as quality measurement and evaluation.

What MT cannot offer at this linguistic level, or can only offer to a very limited extent, is error tolerance: guessing, interpreting, assuming, associating, anticipating or even paraphrasing, reading between the lines, processing non-translatables (correctly) or even distinguishing what makes sense from what does not. Without human knowledge of the world and without (artificial) intelligence, the software can neither learn independently, nor make decisions. Therefore, the "garbage in, garbage out" principle, only too familiar to all translators, takes its full toll here.

What this means is that there cannot be a rapid, one-size-fits-all solution when using machine translation. Every language direction, every text type and every field (domain), and possibly even every target group requires individual MT solutions.

What does it take to use MT successfully?

Machine translation is no more or less than a tool, which can be useful to translators if they know how to use it correctly. As a result, the right workflow must be developed for the task specified accordingly by/with the client (e.g. light or full post-editing, pre-editing). This includes identifying and assessing resources (e.g. terminology, MT system type, training corpuses). We must also determine whether the use of MT even makes sense for a specific field and a given text type. Key factors in this decision can include whether the translation is to be indiscernible and whether linguistic finishing touches and knowledge of the target market are essential, e.g. for marketing texts, or whether the translation is more repetitive in nature, such as internal minutes of routine meetings. These are just a few rough guidelines to give an impression of the number of parameters influencing successful use of MT. The relevant chapters of this book will go into more detail on these subjects.

Dr. Aljoscha Burchardt Jörg Porsiel

April 2017

Translated by Brendan Bleheen

About the Authors

Dr. Aljoscha Burchardt is lab manager at the Language Technology Lab of the German Research Center for Artificial Intelligence (DFKI GmbH). Burchardt has directed several major projects in the area of machine translation (MT) and is active in the central management of META-NET, a European Network of Excellence. His interests include the evaluation of (machine) translation quality and the inclusion of language professionals in the MT R&D workflow. Burchardt is co-developer of the MQM framework for measuring translation quality. He has a background in semantic language technology.

Jörg Porsiel studied English and French, with a minor in Law, at the Institute of Translation and Interpreting (IÜD) at Heidelberg University. He spent semesters abroad at Université Paul Verlaine in Metz, at the Institut Supérieur de Traducteurs et Interprètes (ISTI) in Brussels and Heriot-Watt University in Edinburgh. He graduated from Heidelberg University with a degree in translation in 1992. Since then, he has worked in the terminology management and international technical communication sectors. From 2002 on, he specialised in controlled language at Volkswagen AG in Wolfsburg, and has been responsible for deployment of machine translation at VW since 2005.

Ricarda Essrich: Positionierung als freiberuflicher Übersetzer – Spezialisierung oder Diversifikation?, Umfang: 120 Seiten, ISBN: 978-3-938430-87-3, Erscheinungsjahr: 2017, Preis: 27,00 €

Der Weg zum Erfolg als freiberuflicher Übersetzer führt über die Spezialisierung", behaupten die einen. „Ohne Diversifikation geht es nicht", die anderen. Doch wer hat Recht? Sollte man sich auf ein möglichst winziges exotisches Fachgebiet konzentrieren oder lieber möglichst viel aus möglichst unterschiedlichen Sparten anbieten, um viele Kunden bedienen zu können? Oder liegt die Lösung vielleicht irgendwo dazwischen? Die Autorin macht sich in ihrem Buch auf die Suche nach Antworten auf diese Fragen. Dabei lässt sie in eingestreuten Interviews einige erfolgreiche Kolleginnen und Kollegen zu Wort kommen, die erzählen, wie sie sich als Übersetzer positioniert haben. Die Autorin ist der Überzeugung, dass Übersetzer nur durch eine geschickte Positionierung, die meist auch eine Spezialisierung bedeutet, langfristig wirtschaftlich erfolgreich sein können. Sie zeigt in diesem Buch, wie man sich spezialisiert, was Spezialisierung bedeutet und wie man trotz Spezialisierung das eigene Portfolio verbreitert, um Risiken zu reduzieren und Wachstumspotenziale zu nutzen. Die Konzentration auf bestimmte Fachgebiete oder Fähigkeiten könnte für etliche Sprachmittler der Weg zu deutlich mehr Erfolg am Markt sein. Wie das gehen kann, zeigt vorliegendes Buch.

Angelika Ottmann (Hrsg.): Best Practices – Übersetzen und Dolmetschen. Ein Nachschlagewerk aus der Praxis für Sprachmittler und Auftraggeber, Umfang: 398 Seiten, ISBN: 978-3-938430-85-9, Erscheinungsjahr: 2017, Preis: 31,00 €

Best Practices sind bewährte Verfahrensweisen, d. h. Methoden, Prozesse, Arbeitsweisen und Modelle, die sich in der Praxis bewährt haben und von einem Großteil der Praktiker angewendet und unterstützt werden. In diesem Sinne wendet sich das vorliegende Werk an alle, die mit den Dienstleistungen Übersetzen und Dolmetschen befasst sind, sei es als Ausführende (Übersetzer, Dolmetscher), als Vermittler (Agenturen) oder als Auftraggeber (Unternehmen, Behörden, Institutionen). Es ist ein Nachschlagewerk und Referenzhandbuch für alle, die Antworten auf ihre speziellen Fragen suchen, für alle, die wissen möchten, welche Anforderungen ihr Gegenüber hat, für alle, die über den Tellerrand ihrer Tätigkeit hinausblicken und sehen wollen, wie andere es machen. Das vorliegende Buch ist kein Existenzgründungsleitfaden – die Best Practices legen vielmehr die Standards der Branche zu Vorgehensweisen beim Übersetzen und Dolmetschen dar.

Das Buch enthält eine Vielzahl von Checklisten, die unten in einer ZIP-Datei zusammengestellt sind, heruntergeladen und für den eigenen Gebrauch angepasst werden dürfen.

Jörg Porsiel (Ed.)

Machine Translation

What Language Professionals Need to Know

This anthology is the first in the German-speaking world to offer a comprehensive overview of all key aspects of "Machine Translation". **The articles were written by experts from the fields of research and development, academics and consulting, as well as lawyers and translators.** It is the perspective of translators in particular, who make up half of the authors, that give the book its strong practical focus. In addition, **two articles by lawyers** cover the extremely important topics of data and information security and copyright, which are largely unknown even among specialists. Finally, **case studies from five countries** round off the book.

Jörg Porsiel studied English and French, with a minor in Law, at the Institute of Translation and Interpreting (IÜD) at Heidelberg University. He spent semesters abroad at Université Paul Verlaine in Metz, at the Institut Supérieur de Traducteurs et Interprètes (ISTI) in Brussels and Heriot-Watt University in Edinburgh. He graduated from Heidelberg University with a degree in translation in 1992. Since then, he has worked in the terminology management and international technical communication sectors. From 2002 on, he specialised in controlled language at Volkswagen AG in Wolfsburg, and has been responsible for deployment of machine translation at VW since 2005.

ISBN: 978-3-938430-94-1 49,00 € [D]

9 783938 430941

www.bdue-fachverlag.de

Ratgeber | Tagungsbände | Gesetze und Normen | Fachkommunikation | Lernen und Lehren | Fachterminologie

Weiterbildungs- und Fachverlagsgesellschaft **Fachverlag**

BDÜ